Prioritizing persistent microbiome members in the common bean rhizosphere: an integrated analysis of space, time, and plant genotype

Nejc Stopnisek[1,2,3] and Ashley Shade[1,2,3,4,5]

1. Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48840 USA
2. The Plant Resilience Institute, Michigan State University, East Lansing, MI 48840
3. The DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48840
4. Program in Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, MI 48840
5. Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48840

Corresponding author: shadeash@msu.edu

Keywords: plant-microbiome, Phaseolus vulgaris, plant health, neutral model, co-occurrence network, abundance-occupancy, 16S rRNA, ITS, hub taxa, sustainable agriculture, core microbiome

Abstract

The full potential of managing microbial communities to support plant health is yet-unrealized, in part because it remains difficult to ascertain which members are most important for the plant. However, microbes that consistently associate with a plant species across varied field conditions and over plant development likely engage with the host or host environment. Here, we applied abundance-occupancy concepts from macroecology to quantify the core membership of bacterial/archaeal and fungal communities in the rhizosphere of the common bean (*Phaseolus vulgaris*). Our study investigated the microbiome membership that persisted over multiple dimensions important for plant agriculture, including major growing regions, plant development, annual plantings, and divergent genotypes, and also included re-analysis of public data. We found 48 core bacterial taxa that were consistently detected in all samples, inclusive of all datasets and dimensions. This suggests reliable enrichment of these taxa to the plant environment and time-independence of their association with the plant. More generally, this work provides a robust approach for systematically prioritizing core microbiome memberships in any host or system.

36 **Introduction**

37 Agriculture requires more efficient use of available resources, and the naturally occurring,
38 soil-dwelling microbiota offers potential to contribute to the responsible intensification of
39 agriculture. Selection and breeding of plants for their beneficial associations with microbiota has
40 promise to deliver a new generation of microbe-improved plants [1–6]. The ideal outcome of such
41 efforts would achieve a balance of sustainable agriculture with food security. To achieve this, we
42 must understand the relationships between plants and their associated microbiomes, including the
43 differentiation of key or "core" members that engage with the plant directly from transient or
44 opportunistic members that do not.

45 Common bean (*Phaseolus vulgaris* L.) is the most important food legume grown
46 worldwide, and especially for developing economies in South America, Africa and Asia [7]. The
47 origin of common bean is central Mexico, and from there it spread to central and to south America
48 around 165,000 years ago [8]. This resulted in the development of two major and eco-
49 geographically distinct common bean gene pools with partial reproductive isolation [9–11]. The
50 Mesoamerican gene pool was distributed from northern Mexico to Colombia and the Andean gene
51 pool ranged from southern Peru to northwestern Argentina. Since 8,000 years ago, each pool was
52 separately and selectively bred, leading to further diversification between them [8, 12, 13]. Because
53 of pre-existing genetic differences in each gene pool followed by divergent breeding history,
54 common bean presently offers a distinctive opportunity for understanding how the host and the
55 environment contribute to rhizosphere microbiome assembly.

56 The objective of this study was to apply approaches from macroecology to prioritize the
57 persistent members of a core bacterial, archaeal, and fungal root microbiome inclusive of multiple
58 gradients and categories of drivers expected to be important for plant agriculture. With the
59 cooperation of the U.S. Bean coordinated agricultural project (Bean CAP), we executed a first-of-

60 its kind study of two divergent bean genotypes grown under field conditions in five major North
61 American bean growing regions, including Michigan, Nebraska, Colorado and Washington. These
62 two genotypes belong to the Mesoamerican (Eclipse genotype) and Andean (California Early Light
63 Red Kidney, CELRK genotype) gene pools that represent the major divergences from the wild
64 bean ancestor [8]. For these two divergent genotypes, we also assessed how core members of the
65 root microbiome changed over plant development and root compartment. We also used public data
66 to perform a comparative analysis of the bacterial microbiome found in our study with microbiome
67 members detected in other bean genotypes grown in South America [14, 15]. From our effort that
68 was inclusive of both broad biogeography (including the U.S. and Colombia), plant development,
69 and inter-annual plantings, we discovered a core bean rhizosphere microbiome of 48 members that
70 persistently associated with this nutritionally, agronomically, and economically important crop.
71 This core was discovered in spite of apparent microbiome differences that were attributable to local
72 soil conditions and management. However, we did not detect an influence of plant genotype,
73 suggesting that this core membership supersedes it.

74 **Material and Methods**

75 *Study design, sampling and soil physicochemical analysis*

76 We designed a biogeography study of two divergent bean genotypes (from Mesoamerican
77 and Andean gene pools [16, 17]), both grown in the field in the summer of 2017 at five research
78 and extension farms that represent major U.S. bean production regions (**Table 1**) [18]. The research
79 and extension farms were: Saginaw Valley (SVERC), Michigan (MI), Montcalm county (MRF),
80 MI, Scott Bluff county, Nebraska (NE), Fort Collins, Colorado (CO), and Othello, Washington
81 (WA).

82 Triplicate bean plants were grown in each of three (MI, WA, CO), or four (NE) plots,
83 totaling 9 or 12 plants per growing location. Plants were harvested at flowering stage (mid to late
84 July) because we wanted to analyze mature microbial communities and control for potential
85 differences in the microbiome over early plant development.

86 We selected two common bean cultivars with very distinct genotypes: Eclipse [16] of
87 Mesoamerican origin and California Early Light Red Kidney (CELRK), an old kidney bean
88 landrace of Andean origin [17]. Even though we included only two bean genotypes, we selected
89 highly divergent genotypes from far ends of the spectrum of the common bean genetic diversity;
90 these lineages diversified after a biogeographic bottleneck that separated them between Central and
91 South America [19]. We hypothesized that if there was an effect of bean genotype on the root
92 microbiome, it should be measurable when comparing these genotypes from divergent lineages.
93 Both bean genotypes were grown in each growing location, except in Michigan, where CELRK
94 was grown at Saginaw Valley and Eclipse at Montcalm county.

95 Plant roots with attached soil were packed into bags, stored on ice and shipped immediately
96 to Michigan State University for processing. From each sample, root-attached soil was collected by
97 gently shaking the roots and designated as rhizosphere soil; unattached root-zone soil was also

98 collected and used as bulk soil. It was expected that the rhizosphere soil contains a subset of
99 microbiome diversity that is recruited from the bulk soil [20] but we note that these soils were not
100 "bulk" in the traditional definition, in that they were collected from the agricultural fields and
101 expected to be proximal to and influenced by the roots of previous crops. Soils were sieved through
102 4 mm sieves to remove any plant debris and larger soil minerals before soil physical-chemical
103 analysis. Soil analysis was done at the Michigan State Soil Plant and Nutrient Laboratory on root-
104 zone soil samples pooled by plant genotype and plot (see Supporting Text for details).
105 We designed a second temporal study to assess the dynamics of the core taxa over plant
106 development. The two bean genotypes, CELRK and Eclipse, were grown at the same two sites in
107 Michigan, U.S., Montcalm county and Saginaw Valley, in the summer of 2018. Plants root systems
108 were harvested at 5 growth stages, including stage 1: V2 (appearance of second trifoliate), stage 2:
109 V5 (appearance of fifth trifoliate), stage 3: flowering, stage 4: pod filling, and stage 5:
110 senescence/drying). At each sampling time, roots were collected and transported on ice to the
111 laboratory for immediate processing. Rhizosphere soil was collected by gently shaking roots, as
112 described above. Any remaining, tightly-attached soil was designated as rhizoplane soil, and was
113 collected by first vortexing roots in 1x phosphate buffer solution (PBS) for 4 min and finally
114 removing the supernatant after centrifugation for 10 min at 8000 *g* and 4°C. Collected soil was
115 immediately frozen in liquid nitrogen and stored at -80°C.
116
117 *Microbiome sequencing and analysis*
118 There were 31 rhizosphere (attached to the root and pooled by plot) and 8 root-zone/bulk
119 (one pooled sample per growing location and plant genotype) soil samples sequenced for
120 microbiome analysis from the biogeography study, and 125 rhizosphere and 127 rhizoplane soil
121 (individual plants) samples sequenced from the plant development study. DNA extractions,

122 including negative controls, were performed as per standard soil protocols (see Supporting Text for
123 details). 16S rRNA gene amplicon and ITS amplicon sequencing was performed at the Michigan
124 State Genomics Core Research Support Facility. We processed the 16S rRNA gene amplicons
125 using an open-reference clustering first against the SILVA v128 database [21, 22]. First, raw reads
126 were merged, quality filtered, dereplicated, and clustered into 97% identity operational taxonomic
127 units (OTUs) using the UPARSE pipeline (version 11, [23]). Reads not matching to the SILVA
128 database were used for de-novo clustering at 97% sequence identity. Reference picked and de-novo
129 reads were combined before taxonomy was assigned [21]. Taxonomic annotations for 16S rRNA
130 gene OTU representative sequences were assigned in the QIIME 1.19 environment [24] using
131 SILVA database [22]. ITS OTU representative sequences were taxonomically annotated using the
132 CONSTAX tool [25] with the UNITE database version 7.2 [26]. OTUs with unassigned taxonomy
133 at the domain level and OTUs annotated as mitochondria or chloroplasts were removed.
134 Contaminant OTUs were removed using decontam package in R [27]. Additionally, we performed
135 zero-radios OTU analysis (aka ZOTUs that are clustered at 100% sequence identity) for reads
136 detected within the core OTUs. In short, first we subset all reads that were included within the core
137 OTUs and processed them through the UNOISE3 pipeline [28] to create ZOTUs. Then we
138 determined each ZOTU's affiliation to their originating core OTU by matching the read identities
139 between ZOTU and OTU clusters using a customized code (see GitHub repository).
140 Statistical analysis, including abundance-occupancy analysis to detect a core microbiome
141 [29], and data visualization were performed in R (see Supporting Text). The co-occurrence network
142 and global network properties were calculated using the Molecular Ecological Network Analysis
143 Pipeline (MENAP) [30] and visualized with Cytoscape v.3.5.1 [31]. Comparative analyses with
144 published datasets from Pérez-Jaramillo et al. [15] (BioProject ID PRJEB26084). was performed
145 by downloading study raw reads from NCBI and processing as per our data pipeline to analyze

146 both sets identically. An expanded description of all ecological statistics, MENAP parameters, and
147 our combined core microbiome analysis with public data are provided in Supporting Text.
148
149 *Data and code availability*
150 The raw sequence data are deposited in the NCBI Sequence Read Archive (BioProject ID
151 PRJNA524532). All read processing steps, bioinformatic workflows, R code, and custom scripts
152 are available on GitHub
153 (https://github.com/ShadeLab/PAPER_Stopnisek_2019_BeanBiogeography).
154
155 **Results**
156 *Sequencing summary and microbial diversity across growing regions*
157 There were 31,255 to 506,166 and 22,716 to 252,810 reads per sample for 16S rRNA and
158 ITS biogeography datasets, respectively. We rarefied samples to 31,255 reads for 16S rRNA gene
159 amplicons and to 22,716 for ITS. With these thresholds, we achieved richness asymptotes for both
160 datasets, suggesting that sequencing efforts were sufficient to capture comparative dynamics and
161 diversity (**Fig. S1**). The total richness observed at this rarefaction depth was 1,505 fungal and
162 23,872 bacterial and archaeal OTUs.
163 As reported in other rhizosphere studies, the total fungal diversity was lower than bacterial
164 /archaeal diversity in the rhizosphere of the common bean [32–34]. Richness varied by growing
165 location (ANOVA, F value=12.4, p-value<0.0001 and F value=13.1, p-value<0.0001 for 16S
166 rRNA and ITS data, respectively, **Fig. S2**) but was highest at the Montcalm Research Farm
167 (Michigan, US) for both, bacteria/archaea and fungi. An analysis of community beta diversity
168 revealed strong biogeographic patterns in community structure explained by location, soil pH and
169 fertilization, in agreement with other literature [35–38] (**Fig. S3**, see **Supplemental Text**).

170 However, an influence of plant genotype was either weak or not detected, which we partially
171 attribute to plant growth in field conditions, following observations made in other plant species
172 grown under field conditions [35, 36, 39] (**Fig. S2BD**, see **Supplemental Text).**
173 The common bean rhizosphere microbiome included the major expected lineages for both
174 bacteria and fungi (**Fig. S4AB**), in agreement with other plant rhizosphere studies [40–45].
175 Together with previous studies, these data provide more evidence that root-associated microbial
176 taxa are phylogenetically and potentially functionally conserved [46]. Proteobacteria,
177 Acidobacteria, Bacteroidetes, and Actinobacteria collectively comprised on average 73.5% of the
178 bacteria/archaeal community, Ascomycota dominated the fungal community with a mean total
179 relative abundance of 53% with notable sample-to-sample variance (range from 16.5% to 84.5%).
180
181 *A core rhizosphere microbiome is detected across U.S. bean growing regions*
182 We noticed a large number of OTUs that were shared among all growing locations them for
183 the bacterial/archaeal communities (2,173 taxa, mean 31.5%, range 29.5% to 34.7%). There was a
184 smaller but notable overlap for the fungal communities (70 taxa or mean 4.5%, range from 0.9% to
185 17.9%; **Fig. S4CD**). These data suggested that, despite measured edaphic differences across
186 growing locations and strong biogeographic signal, the common bean rhizosphere recruited many
187 similar taxa that could be functionally important for the bean. Therefore, we explored abundance-
188 occupancy distributions of taxa ([47, 48] and references therein) to infer the core bean microbiome
189 of taxa with an occupancy of 1 (i.e., found in all soil samples, all plots and across all growing
190 locations; **Fig. 1AB**). Among bacteria and archaea, 258 phylogenetically diverse taxa were
191 cosmopolitan in the dataset (**Fig. 1C**), including numerous and abundant Proteobacteria (117
192 OTUs) with a dominant taxon classified as *Arthrobacter* sp. (FM209319.1.1474, mean relative
193 abundance of 1.43%). The bacterial/archaeal core also contained taxa of interest for potential plant

194 benefits (e.g. *Sphingomonas*, *Rhizobium*, *Bacillus, Streptomyces*), as well as some genera that can
195 be associated with disease (e.g. *Ralstonia*). There were 13 taxa in the fungal core (**Fig. 1D**), and
196 these were largely composed of Ascomycota (10 OTUs), with dominating taxon OTU823 from the
197 *Phaeosphaeriaceae* family (mean relative abundance 10.1%). Notably, taxa that were unique to
198 either bean genotype were relatively rare and inconsistently detected (**Fig. 1**, orange and black
199 points). Together, these results suggest that common bean consistently recruits particular
200 microbiome taxa.

201 Next we wanted to investigate if the prioritized core taxa are indeed selected by the plant
202 environment or assembled through neutral processes by applying the Sloan neutral model [49, 50].
203 The neutral expectation of abundance-occupancy distributions is that very abundant taxa will have
204 high occupancy, while rare taxa will have low [48–51]. Taxa that deviate from the neutral
205 expectation are more strongly influenced by deterministic factors, like environmental conditions,
206 than by stochastic factors, like drift and dispersal. The neutral model fit of the abundance-
207 occupancy distribution (solid line, **Fig. 1AB**) identified several taxa that had frequencies either
208 above or below the 95% confidence intervals of the model (dashed lines). Specifically, 13.7% of
209 the bacterial/archaeal and 30.4% of fungal taxa, deviated from the neutral expectation (**Table S3**).
210 One hundred and seventy-one core taxa were predicted above the neutral model partition; these
211 deterministically-selected taxa are prime candidates for follow-up studies of interactions with the
212 host plant. Overall, the bacteria/archaea community had better fit to the neutral expectation than
213 fungal (R^2 of 0.74 and 0.34, and migration rates (m) of 0.301 and 0.003, respectively), suggesting
214 that dispersal was relatively less limiting for the bacteria/archaea than for the fungi. This finding
215 agrees with other work suggesting that fungi are more sensitive to local climate or more dispersal
216 limited than bacteria [52–55].
217

218 *A core rhizosphere microbiome is detected for common bean grown on different continents*

219 We wanted to better understand if these U.S. core taxa were associated with the bean
220 rhizosphere across a larger geographical scale, which would suggest the potential for selective plant
221 recruitment and cosmopolitan distribution of core taxa. Therefore, we compared our U.S. data to a
222 recently published study of rhizosphere bacteria and archaea from common beans grown in
223 Colombian agricultural soil [14]. The Colombian study offered a key contrast because it included
224 eight divergent bean lineages, including wild (n=2), landrace (n=1), and cultivated genotypes
225 (n=5), grown in soil from a different continent that has starkly different climate and management
226 from the U.S. growing regions. To enable direct comparison, we re-analyzed raw reads and
227 compared the datasets by matching to either the same taxon identifiers when clustered to SILVA
228 database, or 100% identity by BLAST to *de novo* clustered reads (see Materials and Methods).
229 Surprisingly, 39.6% (3,359 OTUs) of rhizosphere taxa from the Colombian-grown beans were also
230 shared with the U.S. dataset (**Fig. 2**). Both datasets included taxa that were highly represented in
231 the other: 62% of U.S. core (159 out of 258) were found also in Colombia, and 51% of Colombian
232 core (433 out of 848) were shared with the U.S. (**Fig. 2A**). Core taxa were again defined stringently
233 with an occupancy of 1, and 48 taxa were found across all samples, inclusive of both datasets. We
234 refer to this as the "global" core to distinguish the subset from the larger group of core taxa
235 inclusive to the US only (though note that this descriptor is for simplicity and that this does not
236 include global representation of bean root samples on Earth). These global core taxa were
237 composed of many Proteobacteria, with *Rhizobiales* showing the most consistent relative
238 abundance between the studies (**Fig. 2B,** e.g. 0.187% and 0.138% in Colombia and U.S. dataset,
239 respectively). Notably, none of this global core taxa were universally detected in very high
240 abundance, and all but two OTUs (a U.S. *Arthrobacter* sp. and a Colombian *Austrofundulus*
241 *limnaeus* with mean relative abundances of 1.43% and 1.01%, respectively) would be classified as

242 rare by a typical mean relative abundance threshold below 1%, hinting to a potential role of rare
243 taxa in providing key functions. A similar observation was made with rhizosphere microbiota of 19
244 herbaceous plant species, in which taxa of low abundance were among those significantly enriched
245 in the rhizosphere as compared to the bulk soil [56]. Notably, only 48% of these global core taxa
246 have genus classification, suggesting that most of them are under-described in their functional
247 potential and interactions with plants (**Table S4**).

248 We also analyzed reads associated with the global core taxa with the UNOISE3 pipeline
249 [28] to generate predicted biological sequences (zero-radius OTUs – ZOTUs with100% sequence
250 identity) and provide the maximal possible biological resolution. We found that the global core
251 taxa consisted of 422 ZOTUs, and that there was a range of 2 to 35 ZOTUs identified within each
252 OTU (**Fig. S5**). With one exception (HQ597858.1.1508), all of core OTUs (clustered at 97%
253 sequence identity) contained at least one ZOTU that also had an occupancy of 1. In addition, all of
254 the ZOTUs with an occupancy of 1 were also the most abundant ZOTUs within each OTU (**Fig.**
255 **S5**). This result suggests that the same members constitute the core even with increased taxonomic
256 resolution.

257 A recent study considered the effect of the common bean domestication history on the root
258 microbiome and identified a core set of microbial taxa that are consistently present with these
259 diverse bean genotypes, including bacterial taxa recruited from agricultural and natural soil from
260 Colombia [15]. This core also had high representation of *Proteobacteria*, *Acidobacteria*,
261 *Actinobacteria* and *Verrucomicrobia*, similar to those observed in the present study. We also re-
262 analyzed these data and show that 46 (out of 48) of the global core taxa identified here have
263 occupancy>0.9 across all three included studies (the present study, Pérez-Jaramillo et al. [14] and
264 Pérez-Jaramillo et al. [15]). Forty-two of these taxa had the highest possible occupancy of 1, but
265 only within rhizosphere samples from agricultural soils (**Table S4**). When incorporating data from

266 beans grown in forest soils, the average occupancy decreases to 0.77 (median 0.75, min = 0.53).
267 However, 14 taxa still had an occupancy of 1 when including beans grown in forest soils (**Table**
268 **S4**).

269 In summary, these results show that common bean can associate with a core set of
270 rhizosphere microbiome members at taxon and ecotype levels, across diverse bean genotypes and
271 across continents. Additionally, these core taxa are likely enriched by the host in managed soils, as
272 suggested by their higher occupancy in agroecosystems.

273

274 *Core taxa are enriched in the rhizoplane and are consistently detected across bean development*

275 To identify the common bean core taxa over space (a biogeographic core) we sampled
276 plants across growing locations at the same growth stage (flowering) and focused on the
277 rhizosphere compartment. However, the question remained whether these core taxa are detected
278 beyond that particular plant development stage. To answer this question, we conducted a field
279 experiment to assess the core taxa over time in plant development stage. In the next growing season
280 (2018), we used the same divergent bean genotypes grown at both Michigan, U.S, locations
281 (Montcalm and Saginaw Valley; see Material and Methods). We harvested root systems at 5
282 different plant development stages including flowering stage. We investigated the relative
283 abundance of the global core taxa and the U.S.-specific core taxa in both rhizosphere (soil that
284 could be removed from the root after shaking, n=125 samples) and rhizoplane (soil adhered to the
285 root tissue and removed via vortex in buffer, n=127 samples) compartments to determine their
286 ability to closely associate with the plant tissue. The range of the rhizosphere sequencing depth was
287 7,905-78,436 reads per sample, and for the rhizoplane it was 32-189,433 reads per sample. We
288 rarefied to 15,000 reads per samples (i.e. samples reaching richness asymptote), resulting in loss of

289 3 rhizosphere and 7 rhizoplane samples for a final dataset of 122 rhizosphere and 120 rhizoplane
290 samples. The total richness observed was 36,022 bacterial and archaeal OTUs.

291 From this development time series, we found that all 48 global core taxa were detected a
292 year later on these two Michigan farms. The collective relative abundances of the global core taxa
293 were significantly higher in the rhizoplane as compared to the rhizosphere irrespective of the plant
294 development stage, bean genotype and growing location (**Fig. 3A**). Interestingly, the remaining US
295 core taxa that were found exclusively in the U.S. dataset at an occupancy of 1 were equally
296 abundant in the rhizosphere and rhizoplane at both Michigan growing locations (**Fig. 3A**).

297 We next asked whether there were enrichments of particular core taxa by plant development
298 stage, root compartment, or growing location (**Fig. 3B**). On balance, almost all core taxa showed
299 some growth stage preference, but these trends were specific to each growing location and root
300 compartment. Despite these nuances, all core taxa were consistently found with high occupancy
301 inclusive of the plant development series (**Fig. S6**).

302 Together these results suggest that these core taxa are selected by the plant early in the
303 development stage and maintained. Enrichment of the core taxa in the rhizoplane further supports
304 the hypothesis that these core taxa engage closely with the host plant.

305

306 *Core taxa are not hub or connector taxa in an inter-domain microbiome network*

307 Network analysis has been proposed to be a useful method to identify important members
308 of the plant microbiome with beneficial traits [57–59]. Hub taxa, identified by their high
309 connectivity with many members of the community, are regarded as the most important part of the
310 community and influence network structure and community stability [60, 61]. We applied this
311 method to ask if any of the core taxa that we identified using abundance-occupancy were also key
312 for co-occurrence network structure. Additionally, we were interested in identifying fungal-

313 bacterial co-occurrences because of reports of their potential benefits for the plant [62, 63]. To
314 explore these patterns we applied the molecular ecology network analysis pipeline (MENAP)
315 which constructs ecological association network through random matrix theory (RMT) [30]. We
316 merged rarified 16S and ITS rhizosphere datasets, filtered the datasets to include taxa with
317 occupancy greater than or equal to 50%, and considered only interactions significant at *p*-
318 value<0.05 and RMT threshold of 0.88. The resulting network included 572 taxa (nodes) and 1,857
319 statistically significant co-occurrences (edges) structured among 52 modules. Most of the modules
320 were relatively small, with only six including more than 10 nodes (**Fig. 4**). The network was scale-
321 free (i.e. characteristics of the network are independent of the size of the network) and had small-
322 world characteristics (i.e. highly clustered) as indicated by the node degree distribution fitting to
323 the power law model (R^2= 0.993), and also had significant deviation of the modularity, length and
324 clustering coefficients from those calculated from random network (i.e. same number of nodes and
325 edges), respectively (**Table S5**).

326 The topological role of each taxon within the network was determined by the relationship
327 between their within-module (Zi) and among-module (Pi) connectivity scores as described in [64].
328 Based on this, the majority of taxa were peripheral (potentially, specialists; 563 nodes) (**Fig. 4A**).
329 There were also 3 connectors and 6 module hubs, but no network hubs. Indeed, in agreement with
330 the beta-diversity analysis (**Fig. S3**, Supporting Materials), there was a strong geographic signal in
331 the largest four modules, and these were comprised mostly of bacterial-bacterial (rather than
332 bacterial-fungal or fungal-fungal) hypothesized interactions (**Fig. 4B**). We note that cross-domain
333 edges constituted only a small fraction of all co-occurrences (n=168; bacteria-fungi=156, archaea-
334 bacteria=6 and archaea-fungi=6).

335 The analysis identified 26 co-occurrences between core taxa. However, we were surprised
336 to find that only two bacterial core and no fungal core taxa were also classified as network hubs

337 (taxa that connect to many other taxa within a modules) or connectors (taxa that connect across
338 modules). As exceptions, core *Chitinophagaceae* taxon FR749720.1.1492 was a module hub node
339 and a *Nitrobacter* sp. GDHX01215817.4.1477 was a connector. Our results, inclusive of a dataset
340 of divergent plant genotypes and broad biogeography, suggest that while hub and connector taxa
341 may be important for the maintenance of the root microbiome, these taxa are not consistently
342 detected in the common bean rhizosphere and, by deduction, could not be of universal importance
343 for the host plant. Our study cannot speak to the potential for functional redundancy among hub or
344 connector taxa, which could ultimately suggest a functional core among phylogenetically diverse
345 taxa [65]. Taken together, these results suggest that core taxa likely are important for the plant,
346 while hub and connector taxa are important for the integrity of the soil microbial community and its
347 responses to the local environment.
348

349 **Discussion**

350 Numerous studies have aimed to detect core microbiome members for various plants and
351 animals (see references within [29, 66]). Because methods are inconsistent across studies and
352 because the parameters used to define a core are often arbitrary, there is some question as to how
353 robust or useful such studies may be. For these reasons, it has become easy to overlook new studies
354 that claim to have discovered a core microbiome for their system. If every study or design results in
355 a different core microbiome, how can the research build to move forward?

356 Here, we applied concepts from macroecology to prioritize, for the first time, a core plant
357 microbiome that persists at across growing locations on two continents, over plant development,
358 over two years of annual plantings at two different farms, across highly divergent plant genotypes,
359 and also across datasets collected by different research groups. Furthermore, core members were
360 enriched in the rhizoplane, and not just the rhizosphere. These multiple and consistent lines of

361 evidence provide strong support that these taxa likely engage with the plant and also demonstrate
362 the robustness of the abundance-occupancy method to discover core members.

363 A majority of the 48 core bacterial taxa among these common bean rhizospheres are under-
364 described or largely unknown, as only 23 out of 48 have a genus-level classification. But, due to
365 their ubiquitous association with the bean over space and time, we hypothesize that these taxa
366 provide functions that are crucial for common bean health and should be targeted for microbiome
367 management of in support of common bean productivity and wellness. To test this hypothesis
368 further research is needed, but plant-beneficial traits have been previously reported for members of
369 the genera that we identified as part of the bean core microbiome. For example, members
370 belonging to the genera *Mesorhizobium* and *Rhizobium* are known symbiotic nitrogen fixers of
371 legumes [67, 68], *Ramlibacter* sp. have the ability to promote P mobilization [69], and *Variibacter*,
372 *Novosphingobium* and *Sphingomonas* sp. harbor many specialized genes indicating their
373 relationship with plant hosts [70–72].

374 Discovering these core taxa is a first step in a rich line of inquiry to understand host
375 engagement with them. The next steps are to understand functions associated with these taxa and to
376 determine how they contribute to plant health and productivity under different growth conditions,
377 such as drought or with particular management strategies [39, 73]. These steps will include
378 cultivation dependent and independent approaches aimed to enrich and isolate core members,
379 assemble or bin genomes from isolates and metagenomes, annotate functional genes on both
380 chromosomes and plasmids, link functions and activities through transcript or metabolome
381 analyses, and perform experiments with constructed communities of core members to test
382 hypotheses about microbiome engagement with and benefits to the plant.

To conclude, this work provides robust approaches and general insights for prioritizing core microbiome members, and that also advance goals in plant-microbiome management and microbe-improved crops by providing insights into core member identities and dynamics.

Acknowledgments

This work was supported by the Plant Resilience Institute at Michigan State University. We are grateful to Dr. James Kelly, Dr. Phillip Miklas, Dr. James J. Heitholt, Dr. Carlos Urrea Florez, Dr. Mark A. Brick, Dr. Juan M. Osorno, Dr. Thomas H Smith and other supporting staff at the research extension farms for their partnership in this study. We would also like to thank the Michigan State Genomics Core Research and Technology Support Facility and the Institute for Cyber-Enabled Research High Performance Computing Center for excellent support and service. AS acknowledge support from Michigan State AgBioResearch (Hatch).

Competing Interests

The authors declare no conflict of interest.

Author contributions

N.S. and A.S. designed research, analyzed data and wrote the paper; and N.S. performed research.

Supporting Text, Figures, and Tables accompany this manuscript.

405 **References**

406 1. Dessaux Y, Grandclément C, Faure D. Engineering the Rhizosphere. *Trends Plant Sci* 2016;
407 **21**: 266–278.
408 2. Gopal M, Gupta A. Microbiome Selection Could Spur Next-Generation Plant Breeding
409 Strategies. *Front Microbiol* 2016; **7**: 1971.
410 3. Compant S, Samad A, Faist H, Sessitsch A. A review on the plant microbiome: Ecology,
411 functions, and emerging trends in microbial application. *J Adv Res* 2019; **19**: 29–37.
412 4. Wille L, Messmer MM, Studer B, Hohmann P. Insights to plant-microbe interactions
413 provide opportunities to improve resistance breeding against root diseases in grain legumes.
414 *Plant Cell Environ* 2019; **42**: 20–40.
415 5. Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, et al. Research
416 priorities for harnessing plant microbiomes in sustainable agriculture. *PLOS Biol* 2017; **15**:
417 e2001793.
418 6. Kroll S, Agler MT, Kemen E. Genomic dissection of host–microbe and microbe–microbe
419 interactions for advanced plant breeding. *Curr Opin Plant Biol* 2017; **36**: 71–78.
420 7. Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J. Beans (Phaseolus
421 spp.) – model food legumes. *Plant Soil* 2003; **252**: 55–128.
422 8. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference
423 genome for common bean and genome-wide analysis of dual domestications. *Nat Genet*
424 2014; **46**: 707–713.
425 9. Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A, Capella-Gutierrez
426 S, Ordaz-Ortiz JJ, et al. Genomic history of the origin and domestication of common bean
427 unveils its closest sister species. *Genome Biol* 2017; **18**: 60.
428 10. Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, et al. Beans (Phaseolus

429 ssp.) as a Model for Understanding Crop Evolution. *Front Plant Sci* 2017; **8**: 722.

430 11. Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, et al. Mesoamerican origin
431 of the common bean (Phaseolus vulgaris L.) is revealed by sequence data. *Proc Natl Acad*
432 *Sci* 2012; **109**: 788–796.

433 12. Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools of common
434 bean (Phaseolus vulgaris L., Fabaceae). *Theor Appl Genet* 2009; **118**: 979–992.

435 13. Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al. Molecular
436 analysis of the parallel domestication of the common bean (Phaseolus vulgaris) in
437 Mesoamerica and the Andes. *New Phytol* 2013; **197**: 300–313.

438 14. Perez-Jaramillo JE, Carrion VJ, Bosse M, Ferrao LF V, de Hollander M, Garcia AAF, et al.
439 Linking rhizosphere microbiome composition of wild and domesticated Phaseolus vulgaris
440 to genotypic and root phenotypic traits. *ISME J* 2017.

441 15. Pérez-Jaramillo JE, de Hollander M, Ramírez CA, Mendes R, Raaijmakers JM, Carrión VJ.
442 Deciphering rhizosphere microbiome assembly of wild and modern common bean
443 (Phaseolus vulgaris) in native and agricultural soils from Colombia. *Microbiome* 2019; **7**:
444 114.

445 16. NDSU Agriculture. Eclipse Black Bean. 2004. Fargo.

446 17. McClean P, Myers J. Pedigrees of dry bean cultivars, lines and PIs. *Annu Rep Bean Improv*
447 *Coop* 1990; **33**: 23–25.

448 18. USDA. United States Department of Agriculture National Agricultural Statistics Service
449 Crop Production 2016 Summary. 2017.

450 19. Cichy KA, Porch TG, Beaver JS, Cregan P, Fourie D, Glahn RP, et al. A Diversity Panel for
451 Andean Bean Improvement. *Crop Sci* 2015; **55**: 2149.

452 20. Sasse J, Martinoia E, Northen T. Feed Your Friends: Do Plant Exudates Shape the Root

453 Microbiome? *Trends Plant Sci* 2017.

454 21. Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, et al.
455 Subsampled open-reference clustering creates consistent, comprehensive OTU definitions
456 and scales to billions of sequences. *PeerJ* 2014; **2**: e545.

457 22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal
458 RNA gene database project: improved data processing and web-based tools. *Nucleic Acids*
459 *Res* 2013; **41**: D590-6.

460 23. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat*
461 *Methods* 2013; **10**: 996–998.

462 24. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al.
463 QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;
464 **7**: 335–336.

465 25. Gdanetz K, Benucci GMN, Vande Pol N, Bonito G. CONSTAX: a tool for improved
466 taxonomic resolution of environmental fungal ITS sequences. *BMC Bioinformatics* 2017;
467 **18**: 538.

468 26. UNITE Community. UNITE USEARCH/UTAX release. 2017.

469 27. Davis NM, Proctor DiM, Holmes SP, Relman DA, Callahan BJ. Simple statistical
470 identification and removal of contaminant sequences in marker-gene and metagenomics
471 data. *Microbiome* 2018; **6**: 226.

472 28. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
473 sequencing. *bioRxiv* 2016; 081257.

474 29. Shade A, Stopnisek N. Abundance-occupancy distributions to prioritize plant core
475 microbiome membership. *Curr Opin Microbiol* 2019; **49**: 50–58.

476 30. Deng Y, Jiang Y-H, Yang Y, He Z, Luo F, Zhou J. Molecular ecological network analyses.

477 *BMC Bioinformatics* 2012; **13**: 113.

478 31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A
479 Software Environment for Integrated Models of Biomolecular Interaction Networks.
480 *Genome Res* 2003; **13**: 2498–2504.

481 32. Singer E, Bonnette J, Kenaley SC, Woyke T, Juenger TE. Plant compartment and genetic
482 variation drive microbiome composition in switchgrass roots. *Environ Microbiol Rep* 2019;
483 **11**: 185–195.

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

487 34. De Souza RSC, Okura VK, Armanhi JSL, Jorrín B, Lozano N, Da Silva MJ, et al. Unlocking
488 the bacterial and fungal communities assemblages of sugarcane microbiome. *Sci Rep* 2016;
489 **6**: 1–15.

490 35. Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, et al. Large-scale
491 replicated field study of maize rhizosphere identifies heritable microbes. *Proc Natl Acad Sci*
492 *U S A* 2018; **115**: 7368–7373.

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

496 37. Xu J, Zhang Y, Zhang P, Trivedi P, Riera N, Wang Y, et al. The structure and function of
497 the global citrus rhizosphere microbiome. *Nat Commun* 2018; **9**: 4894.

498 38. Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, et al.
499 Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl*
500 *Acad Sci U S A* 2015; **112**: E911-20.

501 39. Edwards JA, Santos-Medellín CM, Liechty ZS, Nguyen B, Lurie E, Eason S, et al.
502 Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life
503 cycle in field-grown rice. *PLOS Biol* 2018; **16**: e2003862.
504 40. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. Assembly
505 and ecological function of the root microbiome across angiosperm plant species. *Proc Natl*
506 *Acad Sci U S A* 2018; **115**: E1157–E1165.
507 41. Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, et al. Field study reveals core
508 plant microbiota and relative importance of their drivers. *Environ Microbiol* 2018; **20**: 124–
509 140.
510 42. Gomes EA, De Sousa SM, Oliveira CA. Root-Associated Microbiome of Maize Genotypes
511 with Contrasting Phosphorus Use Efficiency. 2018; **2**: 129.
512 43. Estendorfer J, Stempfhuber B, Haury P, Vestergaard G, Rillig MC, Joshi J, et al. The
513 Influence of Land Use Intensity on the Plant-Associated Microbiome of Dactylis glomerata
514 L. *Front Plant Sci* 2017; **8**: 930.
515 44. Leff JW, Lynch RC, Kane NC, Fierer N. Plant domestication and the assembly of bacterial
516 and fungal communities associated with strains of the common sunflower, Helianthus
517 annuus. *New Phytol* 2017; **214**: 412–423.
518 45. Lundberg DS, Lebeis SL, Herrera Paredes S, Yourstone S, Gehring J, Malfatti S, et al.
519 Defining the core Arabidopsis thaliana root microbiome. *Nature* 2012; **488**.
520 46. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, et al.
521 Genomic features of bacterial adaptation to plants. *Nat Genet* 2018; **50**: 138–150.
522 47. Grady KL, Sorensen JW, Stopnisek N, Guittar J, Shade A. Assembly and seasonality of core
523 phyllosphere microbiota on perennial biofuel crops. *Nat Commun* 2019; **10**: 4135.
524 48. Shade A, Dunn RR, Blowes SA, Keil P, Bohannan BJM, Herrmann M, et al. Macroecology

525 to Unite All Life, Large and Small. *Trends Ecol Evol* 2018; **33**: 731–744.

526 49. Woodcock S, Van Der Gast CJ, Bell T, Lunn M, Curtis TP, Head IM, et al. Neutral
527 assembly of bacterial communities. *FEMS Microbiol Ecol* 2007; **62**: 171–180.

528 50. Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, et al. Contribution
529 of neutral processes to the assembly of gut microbial communities in the zebrafish over host
530 development. *ISME J* 2016; **10**: 655–664.

531 51. Gaston KJ, Blackburn TM, Greenwood JJD, Gregory RD, Quinn RM, Lawton JH.
532 Abundance-occupancy relationships. *J Appl Ecol* 2000; **37**: 39–59.

533 52. Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. Drivers of bacterial beta-
534 diversity depend on spatial scale. *Proc Natl Acad Sci U S A* 2011; **108**: 7850–4.

535 53. Gumiere T, Durrer A, Bohannan BJM, Andreote FD. Biogeographical patterns in fungal
536 communities from soils cultivated with sugarcane. *J Biogeogr* 2016; **43**: 2016–2026.

537 54. Talbot JM, Bruns TD, Taylor JW, Smith DP, Branco S, Glassman SI, et al. Endemism and
538 functional convergence across the North American soil mycobiome. *Proc Natl Acad Sci*
539 2014; **111**: 6341–6346.

540 55. Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, et al. Global
541 diversity and geography of soil fungi. *Science (80-)* 2014; **346**.

542 56. Dawson W, Hör J, Egert M, van Kleunen M, Pester M. A Small Number of Low-abundance
543 Bacteria Dominate Plant Species-specific Responses during Rhizosphere Colonization.
544 *Front Microbiol* 2017; **8**: 975.

545 57. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim S-T, Weigel D, et al. Microbial Hub Taxa Link
546 Host and Abiotic Factors to Plant Microbiome Variation. *PLOS Biol* 2016; **14**: e1002352.

547 58. Trivedi P, Delgado-Baquerizo M, Trivedi C, Hamonts K, Anderson IC, Singh BK. Keystone
548 microbial taxa regulate the invasion of a fungal pathogen in agro-ecosystems. *Soil Biol*

549 *Biochem* 2017; **111**: 10–14.

550 59. Shi S, Nuccio EE, Shi ZJ, He Z, Zhou J, Firestone MK. The interconnected rhizosphere:
551 High network complexity dominates rhizosphere assemblages. *Ecol Lett* 2016; **19**: 926–936.

552 60. Carlström CI, Field CM, Bortfeld-Miller M, Müller B, Sunagawa S, Vorholt JA. Synthetic
553 microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nat*
554 *Ecol Evol* 2019; **3**: 1445–1454.

555 61. Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, et al. Core microbiomes
556 for sustainable agroecosystems. *Nat Plants* 2018; **4**: 247–257.

557 62. Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, et al. Microbial
558 Interkingdom Interactions in Roots Promote Arabidopsis Survival. *Cell* 2018; **175**: 973-
559 983.e14.

560 63. Wagg C, Schlaeppi K, Banerjee S, Kuramae EE, van der Heijden MGA. Fungal-bacterial
561 diversity and microbiome complexity predict ecosystem functioning. *Nat Commun* 2019; **10**:
562 1–10.

563 64. Guimerà R, Nunes Amaral LA. Functional cartography of complex metabolic networks.
564 *Nature* 2005; **433**: 895–900.

565 65. Lemanceau P, Blouin M, Muller D, Moënne-Loccoz Y. Let the Core Microbiota Be
566 Functional. *Trends Plant Sci* . 2017. Elsevier Ltd. , **22**: 583–595

567 66. Risely A. Applying the core microbiome to understand host-microbe systems. *J Anim Ecol*
568 2020.

569 67. Andrews M, Andrews ME, Andrews M, Andrews ME. Specificity in Legume-Rhizobia
570 Symbioses. *Int J Mol Sci* 2017; **18**: 705.

571 68. Wang Q, Liu J, Zhu H. Genetic and molecular mechanisms underlying symbiotic specificity
572 in legume-rhizobium interactions. *Front Plant Sci* . 2018. Frontiers Media S.A. , **9**

573 69. Skouri-Panet F, Benzerara K, Cosmidis J, Férard C, Caumes G, De Luca G, et al. In vitro
574 and in silico evidence of phosphatase diversity in the biomineralizing bacterium Ramlibacter
575 tataouinensis. *Front Microbiol* 2018; **8**: 2592.
576 70. Lee JS, Lee KC, Kim KK, Lee B. Complete genome sequence of the Variibacter
577 gotjawalensis GJW-30T from soil of lava forest, Gotjawal. *J Biotechnol* 2016; **218**: 64–65.
578 71. Khan AL, Waqas M, Kang SM, Al-Harrasi A, Hussain J, Al-Rawahi A, et al. Bacterial
579 endophyte Sphingomonas sp. LK11 produces gibberellins and IAA and promotes tomato
580 plant growth. *J Microbiol* 2014; **52**: 689–695.
581 72. Aylward FO, McDonald BR, Adams SM, Valenzuela A, Schmidt RA, Goodwin LA, et al.
582 Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and
583 biodegradative capabilities. *Appl Environ Microbiol* 2013; **79**: 3724–33.
584 73. Xu L, Naylor D, Dong Z, Simmons T, Pierroz G, Hixson KK, et al. Drought delays
585 development of the sorghum root microbiome and enriches for monoderm bacteria. *Proc*
586 *Natl Acad Sci* 2018; **115**: E4284–E4293.
587
588

589 **Figure Legends**

590 **Fig. 1**: Abundance-occupancy distributions were used to identify core members of the rhizosphere
591 microbiome for bacteria/archaea (A) and fungi (B). Taxa exclusive to a genotype are indicated in
592 orange (CELRK) or black (Eclipse), and taxa shared across both genotypes are white. The solid
593 line represents the fit of the neutral model, and the dashed line is 95% confidence around the model
594 prediction. Taxa with an occupancy of 1 (i.e., detected in all samples) were considered members of
595 the core. Relative abundance of these taxa is represented as boxplots, grouped by order and number
596 of taxa therein (C, D). Panels C and D are color-coded by phyla level.

597

598 **Fig. 2**: A global core rhizosphere microbiome. There were 3361 shared bacterial/archaeal taxa
599 across U.S. and Colombia rhizosphere samples, suggesting highly similar recruitment across
600 continental scales. Forty-eight taxa were detected in all samples of both datasets, as depicted by the
601 Venn diagram (A), and included many Proteobacteria (B). Relative abundance of the 48 U.S.-
602 Colombia core taxa is represented as boxplots (left panel), grouped by order and dataset
603 (Colombia/U.S.). Number of taxa per order is represented as bars (right panel). Labels on the y-axis
604 and bars are color-coded by phylum level.

605

606 **Fig. 3:** Relative abundance of core taxa in the root system of the common bean during plant
607 development by root compartment and growing location. The combined relative abundance of 48
608 core taxa is significantly higher in the rhizoplane (green) compared to the rhizosphere (orange) and
609 does not show plant development dependence (A – upper panel). The combined relative abundance
610 of remaining core taxa specific to the U.S. (n=210), tend to be high throughout the plant
611 development but are equally abundant in both compartments (A - lower panel). Stars above box
612 plots represent statistical significances as determined by Wilcoxon test (**** ≤ 0.0001, *** ≤

613 0.001, ** ≤ 0.01, * ≤ 0.05). Z-score normalized relative abundance of the 48 core taxa across plant
614 development, compartments and growing location (B). Taxa on the y-axis are arranged by their
615 classification at the phylum level.
616
617 **Fig. 4**: Network co-occurrence analysis shows that core rhizosphere microbiome members are
618 predominantly classified as peripheral taxa that are weakly connected, and generally clustering by
619 the growing location. The network depicts only clusters of modules that were connected by more
620 than 6 nodes (A). Nodes shape is representing domain associations (archaeal, bacterial or fungal)
621 and node size is proportional to its total number of connecting edges (A). The four largest modules
622 are generally reflective of community biogeography and distinguished by color (A). The within-
623 (Zi) and among- (Pi) module connectivity plot was used to identify module (Pi<0.62, Zi>0.2.5) or
624 network hub taxa (Pi>0.62, Zi>2.5), as well as connector (Pi>0.62, Zi<2.5) and peripheral taxa (B).
625 The density plots surrounding the Zi-Pi plot represent core (green) and non-core taxa (black) (B).
626 The relative abundances of taxa within each module is represented in box plots (C).
627
628 **Supporting Figure legends**
629 **Fig. S1**: Sequencing depth and rarefaction curves for 16S rRNA (A,B) and ITS (C, D) dataset from
630 samples collected in 2017. The red splitted line represents the rarefaction threshold. Note that we
631 submitted amplicons for sequencing from the 2017 sampling effort twice which resulted in the
632 sample to sample variation in read depth (rhizosphere 16S rRNA samples sequenced first followed
633 by the ITS amplicons with addition of the root-associated samples).
634

635 **Fig. S2**: Alpha diversity indices for the 16S RNA (A, B) and ITS (C, D) datasets. Represented are
636 rich-ness, Shannon and Pielou indices measured by growing location (A, C) and bean genotype (B,
637 D). For statistical comparison of the pairs we used ANOVA (A, C) or Wilcoxon test (B, D).
638
639 **Fig. S3**: Growing location drives bacterial/archaeal (A) and fungal (B) microbiome structure of the
640 common bean rhizosphere. The principal coordinate analysis (PCoA) is based on Bray-Curtis
641 distances. Growing location is indicated by color and plant genotype is indicated by shape shapes
642 (diamond=CELRK, circle=Eclipse, square=root zone soil). The strength of statistically significant
643 (p-value < 0.01) explanatory variables are shown as the length of fitted vectors.
644
645 **Fig S4**: Community composition (A, C) and number of shared taxa between sites represent as Venn
646 diagrams (B, D). The 16S rRNA dataset is represented in the top panels (A, B) and ITS in the lower
647 panels (C, D). The bar charts are colored based on the phylum association (phyla represented by
648 relative abundance < .05 are grouped and labelled as other). For Venn diagrams, samples were
649 grouped by the growing location and root zone samples were removed.
650
651 **Fig. S5**: Analysis of ZOTUs represented by each identified core OTU. 48 core OTUs were
652 represented by as few as 4 ZOTUs and by up to 35 ZOTUs. For every OTU we found at least one
653 ZOTU with occupancy = 1 and all of them, except of 2 ZOTUs, had also the highest relative
654 abundance among them. Points are color coded by their presence, red representing those with
655 occupancy < 1 and blue for ZOTUs with occupancy of 1. The OTUs on the x-axis are ordered
656 alphabetically.
657

658 **Fig. S6**: Occupancy of core OTUs in the development study. Occupancy is represented by color
659 and size.
660
661 **Fig. S7**: Comparison of relative abundance of the 48 core OTUs between the DNA isolation
662 methods used in the development study (G=Griffith, P=PowerSoil). Statistical difference,
663 determined by Wilcoxon test, is represented as star symbol (*<0.05, **<0.01, ***<0.001). The
664 pints are color coded by sample they derived from.
665
666 **Fig. S8**: The effect of isolation method on alpha and beta diversity. For alpha diversity richness,
667 Shannon and Pielou indices are presented (A). For the principal coordinates analysis, Bray-Curtis
668 distance matrix was used. Symbols are colored by samples. Wilcoxon test was used to determine
669 statistical differences between isolation methods for the alpha diversity metrices (*<0.05).
670 PERMANOVA was used to determine the effect of isolation method on community structure.
671
672

673 **Tables**

674 **Table 1**: Description of geographical, managemental and soil properties of the common bean

675 growing locations included in the study.

Sample same	Growing location	Elevation	Bean genotype	Weather (T, precipitation)*	Rotation history	Fertilization (per acre)	Irrigation	pH	Nitrogen (%)	Organic matter (%)
SVREC	Saginaw Valley Research and Extension Center, MI, US	190 m	CELRK	24.5ºC, 5.7 mm	Common bean, wheat, maize	Synthetic (400lbs of 15N-5 P-13K)	None	7.7 (± 0.1)	0.13 (± 0.00)	2.33 (± 0.06)
MRC	Montcalm Research Center, MI, US	280 m	Eclipse	24ºC, 7.4 mm	Common bean, maize, potato	Synthetic (200lbs of 19N-0P-19K)	Yes	5.9 (± 0.4)	0.10 (± 0.01)	2.13 (± 0.06)
NE	Scottsbluff, NE, US	1200 m	CELRK and Eclipse	27ºC, 2.18 mm	Common bean, maize, common bean	None	Yes	7.9 (± 0.1)	0.07 (± 0.01)	1.39 (± 0.18)
CO	ARDEC, Fort Collins, CO, US	1536 m	CELRK and Eclipse	31ºC, 1.4 mm **	Common bean, maize, barley	Organic (220lbs urea (2016), 3 tons manure (2015))	Yes	8.1 (± 0.2)	0.11 (± 0.00)	1.7 (± 0.15)
WA	WSU, Othello, WA, US	320 m	CELRK and Eclipse	26.3ºC, 0.2 mm	Common bean, maize, wheat	Synthetic (40lbs N/20lbs P/10lbs ZN2O5)	Yes	6.1 (± 0.4)	0.08 (± 0.01)	1.78 (± 0.10)

676 * Temperature (as max air temperature) and precipitation are averaged for the period between May

677 1st and July 31st 2017.

678 ** Data available only for the period between Jun 28th and Jul 31st 2017.

679

680 **Supporting Table legends**

681 **Table S1:** PERMANOVA results for the 16S rRNA and ITS data. Highly correlated or/and

682 statistically significant values are highlighted in bold.

683

684 **Table S2**: 20 differentially abundant OTUs between the two plant genotypes as identified by using

685 DESeq2 (Love et al. 2014).

686

687 **Table S3**: Sloan neutral model summary.

688

689 **Table S4**: List of 48 core taxa and their taxonomic classification.

690

691 **Table S5**: Occupancy of core OTUs in agricultural, natural (forest) soils or when combined.
692 Results are based on the re-analysis of the data from the Pérez-Jaramillo et al. 2019. OTUs with
693 occupancy of 1 in both soils are highlighted in orange. Additional to the number we used green
694 gradient shading to represents the occupancy of each OTU.

695

696 **Table S6**: Summary of network properties of actual and random network.

697

698

Figures

Fig. 1: Abundance-occupancy distributions were used to identify core members of the rhizosphere microbiome for bacteria/archaea (A) and fungi (B). Taxa exclusive to a genotype are indicated in orange (CELRK) or black (Eclipse), and taxa shared across both genotypes are white. The solid line represents the fit of the neutral model, and the dashed line is 95% confidence around the model prediction. Taxa with an occupancy of 1 (i.e., detected in all samples) were considered members of the core. Relative abundance of these taxa is represented as boxplots, grouped by order and number of taxa therein (C, D). Panels C and D are color-coded by phyla level.

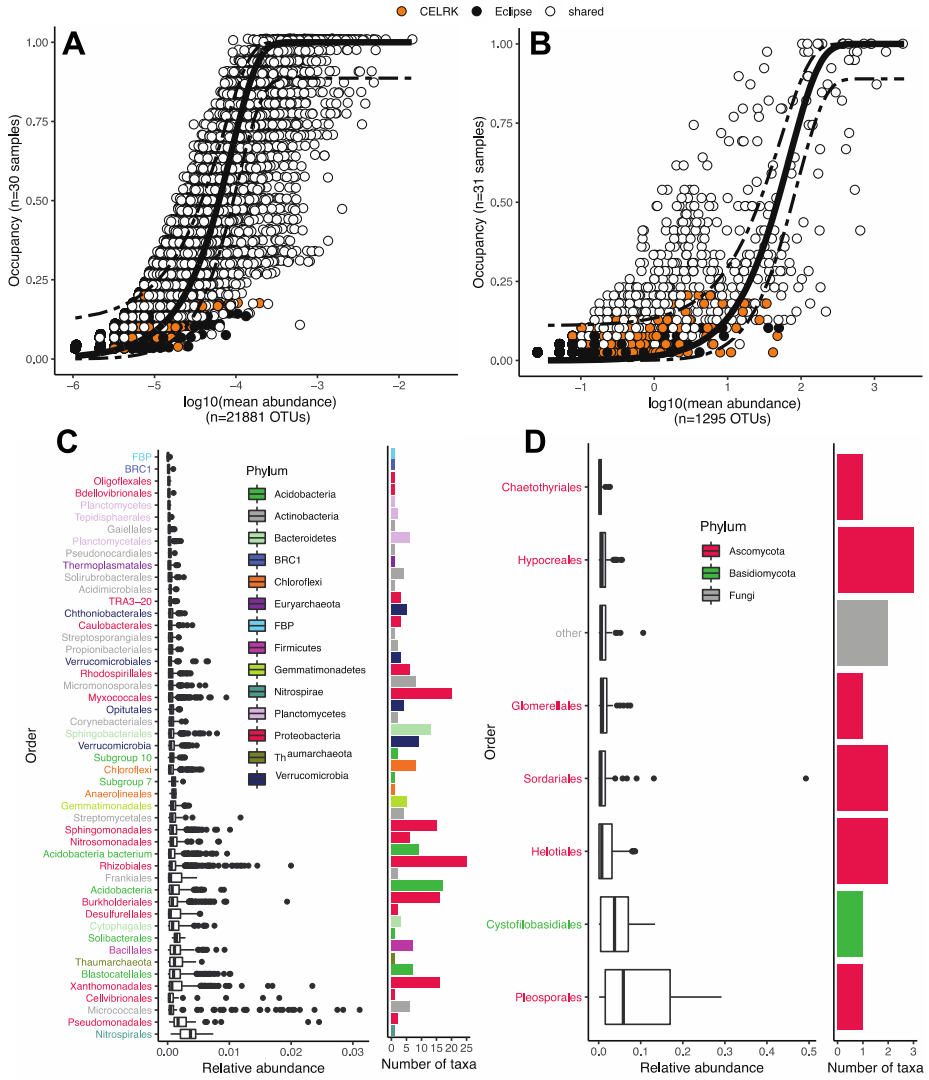

Fig. 2: A global core rhizosphere microbiome. There were 3361 shared bacterial/archaeal taxa across U.S. and Colombia rhizosphere samples, suggesting highly similar recruitment across continental scales. Forty-eight taxa were detected in all samples of both datasets, as depicted by the Venn diagram (A), and included many Proteobacteria (B). Relative abundance of the 48 U.S.-Colombia core taxa is represented as boxplots (left panel), grouped by order and dataset (Colombia/U.S.). Number of taxa per order is represented as bars (right panel). Labels on the y-axis and bars are color-coded by phylum level.

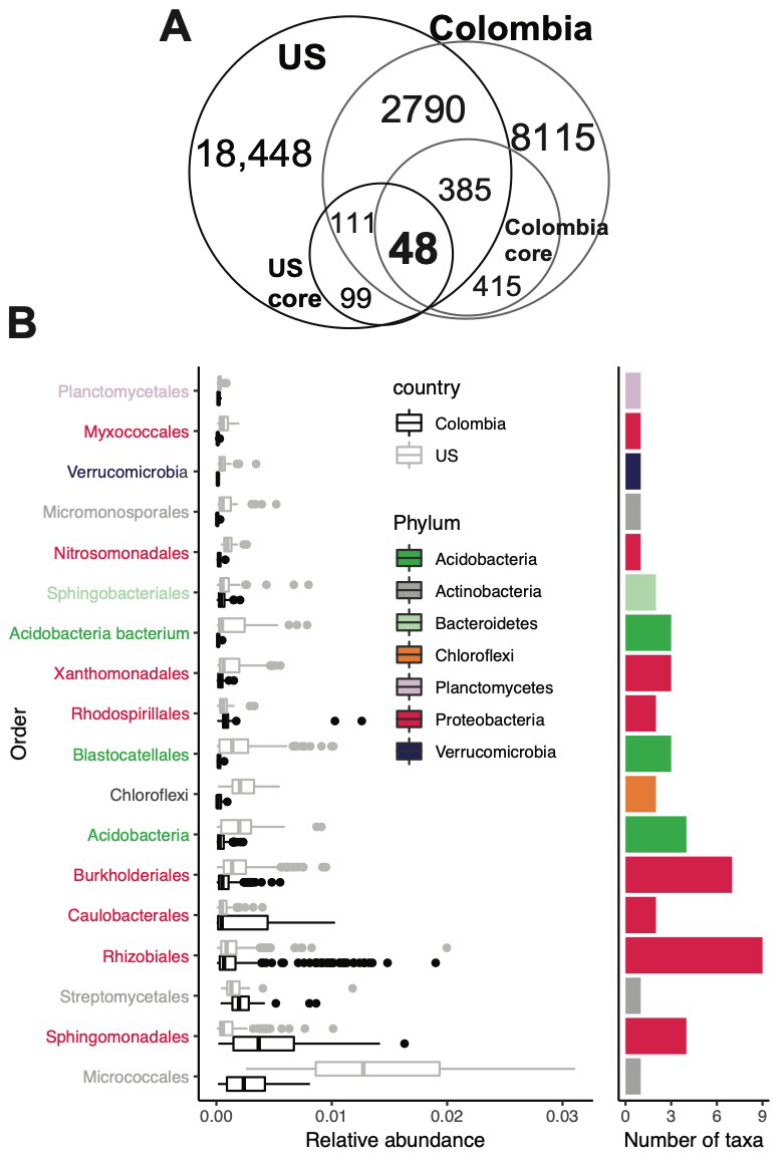

Fig. 3: Relative abundance of core taxa in the root system of the common bean during plant development by root compartment and growing location. The combined relative abundance of 48 core taxa is significantly higher in the rhizoplane (green) compared to the rhizosphere (orange) and does not show plant development dependence (A – upper panel). The combined relative abundance of remaining core taxa specific to the U.S. (n=210), tend to be high throughout the plant development but are equally abundant in both compartments (A - lower panel). Stars above box plots represent statistical significances as determined by Wilcoxon test (**** ≤ 0.0001, *** ≤ 0.001, ** ≤ 0.01, * ≤ 0.05). Z-score normalized relative abundance of the 48 core taxa across plant development, compartments and growing location (B). Taxa on the y-axis are arranged by their classification at the phylum level.

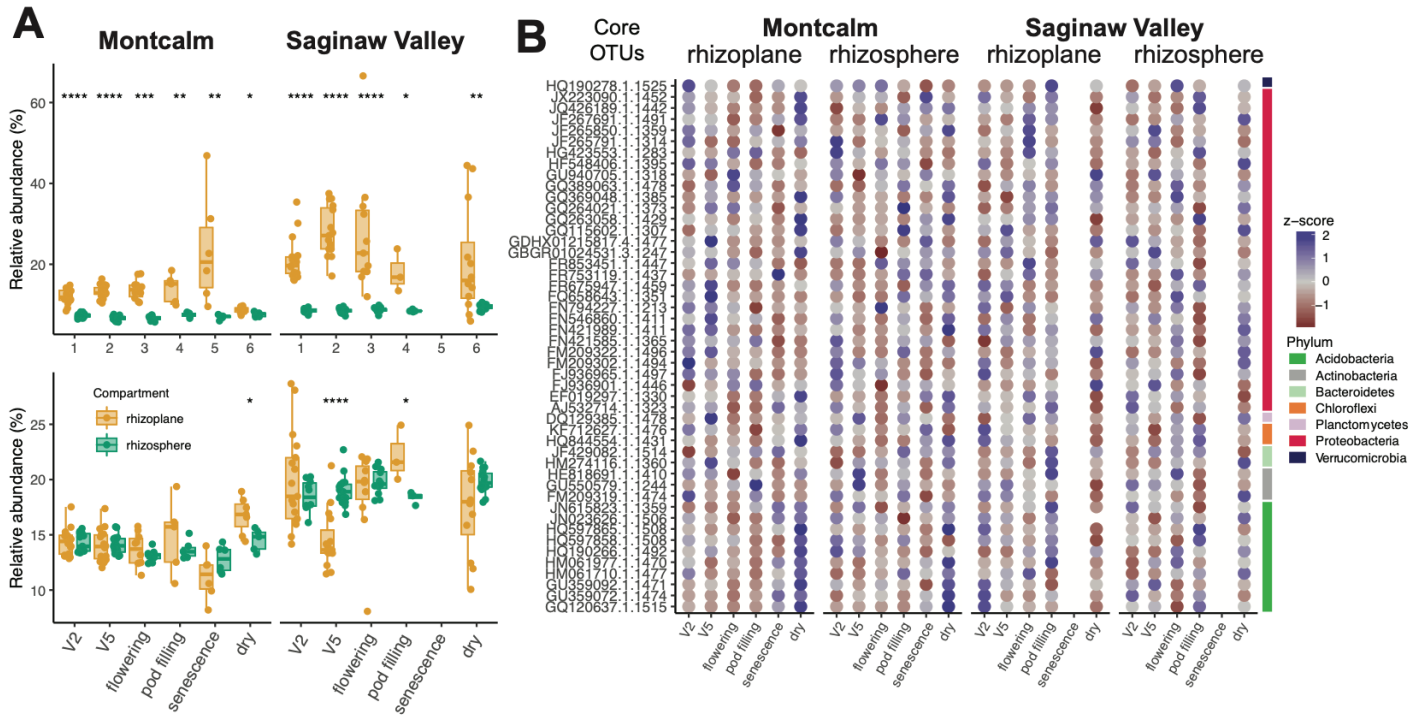
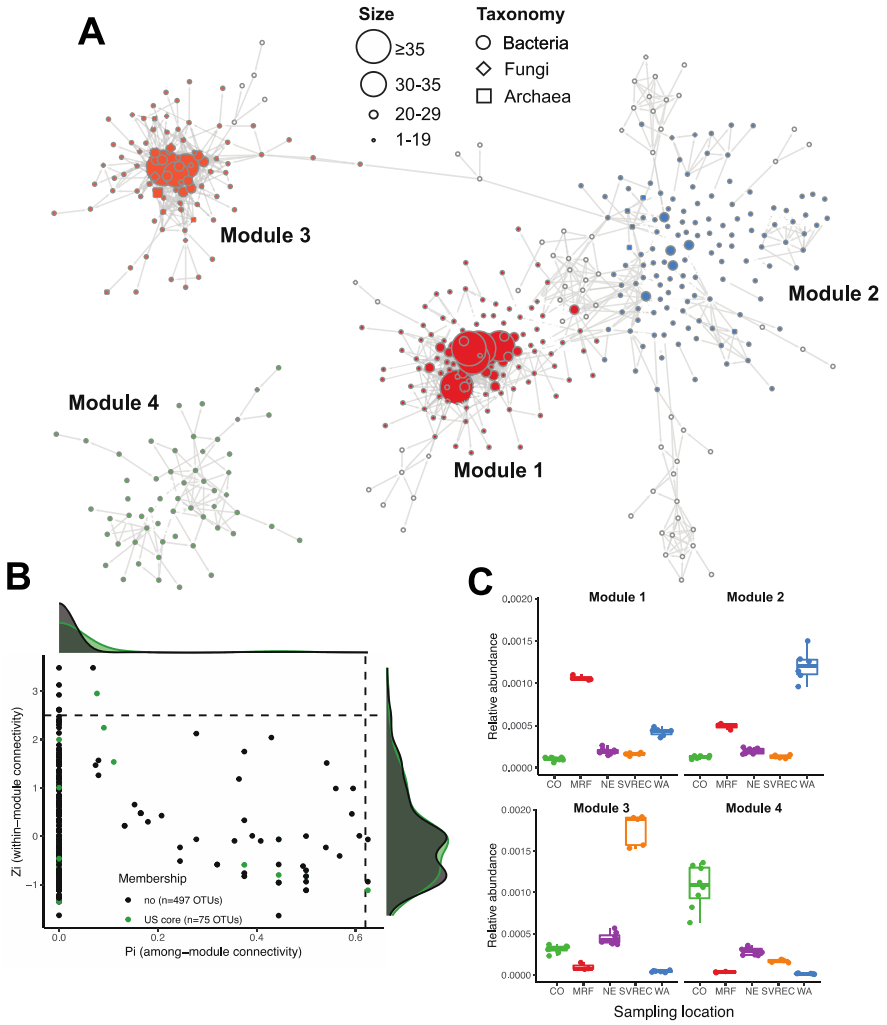
Fig. 4: Network co-occurrence analysis shows that core rhizosphere microbiome members are predominantly classified as peripheral taxa that are weakly connected, and generally clustering by the growing location. The network depicts only clusters of modules that were connected by more than 6 nodes (A). Nodes shape is representing domain associations (archaeal, bacterial or fungal) and node size is proportional to its total number of connecting edges (A). The four largest modules are generally reflective of community biogeography and distinguished by color (A). The within- (Zi) and among- (Pi) module connectivity plot was used to identify module (Pi<0.62, Zi>0.2.5) or network hub taxa (Pi>0.62, Zi>2.5), as well as connector (Pi>0.62, Zi<2.5) and peripheral taxa (B). The density plots surrounding the Zi-Pi plot represent core (green) and non-core taxa (black) (B). The relative abundances of taxa within each module is represented in box plots (C).