AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

PLANT MICROBIOLOGY

The Influence of the Host Plant Is the Major Ecological Determinant of the Presence of Nitrogen-Fixing Root Nodule Symbiont Cluster II *Frankia* Species in Soil

Kai Battenberg,[a] Jannah A. Wren,[a] Janell Hillman,[b] Joseph Edwards,[c] Liujing Huang,[d] Alison M. Berry[a]

Department of Plant Sciences, University of California, Davis, California, USA[a]; Santa Clara Valley Water District, San Jose, California, USA[b]; Department of Plant Biology, University of California, Davis, California, USA[c]; College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, China[d]

ABSTRACT The actinobacterial genus *Frankia* establishes nitrogen-fixing root nodule symbioses with specific hosts within the nitrogen-fixing plant clade. Of four genetically distinct subgroups of *Frankia*, cluster I, II, and III strains are capable of forming effective nitrogen-fixing symbiotic associations, while cluster IV strains generally do not. Cluster II *Frankia* strains have rarely been detected in soil devoid of host plants, unlike cluster I or III strains, suggesting a stronger association with their host. To investigate the degree of host influence, we characterized the cluster II *Frankia* strain distribution in rhizosphere soil in three locations in northern California. The presence/absence of cluster II *Frankia* strains at a given site correlated significantly with the presence/absence of host plants on the site, as determined by glutamine synthetase (*glnA*) gene sequence analysis, and by microbiome analysis (16S rRNA gene) of a subset of host/nonhost rhizosphere soils. However, the distribution of cluster II *Frankia* strains was not significantly affected by other potential determinants such as host-plant species, geographical location, climate, soil pH, or soil type. Rhizosphere soil microbiome analysis showed that cluster II *Frankia* strains occupied only a minute fraction of the microbiome even in the host-plant-present site and further revealed no statistically significant difference in the α-diversity or in the microbiome composition between the host-plant-present or -absent sites. Taken together, these data suggest that host plants provide a factor that is specific for cluster II *Frankia* strains, not a general growth-promoting factor. Further, the factor accumulates or is transported at the site level, i.e., beyond the host rhizosphere.

IMPORTANCE Biological nitrogen fixation is a bacterial process that accounts for a major fraction of net new nitrogen input in terrestrial ecosystems. Transfer of fixed nitrogen to plant biomass is especially efficient via root nodule symbioses, which represent evolutionarily and ecologically specialized mutualistic associations. *Frankia* spp. (*Actinobacteria*), especially cluster II *Frankia* spp., have an extremely broad host range, yet comparatively little is known about the soil ecology of these organisms in relation to the host plants and their rhizosphere microbiomes. This study reveals a strong influence of the host plant on soil distribution of cluster II *Frankia* spp.

KEYWORDS *Frankia*, metagenomics, nitrogen fixation, plant-microbe interactions, rhizosphere-inhabiting microbes, root nodule, soil microbiology, symbiosis

Received 20 September 2016 **Accepted** 7 October 2016

Accepted manuscript posted online 21 October 2016

Citation Battenberg K, Wren JA, Hillman J, Edwards J, Huang L, Berry AM. 2017. The influence of the host plant is the major ecological determinant of the presence of nitrogen-fixing root nodule symbiont cluster II *Frankia* species in soil. Appl Environ Microbiol 83:e02661-16. https://doi.org/10.1128/AEM.02661-16.

Editor Claire Vieille, Michigan State University

Copyright © 2016 American Society for Microbiology. All Rights Reserved.

Address correspondence to Alison M. Berry, amberry@ucdavis.edu.

Biological nitrogen fixation as a bacterial process in symbiosis with plants accounts for a major fraction of nitrogen input in terrestrial ecosystems (1). Plants engage in symbiotic associations with microorganisms that vary in their degree of association. The spectrum ranges from obligate endosymbioses, such as arbuscular mycorrhizal (AM) fungi (2), to generally facultative nitrogen-fixing symbioses; the latter, including the rhizobia-legume nodules (3), *Frankia*-actinorhizal nodules (4), and cyanobacterium-*Gunnera* associations (5).

Nitrogen-fixing root nodule symbiosis (RNS) began with one evolutionary event that occurred in the common ancestor of a clade of angiosperms known as the nitrogen-fixing clade (NFC) (6) approximately 100 million years ago (7). The *Frankia*-actinorhizal symbioses are of particular interest in the interpretation of the evolutionary underpinnings of RNS because these symbioses developed, probably independently, within several distinct lineages across the NFC (4). Genetically distinct subgroups within the genus *Frankia* (cluster I, II, and III) form compatible associations with different lineages of plants within the NFC (6), while cluster IV includes atypical nonsymbiotic and/or non-nitrogen-fixing strains (8). Of these subgroups, cluster II *Frankia* strains have the broadest host range among nitrogen-fixing plants: species in four different families in two orders (Coriariaceae, Datiscaceae, Rhamnaceae, and Rosaceae) form RNS with cluster II *Frankia* strains. Representatives of these families are particularly abundant in western North America, where the current study has been carried out.

Comparisons of representative sequenced genomes from these subgroups suggested a general correlation of *Frankia* genomic features with the corresponding host biogeographic distribution (9). That being said, since RNS is first established between the host-plant root and the bacterial symbiont by the two partners coming into physical contact with each other in the rhizosphere, it is to be expected that all of the symbiotic *Frankia* subgroups must have at least some capability to persist in a free-living state in soil, in addition to living in symbiosis. Clusters I, III, and IV are known to have a free-living phase in soil: infective units of clusters I and III are ubiquitously distributed in soils (10, 11). Moreover, infective units of cluster I *Frankia* can be as abundant, or more abundant, in nearby nonhost rhizosphere soil than in hosts (12). Similarly, strains of cluster III *Frankia* are well known to occur in soil independent of host-plant presence (11). Cluster II *Frankia* strains, however, have been characterized as highly host dependent, with only one report of nodule formation from soil devoid of a host, 18 months after inoculation (13). Previous DNA-based surveys of cluster II *Frankia* strain distribution have been conducted using field-collected nodules (14–16). Thus, there is only limited knowledge about the soil ecology of cluster II *Frankia* strains in relation to their host plants, particularly concerning the degree of association with the host presence. Since the cluster II *Frankia* strains have been shown to be phylogenetically early divergent within the genus (17), a better understanding of host association in this subgroup may shed light on the process of evolutionary acquisition of RNS.

The paucity of evidence for cluster II *Frankia* strains living independently of host-plant influences compared with the other subgroups led us to hypothesize that cluster II *Frankia* strains may be unusually dependent on the unique environment provided by the host plant for their long-term persistence (14). To test this, we conducted a comparative survey to determine the soil distribution pattern of free-living cluster II *Frankia* strains in three geographical locations in northern California; in different soil types, including serpentine soils, a depauperate soil type that harbors several actinorhizal hosts; and under different climatic conditions. The distribution of cluster II *Frankia* strains was assessed through detection of DNA sequences unique to cluster II *Frankia* strains in host and nonhost rhizosphere soils, sampled in sites with species of host and nonhost plants. The microbiome of a subset of these soil samples was also analyzed via high-throughput sequencing.

Because of the seasonally arid climate and soil environment, which has been shown to limit nitrogen fixation activity in *Ceanothus* (18), nitrogen-stable isotope analysis was conducted to confirm an active state of nitrogen-fixing symbiosis in each host plant

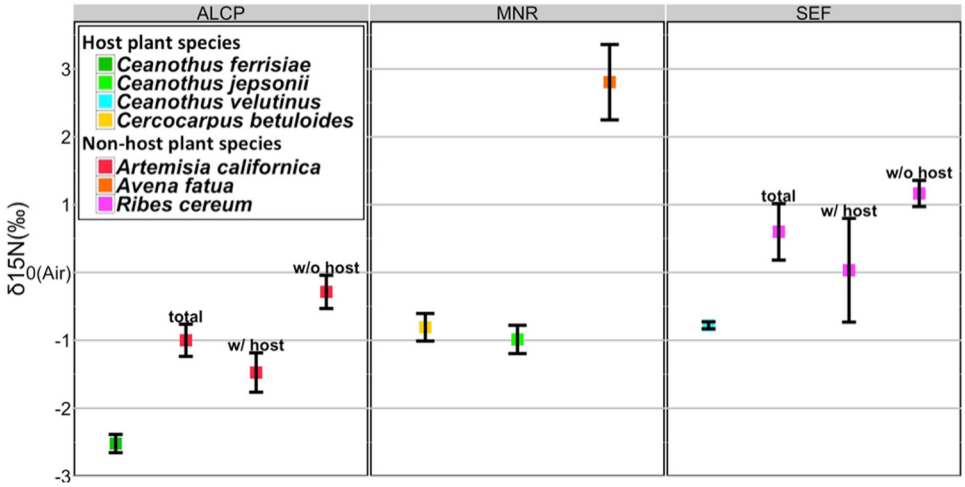


FIG 1 Mean δ^{15}N of sampled plants in 3 locations. *Ceanothus ferrisiae*, *Ceanothus jepsonii*, *Ceanothus velutinus*, and *Cercocarpus betuloides* are known hosts of cluster II *Frankia* strains, while *Artemisia californica*, *Avena fatua*, and *Ribes cereum* are known nonhosts. The error bars show standard errors. The three values for *A. californica* and *R. cereum* represent total samples, samples from sites with hosts, and samples from sites without hosts.

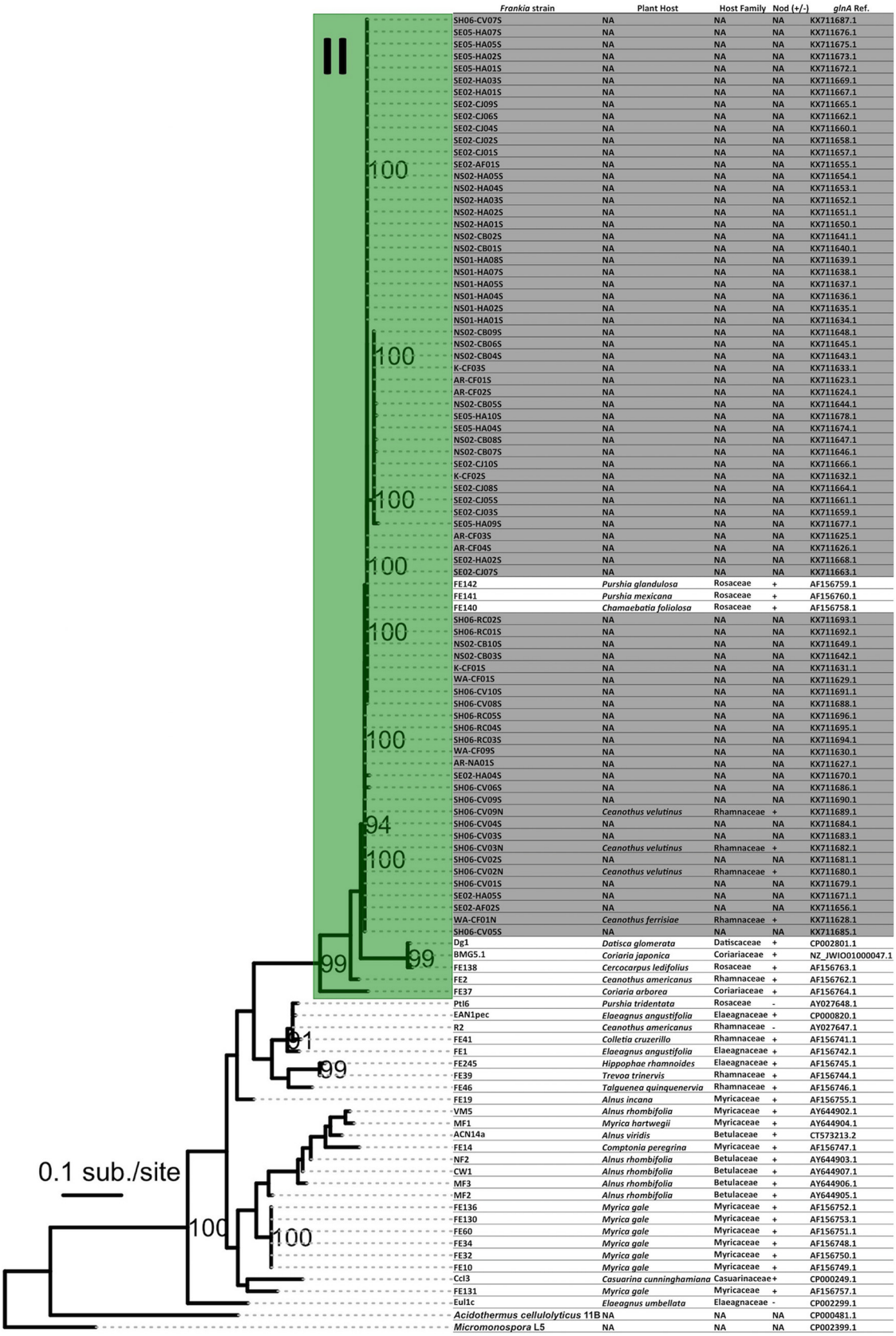
sampled. In addition, stable isotope analysis enabled us to assess the degree of site-level influence of nitrogen-fixing host plants (i.e., beyond the host rhizosphere) on cluster II *Frankia* strain distribution.

RESULTS

Climate and temperature of the three locations. According to the WorldClim database, the annual pattern of precipitation is similar across the three locations (i.e., dry summers and wet winters), but the annual precipitation was lower in Anderson Lake County Park, Santa Clara County, California (ALCP) (<530 mm) compared to the other two locations (>880 mm), and there was more summer (June to August) precipitation in Sagehen Experimental Forest, Truckee, California (SEF) (>50 mm) than in the other locations (<15 mm). The average monthly temperature is more than 5°C lower in SEF than in the two other locations throughout the year (see Fig. S4 in the supplemental material).

Leaf δ^{15}N analysis. Factorial analysis of variance (ANOVA) showed that the host plant species in all three locations had significantly more negative mean δ^{15}N, i.e., enrichment in ^{14}N, than the co-occurring nonhosts in pairwise comparisons (Fig. 1): *Ceanothus ferrisiae* had values significantly more negative than *Artemisia californica* ($P < 0.001$) in host-plant-present sites (AR, WA, and K). *Ceanothus jepsonii* and *Ceanothus betuloides* had values significantly more negative than those of *Avena fatua* ($P < 0.01$ and $P < 0.001$, respectively) in host-plant-present sites (SE02 and SE02, respectively). *Ceanothus velutinus* had values significantly more negative than *Ribes cereum* ($P < 0.05$) in host-plant-present sites (SH06). The significantly more negative δ^{15}N values, all below zero, indicated that the host plants were engaged in nitrogen fixation via RNS (19). Additionally, nonhosts in host-plant-present sites had a significantly more negative δ^{15}N than the same nonhosts in host-plant-absent sites: *A. californica* had a significantly more negative δ^{15}N in host-plant-present sites than this same species in the host-plant-absent site ($P < 0.01$), and *R. cereum* had a significantly more negative δ^{15}N in the host-plant-present site than in the host-plant-absent site ($P < 0.01$).

Presence of cluster II *Frankia* strains correlated with the presence of host-plant species. The most likely phylogeny based on partial *glnA* sequence showed that sequences obtained from soil and nodule samples were all nested within a well-supported clade (bootstrap support of 99) with all the known cluster II *Frankia* strains included in the tree (Fig. 2), demonstrating that all sequences obtained from soil and nodule samples belonged to cluster II. No notable difference was found between sequences obtained from soil samples and nodules, except that five nucleotides



Frankia strain	Plant Host	Host Family	Nod (+/-)	*glnA* Ref.
SH06-CV07S	NA	NA	NA	KX711687.1
SE05-HA07S	NA	NA	NA	KX711676.1
SE05-HA05S	NA	NA	NA	KX711675.1
SE05-HA02S	NA	NA	NA	KX711673.1
SE05-HA01S	NA	NA	NA	KX711672.1
SE02-HA03S	NA	NA	NA	KX711669.1
SE02-HA01S	NA	NA	NA	KX711667.1
SE02-CJ09S	NA	NA	NA	KX711665.1
SE02-CJ06S	NA	NA	NA	KX711662.1
SE02-CJ04S	NA	NA	NA	KX711660.1
SE02-CJ02S	NA	NA	NA	KX711658.1
SE02-CJ01S	NA	NA	NA	KX711657.1
SE02-AF01S	NA	NA	NA	KX711655.1
NS02-HA05S	NA	NA	NA	KX711654.1
NS02-HA04S	NA	NA	NA	KX711653.1
NS02-HA03S	NA	NA	NA	KX711652.1
NS02-HA02S	NA	NA	NA	KX711651.1
NS02-HA01S	NA	NA	NA	KX711650.1
NS02-CB02S	NA	NA	NA	KX711641.1
NS02-CB01S	NA	NA	NA	KX711640.1
NS01-HA08S	NA	NA	NA	KX711639.1
NS01-HA07S	NA	NA	NA	KX711638.1
NS01-HA05S	NA	NA	NA	KX711637.1
NS01-HA04S	NA	NA	NA	KX711636.1
NS01-HA02S	NA	NA	NA	KX711635.1
NS01-HA01S	NA	NA	NA	KX711634.1
NS02-CB09S	NA	NA	NA	KX711648.1
NS02-CB06S	NA	NA	NA	KX711645.1
NS02-CB04S	NA	NA	NA	KX711643.1
K-CF03S	NA	NA	NA	KX711633.1
AR-CF01S	NA	NA	NA	KX711623.1
AR-CF02S	NA	NA	NA	KX711624.1
NS02-CB05S	NA	NA	NA	KX711644.1
SE05-HA10S	NA	NA	NA	KX711678.1
SE05-HA04S	NA	NA	NA	KX711674.1
NS02-CB08S	NA	NA	NA	KX711647.1
NS02-CB07S	NA	NA	NA	KX711646.1
SE02-CJ10S	NA	NA	NA	KX711666.1
K-CF02S	NA	NA	NA	KX711632.1
SE02-CJ08S	NA	NA	NA	KX711664.1
SE02-CJ05S	NA	NA	NA	KX711661.1
SE02-CJ03S	NA	NA	NA	KX711659.1
SE05-HA09S	NA	NA	NA	KX711677.1
AR-CF03S	NA	NA	NA	KX711625.1
AR-CF04S	NA	NA	NA	KX711626.1
SE02-HA02S	NA	NA	NA	KX711668.1
SE02-CJ07S	NA	NA	NA	KX711663.1
FE142	*Purshia glandulosa*	Rosaceae	+	AF156759.1
FE141	*Purshia mexicana*	Rosaceae	+	AF156760.1
FE140	*Chamaebatia foliolosa*	Rosaceae	+	AF156758.1
SH06-RC02S	NA	NA	NA	KX711693.1
SH06-RC01S	NA	NA	NA	KX711692.1
NS02-CB10S	NA	NA	NA	KX711649.1
NS02-CB03S	NA	NA	NA	KX711642.1
K-CF01S	NA	NA	NA	KX711631.1
WA-CF01S	NA	NA	NA	KX711629.1
SH06-CV10S	NA	NA	NA	KX711691.1
SH06-CV08S	NA	NA	NA	KX711688.1
SH06-RC05S	NA	NA	NA	KX711696.1
SH06-RC04S	NA	NA	NA	KX711695.1
SH06-RC03S	NA	NA	NA	KX711694.1
WA-CF09S	NA	NA	NA	KX711630.1
AR-NA01S	NA	NA	NA	KX711627.1
SE02-HA04S	NA	NA	NA	KX711670.1
SH06-CV06S	NA	NA	NA	KX711686.1
SH06-CV09S	NA	NA	NA	KX711690.1
SH06-CV09N	*Ceanothus velutinus*	Rhamnaceae	+	KX711689.1
SH06-CV04S	NA	NA	NA	KX711684.1
SH06-CV03S	NA	NA	NA	KX711683.1
SH06-CV03N	*Ceanothus velutinus*	Rhamnaceae	+	KX711682.1
SH06-CV02S	NA	NA	NA	KX711681.1
SH06-CV02N	*Ceanothus velutinus*	Rhamnaceae	+	KX711680.1
SH06-CV01S	NA	NA	NA	KX711679.1
SE02-HA05S	NA	NA	NA	KX711671.1
SE02-AF02S	NA	NA	NA	KX711656.1
WA-CF01N	*Ceanothus ferrisiae*	Rhamnaceae	+	KX711628.1
SH06-CV05S	NA	NA	NA	KX711685.1
Dg1	*Datisca glomerata*	Datiscaceae	+	CP002801.1
BMG5.1	*Coriaria japonica*	Coriariaceae	+	NZ_JWIO01000047.1
FE138	*Cercocarpus ledifolius*	Rosaceae	+	AF156763.1
FE2	*Ceanothus americanus*	Rhamnaceae	+	AF156762.1
FE37	*Coriaria arborea*	Coriariaceae	+	AF156764.1
Ptl6	*Purshia tridentata*	Rosaceae	-	AY027648.1
EAN1pec	*Elaeagnus angustifolia*	Elaeagnaceae	+	CP000820.1
R2	*Ceanothus americanus*	Rhamnaceae	-	AY027647.1
FE41	*Colletia cruzerillo*	Rhamnaceae	+	AF156741.1
FE1	*Elaeagnus angustifolia*	Elaeagnaceae	+	AF156742.1
FE245	*Hippophae rhamnoides*	Elaeagnaceae	+	AF156745.1
FE39	*Trevoa trinervis*	Rhamnaceae	+	AF156744.1
FE46	*Talguenea quinquenervia*	Rhamnaceae	+	AF156746.1
FE19	*Alnus incana*	Myricaceae	+	AF156755.1
VM5	*Alnus rhombifolia*	Myricaceae	+	AY644902.1
MF1	*Myrica hartwegii*	Myricaceae	+	AY644904.1
ACN14a	*Alnus viridis*	Betulaceae	+	CT573213.2
FE14	*Comptonia peregrina*	Myricaceae	+	AF156747.1
NF2	*Alnus rhombifolia*	Betulaceae	+	AY644903.1
CW1	*Alnus rhombifolia*	Betulaceae	+	AY644907.1
MF3	*Alnus rhombifolia*	Betulaceae	+	AY644906.1
MF2	*Alnus rhombifolia*	Betulaceae	+	AY644905.1
FE136	*Myrica gale*	Myricaceae	+	AF156752.1
FE130	*Myrica gale*	Myricaceae	+	AF156753.1
FE60	*Myrica gale*	Myricaceae	+	AF156751.1
FE34	*Myrica gale*	Myricaceae	+	AF156748.1
FE32	*Myrica gale*	Myricaceae	+	AF156750.1
FE10	*Myrica gale*	Myricaceae	+	AF156749.1
Ccl3	*Casuarina cunninghamiana*	Casuarinaceae	+	CP000249.1
FE131	*Myrica gale*	Myricaceae	+	AF156757.1
Eul1c	*Elaeagnus umbellata*	Elaeagnaceae	-	CP002299.1
Acidothermus cellulolyticus 11B	NA	NA	NA	CP000481.1
Micromonospora L5	NA	NA	NA	CP002399.1

FIG 2 Phylogeny of *Frankia* strains based on partial *glnA* sequences. A maximum likelihood tree was generated using RAxML with 100 bootstrap replicates. All nodes with bootstrap support of ≥50 are indicated on the tree. The clade that includes all cluster II strains is indicated by a green box. Sequences obtained in this study are shaded in gray.

TABLE 1 Fisher's exact test results and distances from the closest host plants[a]

Location and site(s)	No. of soil samples		*P* (Fisher's exact test)	Distance from the closest host plant (m)
	With host	Without host		
ALCP				
AR, WA, and K	10/10			
CC		0/8	<0.001	>500
MNR				
NS02	15/15			
NS01		6/10	<0.05	>14
SE02	17/17			
SE05		7/10	<0.05	>25
NS02 and SE02	32/32			
NS01 and SE05		13/20	<0.001	
SEF				
SH06	15/15			
SH01		0/10	<0.001	>7

[a]For the number of soil samples taken, the numerator describes the number of samples where cluster II *Frankia* strains were detected, and the denominator describes the total number of samples collected.

consistently showed mixed signals in samples from soil, but these nucleotides were never ambiguous in the nodule samples. Fisher's exact test showed a significant difference in the rate of occurrence of cluster II *Frankia* strains between host-plant-present sites and host-plant-absent sites in all four comparative pairs of sites (Table 1). This correlation was particularly strong in the pairs in ALPC and SEF where cluster II *Frankia* strains were not detected from any samples collected from host-plant-absent sites (both $P < 0.001$). This correlation, while still significant, was weaker in the two pairs in McLaughlin Natural Reserve, Lower Lake, California (MNR) ($P < 0.05$) where cluster II *Frankia* strains were detected in 65% of the samples collected from host-plant-absent sites. Distances from host-plant-absent site to the nearest host plant varied among the three sampling locations: >500 m in CC (host-plant-absent site); however, only >14 m and >25 m in NS01 and SE05, respectively; and >7 m in SH01.

Soil pHs did not differ within location but differed between locations. The mean soil pH in each of the three locations in this study ranged from 5.4 to 7.0 (Table 2). ANOVA showed that overall soil pH significantly differed across different locations ($P < 0.001$). However, within a location, comparison of sites with and without host-plant species showed that there was no significant difference in soil pH.

Presence of host plants did not have a major impact on the soil microbiome. The principal-coordinate analysis (PCoA) did not show any apparent clustering of rhizosphere microbiomes according to the site, plant species, or presence of cluster II *Frankia* strains (Fig. 3A, B, and C). The permutational multivariate analysis of variance (MANOVA) did not find any of these factors to have significant effects either. The microbiomes of the two sites combined clustered more tightly with each other than with any of the microbiomes collected from *Oryza sativa* rhizospheres as an outgroup (Fig. 3D). The phylum-level microbiome composition across samples did not show a clear difference, and the α-diversity was not significantly affected by the site, plant species, or the presence/absence of cluster II *Frankia* strains either (Fig. 4).

Presence of host plants significantly increased relative abundance of cluster II *Frankia* strains. Phylogenetic analysis separated the 21 *Frankia* operational taxonomic units (OTUs) within the Greengenes database into all subgroups within the genus: clusters I, II, III, and IV. OTUs that did not form a clade with strains of known cluster with ≥50 bootstrap support were considered indeterminate strains (see Fig. S3 in the supplemental material). Of these 7 OTUs, 4 were most closely related to a known cluster I, and 3 were most closely related to a known cluster III.

Of the 21 OTUs, 8 OTUs (11492, 11620, 36926, 42219, 99143, 1780415, 2484834, 4477681) were present among the 10,938 OTUs that were found in at least one rhizosphere microbiome from this study. These 8 OTUs included members of clusters I,

TABLE 2 Site descriptions

Location	Site	Host plant species	Species sampled	Soil type	Soil pH (range)	Vegetation macro group
ALCP						
AR, WA, and K vs CC	AR	*Ceanothus ferrisiae*	*Ceanothus ferrisiae*	Serpentine	6.8 (6.6–7.0)	California coastal scrub
	WA	*Ceanothus ferrisiae*	*Ceanothus ferrisiae*	Nonserpentine	7.0 (6.7–7.2)	California coastal scrub
	K	*Ceanothus ferrisiae*	*Ceanothus ferrisiae*	Serpentine	7.0 (6.8–7.1)	California coastal scrub
	CC		*Arctostaphylos glauca*, *Artemisia californica*, *Pickeringia montana*	Serpentine	6.5 (6.0–6.9)	California coastal scrub
MNR						
NS02 vs NS01	NS01		*Heteromeles arbutifolia*	Nonserpentine	5.9 (5.8–6.0)	California forest and woodland
	NS02	*Cercocarpus betuloides*	*Cercocarpus betuloides*, *Heteromeles arbutifolia*	Nonserpentine	5.8 (5.5–6.1)	California forest and woodland
SE02 vs SE05	SE02	*Ceanothus jepsonii*	*Ceanothus jepsonii*, *Heteromeles arbutifolia*	Serpentine	6.6 (6.6–6.7)	California chaparral
	SE05		*Adenostoma fasciculatum*, *Heteromeles arbutifolia*	Serpentine	6.7 (6.6–7.0)	California chaparral
Unpaired	SE06			Serpentine	NA[a]	California annual and perennial grassland
SEF						
SH06 vs SH01	SH01		*Ribes cereum*	Nonserpentine	5.4 (5.0–6.0)	Western Cordilleran montane-boreal wet meadow
	SH06	*Ceanothus velutinus*	*Ceanothus velutinus*, *Ribes cereum*	Nonserpentine	5.8 (5.8–5.9)	Californian-Vancouverian montane and foothill forest

[a]NA, not applicable.

II, III, IV, and indeterminate strains. All reads of cluster II strain were found in the host-plant-present site (SE02), although in very low abundance (10 total reads, ≤3 reads, and <0.05‰ per sample). All of the reads that mapped to indeterminate OTUs were more than 1% different from any of the cluster II OTUs, indicating that the results for cluster II OTUs are coherent. Within that host-plant-present site, more reads were found in *Heteromeles arbutifolia* (nonhost) rhizosphere samples than in *C. jepsonii* (host at the site), but this difference was not statistically significant.

When relative abundance was compared between the host-plant-present site and host-plant-absent sites, factorial ANOVA showed a significant increase in cluster II *Frankia* strain relative abundance in host-plant-present sites ($P < 0.001$). On the other hand, factorial ANOVA found that the relative abundances of clusters I and IV strains were not significantly affected by host-plant presence but showed an inverse trend compared with that of cluster II ($P < 0.07$ and $P < 0.09$, respectively), i.e., more abundant in the host-plant-absent site than in the host-plant-present site. The abundance of cluster III *Frankia* was not affected significantly by host-plant presence ($P < 0.38$) (Fig. 5), but the indeterminate *Frankia* sequences were significantly more abundant in the host-plant-absent site than in the host-plant-present site ($P < 0.001$) (Fig. 5). Thus, only cluster II *Frankia* increased in relative abundance in host-plant-present sites, while the other groups either showed a negative correlation with host-plant-present sites or no correlation.

DISCUSSION

Cluster II *Frankia* strain persistence in soil depends on the specialized environment that the host plant provides. Our study has demonstrated, based both on presence/absence of *glnA* sequences and relative abundance of OTUs in rhizosphere soil, a significant correlation between the presence of cluster II *Frankia* strains and the presence of host plants in a particular site, that is consistent across all three experimental locations. There was complete correlation between cluster II strains and host plant presence in two of the three locations (ALCP and SEF). Even though a high proportion of host-plant-absent site samples in MNR were positive for cluster II *Frankia*

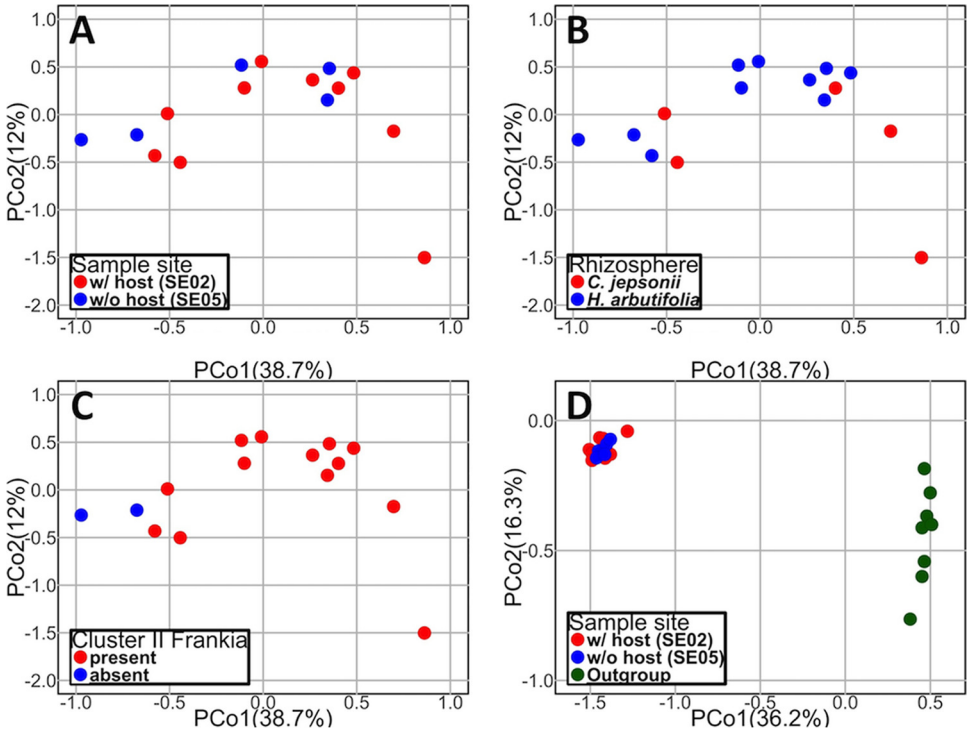


FIG 3 Principal-coordinate analysis of the rhizosphere microbiomes. (A, B, and C) Plots of the principal-coordinate analysis of the 15 rhizosphere microbiomes, except categorized according to collection site (A), plant species (B), and presence or absence of cluster II *Frankia* strains (C). Panel D is also categorized according to collection site but includes 8 *Oryza sativa* rhizosphere soil samples as an outgroup comparison. The percent variability explained by each axis is indicated on the axis.

strains (65% of samples), this was nevertheless statistically significantly different from the host-plant-present site samples (100% of samples). This confirms and expands upon previous observations of increased abundance of cluster II *Frankia* strains in soil of one host-plant-present site relative to the soil of an adjacent host-plant-absent site, using a bait plant bioassay (20). While the effect of host-plant presence was consistent, the

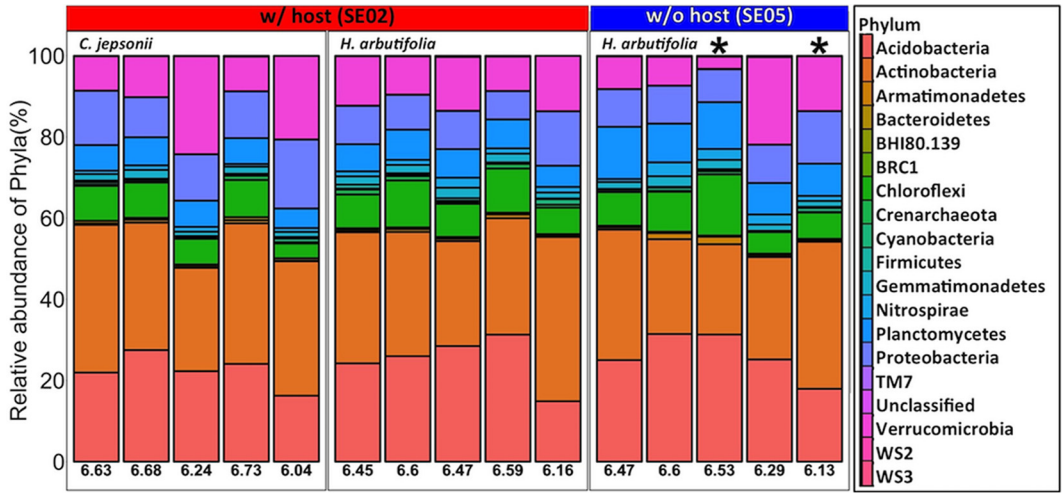


FIG 4 Relative abundances of bacterial phyla in each sample. The bacterial phylum composition of 15 rhizosphere soil microbiomes is compared. The red box above the graph indicates sites with host; the blue box indicates sites without host. Samples from the host-plant-present site are further separated by plant species, *C. jepsonii* and *H. arbutifolia*, whereas all samples from the host-plant-absent site was collected from *H. arbutifolia*. The two samples where cluster II *Frankia* strains was absent according to *glnA* are indicated with an asterisk above the bar. The α-diversity of each sample is indicated below each bar graph.

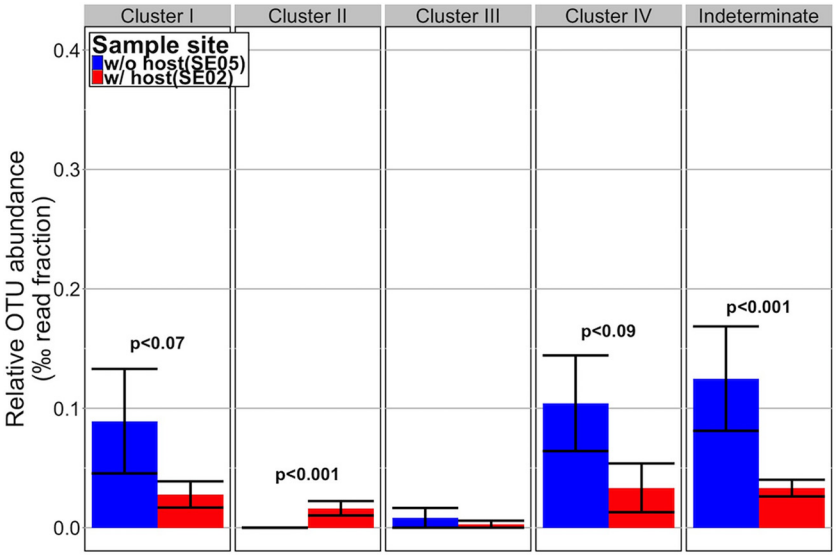


FIG 5 Mean relative abundances of all identified *Frankia* OTUs. The relative read fraction of each *Frankia* OTU in each subgroup is compared across sites with and without host plants. Error bars indicate standard errors. The clusters with relative abundances that are significantly different or those showing a trend between two sites ($P < 0.1$) have their P values indicated above the bar.

following ecological factors did not correlate with cluster II *Frankia* strain presence/absence across the three locations: (i) host species present (*C. ferrisiae*, *C. jepsonii*, *C. velutinus*, and *Cercocarpus betuloides*), (ii) soil type (serpentine versus nonserpentine), (iii) elevation, (iv) geographic location, (v) summer or annual precipitation, (vi) average monthly temperature, or (vii) soil pH. The consistency of the host influence is evidence that a host-plant-derived factor present in the soil of host-plant-present sites is the major determinant of cluster II *Frankia* distribution in these ecosystems.

We found that pH was not a significant ecological factor determining presence/absence of cluster II *Frankia* strains. The pH differed significantly among the three locations, but within each location, the pattern was consistent between sites, regardless of host presence or absence. Earlier studies found pH to have an influence on *Frankia* nodulation units in cluster I strains (21), but, on the other hand, both cluster I *Frankia* and cluster II *Frankia* strains have been reported to nodulate successfully in a wide range of pH between 3.5 to 8.0 (21) and 5.0 to 9.0 (22), respectively.

The range of soil pH in our study was between 5.0 and 7.2. Although no studies to date have examined the effect of pH directly on cluster II *Frankia* strain populations in soil, in culture, the optimal growing condition seems to be neutral to basic. The cultured cluster II *Frankia* strain BMG5.1 was shown to grow optimally at pH between 9.0 and 10.0 (23), and in another study, *Coriaria myrtifolia* had optimal growth and nodulation at pH between 7.0 and 9.0 (22). It seems, therefore, that in general *Frankia* spp. can tolerate a range of pH while having a more specific optimum. It is possible that at a more basic pH, where cluster II *Frankia* spp. are known to grow optimally, there will be a difference in cluster II *Frankia* abundance; however, *Frankia* presence/absence is not affected.

Host-plant influence extends to the site level. ANOVA of the rhizosphere microbiomes showed that while host-plant presence/absence at a given site significantly correlated with the relative abundance of cluster II *Frankia* spp., this correlation did not differ significantly between the rhizospheres of *C. jepsonii* (host plant) and *H. arbutifolia* (nonhost plant), when *H. arbutifolia* was growing in the host-plant-present site. This suggests that the effect of the host plant is not limited to its rhizosphere but can extend to the site level.

A site-level influence was also observed for nitrogen fixation, determined by the ANOVA of δ^{15}N. While host plants all showed significantly more negative δ^{15}N values

than the co-occurring nonhost plants, indicating active engagement in nitrogen fixation via RNS (19), the nonhost species in host-plant-present sites also showed significantly more negative $\delta^{15}N$ values in host-plant-present sites than in the same species in host-plant-absent sites. This pattern of ^{14}N stable isotope enrichment at the site level is similar to the facilitative effect of nitrogen-fixing plants on primary succession, reported in other nitrogen-fixing ecosystems (24). It was shown previously that tissue extracts from both roots and shoots of a host plant *Casuarina cunninghamiana* (cluster I *Frankia* host) enhanced growth of a range of *Frankia* spp. grown in culture but particularly enhanced the *in vitro* growth of the *Casuarina*-compatible strains (25). This suggests that some factor derived from the host plant, either as leaf litter or root exudate, can explain the site-level effect observed in our experiments.

Strength of host plant influence determines the strength of the correlation. Our findings suggest that the abundance of cluster II *Frankia* spp. in a given site is associated with the strength of the influence that a host plant exerts. In the case of MNR, where the host-plant-absent site had a relatively close proximity to the nearest host plant growing in a similar vegetation type (14 to 25 m), cluster II *Frankia* spp. were detectable; however, no cluster II *Frankia* spp. were detectable in the CC site in ALCP where no host plant was proximate (>500m), despite similarities in vegetation types. At the SEF site, cluster II *Frankia* spp. were not detected in the wetland site (SH01) despite the high proximity to the nearest host plants (7 m); however, the wetland site had dramatically different soil conditions and vegetation type than the adjacent montane chaparral. Cluster II *Frankia* itself can survive in moist wetland soil: *Datisca glomerata*, another host for cluster II *Frankia* spp. found in California, Nevada, and Baja California, grows and nodulates in seasonally flooded streambeds or washes throughout its range (26). Rather, it is likely that cluster II *Frankia* spp. were absent in the wetland site at SEF because the more xeric-adapted host plants (*Ceanothus* spp.) could not grow in wetland soil, thus eliminating a host effect.

The strength of the influence that the host plants may have on a given site might be affected by factors other than proximity or vegetation type. Since the cluster II *Frankia* enhancement factor may be able to persist in the soil over time, a legacy effect from host plants that were previously present might play a role. For example, cluster II *Frankia* spp. have been found in sites devoid of host plants for more than 100 years (13); *Elaeagnus*-compatible (cluster III) *Frankia* spp. were present at a site that had not had *Elaeagnus* for many years (11), and *Comptonia peregrina* (cluster I *Frankia* host) seed was viable after being in a pine-dominated forest seed bank for more than 70 years (27). Alternatively, since *Ceanothus* and *Cercocarpus* support mycorrhizal symbiosis (28, 29), the mycorrhizal mycelia may be a means of conveying a chemical signal.

Strain diversity of cluster II *Frankia* spp. There was low *glnA* sequence variability throughout this study with all samples within 0.01 substitution/site from each other; in fact, of the 74 sequences generated in this study, 56 had at least one other sequence that was identical (Fig. 2). Low genetic diversity of cluster II *Frankia* spp. based on DNA from field nodule collections, rather than from rhizosphere soil DNA, has been well documented in Oregon with repetitive sequence PCR (rep-PCR) (14) or PCR-restriction fragment length polymorphism (RFLP) (30) and in California and Nevada (16) using single-sequence analysis (*glnA* or 16S rRNA gene). Interestingly, a *glnA* sequence identical to one found in the current study has been detected in DNA extracted from a nodule of *Dryas drummondii* collected in Alaska (K. Battenberg, unpublished data). On the other hand, DNA from nodules collected in the southeastern United States (Tennessee) showed a substantial divergence, based on RFLP, from that of nodules from western North America (30). A study of nodules collected in a circumscribed area in southern coastal California found diversity of cluster II *Frankia* at the species level, using rep-PCR (15). The sampling area in the southern coastal California study (31) corresponds to a zone of biodiversity for the genus *Ceanothus* as documented previously (32). Taken together, these findings suggest that a suite of closely related strains of

cluster II *Frankia* is dominant in major portions of western North America, with some localized diversity.

Various factors have been suggested to account for the observed strain diversity within cluster II, however low it may be. Previously, a weak geographic pattern of strain diversity by *Ceanothus* collection sites was found (30). In reference 15, diversity was attributed to two ecological groupings of the host plants (*Ceanothus* spp.), but the sequence groupings were separated by $<1\%$. Both of these findings reflect the generally low cluster II *Frankia* strain sequence diversity in western North America.

Whole-genome sequencing has been carried out on strains of cluster II *Frankia* from Pakistan (33) and Japan (23); additional whole-genome sequencing from a greater number of sources will be valuable for assessing the degree of global geographic and genetic diversity within this clade.

Persistence and enrichment of cluster II *Frankia* spp. in soil. It has been postulated that cluster II *Frankia* is an obligate symbiont, based on the difficulty of isolating any strain in this group and the difficulty of detecting it in soils devoid of host plants (13). There has been only one report of nodulation of a compatible host from soil that was devoid of native host plants (13), although only after 18 months of bait plant incubation. Recently, a cluster II *Frankia* strain (BMG5.1) was isolated in pure culture from a nodule of *Coriaria japonica* collected in Japan (23). These two examples indicate that not all cluster II *Frankia* strains are obligate symbionts. That being said, it seems clear that the cluster II *Frankia* strains form a strongly restrictive association with their host plants. Our study adds to the complex nature of cluster II *Frankia* strain association with their hosts.

Two degrees of ecological association between a microsymbiont and its host may be considered: persistence (survival) and enrichment. Members of cluster I and III *Frankia* spp. have been known to persist in soil devoid of host plants for many years and to readily nodulate their respective hosts (10, 11, 21). Moreover, the number of infective units of cluster I *Frankia* spp. has been known to increase under *Alnus* host plants up to 30-fold, which makes the presence of the host plants a major factor in amplifying *Frankia* populations (21). Interestingly, the rhizosphere of *Betula* sp., a nonhost closely related to the genus *Alnus*, showed enrichment of cluster I *Frankia* spp. compared to the rhizosphere of *Alnus* spp. at a host-plant-present site (12) and abundance comparable to *Alnus* spp. at a host-plant-absent site (34).

In our study, as evidenced in the microbiome analysis, an enrichment effect of cluster II *Frankia* was present but was much subtler. Cluster II *Frankia* spp. were detected in the host-plant-present sites and were also detected in rhizosphere soil of a nonhost, *H. arbutifolia*, in the host-plant-present site, but no cluster II *Frankia* spp. were detected in the host-plant-absent site. In contrast, strains of the other clusters of *Frankia* were detected in both host-plant-present and -absent sites.

Relative abundance of different *Frankia* clusters in soil. The relatively greater abundance of cluster I *Frankia* spp. among the typical symbiotic and nitrogen-fixing subgroups (clusters I, II, and III) that we observed is in congruence with results of previous studies (34, 35). On the other hand, we found that cluster III was the least abundant: 6 reads total and only two samples with any reads detected. This contrasts with previous findings where cluster III *Frankia* spp. was the dominant or codominant subgroup in a noncompatible host rhizosphere (34, 35). This difference may be due to the fact that the soil conditions among these studies are at the opposite end of the ecological spectrum. In reference 35 the soil was moist and in a university campus arboretum, whereas our MiSeq samples were collected in dry nutrient-poor serpentine soil in summer. It is possible that there was a specific effect of serpentine soil conditions on cluster III *Frankia* spp. For example, Oline (36) showed that while microbiomes from serpentine soils were similar to those of nearby nonserpentine soils at the phylum level, distinct subgroups were adapted to the specialized environment of serpentine soil.

Not very much is known about the population distribution of the atypical cluster IV *Frankia* strains in soil. However, since they comprise the most abundant subgroup in

our samples, it is clear that at least some strains of cluster IV *Frankia* are adapted to serpentine soil. The observed trend of decreasing clusters I and IV *Frankia* spp. in host-plant-present sites compared to that in host-plant-absent sites ($0.05 < P < 0.1$) (Fig. 5) suggests that the host plant may be influencing the population sizes of strains of clusters I and IV as well as of cluster II *Frankia*. This may be related to an inhibitory factor in the host soil ecosystem, as discussed below.

Nearly 34% of the reads that mapped to the genus *Frankia* mapped to OTUs of indeterminate identity. Thus, depending on the true identity of these OTUs, our results are subject to change. However, all of the reads that mapped to indeterminate OTUs were >1% different from any of the cluster II OTUs; with the mapping threshold set to 99%, no reads that mapped to these OTUs would map to cluster II OTUs, and all indeterminate OTUs are more closely related to known cluster I or III strains than to cluster II strains.

Nature of the host-dependent factor. While the relative abundance of cluster II *Frankia* spp. in the microbiome was significantly higher in the host-plant-present site, it is still clear that this group of OTUs is quite rare in the microbiome, representing only <0.05‰ in any sample. This suggests that the host factor is not a source of energy to sustain a large population; i.e., it is not promoting the proliferation of cluster II *Frankia* outside the host, in contrast to some strains within the cluster I *Frankia* that can utilize a host-derived soil carbon source (37).

Further, permutational MANOVA and factorial ANOVA both showed that the presence of the host plant did not significantly affect the relative abundances of all OTUs in the microbiome overall nor the α-diversity; in fact PCoA with an outgroup showed that the microbiomes of the samples from the host-plant-present and host-plant-absent sites (SE02 and SE05) were very similar to each other. This minimal influence of host/nonhost on the overall microbiome together with the trend observed for clusters I and IV *Frankia* OTUs suggests that a host-derived factor has an effect specifically on cluster II *Frankia* OTUs.

Cluster II *Frankia* OTUs were found in the rhizosphere soil of the nonhost *H. arbutifolia* in host-plant-present sites, but not in the rhizosphere of *H. arbutifolia* in host-plant-absent sites. Thus, the host-plant effect was not limited to the host-plant rhizosphere but rather extended to the level of the site where the host plant was present. Extracts from host-plant roots and shoots have also been found to enhance growth of *Frankia* strains in culture particularly strongly if the strain is compatible (25). Taking this into account, we propose that possible mechanisms for the dispersal of this host-dependent factor might be (i) diffusion of a compound or compounds originating as root exudates, (ii) leaf litter decomposition and leaching into the soil profile, or (iii) chemical signals dispersed via mycorrhizal networks (38).

A plant factor with specific targets that might have a strong impact on *Frankia* strain presence in the rhizosphere might be a plant-derived secondary compound related to nodulation signaling, such as a flavonoid. Flavonoids are known signaling molecules in rhizobial symbioses (39, 40). Additionally, in *Myrica* sp. symbioses, a suite of hydroxy-chalcones, termed myrigalones, has been shown to promote infective *Frankia* and inhibit noninfective *Frankia* strains (41). Alternatively, this factor might be a terpenoid with effects similar to strigolactone, known to stimulate spore germination and limited hyphal growth of fungal symbionts in the rhizosphere in arbuscular mycorrhizal endosymbioses (42). The host signaling pathway that initiates AM symbiosis is ancestral to signal exchange in RNS; and the microsymbiont signal molecules are also similar between these two types of endosymbiosis (3). Two cluster II *Frankia* genomes contain close homologs of rhizobial *nodABC* genes (14, 59). These *nod* gene homologs have been shown to be expressed in developing root nodules of *D. glomerata* (17).

In summary, our data demonstrate the existence of a host site enrichment of cluster II *Frankia* strains that is independent of host plant species, vegetation type, geographical location, climate, soil type, and soil pH. The dependency of *Frankia* spp. on their

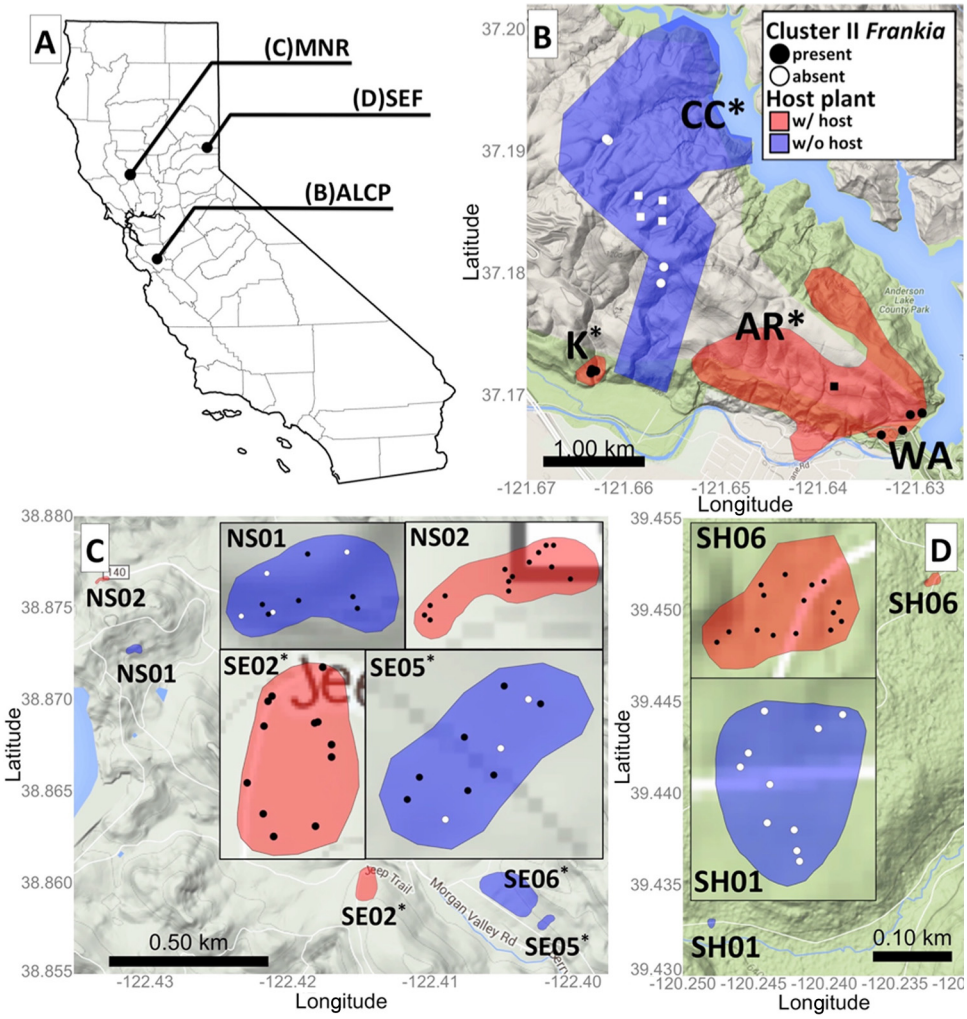


FIG 6 Soil sampling scheme. (A) The three soil sampling locations are indicated on the map of California. Individual sites within each location are as follows: within Anderson Lake County Park (ALCP) (B), within McLaughlin Natural Reserve (MNR) (C), and within Sagehen Experimental Forest (SEF) (D). Red and blue shading indicate sites with and without host plants. Serpentine soil sites are indicated by an asterisk. The open or filled dots indicate a location where cluster II *Frankia* strains were absent (open) or present (filled), respectively. In panel B, squares indicate locations where GPS coordinates were not recorded. (A) U.S. Census Bureau map (https://commons.wikimedia.org/wiki/File:California_counties_outline_map.svg). (B to D) Map data ©2016 Google.

hosts for survival and persistence has been postulated, based on *in vitro* experiments (21). Host and nonhost rhizospheres were recently compared with respect to cluster I *Frankia* strains in one location (34), showing no significant host rhizosphere effect. Cluster III *Frankia* strains were shown to have a distribution independent of the host rhizophere, implying a broad physiological adaptation in soil (37). The current report is the first comprehensive study showing a significant host plant influence on cluster II *Frankia* strains in natural soil environments, an effect that is distinct from that of other ecological parameters. Identification and testing of potential specific host-plant compounds and transmission pathways remain to be carried out in more controlled-environment experiments.

MATERIALS AND METHODS

Sampling sites. All soil, leaf, and root nodule samples were collected from three locations in northern California: in and near Anderson Lake County Park, Santa Clara County, California (ALCP), McLaughlin Natural Reserve, Lower Lake, California (MNR), or Sagehen Experimental Forest, Truckee, California (SEF), as shown in Fig. 6 (see Table S1 in the supplemental material for the exact coordinates). Maps of these locations were generated with R (v3.0.2) with the ggmap package (Fig. 6B to D). These locations varied considerably in altitude, monthly precipitation, monthly temperature, and soil type. The

altitudes were approximately 200 m, 650 m, and 1,950 m above sea level in ALCP, MNR, and SEF, respectively.

Multiple sampling sites were selected within each location: ALCP, $n = 4$ sampling sites; MNR, $n = 5$ sampling sites; and SEF, $n = 2$ sampling sites. The main factors that were tested across these sites were (i) host plant presence/absence (i.e., whether a plant that has been reported to form RNS with *Frankia* strains is present on the site or not (see Fig. S1 in the supplemental material for an example of a host-plant-present site), (ii) host plant species (if present), (iii) soil type (serpentine versus nonserpentine), (iv) soil pH, (v) vegetation type, and (vi) climate type. Factors 1 to 5 are described in Table 2. The soil type in each site was determined by methods described below, and the vegetation types were determined to the macro group level according to the U.S. National Vegetation Classification (http://usnvc.org/explore-classification/) using *A Manual of California Vegetation* (43). The annual patterns of temperatures and precipitation at each location were determined according to the WorldClim database (44).

Sites within each location were paired according to host-plant presence/absence for comparison. In ALCP, AR, WA, and K (host-plant-present sites) were compared with CC (host-plant-absent site) on serpentine (AR and K) and nonserpentine soil (WA). In MNR, SE02 (host-plant-present site) was compared with SE05 (host-plant-absent site), both on serpentine soil, and NS02 (host-plant-present site) was compared with NS01 (host-plant-absent site), both on nonserpentine soil. In SEF, SH06 (host-plant-present site) was compared with SH01 (host-plant-absent site), both on nonserpentine soil. Vegetation composition was similar for pairs in ALCP (45) (Table 2) and in MNR (46) (Table 2). In the Sagehen Creek site, the pairing was between sites with dissimilar but adjacent vegetation types (Table 2). In ALCP, the host plant *Ceanothus ferrisiae* (Rhamnaceae; federally listed as endangered) was present in AR, WA, K but absent in CC. In MNR, the host plant *Ceanothus jepsonii* (Rhamnaceae) was present in SE02, and the host plant *Cercocarpus betuloides* (Rosaceae) was present in NS02. In SEF, *Ceanothus velutinus* (Rhamnaceae) was present in SH06. Site SE06 in MNR (host-plant-absent site) was used to collect leaf samples of *Avena fatua* (Poaceae) for δ^{15}N comparison (Table 2).

Soil and nodule sampling. Rhizosphere soil samples were collected from both hosts and nonhosts in all three locations for DNA analysis and placed on ice for transport to University of California (UC) Davis. For ALCP, samples were collected from 4 sites: AR ($n = 5$), K ($n = 3$), and WA ($n = 2$), host plant present; and CC ($n = 8$), host plant absent. In ALCP, samples were collected from the rhizospheres of host species *C. ferrisiae* (host-plant-present sites), and from CC (host-plant-absent site), samples were collected from the rhizospheres of *Artemisia californica* (Asteraceae), *Arctostaphylos glauca* (Ericaceae), and *Pickeringia montana* (Fabaceae), a nonnodulating legume (47). One soil sample in the nonhost site was collected away from any plant (Table 2 and Table S1).

For MNR, samples were collected from 4 sites: NS02 ($n = 15$) and SE02 ($n = 17$), host-plant-present sites; and NS01 ($n = 10$) and SE05 ($n = 10$), host-plant-absent sites. In MNR, samples were collected from the rhizospheres of host species *C. jepsonii* or *C. betuloides* or nonhost species *Adenostoma fasciculatum* (Rosaceae) or *Heteromeles arbutifolia* (Rosaceae).

For SEF, samples were collected from two sites: SH06 ($n = 15$), host-plant-present site; and SH01 ($n = 10$), host-plant-absent site. In SEF, samples were collected from the rhizospheres of host species *C. velutinus* or nonhost species *Ribes cereum* (Grossulariaceae).

Soil samples were collected by first removing the free surface organic material and an additional >15 cm of the soil profile before collecting the soil below. Root nodules were searched for every time that soil was excavated and collected when found. Nodules were collected from one individual of *C. ferrisiae* at the WA site and three individuals of *C. velutinus* from SH06.

Plant tissue sampling for δ^{15}N stable isotope analysis. Tissue samples were collected from seven plant species on six separate occasions between November 2011 and August 2015 (Table 3) and placed on ice for transport to UC Davis. In ALCP, *C. ferrisiae* was collected from the host-plant-present sites and *A. californica* was collected from host-plant-present and host-plant-absent sites. In MNR, *C. betuloides* and *C. jepsonii* were each collected from one host-plant-present site, and *A. fatua* was collected from one host-plant-absent site. In SEF, *C. velutinus* was collected in one host-plant present site, and *R. cereum* was collected from host-plant-present and host-plant-absent sites.

For *C. betuloides*, *C. ferrisiae*, *C. jepsonii*, *C. velutinus*, and *R. cereum*, the youngest mature leaves were collected; for *A. californica*, whole shoot tips, including the youngest mature leaves, were collected; and for *A. fatua*, the entire mature inflorescences were collected.

Soil type and pH determination. Soil chemical analysis was carried out to determine soil pH and to determine whether soils were serpentine or not. In MNR and SEF, rhizosphere soil collected for DNA analysis was also used for chemical analysis. Equal amounts of three rhizosphere soil samples were mixed into one pooled sample, and three pooled samples were made for each site. Samples were analyzed for pH by the UC Davis Analytical Laboratory.

In ALCP, a second set of soil samples (AR, $n = 3$; K, $n = 5$; WA, $n = 2$; and CC, $n = 4$) was collected for chemical analysis. For each sample, three subsamples of equal amounts were collected and subsequently mixed. Soil samples were analyzed by the Soil & Plant Laboratory (San Jose, CA) for pH and to test the calcium-to-magnesium (Ca/Mg) ratio for indication of serpentine soil (Ca/Mg of <1).

In MNR, the soil types on sites NS01 and NS02 were previously reported as nonserpentine, whereas that of site SE02 was reported as serpentine (46). The soil types in SE05 and SE06 were determined to be serpentine based on the USDA National Cooperative Soil Survey (NCSS) (http://websoilsurvey.nrcs.usda.gov/, and personal communication with the resident directors of MNR, P. Aigner and C. Koehler. Likewise in SEF, soil types at SH01 and SH06 were determined to be nonserpentine based on the NCSS.

TABLE 3 Leaf sampling scheme

Collection occasion and species	No. of leaf samples collected at each location and site											
	ALCP				MNR					SEF		
	AR	K	WA	CC	NS01	NS02	SE0	SE05	SE06	SH01	SH06	Total
February 2012												
Ceanothus ferrisiae	6	5	0	0								11
Artemisia californica	2	2	0	3								7
February 2013												
Ceanothus ferrisiae					3	0	4	0				7
Artemisia californica					4	0	4	5				13
June 2014												
Cercocarpus betuloides					0	5	0	0	0			5
March 2015					0	0	0	5	0			5
Ceanothus jepsonii												
August 2015												
Avena fatua					0	0	0	0	6			6
June 2015												
Ceanothus velutinus										0	10	10
Ribes cereum										10	5	15

DNA extraction, amplification, and sequencing. Total DNA was extracted from rhizosphere soil samples using the Power Soil DNA isolation kit (Mo Bio Laboratories Inc., CA, USA) according to the manufacturer's protocol. Total DNA was extracted from the root nodule samples using a DNeasy plant minikit (Qiagen, CA, USA) according to the manufacturer's protocol.

For each DNA extract, PCR was carried out, targeting either glutamine synthetase A (*glnA*) or the 16S rRNA gene, as in Table 4. For *glnA*, a nested PCR was carried out using two sets of cluster II *Frankia* strain-specific primers, modified from reference 6. Use of the primer pairs from reference 6 usually resulted in a mixed sequence that showed similarity to a range of actinobacterial taxa (see Fig. S2 in the supplemental material for an example of an electropherogram), but after numerous attempts, we obtained one sequence without mixed peaks (see Fig. S2 in the supplemental material for the electropherogram) that showed 100% identity to a partial *glnA* gene sequence of a published cluster II *Frankia* strain, FE141 (GenBank accession no. AF156760.1). The two reverse-direction primers (external and internal) were modified to perfectly complement FE141. Because the FE141 partial *glnA* sequence was too short to cover the range for the forward-direction primers and because the primers needed a wider target range than a single strain, the two forward-direction primers were modified to perfectly complement a different published cluster II *Frankia* strain, Dg1 (GenBank accession no. CP002801.1). The specificity of the *glnA* primer pairs was tested using Primer-BLAST from NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) by searching for potential amplicons within the nonredundant (nr) database. While the external pair had multiple potential amplicons, the internal pair only had a single match to Dg1.

PCR was conducted using a *Taq* PCR core kit (Qiagen, CA, USA) according to the manufacturer's protocol. For the 16S rRNA gene, universal 16S rRNA gene primers were used (48).

For *glnA* nested PCR, 10 μl of template DNA was used for each 50-μl reaction to compensate for the low level of DNA extracted from soil (often $<$20 ng/μl) for the external PCR, and 3 μl of template DNA was used for each 50-μl reaction for the internal PCR. Both the external and internal PCRs were run on a thermal cycler programmed for a hot start (95°C for 120 s), 35 cycles (94°C for 30 s, 57°C for 45 s, and 72°C for 60 s), and final extension (72°C for 420 s).

TABLE 4 Primers used in this study

Name	Target	Direction	Sequence
E683-700	16S rRNA	Forward	5′-GTGTAGMGGTGAAATKCG-3′
E1052-1072	16S rRNA	Reverse	5′-CGAGCTGACGACARCCATGCA-3′
DB41	*glnA* (external)	Forward	5′-TTCTTCATCCACGACCCG-3′
DB41M	*glnA* (external)	Forward	5′-TTCTTCATCCATGACCCGA-3′
FRKGSrev	*glnA* (external)	Reverse	5′-GTATTTGTAGAGCATCAGGTTGTC-3′
FRKGSrevM	*glnA* (external)	Reverse	5′-TTATGTATTTGTAGAGCATCAGGTTGTC-3′
FRKGS2F	*glnA* (internal)	Forward	5′-CATCCACGACCCGATCACCCG-3′
FRKGS2FM	*glnA* (internal)	Forward	5′-CATCCATGACCCGATCACTCGG-3′
FRKGS2R	*glnA* (internal)	Reverse	5′-CTCGACGGTGATACCGGCCCTGATC-3′
FRKGS2RM	*glnA* (internal)	Reverse	5′-CTGATCAGTGCCCGGCTCATCTCCGAC-3′

For 16S rRNA gene PCR, 1 μl of template DNA was used for each 50-μl reaction. Reactions were run on a thermal cycler programmed for a touchdown PCR with annealing temperatures from 63 to 59°C with 1°C increments for 2 cycles each (94°C for 30 s, 63 to 59°C for 45 s, and 72°C for 60 s), followed by 25 cycles (94°C for 30 s, 58°C for 45 s, and 72°C for 60 s) and final extension (72°C for 420 s).

The presence of amplicons of the appropriate sizes was assessed on 0.8% (wt/vol) agarose gel run at 100 V in Tris-borate-EDTA (TBE) buffer with ethidium bromide. The amplified fragments of *glnA* were then purified using a QIAquick gel extraction kit (Qiagen, CA, USA) according to the manufacturer's protocol. Purified DNA fragments were sequenced in both directions on an ABI Prism 3730 genetic analyzer in the DNA Sequencing Facility at UC Davis.

Phylogenetic analysis. The forward and reverse sequences of *glnA* obtained from each sample were proofread and assembled using Sequencher v5.1 (Gene Codes Corporation, MI, USA) to obtain a consensus sequence. A total of 74 sequences were generated. These sequences were aligned with those from 34 *Frankia* strains and 2 closely related outgroups available in GenBank. Sequences were aligned using MUSCLE (49) through MEGA6 (50) and manually adjusted to improve alignment. The resulting 286-bp alignment was then used to select the most likely evolutionary model via jModelTest2 (51). Values for model parameters were estimated by RAxML (52), which was used to reconstruct the phylogeny based on the maximum likelihood method. To avoid having the tree lodge onto a local optimum, 4 parallel searches were conducted to determine the best tree. Bootstrap values were based on 100 replicates.

MiSeq analysis. A subset of the DNA extracts from host and nonhost rhizosphere samples in two sites in MNR, SE02 ($n = 10$) and SE05 ($n = 5$), were further analyzed for their microbiome biodiversity based on the 16S rRNA gene. In SE02 (host-plant-present site), samples were taken from rhizospheres of *C. jepsonii* ($n = 5$) and *H. arbutifolia* ($n = 5$). Samples from SE05 (host-plant-absent site) were taken from the rhizosphere of *H. arbutifolia*. The V4 region of the 16S rRNA gene was amplified (primer sequences are available in Table S2 in the supplemental material) and sequenced via the Illumina MiSeq platform according to the methods described previously (53) (NCBI BioProject accession number PRJNA338518). Contiguous reads were generated using the PandaSeq pipeline (54). Of the raw sequences obtained from the MiSeq analysis, any that were >275 bp or that contained any ambiguous sites were removed. The remaining sequences were aligned against the Greengenes 16S rRNA database (13_8 release) (55) to obtain the number of sequence reads that belonged to each operational taxonomic unit (OTU), based on 99% pairwise identity, according to reference 53). The Greengenes database was chosen as our reference database based on its wide range of coverage that was suited for our aim to survey the overall microbiome. Organellar sequences were removed from any subsequent analyses (see Table S3 in the supplemental material).

δ^{15}N analysis. Plant tissue samples were dried at 60°C for >72 h and then were ground to powder with an agate mortar and pestle. For *A. fatua*, the glume was separated from the rest of the inflorescence, and for *A. californica*, the dried leaves were removed from stems prior to grinding. Powdered samples were then passed through a soil sieve number 35 (0.5-mm^2 grids). Approximately 5 mg per sample was weighed, loaded into tin capsules, and analyzed in a PDZ Europa 20:20 isotope ratio mass spectrometer (Sercon Ltd., Cheshire, UK) in the Stable Isotope Facility, UC Davis, for δ^{15}N natural abundance.

Statistical analysis. All statistical analyses were conducted using R (v3.0.2) with the Lme4 package (56) and Vegan package (57), in addition to the base package.

For the δ^{15}N data, a series of factorial ANOVA were conducted using anova() in R to determine if a particular factor, site, and/or plant species had a significant effect on the values. All factors were assumed to have a fixed effect rather than a random effect. When any factor was found to have a significant effect, *post hoc* comparisons were made by fitting the values to a linear model using lm() in R to test if a category within a factor (e.g., *C. ferrisiae* or *A. californica* within plant species) resulted in significantly higher or lower values by calculating the least-squares mean.

The correlation between the presence of strains of cluster II *Frankia* in a soil and their host plants were tested by conducting a one-sided Fisher's exact test using fisher.test() in R. These analyses were conducted on all comparative pairs of sites independently.

Based on the MiSeq data of rhizosphere soil microbiomes, principal-coordinate analysis (PCoA) was conducted. Of the 10,938 OTUs that were identified from the samples, those that had <30 total reads across all 15 samples were removed for the downstream analysis. The relative abundance of the retained 2,840 OTUs was calculated by dividing the individual reads by the total remaining reads of each sample (see Table S1). PCoA was conducted based on the Bray-Curtis dissimilarity of log-transformed relative abundance using pcoa() in R. As an outgroup comparison, this data set was analyzed by comparing it with 8 rhizosphere soil samples from *Oryza sativa* growing in a conventionally cultivated paddy near Sacramento, California (53site SFT20A) (53).

Permutational MANOVA was conducted using adonis() in R with sampling site, plant species, and presence of cluster II *Frankia* strains as factors. The relative abundances of the 2,840 OTUs were used to calculate the α-diversity (Shannon index) at the phylum level for each sample using diversity() in R.

To determine the specific cluster identity of each *Frankia* OTU within the Greengenes database, all OTUs that were classified as family *Frankiaceae* within the database were collected, and a phylogenetic tree of the full-length 16S rRNA gene was constructed together with sequences from other known *Frankia* strains (data not shown). This tree confirmed that only the 21 OTUs classified as genus *Frankia* in the database formed a clade together with the other known *Frankia* strains. A second phylogenetic tree was then constructed with the members within this clade with these 21 OTUs using the same methods described above for analyzing *glnA* sequences (see Fig. S3 in the supplemental material). The

purpose of the phylogenetic reconstructions with *glnA* (Fig. 2) or 16S (Fig. S3) markers was to assign each sequence to a cluster within *Frankia*. The topology of the reconstructed phylogeny among the clusters has been shown to vary depending on the marker used, but members within each cluster are stable (8, 17, 58).

Another series of factorial ANOVA were conducted to test whether (i) relative abundance of each *Frankia* cluster, (ii) soil pH per site and per location, or (iii) α-diversity significantly differed according to the experimental variables: site, presence/absence of host-plant species, and plant species. The same methods as for the δ^{15}N data were used, except that the relative abundance of a *Frankia* cluster was fitted to a generalized linear model using glm() in R with its dispersion parameter set according to the quasi-Poisson model.

Accession number(s). DNA sequences were submitted to NCBI. The GenBank accession numbers for *glnA* sequences are KX711623.1 through KX711696.1. The MiSeq nucleotide sequences are found in NCBI Sequence Read Archive (SRA) BioProject accession number PRJNA338518.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02661-16.

TEXT S1, PDF file, 0.3 MB.
DATASET S2, XLSX file, 0.02 MB.
DATASET S3, XLSX file, 0.05 MB.
DATASET S4, XLSX file, 3.0 MB.

ACKNOWLEDGMENTS

We thank Isaac Gifford and Daniel Potter for insightful discussions and Cristina Covblic for assisting with experimental procedures.

Author contributions: A.M.B. and K.B. designed the experiments, with guidance by J.H.; field work was carried out by K.B., A.M.B., L.H., J.A.W., and J.H.; L.H. assisted in data analysis of the pilot study; J.A.W. carried out laboratory analytical procedures; J.E. contributed to design and data analysis for the microbiome study; the manuscript was written by K.B. and A.M.B., with contributions by J.A.W. and other coauthors.

This work was supported by USDA-NIFA-CA-D-PLS-2173-H (to A.M.B., University of California Agricultural Experiment Station). Modest funding (<$15,000) was provided by the Santa Clara Valley Water District; Kai Battenberg and Jannah A. Wren received student grants ($1,000 each) from the California Native Plant Society.

REFERENCES

1. Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, Vitousek P, Leach A, Bouwman AF, Butterbach-Bahl K, Dentener F, Stevenson D, Amann M, Voss M. 2013. The global nitrogen cycle in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 368:20130164. https://doi.org/10.1098/rstb.2013.0164.
2. Harrison MJ. 2005. Signaling in the arbuscular mycorrhizal symbiosis. Annu Rev Microbiol 59:19–42. https://doi.org/10.1146/annurev.micro.58.030603.123749.
3. Oldroyd GED. 2013. Speak, friend, and enter: signaling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol 11:252–263. https://doi.org/10.1038/nrmicro2990.
4. Pawlowski K, Demchenko KN. 2012. The diversity of actinorhizal symbiosis. Protoplasma 249:967–979. https://doi.org/10.1007/s00709-012-0388-4.
5. Johansson C, Bergman B. 1993. Reconstitution of the symbiosis of *Gunnera manicata* Linden: cyanobacterial specificity. New Phytol 126:643–652.
6. Werner GDA, Cornwell WK, Sprent JI, Kattge J, Kiers ET. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. Nat Commun 5:4087.
7. Clawson ML, Bourret A, Benson DR. 2004. Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences. Mol Phylogenet Evol 31:131–138. https://doi.org/10.1016/j.ympev.2003.08.001.
8. Hahn D. 2008. Polyphasic taxonomy of the genus *Frankia*, p 25–47. *In* Pawlowski K, Newton WE (ed), Nitrogen-fixing actinorhizal symbioses, 1st ed. Springer, Dordrecht, Netherlands.
9. Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, Bassi CA, Berry AM, Bickhart DM, Choisne N, Couloux A, Cournoyer B, Cruveiller S, Daubin V, Demange N, Francino MP, Goltsman E, Huang Y, Kopp OR, Labarre L, Lapidus A, Lavire C, Marechal J, Martinez M, Mastronunzio JE, Mullin BC, Niemann J, Pujic P, Rawnsley T, Rouy Z, Schenowitz C, Sellstedt A, Tavares F, Tomkins JP, Vallenet D, Valverde C, Wall LG, Wang Y, Medigue C, Benson DR. 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Res 17:7–15.
10. Paschke MW, Dawson JO. 1992. The occurrence of *Frankia* in tropical forest soils of Costa Rica. Plant Soil 142:63–67.
11. Nalin R, Normand P, Domenach A. 1997. Distribution and N_2-fixing activity of *Frankia* strains in relation to soil depth. Physiol Plant 99:732–738. https://doi.org/10.1111/j.1399-3054.1997.tb05378.x.
12. Smolander A. 1990. *Frankia* populations in soils under different tree species-with special emphasis on soils under *Betula pendula*. Plant Soil 121:1–10. https://doi.org/10.1007/BF00013091.
13. Nouioui I, Sbissi I, Ghodhbane-Gtari F, Benbrahim FK, Normand P, Gtari M. 2013. First report on the occurrence of the uncultivated cluster II *Frankia* microsymbionts in soil outside the native actinorhizal host range area. J Biosci 38:695–698. https://doi.org/10.1007/s12038-013-9366-z.
14. Jeong S, Myrold DD. 1999. Genomic fingerprinting of *Frankia* microsymbionts from *Ceanothus* copopulations using repetitive sequences and polymerase chain reactions. Can J Bot 77:1220–1230. https://doi.org/10.1139/b99-069.
15. Oakley B, North M, Franklin JF, Hedlund BP, Staley JT. 2004. Diversity and distribution of *Frankia* strains symbiotic with *Ceanothus* in California.

Appl Environ Microbiol 70:6444–6452. https://doi.org/10.1128/AEM.70.11.6444-6452.2004.
16. Vanden Heuvel BD, Benson DR, Bortiri E, Potter D. 2004. Low genetic diversity among *Frankia* spp. strains nodulating sympatric populations of actinorhizal species of Rosaceae, *Ceanothus* (Rhamnaceae) and *Datisca glomerata* (Datiscaceae) west of the Sierra Nevada (California). Can J Microbiol 50:989–1000. https://doi.org/10.1139/w04-079.
17. Persson T, Battenberg K, Demina IV, Vigil-Stenman T, Vanden Heuvel B, Pujic P, Facciotti MT, Wilbanks EG, O'Brien A, Fournier P, Cruz Hernandez MA, Herrera AM, Médigue C, Normand P, Pawlowski K, Berry AM. 2015. *Candidatus* Frankia datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes *nodABC* in symbiosis with its host plant. PLoS One 10:e0127630. https://doi.org/10.1371/journal.pone.0127630.
18. Kummerow J, Alexander JV, Neel JW, Fishbeck K. 1978. Symbiotic nitrogen fixation in *Ceanothus* roots. Am J Bot 65:63–69. https://doi.org/10.2307/2442555.
19. Boddey RM, Peoples MB, Palmer B, Dart PJ. 2000. Use of the ^{15}N natural abundance technique to quantify biological nitrogen fixation by woody perennials. Nutr Cycling Agroecosyst 57:235–270. https://doi.org/10.1023/A:1009890514844.
20. Jeong S, Myrold DD. 2001. Population size and diversity of *Frankia* in soils of *Ceanothus velutinus* and Douglas-fir stands. Soil Biol Biochem 33:931–941. https://doi.org/10.1016/S0038-0717(00)00241-8.
21. Benson DR, Silvester WB. 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol Rev 57:293–319.
22. Cañizo A, Miguel C, Rodriguez-Barrueco C. 1979. The effect of pH on nodulation and growth of *Coriaria myrtifolia* L. Plant Soil 49:195–198.
23. Gtari M, Ghodhbane-Gtari F, Nouioui I, Ktari A, Hezbri1 K, Mimouni W, Sbissi I, Ayari Yamanaka T, Normand P, Tisa LS, Boudabous A. 2015. Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Sci Rep 5:13112. https://doi.org/10.1038/srep13112.
24. Chapin FS, Walker LR, Fastie CL, Sharman LC. 1994. Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. Ecol Monogr 64:149–175. https://doi.org/10.2307/2937039.
25. Zimpfer JF, Igual JM, McCarty B, Smyth C, Dawson JO. 2004. *Casuarina cunninghamiana* tissue extracts stipulate the growth of *Frankia* and differentially alter the growth of other soil microorganisms. J Chem Ecol 30:439–452. https://doi.org/10.1023/B:JOEC.0000017987.19225.86.
26. Stone WJ. 2012. *In* Hickman JC (ed). The Jepson manual: vascular plants of California. University of California Press, Berkeley, CA.
27. Del Tredici P. 1977. The buried seeds of *Comptonia peregrina*, the sweet fern. Bull Torrey Bot Club 104:270–275. https://doi.org/10.2307/2484309.
28. Rose SL. 1979. Mycorrhizal associations of some actinomycete nodulated nitrogen-fixing plants. Can J Bot 58:1449–1454.
29. McDonald KR, Pennell J, Frank JL, Southworth D. 2010. Ectomycorrhizas of *Cercocarpus ledifolius* (Rosaceae). Am J Bot 97:1867–1872. https://doi.org/10.3732/ajb.0900357.
30. Ritchie NJ, Myrold DD. 1999. Geographic distribution and genetic diversity of *Ceanothus*-infective *Frankia* strains. Appl Environ Microbiol 65:1378–1383.
31. Murry MA, Konopka AS, Pratt SD, Vandergon TL. 1997. The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus*. Physiol Plant 99:714–721. https://doi.org/10.1111/j.1399-3054.1997.tb05376.x.
32. Burge DO, Erwin DM, Islam MB, Kellermann J, Kembel SW, Wilken DH, Manos PS. 2011. Diversification of *Ceanothus* (Rhamnaceae) in the California Floristic Province. Int J Plant Sci 172:1137–1164. https://doi.org/10.1086/662028.
33. Persson T, Benson DR, Normand P, Vanden Heuvel BD, Pujic P, Chertkov O, Teshima H, Bruce DC, Detter C, Tapia R, Han S, Han J, Woyke T, Pitluck S, Pennacchio L, Nolan M, Ivanova N, Pati A, Land ML, Pawlowski K, Berry AM. 2011. Genome sequence of "*Candidatus* Frankia datiscae" Dg1, the uncultured microsymbiont from nitrogen-fixing root nodules of the dicot *Datisca glomerata*. J Bacteriol 193:7017–7018. https://doi.org/10.1128/JB.06208-11.
34. Samant S, Huo T, Dawson JO, Hahn D. 2016. Abundance and relative distribution of *Frankia* host infection groups under actinorhizal *Alnus glutinosa* and non-actinorhizal *Betula nigra* trees. Microb Ecol 71:473–481. https://doi.org/10.1007/s00248-015-0643-2.
35. Rodriguez D, Guerra TM, Forstner MRJ, Hahn D. 2016. Diversity of *Frankia* in soil assessed by Illumina sequencing of *nifH* gene fragments. Syst Appl Microbiol 39:391–397. https://doi.org/10.1016/j.syapm.2016.06.007.
36. Oline DK. 2006. Phylogenetic comparisons of bacterial communities from serpentine and nonserpentine soils. Appl Environ Microbiol 72:6965–6971. https://doi.org/10.1128/AEM.00690-06.
37. Samant S, Dawson JO, Hahn D. 2015. Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. Syst Appl Microbiol 38:501–505. https://doi.org/10.1016/j.syapm.2015.07.005.
38. Bennett AE, Daniel TJ, Öpik M, Davison J, Moora M, Zobel M, Selosse Evans D. 2013. Arbuscular mycorrhizal fungal networks vary throughout the growing season and between successional stages. PLoS One 8:e83241. https://doi.org/10.1371/journal.pone.0083241.
39. Peters NK, Frost JW, Long SR. 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science 233:977–980. https://doi.org/10.1126/science.3738520.
40. Kapulnik Y, Joseph C, Phillips D. 1987. Flavone limitations to root nodulation and symbiotic nitrogen fixation in alfalfa. J Plant Physiol 84:1193–1196. https://doi.org/10.1104/pp.84.4.1193.
41. Popovici J, Comte G, Bagnarol E, Alloisio N, Fournier P, Bellvert F, Bertrand C, Fernandez MP. 2010. Differential effects of rare specific flavonoids on compatible and incompatible strains in the *Myrica gale-Frankia* actinorhizal symbiosis. Appl Environ Microbiol 76:2451–2460. https://doi.org/10.1128/AEM.02667-09.
42. Parniske M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat Rev Microbiol 6:763–775. https://doi.org/10.1038/nrmicro1987.
43. Sawyer JO, Keeler-Wolf T, Evens JM. 2009. A manual of california vegetation, 2nd ed. California Native Plant Society Press, Sacramento, CA.
44. Humans RJ, Cameron SE, Parra JL, Jonesv PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978. https://doi.org/10.1002/joc.1276.
45. US Fish and Wildlife Service. 2011. *Ceanothus ferrisiae* (Coyote ceanothus) 5-year review: summary and evaluation. US Fish and Wildlife Service, Washington, DC. http://ecos.fws.gov/ecp0/profile/speciesProfile?spcode=Q1U3.
46. Anacker BL, Harrison SP. 2012. Historical and ecological controls on phylogenetic diversity in Californian plant communities. Am Nat 180:257–269. https://doi.org/10.1086/666650.
47. Sprent JI, Ardley JK, James EK. 2013. From north to south: a latitudinal look at legume nodulation processes. South Afr J Bot 89:31–41. https://doi.org/10.1016/j.sajb.2013.06.011.
48. Wang Y, Qian P. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS One 4:e7401. https://doi.org/10.1371/journal.pone.0007401.
49. Edgar CR. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
50. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.
51. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109.
52. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. https://doi.org/10.1093/bioinformatics/btl446.
53. Edwards J, Johnson C, Santos-Medellína C, Luriea E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci U S A 112:E911–E920. https://doi.org/10.1073/pnas.1414592112.
54. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl Environ Microbiol 77:3846–3852. https://doi.org/10.1128/AEM.02772-10.
55. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.
56. Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48.
57. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB,

Simpson GL, Solymos P, Henry Stevens H, Wagner H. 2015. Vegan: community ecology package. https://CRAN.R-project.org/package=vegan.
58. Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P. 2014. Phylogeny of the class *Actinobacteria* revisited in the light of complete genomes. The orders '*Frankiales*' and *Micrococcales* should be split into coherent entities: proposal of *Frankiales* ord. nov., *Geodermatophilales* ord. nov., *Acidothermales* ord. nov and *Nakamurellales* ord nov. Int J Syst Evol Microbiol 64:3821–3832. https://doi.org/10.1099/ijs.0.063966-0.
59. Nguyen TV, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, Kalinowski J, Pawlowski K. 2016. An assemblage of *Frankia* cluster II strains from California contains the canonical *nod* genes and also the sulfotransferase gene *nodH*. BMC Genomics 17:796. https://doi.org/10.1186/s12864-016-3140-1.