PLANT EFFECTS ON THE DYNAMICS OF *FRANKIA* POPULATIONS IN SOIL

By

Seifeddine Ben Tekaya, BS., MS

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a Major in Aquatic Resources
May 2018

Committee Members:

Dittmar Hahn, Chair

Robert J. C. McLean

David Rodriguez

Jeffrey O. Dawson

Mark Paschke

COPYRIGHT

by

Seifeddine Ben Tekaya

2018

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

To Mom and Dad: Leila and Fathi. To the memory of a day at the Ministry of Higher Education, where you stood against everyone in your belief I could become a scientist. May you find in this work the expression of my eternal gratitude.

Thank you from the bottom of my heart.

ACKNOWLEDGEMENTS

A dissertation is a challenge that involves learning, patience and perseverance. I do deeply realize that if not for the many people around me, this piece of work would have never been the same. Expressing thanks is a very hard task, as it is impossible to do justice to everyone. First of all, I want to express my deepest gratitude to Dr Dittmar Hahn, Professor and chair, who offered me the opportunity to be part of his lab at a very critical moment of my career. He revived the desire in me to do experiments, and the joy to be a part of an active and strong laboratory team. During my journey in his lab, he made sure to guide me through the right directions, he constantly helped me in every single step of my project, and was always available to answer my questions. Working with Dr Hahn was my most wonderful academic experience. I would like to thank Dr David Rodriguez, for being part of my dissertation committee, and being always helpful on resolving experimental questions and complex data analysis. I am also thankful to Dr. Jeffrey O. Dawson (University of Illinois), for being on my dissertation committee and for providing soil samples for my research work. I want to expand my gratitude to Dr. Mark W. Paschke (Colorado State University) , and Dr Robert J. C. McLean for agreeing to be on my dissertation committee. A special thanks to Trina, who mentored me during my first steps in the lab, and for continuously helping me at every step of my dissertation.

I would like to thank my past and current lab members Abi, Adam, Adriana, Anna, Christophe, Diarra, Elise, Fritzina, Jackie, Jared, Phuong, Rebecca, Spandana, Stephanie and Suvi. I am cordially thankful to the Texas State University, Department of Biology, College of Science, Graduate College for providing financial support for my PhD.

I am deeply grateful to my partner and soulmate Hayfa Chergui, for her continuous support, presence and positive vibes at all moments. I am certainly the most blessed for having her on my side. Last but not least, my sincere word of thanks goes to my sister Abir and brother Fakhri for their love, wishes and prayers.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

Frankia are slow growing actinobacteria that are able to form root nodules with some woody non-leguminous plants. Studies on the ecology of these bacteria are hampered by difficulties to isolate them into pure cultures which was a prerequisite for quantitative analyses in the past. We have therefore focused on the development of molecular approaches that allowed us to retrieve quantitative data from environmental samples unbiased by the limitations of culturability. A first objective of the current study was to develop *q*PCR based methods to distinguish groups within the genus and quantify their populations in soil. Additional attempts were made to distinguish and quantify typical, nitrogen-fixing frankiae from atypical, generally non-nitrogen fixing frankiae. Both SybrGreen- and Taqman-based *q*PCR methods were subsequently evaluated for the quantification of these populations in different soils. Methods evaluation resulted in a restrained diversity within soils vegetated with *Alnus* and *Betula nigra* trees, where usually one group of *Frankia* was dominant, while prairie soils reflected more diverse distribution represented by all four clusters. These methods are then used to study long term effects of agricultural management practices on abundance and diversity of frankiae. Data from these analyses were contrasted with Illumina sequencing data targeting both 16s rRNA and *nif*H gene amplicons. Resulting Illumina data revealed homogenous distribution in soils under agricultural treatments, mainly dominated by members of cluster 1 and 3 *Frankia*, while prairie soils were more diverse showing *Frankia* representatives of all four clusters. *q*PCR analysis on the other hand reflected diverse pattern in the prairie soils with high abundance of members of cluster 2 and cluster 4, while the agricultural treatments exhibited patterns mostly dominated by cluster 1b and low numbers of cluster 2 and 4. Both *q*PCR and Illumina sequencing methods were also applied in analyses of microcosm experiments aiming to investigate the effects of plants species on indigenous populations of *Frankia* and relate abundance/diversity in soils to root nodule populations. Investigation on the root nodules of six host plants revealed

Frankia populations represented by eight different sequences of *nif*H gene fragments. two of these sequences characterized frankiae in *S. argentea* nodules, and three frankiae in *A.*

glutinosa nodules. frankiae in *A. cordata* nodules were represented by five sequences, one of which was also found in nodules *A. glutinosa* and *C. equistifolia*, while another one was detected in nodules from *A. glutinosa*. *A. viridis* and *Hippophaë rhamnoides* did not nodulate. Quantitative PCR assays showed vegetation generally increased the abundance of frankiae in soil, independent of the target gene (i.e. the *nif*H or the 23S rRNA gene). Targeted Illumina sequencing of *Frankia*-specific *nif*H gene fragments detected 24 unique sequences from all rhizosphere soils four of which were also found in nodules, while the remaining four sequences in nodules were not found in soils. Seven of the 24 sequences from soils represented more than 90% of the reads obtained in most samples, with the two most abundant sequences not found in root nodules and only two of these sequences detected in nodules. In the last chapter, both Illumina and quantitative PCR targeting 23S rRNA gene fragments were used to investigate *Frankia* diversity in soils from five different continents. Illumina sequencing resulted in a generally low diversity of *Frankia* populations, with only 18 distinct reads obtained from all soils, and with few sequences identical or closely related to those of cultured relatives. *Frankia* populations in individual soils were generally represented by only one or two abundant reads, with additional reads often very similar. *q*PCR analysis detected representatives of all clusters in soils from Rwanda, Hungary and Japan, while that from Peru harbored cluster 1a, 2 and 4 frankiae and Alaskan soil cluster 1b frankiae only. Meta-analyses including results from bioassays and clone libraries revealed large quantitative, but also qualitative differences, suggesting the presence of methodological biases such as selective nodulation, PCR amplification artifacts, or short reads length biases that could affect taxonomical assignments.

To conclude, this study made it possible to investigate *Frankia* dynamics in soils using different methodological tools, from SybrGreen and Taqman-based *q*PCR that

target all clusters to next generation sequencing, that has been proven successful in assessing nitrogen-fixing *Frankia* diversity in soils. Furthermore, the current study shows that different methodological approaches to analyze *Frankia* diversity in soils qualitatively and quantitatively might retrieve considerably different diversity patterns, impacted by biases and limitations of each of the approaches used. For future purposes,

developing Illumina sequencing methods that target specific rRNA *Frankia* fragments, such as 23S rRNA gene fragments, would offer a more comparable tool to the actual *nif*H gene based protocol. In addition, as digital PCR is emerging as a state of the art quantitative tool, it could offer a more suitable method for *Frankia* quantification generating absolute data rather than the relative numbers delivered by *q*PCR.

CHAPTER I

GENERAL INTRODUCTION

Ever since Robert Koch established his postulates defending the hypothesis of linking an existing disease to a causative microbe, isolating microorganisms into pure culture became a basic pillar in the discipline of microbiology. Since it is widely accepted that more than 99.8% of the Prokaryotes have not yet been obtained in pure culture (Pham & Kim, 2012), finding tools to allow exploration of the microbial kingdom unbiased by the limitations of culturability were essential. Within the last two decades, molecular detection techniques have increasingly been used to analyze microbes in environmental samples, with qualitative techniques such as T-RFLP, DGGE, and TGGE studying diversity, to quantitative techniques such as *in situ* hybridization or quantitative polymerase chain reaction (*q*PCR) used to enumerate specific target organisms in environmental samples. In addition, new sequencing technologies, i.e. Next Generation Sequencing (NGS), allow for the fast retrieval of tremendous numbers of sequences from environmental samples, and thus provide a means to decipher the large diversity of countless microbes in the environment.

Some of these microbes are represented by the genus *Frankia*. Members of the genus *Frankia* are difficult to isolate, and thus molecular tools have been developed to analyze structure and dynamics of populations of this genus in environmental samples such as soils. Members of the genus *Frankia* are generally characterized as nitrogen-fixing symbionts, though atypical strains unable to fix nitrogen have been isolated. Nitrogen is often a limiting nutrient for plant growth and development. Nitrogen-fixation is a process only accomplished by members of the Domain Bacteria and by some Archaea. A selected group of angiosperms, actinorhizal plants, that form root nodules with members of the genus *Frankia,* therefore has growth advantages over other plant species in environments with limited nitrogen availability.

Frankia are slow growing actinobacteria that are able to form root nodules with some woody non-leguminous plants (Dawson, 1986, Benson, 1988, Huss-Danell, 1997, Benson & Dawson, 2007). The relationship between the bacterium and its host plant is

described as a facultative mutualistic relationship, which means that both organisms can live separately but their association involves reciprocal trophic trade (Pozzi *et al.*, 2015). The first successful and confirmed isolation in pure culture occurred in 1978 (Callaham *et al.*, 1978). Prior trials failed mostly because of the particularly slow growth rate of the organism, resulting in competitive exclusion by fast growing contaminants (Normand *et al.*, 2007).

Frankia strains produce three types of cell differentiations: hyphae, vesicles and spores. Depending on both host plant and *Frankia* strain, all of those structures could occur *in planta*. *Frankia*'s hyphal structure, chains of cells that are also referred to as filaments or pseudo-hyphae, are smaller than the conventional hyphal structures encountered among fungi (Callaham *et al*., 1978). Vesicles are the sites of nitrogen fixation, containing nitrogenase enzymes that are produced during nitrogen deprivation conditions. Spores are dormant and resistant forms, which are produced in sporangia that are terminally or intercalary located within the hyphae (Callaham *et al*., 1978).

Frankia strains are considered heterotrophic aerobes, with average generation times of about 15 hours. This slow growth is mainly attributed to the hyphal formation. After transfer, the kinetic growth consists of a stationary phase followed by a short exponential phase (Horriere *et al*., 1983; Newcomb & Wood, 1987). Biomass increase is slow over time and in order to avoid floc formation, or nutrient gradients across mycelia, cultivated strains must be shaken permanently (Benson & Schultz, 1990). Generally, maintenance media are simple, consisting of basal required nutrients and some Calvin cycle derivatives such as propionate and pyruvate. It has been shown, however, that various strains react differently to these nutrients. For example, strain ArI3 grown on propionate induced an activity that is 7 to 26 times the activity with other carbon sources (Baker, 1987; Lechevalier & Lechevalier, 1990). CcI2, on the other hand grew best on a pyruvate added medium (Young & Torrey, 1986). Most strains ideally grow in liquid media after seven to ten days of incubation. However, growth on solid plates is harder because of the inhibitory nature of many types of agar (Benson & Schultz, 1990). Also, conventional components that most bacteria require for growth, such as yeast extract, exercise an inhibitory effect on *Frankia* growth (Lechevalier & Lechevalier, 1990). Because of its slow growth and difficult maintenance requirements, isolation of *Frankia*

from soil is difficult. Serial dilution could not be applied as this will result in overgrowth by faster growing taxa. Thus, direct isolation from disinfected root nodules, a natural locale of enrichment, emerged as the most prominent method of *Frankia* cultivation.

Comparative analyses of ribosomal RNA genes or fragments, nitrogenase encoding genes (e.g. *nif*H) as well as many others were used to classify *Frankia* within the Domain Bacteria and to characterize groups within the genus (Benson & Clawson, 2000). Four major groups, referred to as clusters (Normand *et al.*, 1996), could be distinguished by comparative sequence analyses, but also largely separating *Frankia* strains on the basis of their morphology, cultivability and infectivity (Ghodhbane-Gtari *et al.*, 2010). Each group encompasses strains that share the same host plant specificity though in some cases, specificity could overlap between different groups. In each cluster, strains fall into subgroups identified through DNA-DNA homology (Normand *et al.*, 1996). *Frankia* of cluster 1 can form nodules within the Hamamelidae, order of Fagales that includes the Betulaceae, Myricaceae and Casuarinaceae (Simonet *et al*., 1999). Thus, the very strains that would infect the Casuarinaceae family will form one distinguished subgroup. Another subgroup consists of strains that infect alder species (Vanden Heuvel *et al*., 2004). Again, some of these species could also infect some Myricaceae members with variable effectiveness (Benson & Dawson, 2007). Members of cluster 2 *Frankia* infect host plants of the Coriariaceae, Datiscaceae, Rosaceae and Rhamnaceae. These strains pose more than one mystery because of the difficulties encountered with their isolation into pure culture. Up to date, there is one single report of a successful isolation of a cluster 2 strain (Gtari *et al.*, 2015). The cluster 3 strains successfully establish root nodules with members of the Myricaceae, Rhamnaceae, Elaeagnaceae and *Gymnostoma*, which is a genus belonging to the Casuarinaceae family. Some strains have been reported successfully infective but not effective in members of Betulaceae, non *Gymnostoma* Casuarinaceae, Rosaceae and *Ceanothus* (Benson & Clawson, 2000). Aside from these three clusters, that are representing nitrogen-fixing *Frankia* strains, it is important to mention the existence of a fourth cluster, embracing atypical *Frankia* strains that are generally unable to fix nitrogen and/or unable to infect plants. Some non-nitrogen-fixing members of cluster 4 were isolated from root nodules of *Alnus*, *Coriaria*, *Datisca*, *Ceanothus* and *Purshia* (Kohls *et al*., 1994).

Actinorhizal plants emerged in the late cretaceous and their distribution became ubiquitous as they adapted to a large set of environments (Benson & Dawson, 2007). They produce leaf litter with small C/N ratio and thus amend soil with nitrogen resources (Normand *et al.*, 2007). Alder associated *Frankia* are ubiquitous and their distribution does not seem to be strictly related to the host range (Clawson *et al*., 1998). Cluster 1 *Frankia* members were observed in root nodules of alder trees planted outside their geographic range (Benson & Dawson, 2007). The genus *Alnus* is the only actinorhizal representative of the family Betulaceae, with its subgenus Alnaster as the closest specimen to the ancestral alder form. Most of the *Alnus* species occur in boreal temperate regions of the northern and southern hemispheres (Miki, 1977). The Casuarinaceae family is presumably related to the Betulaceae as it could be infected by cluster 1 *Frankia* strains (Diem & Dommergues, 1990; Simonet *et al*., 1999). However, *Casuarina*-infective strains are not as ubiquitous as those infective of *Myrica* and *Alnus* (Normand *et al*., 2007). It has been established that *Casuarina*-infective *Frankia* occurrence in soil is intimately related to its host (Benson & Dawson, 2007). Thus, its diversity would be generally greater within the native range of the host. Yokohama (Yokoyama *et al.*, 2000) suggested that the *Coriaria* origin might be Eurasia or North America, as the basal members of the genus occur in Asia and Central America. *Coriaria* sp. are infected by cluster 2 *Frankia* members that are also found in *Datisca* and *Ceanothus* root nodules. *Datisca* is considered the only genus of its family Datiscaceae with two species: *Datisca cannabina* and *Datisca glomerata*. The particularity of the genus *Datisca* is that they are the only actinorhizal plants that do not produce a woody stem. Their distribution is disjunct and their anatomy is well adapted to a Mediterranean climate. Thus, they are obviously abundant in the Mediterranean basin. In addition, *Datisca glomerata* is found on the western slope from Northern California to Baja California (Swensen *et al*., 1994). The richness of strains within those two plant families is relatively low. In some studies covering 42 *Coriaria* nodules, screening of 16S rRNA gene sequence variability revealed only two sequences differing by only one nucleotide (Clawson *et al.*, 1997, Benson & Dawson, 2007).

Unlike members of cluster 2, *Frankia* members of cluster 3 infective of the Elaeagnaceae family, have been reported highly diverse. They are referred to as

cosmopolitan as they are continuously found in different actinorhizal families. As an example, some 16S rRNA gene sequences identified from *Elaeagnus angustifolia* and *Myrica pensylvanica* nodules in Connecticut were also detected in nodules from *Discaria trinervis* in Chile and *Elaeagnus pungens* in New Zealand (Nalin *et al.*, 1997). Since all introduced Elaeagnaceae in America, Europe, Australia and New Zealand form nodules, the role of those hosts to insure the perenniality of cluster 3 *Frankia* was accepted. The cosmopolitan status of cluster 3 is well observed when strains that infect native *Shepherdia* and southern hemisphere Rhamnaceae cross infect *Elaeagnus* introduced plants as well (Caru *et al.*, 2003). The Myricaceae family comprises 4 genera that are *Myrica*, *Morella*, *Comptonia* and *Canacomyrica*, most of them are located in Africa, Asia, North and South America (Benson & Dawson, 2007). Strains that colonize root nodules of these genera are mainly cluster 1 *Frankia* members. The cosmopolitan cluster 3 could be observed in some cases with lower frequency than the first. The diversity of cluster 1 is reported to be generally low. For example, in one study, 3 sequences made up 81% of the total 16S rRNA gene fragments analyzed (Clawson *et al*., 1998). Since Myricaceae could be nodulated by both cluster 1 and cluster 3 members, it seems that the environmental distribution of their related *Frankia* symbionts is very wide. The Rhamnaceae family has six nodulating genera (Richardson *et al*., 2000). Five of those genera are located in the southern hemisphere and are nodulated by the cosmopolitan cluster 3 (Lechevalier & Ruan, 1984). These genera are *Colletia, Discaria, Kentrothamnus, Retanilla* and *Trevoa*. In the northern hemisphere, the genus *Ceanothus* is the most abundant representative of the family. The major group of *Frankia* that infects it are those of cluster 2, however, members of cluster 3 were also found according to some studies (Lechevalier & Ruan, 1984, Vanden Heuvel *et al.*, 2004). The family Rosaceae is distributed worldwide (Evans *et al*., 2000), and holds four actinorhizal genera that could be nodulated with only cluster 2 members (Benson *et al*., 1996).

Frankia is a slow growing bacterium that cannot be isolated reliably from root nodules or from soil, and thus quantitative studies relying on isolation cannot be performed. Thus, in order to study the ecology of *Frankia*, methods were developed that target *Frankia* unbiased by the limitations of culturability (Samant *et al*., 2012; Samant *et al*., 2014). These methods comprise molecular tools that allowed researchers to draw

broader conclusions from gathered data on root nodule and soil *Frankia* as they target one group versus the community as a whole, not a separate cultivar. It is important to mention that most of the studies were targeting *Frankia* populations inside root nodules, a natural locale of enrichment, and thus data on *Frankia* in soils were scarce. In this context, our laboratory is considered a pioneer when it comes to *Frankia* detection in soil. Hahn et al. (Hahn *et al.*, 1993) published their first report on probe design for *in situ* hybridization of *Frankia* rRNA in soil. A SybrGreen based quantitative PCR approach was established and has been proven efficient in detection and quantification of *Frankia* cells in various soil samples (Samant *et al.*, 2012, Samant *et al.*, 2014). The principle consisted of designing primers on the basis of *Frankia* specific conserved regions within 23S rRNA gene fragments as well as the *nif*H nitrogenase encoding gene. Additional primers were also successfully designed to detect specific sub-groups allowing intraspecific analysis of each cluster. These newly designed methods have been proven efficient and allowed examination of *Frankia* directly in its host rhizosphere, and draw interpretations concerning host versus edaphic factors triggering population dynamics in soil (Samant *et al.*, 2015, Samant *et al.*, 2016). It is important to mention, that these established methods fit only clusters 1 and 3.

Objective of the study

Specific aim 1

A prominent aim of this dissertation was to develop *q*PCR based methods to detect cluster 2 and cluster 4 *Frankia* in soil, as a logical continuation of research already done in our lab that successfully established *q*PCR quantification methods for cluster 1 and cluster 3. This goal required the amendment of our current database with 23S rRNA gene sequences of isolates representing cluster 4, and PCR products from uncultured frankiae in root nodules of cluster 2 plant species (e.g. *Datisca*, *Coriaria*, *Ceanothus*). Specific primers were designed for SybrGreen-based detection, and later amended with probes for Taqman-based quantification. Taqman-based quantification was used for the detection of the combination of clusters 1, 2 and 3, together representing the nitrogen-fixing members of the genus *Frankia.* Both SybrGreen- and Taqman-based *q*PCR

methods were subsequently evaluated for the detection on the genus and cluster level, as well as to distinguish typical nitrogen-fixing from atypical, generally non-nitrogen fixing frankiae.

Specific aim 2

A second aim was to assess long-term effects of cultivation (Morrow Plots) on abundance and diversity of the genus *Frankia* and of specific subgroups within the genus using *q*PCR methods developed under specific aim 1. These data were contrasted with Illumina sequencing data for *Frankia* of clusters 1 and 3 using previously developed methods with *nif*H gene fragments as a target. The results were compared to diversity studies using sequence data of 16S rRNA genes generated with universal primers for Bacteria in soils from the Morrow Plots and the original prairie.

Specific aim 3

We were also interested to assess effects of different host plant species (here *Alnus glutinosa*, *A. cordata*, *A. viridis*, and *Shepherdia argentea*, *Hippophaë rhamnoides*, *Casuarina equisetifolia* and *Betula pendula* for comparison) on soil populations of *Frankia* and to relate abundance/diversity to root nodule populations. Plants were grown in microcosms on natural soil for 7 months. DNA was extracted from rhizosphere soil, and was analyzed by *q*PCR and *nif*H gene targeted Illumina sequencing. Nodules were only obtained on *A. glutinosa*, *A. cordata* and *S. argentea*, and on *C. equisetifolia* (just one), but not on *A. viridis*, *H. rhamnoides*, or *B. pendula*. *Nif*H gene sequence analyses of uncultured endophytes in root nodules was used to identify major populations of endophytes in root nodules of host plants, and compared to *nif*H gene sequence diversity in soils obtained by Illumina sequencing.

Specific aim 4

Previous research in our lab contrasted *Frankia* diversity in soils from five different continents, analyzed by comparative sequence analyses of *nif*H gene fragments in root nodules obtained in bioassays with promiscuous host plants of the genus *Morella*,

and *nif*H gene clone libraries from the same soils. The objective of work presented in this chapter was to contrast data on *Frankia* diversity from those studies with diversity analyses using *nif*H gene targeted Illumina sequencing and quantitative PCR approaches on these same soils.

References

Baker DD (1987) Relationships among pure cultured strains of *Frankia* based on host specificity. *Physiol Plant* 70: 245-248.

Benson DR (1988) The genus *Frankia*: actinomycete symbionts of plants. *Microbiol Sci* 5: 9-12.

Benson DR & Schultz NA (1990) Physiology and biochemistry of *Frankia* in culture. *The biology of Frankia and actinorhizal plants*, (Schwintzer CR & Tjepkema JD, eds.), pp. 107-127. Academic Press, Inc., San Diego, USA.

Benson DR, Stephens DW, Clawson ML & Silvester WB (1996) Amplification of 16S rRNA genes from *Frankia* strains in root nodules of *Ceanothus griseus*, *Coriaria arborea*, *Coriaria plumosa*, *Discaria toumatou*, and *Purshia tridentata*. *Appl Environ Microbiol* 62: 2904-2909.

Benson DR & Clawson ML (2000) Evolution of the actinorhizal plant symbioses. In: Triplett EW (ed.) Prokaryotic nitrogen fixation: A model system for analysis of biological process. Horizon Scientific Press, Wymondham, UK pp. 207-224.

Benson DR & Dawson J (2007) Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. *Physiol Plant* 130: 318-330.

Callaham D, Del Tredici P & Torrey JG (1978) Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia*. *Science* 199: 899-902.

Caru M, Mosquera G, Bravo L, Guevara R, Sepulveda D & Cabello A (2003) Infectivity and effectivity of *Frankia* strains from the Rhamnaceae family on different actinorhizal plants. *Plant Soil* 251: 219-225.

Clawson ML, Benson DR, Resch SC, Stephens DW & Silvester WB (1997) Typical *Frankia* infect actinorhizal plants exotic to New Zealand. *New Zealand J Bot* 35: 361-367.

Clawson ML, Caru` M & Benson DR (1998) Diversity of *Frankia* strains of root nodules of plants from the families Elaeagnaceae and Rhamnaceae. *Appl Environ Microbiol* 64: 3539–3543

Dawson JO (1986) Actinorhizal plants: their use in forestry and agriculture. *Outlook on Agriculture* 15: 202-208.

Diem HG, Dommergues YR (1990) Current and potential uses and management of Casuarinaceae in the tropics and subtropics. In: Schwintzer CR & Tjepkema JD (eds). The biology of *Frankia* and Actinorhizal Plants. Academic Press, New York, pp 317–342.

Evans RC, Alice LA, Campbell CS, Kellogg EA & Dickinson TA (2000) The granule-bound starch synthase (GBSSI) gene in Rosaceae: multiple loci and phylogenetic utility. *Mol Phylogenet Evol* 17: 388–400.

Ghodhbane-Gtari F, Nouioui I, Chair M, Boudabous A & Gtari M (2010) 16S-23S rRNA intergenic spacer region variability in the genus *Frankia*. *Microb Ecol* 60: 487-495.

Gtari M, Ghodhbane-Gtari F, Nouioui I, *et al.* (2015) Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. *Sci Rep* 5: 13112; doi: 10.1038/srep13112.

Hahn D, Amann RI & Zeyer J (1993) Whole-cell hybridization of *Frankia* strains with fluorescence- or digoxigenin-labeled, 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* 59: 1709-1716.

Horriere F, Lechevalier MP & Lechavalier HA (1983) *In vitro* morphogenesis and ultrastructure of a *Frankia* sp. ArI3 (Actinomycetales) from *Alnus rubra* and a morphologically similar isolate (AirI2) from *Alnus incana* subsp. *rugosa*. *Can J Bot* 61: 2843-2854.

Huss-Danell K (1997) Actinorhizal symbioses and their N-2 fixation. *New Phytol* 136: 375-405.

Kohls SJ, Thimmapuram J, Bushena CA, Paschke MW & Dawson JO (1994) Nodulation patterns of actinorhizal plants in the family Rosaceae. *Plant Soil* 162: 229–239

Lechevalier MP & Ruan JS (1984) Physiology and chemical diversity of *Frankia* spp. isolated from nodules of *Comptonia peregrina* (L.) Coult. and *Ceanothus americanus* L. *Plant Soil* 78: 15-22.

Lechevalier MP & Lechevalier HA (1990) Systematics, isolation, and culture of *Frankia*. In: Schwintzer CR & Tjepkema JD (eds), The biology of *Frankia* and actinorhizal plants, Academic Press, New York, pp. 35-60.

Miki A (1977) Late Cretaceous pollen and spore floras of northern Japan: composition and interpretation. *J Fac Sci Hokkaido Univ Ser IV* 17: 399–436

Nalin R, Normand P & Domenach AM (1997) Distribution and N-2-fixing activity of *Frankia* strains in relation to soil depth. *Physiol Plant* 99: 732-738.

Newcomb W & Wood S.M (1987) Morphogenesis and fine structure of *Frankia* (Actinomycetales): The microsymbiont of nitrogen-fixing actinorhizal root nodules. *Int Rev Cytol* 109: 1-88.

Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L & Misra AK (1996) Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. *Intl J System Bacteriol* 46: 1-9.

Normand P, Lapierre P, Tisa LS, *et al.* (2007) Genome characteristics of facultatively symbiotic *Frankia* sp strains reflect host range and host plant biogeography. *Genome Res* 17: 7-15.

Pham VHT & Kim J (2012) Cultivation of unculturable soil bacteria. *Trends Biotechnol* 30: 475-484.

Pozzi AC, Bautista-Guerrero HH, Nouioui I, Cotin-Galvan L, Pepin R, Fournier P, Menu F, Fernandez MP & Herrera-Belaroussi A (2015) In-planta sporulation phenotype: a major life history trait to understand the evolution of *Alnus*-infective *Frankia* strains. *Environ Microbiol* 17: 3125-3138.

Richardson JE, Fay MF, Cronk QCB, Bowman D & Chase M (2000) A phylogenetic analysis of Rhamnaceae using rbcL and trnL-F plastid DNA sequences. *Am J Bot* 87: 1309–1324

Samant S, Amann RI & Hahn D (2014) Evaluation of the 23S rRNA gene as target for *q*PCR based quantification of *Frankia* in soils. *System Appl Microbiol* 37: 229-234.

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.

Samant S, Sha Q, Iyer A, Dhabekar P & Hahn D (2012) Quantification of *Frankia* in soils using SybrGreen based *q*PCR. *System Appl Microbiol* 35: 191-197.

Samant SS, Dawson JO & Hahn D (2015) Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. *System Appl Microbiol* 38: 501-505.

Simonet P, Navarro E, Rouvier C, Reddell P, Zimpfer J, Dommergues Y, Bardin R, Combarro P, Hamelin J, Domenach A, Gourbie `re F, Prin Y, Dawson JO & Normand P (1999) Co-evolution between *Frankia* populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal. *Environ Microbiol* 1: 525–533

Swensen SM, Mullin BC & Chase MW (1994) Phylogenetic affinities of the Datiscaceae based on an analysis of nucleotide sequences from the plastid rbcL gene. *System Bot* 19: 157–168

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.

Yokoyama J, Suzuki M, Iwatsuki K & Hasebe M (2000) Molecular phylogeny of *Coriaria*, with special emphasis on the disjunct distribution. *Mol Phylogenet Evol* 14: 11-19.

Young P & Torrey JG (1986) A comparison of carbon source utilization for growth and nitrogenase activity in two *Frankia* isolates. *Can J Microbiol* 32: 353-358.

CHAPTER II

SYBRGREEN- AND TAQMAN-BASED QUANTITATIVE PCR APPROACHES ALLOW ASSESSMENT OF THE ABUNDANCE AND RELATIVE DISTRIBUTION OF *FRANKIA* CLUSTERS IN SOILS

published as

Ben Tekaya, S., Ganesan, A.S., Guerra, T., Dawson, J.O., Forstner, M.R.J., Hahn, D. 2017. SybrGreen and Taqman-based *q*PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. Appl. Environ. Microbiol. **83**:e02833-16. https://doi.org/10.1128/AEM.02833-16.

ABSTRACT

The nodule-forming actinobacterial genus *Frankia* can generally be divided into 4 taxonomic clusters, with clusters 1, 2, and 3 representing nitrogen-fixing strains of different host infection groups and cluster 4 representing atypical, generally non-nitrogen-fixing strains. Recently, quantitative PCR (*q*PCR)-based quantification methods have been developed for frankiae of clusters 1 and 3; however, similar approaches for clusters 2 and 4 were missing. We amended a database of partial 23S rRNA gene sequences of *Frankia* strains belonging to clusters 1 and 3 with sequences of frankiae representing clusters 2 and 4. The alignment allowed us to design primers and probes for the specific detection and quantification of these *Frankia* clusters by either SybrGreen- or TaqMan-based *q*PCR. Analyses of frankiae in different soils, all obtained from the same region in Illinois, USA, provided similar results, independent of the *q*PCR method applied, with abundance estimates of 10×10^5 to 15×10^5 cells (g soil) $^{-1}$ depending on the soil. Diversity was higher in prairie soils (native, restored, and cultivated), with frankiae of all 4 clusters detected and those of cluster 4 dominating, while diversity in soils under *Alnus glutinosa*, a host plant for cluster 1 frankiae, or *Betula nigra*, a related non-host plant, was restricted to cluster 1 and 3 frankiae and generally members of subgroup 1b were dominating. These results indicate that vegetation affects the basic

composition of frankiae in soils, with higher diversity in prairie soils compared to much more restricted diversity under some host and non-host trees.

KEYWORDS: alder, birch, *q*PCR, quantification, saprotrophic, soil

INTRODUCTION

Frankiae are slow-growing actinobacteria that are able to form root nodules with some woody non-leguminous plants (1–3). Root nodule formation is host plant specific and largely, but not exclusively, correlates with assignments of strains to specific clusters derived from comparative analyses of 16S rRNA gene sequences (4). Cluster 1 represents frankiae that form nodules on plants from the genera *Alnus, Morella, Myrica,* and *Comptonia* and includes a subgroup infecting the genera *Casuarina* and *Allocasuarina* (5, 6). Members of cluster 2 represent frankiae nodulating *Dryas, Purshia, Chamaebatia, Cercocarpus, Ceanothus, Datisca,* and *Coriaria,* while members of cluster 3 form nodules on plants that include the genera *Elaeagnus, Hippophaë, Shepherdia, Myrica, Morella, Gymnostoma, Discaria, Trevoa, Retanilla, Kentrothamnus,* and *Colletia* (5, 6). In addition to the typical nitrogen-fixing frankiae, atypical, generally non-nitrogen-fixing and/or non-nodulating frankiae have been identified in cluster 4 (4). Since the first report of an isolation of *Frankia* from root nodules in 1978 (7), a large number of isolates has been obtained for clusters 1 and 3 (8–11) and a few for cluster 4(12, 13). An isolate representing cluster 2, however, has been obtained only recently (14). Consequently, most studies on *Frankia* have been performed on those from clusters 1 and 3 and far less on frankiae of clusters 2 and 4. Information on the ecology of cluster 2 and 4 frankiae therefore is quite limited.

We have recently developed SybrGreen-based quantitative PCR (*q*PCR) methods that used *nifH* or 23S rRNA genes as a target to quantify uncultured *Frankia* populations in different soils (15–17). *nif*H as a target only detected the combination of members of clusters 1 and 3 but not those of clusters 2 and 4, while 23S rRNA genes as targets covered all frankiae on the genus level, i.e., clusters 1, 2, 3, and 4 together. Targeting the 23S rRNA gene also allowed us to distinguish between cluster 1 and 3 frankiae and

subgroups within cluster 1 (i.e., clusters 1a, 1b, and 1c) (15). The sum of individual detections generally equaled those on the genus level with both *nif*H and 23S rRNA genes as targets, indicating that members of clusters 2 and 4 were not present at all or not in detectable numbers in the soils analyzed (15, 17, 18). However, this statement is highly speculative, since direct proof of the presence or absence of cluster 2 and 4 frankiae in these soils has not been provided due to the lack of adequate detection and quantification procedures.

In this study, we have amended our database of sequences of 23S rRNA gene fragments of *Frankia* strains representing clusters 1 and 3 (15) to include sequences of frankiae assigned to clusters 2 and 4.Sequence alignments were used to design primers for the specific detection and quantification of *Frankia* clusters by SybrGreen-based *q*PCR and subsequently for the design of probes differentiating members of clusters 1, 2, and 3 (i.e., presumably all nitrogen-fixing frankiae) from those of cluster 4 (i.e., generally non-nitrogen-fixing frankiae). These probes were then used in combination with the specific primer combinations in TaqMan-based *q*PCR to quantify nitrogen-fixing (i.e., cluster 1, 2, and 3) and non-nitrogen-fixing (i.e., cluster 4) frankiae, as well as frankiae of the individual clusters and subgroups within cluster 1 in soils from different locations.

RESULTS

Comparative sequence analysis

Phylogenetic analysis of trimmed and aligned sequences provided a topology with four major clades that represented the previous assignment of clusters 1 to 4 (Fig. 1). Sequences of the uncultured endophytes from *Datisca, Coriaria,* and *Ceanothus* clustered with sequences of cluster 2 frankiae retrieved from the database, i.e., the endophyte from *Datisca glomerata* (Dg1) and pure-culture BMG5.1, representing one clade as cluster 2, while those of the atypical strains assembled together with cluster 4 strain EuI1c in another clade as cluster 4 (Fig. 1). While cluster 2 frankiae resembled a concise clade with uncorrected p distance values of aligned sequences between 94 and 100%, cluster 4 frankiae were more diverse, with 3 concise subgroups and distance values of aligned sequences between subgroups of 73 to 82% (Fig. 1).

Primer and probe design and evaluation

The amended database of partial 23S rRNA gene sequences of *Frankia* strains was used to design two forward primers, 23Dat1578f, targeting cluster 2 frankiae, and 23NNF1561f, targeting cluster 4 strains. Primer 23Dat1578f was specific with no mismatches to sequences of all cluster 2 frankiae and 3 to 7 mismatches to those of other *Frankia* strains in our database. TestPrime1.0 and TestProbe 3.0 analyses using the SILVA database revealed that primer 23Dat1578f (*Escherichia coli* position 1415) was specific for the two target organisms in the database, i.e., the *Frankia* endophyte from *Datisca glomerata* (Dg1; GenBank accession number CP002801) and strain BMG5.1 (JWIO01000013), and displayed at least 3 mismatches to sequences of non-target frankiae (i.e., 3 mismatches for representatives of clusters 1c, CcI3, BMG5.23 and Thr, and 4 mismatches for cluster 1a and 3 *Frankia* strains), while other organisms (*Prochlorococcus* sp.) displayed at least 5 mismatches. Together with primer 23Fra1769r, specific detection was achieved for the target organisms, i.e., cluster 2 frankiae, represented by the endophyte from *Datisca glomerata* (Dg1) in the database, with at least 4 mismatches to non-target bacteria. Primer 23NNF1561f (*E. coli* position 1415) could not be designed to cover all sequences of strains within cluster 4. It was specific for the subgroup including strains CN4, AgB1.9, AgW1.1, EuI1c, DC12, and CaI1, with 2 mismatches to strains CN3, CN6, CN7, CNm, CNm3, CNm7, 32-61, and PtI4 and 5 mismatches to AgB1.5, AgB1.7, and AgB1.10. TestPrime 1.0 and TestProbe 3.0 analyses of primer 23NNF1561f in the SILVA database only retrieved the sequence of strain EuI1c (CP002299) with no mismatches and confirmed 2 mismatches to that of strain CN3. For the primer target position on the 23S rRNA gene, identified as position 1415 on the reference gene of *E. coli*, non-target sequences remained undetected at the program search limit of 5 mismatches. However, sequences with 3 and 4 mismatches to the sequences of non-target organisms (e.g., *E. coli* and *Francisella tularensis*) were retrieved at a different position, i.e., *E. coli* position 419.

Probe NF1715f was designed to target all *Frankia* strains of clusters 1, 2, and 3 in our database; however, strains of cluster 3 (e.g., EUN1f, BMG5.12, and EAN1pec) displayed one mismatch. TestProbe analyses confirmed these data, with non-target

organisms (e.g., *Streptomyces* sp.) having at least 2 mismatches to probe NF1715f (*E. coli* position 1453). Probe NNF1715f (*E. coli* position 1462), targeting cluster 4 frankiae, was identical to sequences of cluster 4 strains detected by 23NNF1561f; however, they displayed 1 to 3 mismatches to strains of the remaining subgroups. TestProbe analyses retrieved only sequences from strain EuI1c without mismatches, while those of strain CN3 displayed 2 mismatches, while other non-target organisms, such as *Actinoplanes* and *Streptomyces* sp., displayed 3 mismatches.

Annealing temperatures for all new primer combinations were established in the same range as those of our previously designed primers, 62 to 68°C (Table 1). In order to enhance coverage of primer 23NNF1561f to include members of the subgroup depicting 2 mismatches, annealing temperatures of 62°C were used instead of 66°C. Comparative analyses of amplifications using genus-specific detection with primers 1655f/1769r and specific detection with 23NNF1561f/1769r using pure cultures of AgB1.9, CN3, and AgB1.10, representing the three subgroups within cluster 4 and representative frankiae of the remaining clusters, resulted in complete detection of strains AgB1.9 and CN3, while strain AgB1.10 and other frankiae were not detected by the specific primer combination. The use of TaqMan-based *q*PCR using genus-specific primers 1655f/1769r and probes NF1715f and NNF1715f allowed us to circumvent coverage problems for cluster 4, since strains AgB1.9, CN3, and AgB1.10, representing the three subgroups, were detected quantitatively, as were strains representing nitrogen-fixing frankiae of the remaining subgroups.

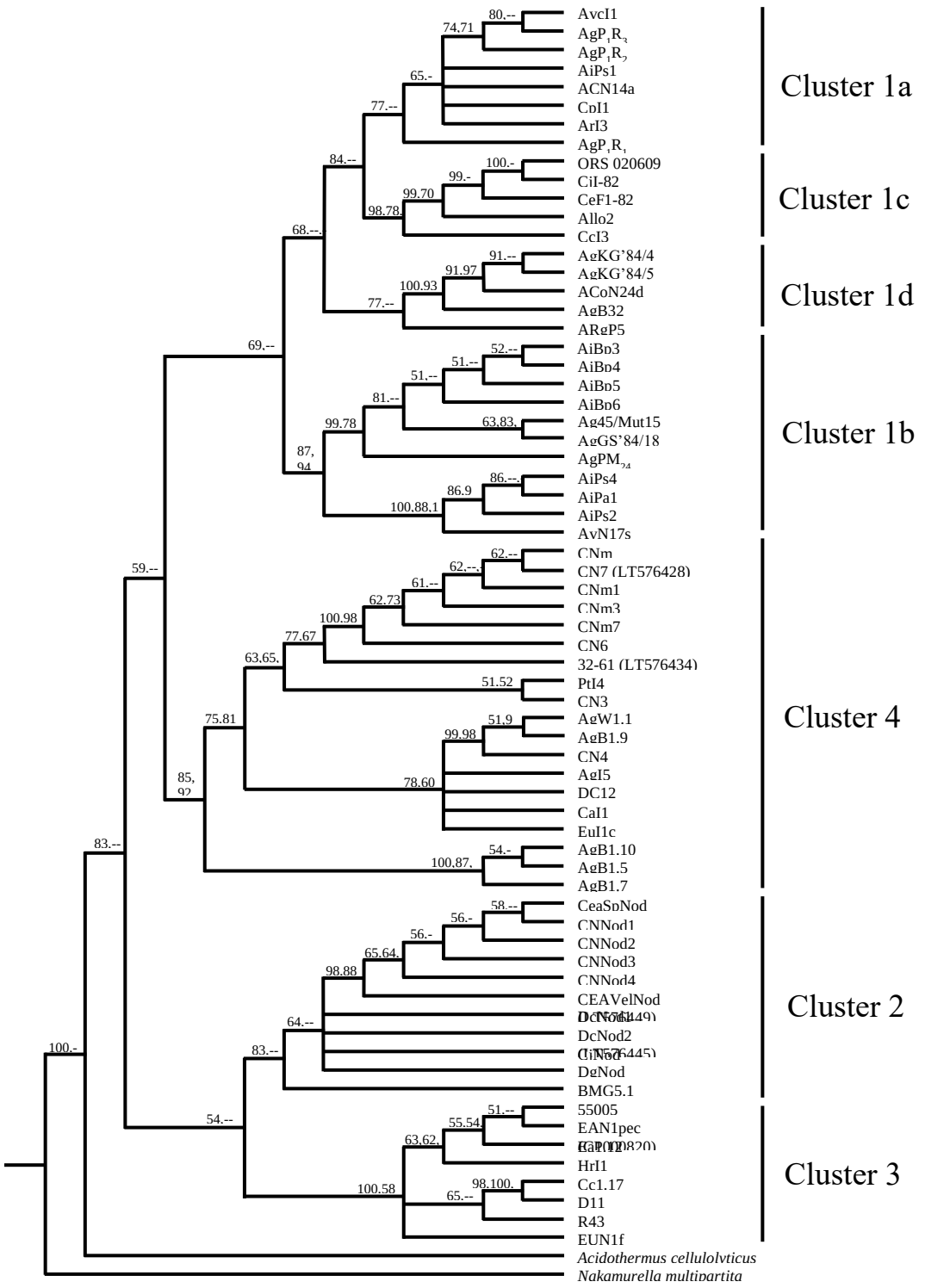

FIG 1 Sequence relationships for selected *Frankia* strains provided to show probe utility. Numbers above the branches represent the bootstrap values from a neighbor-joining (NJ) bootstrap analysis (10,000 replicates) using the HKY85 correction, followed by maximum likelihood bootstrap (1,000 replicates) and Bayesian analysis values, respectively, noted for clades with greater than 50% bootstrap support. All results are plotted on the NJ bootstrap topology. Both the ML and Bayes analyses resolved generally similar topologies, although there are fewer supported tip nodes in the latter two analyses and the larger clades'

basal relationships are either not resolved (ML) or have an alternative arrangement of the main clades (Bayes basal arrangement not depicted).

Method evaluation

Amplicons of 23S rRNA genes of strains ArI3, Ag45/Mut15, CcI3, EAN1pec, AgB1.9, and uncultured frankiae from root nodules of *Coriaria* were specifically detected with primer combinations targeting the respective clusters (Fig. 2). Quantification of individual amplicons resulted in values comparable to those in mixtures of amplicons, independent of primers (i.e., genus or cluster specific) and detection procedure (i.e., SybrGreen- or TaqMan-based *q*PCR) (Fig. 2). One-way analyses of variance (ANOVAs) did not detect statistically significant differences for cluster 1a (P = 0.07), 1b (P = 0.3), 1c (P = 0.9), 2 (P = 0.1), 3 (P = 0.9), and 4 (P = 0.9) across treatments. The sum of these clusters was also not statistically significantly different from genus- or group-specific detections (P = 0.1).

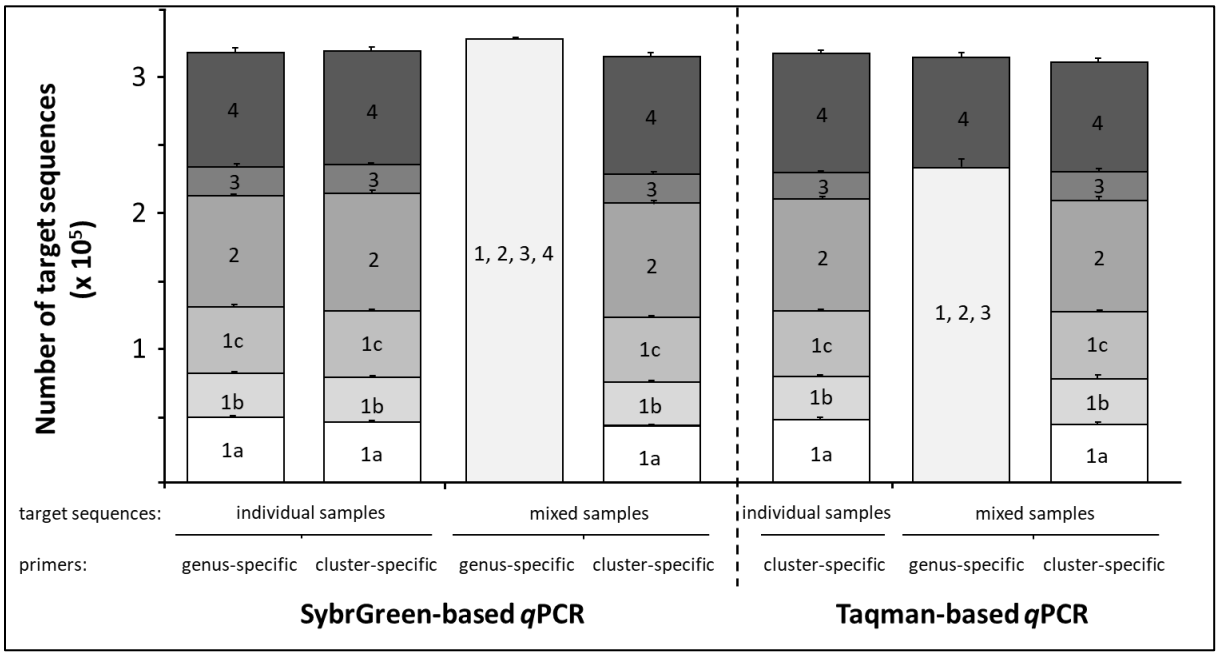

FIG 2 SybrGreen- and TaqMan-based *q*PCR quantification of amplicons from representative pure cultures or endophytes of *Frankia* clusters 1, 2, 3, and 4 and of subgroups a, b, and c within cluster 1. Individual amplicons or mixtures of amplicons were quantified using either genus- or (sub)cluster-specific primer combinations in SybrGreen-based analyses or additional probes targeting either cluster 1, 2, and 3 frankiae or cluster 4 frankiae.

SybrGreen-based quantification using the *nif*H gene as the target to quantify frankiae of clusters 1 and 3 resulted in abundance estimates of 10 x 105 to 15 x 105 cells (g soil)$^{-1}$ depending on the soil (Fig. 3). Except for soil BAHF (P = 0.002), these estimates were not significantly different from those from 23S rRNA gene-based based detection targeting frankiae of all four clusters, or the sum of abundances of specific clusters and subgroups (p values between 0.07 and 0.9). The latter analyses revealed the presence of cluster 1 frankiae, and here especially subgroup 1b, represented by *Frankia* strain Ag45/Mut15, only, or in combination with small numbers of frankiae of cluster 3 (Fig. 2). Frankiae of cluster 2 and 4, as well as cluster 1 frankiae represented by strain ArI3 (subgroup 1a) and CcI3 (subgroup 1c) were not detected. Abundance and diversity data obtained by SybrGreen-based analyses were confirmed in all four soils by Taqman-based analyses of the individual clusters and subgroups, as well as by targeting all nitrogen-fixing frankiae, i.e. clusters 1, 2, and 3 (Fig. 3).

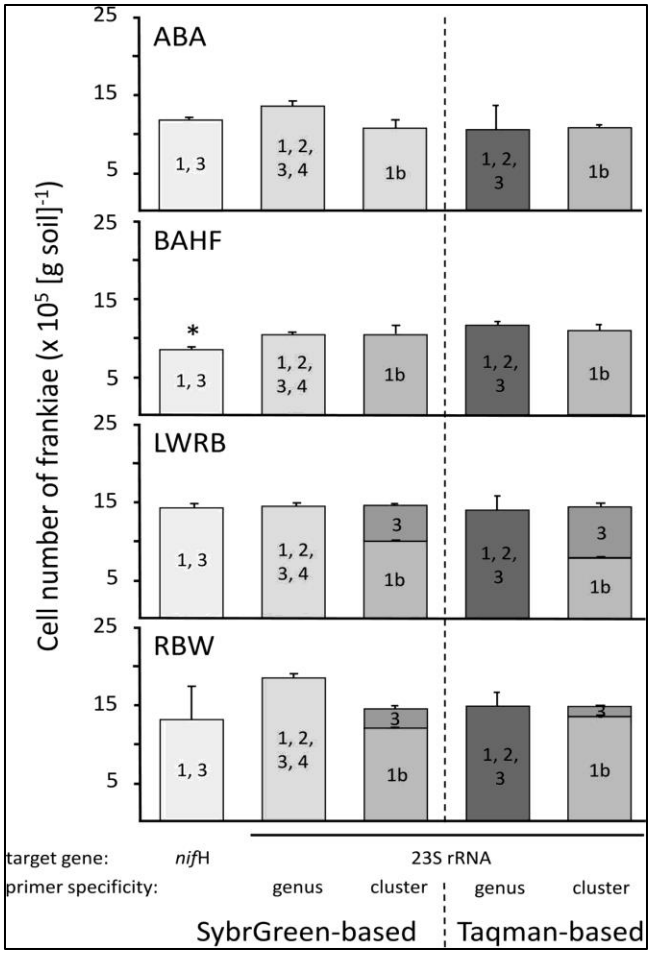

FIG 3 SybrGreen- and TaqMan-based *q*PCR quantification of *Frankia* clusters 1, 2, 3, and 4 in soils vegetated with host trees (*Alnus glutinosa*; soils ABA and BAHF) or non-host trees (*Betula nigra*; soils LWRB and RBW). Quantification (from left to right) used *nif*H gene fragments as a target detecting clusters 1 and 3 or 23S rRNA gene fragments generated with primer combinations detecting the genus *Frankia*, i.e., all clusters 1, 2, 3, and 4 or primer combinations specific for clusters 1a/d, 1b, 1c, 2, 3, and 4 (presented as the sum of the individual clusters and subgroups detected). Only frankiae of subgroups 1b and cluster 3 were detected in these soils, while the remaining clusters and subgroups remained undetected (all using SybrGreen). TaqMan-based detection of 23S rRNA fragments resulted in compositions of frankiae with respect to clusters and subgroups that were similar to those of SybrGreen-based detection, with their sum representing quantification values similar to those of frankiae detected with genus-specific primers and probe NF1715f targeting frankiae of clusters 1, 2, and 3. Statistically significantly different values between treatments of the same sample are highlighted with an asterisk.

For prairie soils, SybrGreen-based detection using *nif*H as target resulted in significantly lower abundance estimates (i.e. 5 to 10 x 10^5 cells [g soil]$^{-1}$) than SybrGreen-based detection of all clusters in the genus (i.e., 15 x 10^5 to 25 x 10^5 cells [g soil]$^{-1}$)($P < 0.001$) (Fig. 4). Frankiae of clusters 1, 2, 3, and 4 were detected in two of the three soils, with the sum of their abundance estimates matching the estimates on the genus level (Fig. 4). TaqMan-based *q*PCR resulted in similar composition of the subgroups and detection of nitrogen-fixing frankiae equal to the sum of subgroups in clusters 1, 2, and 3 (Fig. 4). Subgroup 1a, represented by strain ArI3, and subgroup 1c, represented by strain CcI3, were not detected at all and thus were not found to be abundant

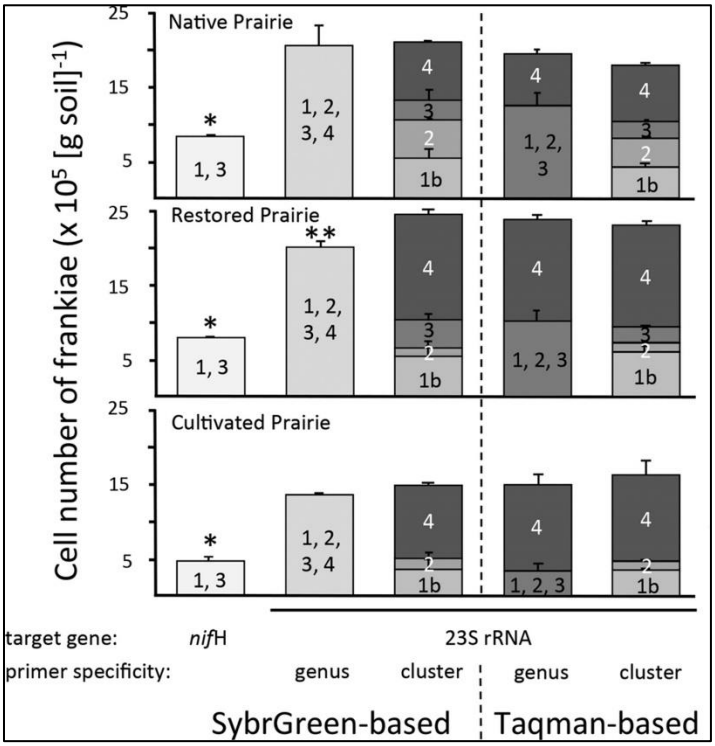

FIG 4 SybrGreen- and TaqMan-based qPCR quantification of Frankia clusters 1, 2, 3, and 4 in prairie soils (native, restored, and cultivated). Quantification (from left to right) used nifH gene fragments as a target detecting clusters 1 and 3 or 23S rRNA gene fragments generated with primer combinations detecting the genus Frankia, i.e., all clusters 1, 2, 3, and 4 or primer combinations specific for clusters 1a/d, 1b, 1c, 2, 3, and 4 (presented as the sum of the individual clusters and subgroups detected). Frankiae of subgroup 1b and clusters 2, 3, and 4 were generally detected in these soils, while subgroups 1a/d and 1c remained undetected (all using Sybr Green). Statistically significantly different values between treatments of the same sample are highlighted with one or two asterisks.

in any of the soils tested. Despite incomplete coverage of cluster 4 *Frankia* strains by both forward primer 23NNF1561f and probe NNF1715f, comparable abundance estimates for cluster 4 frankiae in these soils were obtained by both SybrGreen and TaqMan-based *q*PCR (P values between 0.05 and 0.5) (Fig. 4).

DISCUSSION

Despite the short length of the sequences and thus low or no bootstrap support for major clades, the topology of our phylogenetic analyses confirmed cluster assignments of *Frankia* strains and uncultured endophytes from root nodules derived from comparative analyses of 16S rRNA gene sequences (4). Cluster assignments were also consistent with previous analyses using comparative sequence analyses of the insertion in domain III of the 23S rRNA gene (19), partial gyrase B (*gyr*B), nitrogenase reductase (*nif*H), or glutamine synthetase II (*gln*II) sequence analyses (20), as well as sequence analysis of the 16S-23S rRNA internally transcribed spacer (ITS) (21) or protein mass fingerprints from whole cells (22). While our analyses of uncultured endophytes in root nodules of *Datisca cannabina*, *Coriaria nepalensis*, *C. japonica*, and *Ceanothus* sp. revealed a limited sequence diversity and thus resulted in the assembly of a concise cluster 2, sequences of cluster 4 frankiae were more diverse, with significant sequence diversity establishing 3 subgroups.

Subgroups were previously described for clusters 1 and 3 (23, 24), with those of cluster 1 depicting remarkable physiological differences between each other (25). Physiological characteristics of subgroups of cluster 3 remain to be studied. Within cluster 1, at least three subgroups were established (15), i.e., subgroup 1a, represented by *Frankia* strain ArI3, subgroup 1b, represented by *Frankia* strain Ag45/Mut15 that, in contrast to frankiae of subgroup 1a, was able to grow with leaf litter as the nutrient resource (25), and subgroup 1c, the *Casuarina*-infective strains that require the presence of host plants for growth (26). Our phylogenetic analysis retrieved an additional subgroup 1d (Fig. 1) that, however, could not be distinguished from subgroup 1a with the primer combinations used in our study (Table 1). Using sequences of the insertion in domain III of the 23S rRNA gene as the target for specific detection and quantification of these

subgroups, subgroup 1b was found to be most prominent in many soils from temperate regions (15, 17, 18).

The lack of sequence diversity in our cluster 2 frankiae might be a function of our limited sampling strategy, i.e., the focus on uncultured *Frankia* endophytes in root nodules of a few host plants, i.e., *Datisca*, *Coriaria*, and *Ceanothus* sp. from locations in the United States and in Pakistan only. While previous studies on the diversity of cluster 2 frankiae in root nodules of different host plants also indicated low sequence diversity of these endophytes (27, 28), future studies on cluster 2 frankiae should include a larger diversity of host plants and more locations to retrieve additional information on *Frankia* diversity. Sequences of cluster 4 frankiae were more diverse, with 3 subgroups delineated. Since isolates are available for each of these subgroups, additional studies on physiological properties could provide information on a potential linkage between phenotypic and genotypic characteristics. Ultimately, however, additional assessments on overall diversity of the genus *Frankia* using next-generation sequencing methods in soil samples with and without host plants for all clusters will be required for a more comprehensive analysis of Frankia diversity.

Forward primers were designed identical to sequences of all cluster 2 frankiae (23Dat1578f) and those of one subgroup of cluster 4, represented by *Frankia* strain EuI1c (23NNF1561f). The latter subgroup was selected over the remaining 2 subgroups because isolates had been obtained from a variety of different host plant species, i.e., *Coriaria nepalensis*, *Alnus glutinosa*, *Elaeagnus umbellata*, *Datisca cannabina*, and *Ceanothus americanus* (CN4, AgB1.9, AgW1.1, EuI1c, DC12, and CaI1) (12, 13, 29–31). The reduction in annealing temperature from 66 to 62°C allowed us to increase coverage to two of the three subgroups detected without losing specificity for other *Frankia* strains. TaqMan-based quantification using genus-specific primers and a probe targeting cluster 4 frankiae resulted in the detection and reliable quantification of representative strains of these subgroups and thus might provide an adequate alternative to detection using specific forward primers. The new forward primers 23Dat1578f and 23NNF1561f, as well as the design of probes targeting nitrogen-fixing frankiae (clusters 1, 2, and 3) and cluster 4 frankiae (atypical, generally non-nitrogen-fixing strains), now allows us to expand *q*PCR analyses of frankiae in soils from clusters 1 and 3 and subgroups within

cluster 1 to include clusters 2 and 4 and to distinguish between nitrogen-fixing and non-nitrogen-fixing populations (Table 1).

Previous analyses of soils ABA, BAHF, LWRB, and RBW revealed cell densities of about 10^6 cells (g soil)$^{-1}$, with cluster 1b representing the most prominent *Frankia* population, while cluster 3 frankiae were present in small numbers and clusters 1a, 1d, and 1c were usually absent (17). These results were largely confirmed in our current analyses, where specific analyses retrieved frankiae of cluster 1b only (soils ABA and BAHF), while soils LWRB and RBW harbored small numbers of cluster 3 frankiae as well (Fig. 3). Frankiae of clusters 2 and 4 remained undetected, indicating that these populations are either absent or present in numbers below the detection limit. Cluster 1b frankiae have been detected as major populations in several studies, with absolute numbers depending on the sampling depth, physicochemical conditions, and vegetation (15, 17, 18). These results demonstrated differential effects of environmental conditions, including plant species, carbon resources, and matric potentials on the fate of specific *Frankia* strains in soil. These factors could affect subpopulations of indigenous frankiae of clusters 2 and 4 that were both detected in all three prairie soils but not in soils under *A. glutinosa* or *B. nigra*. Their presence in prairie soils but not in soils under host plant *A. glutinosa* and non-host plant *B. nigra* suggests that vegetation is affecting the abundance of frankiae of these clusters. Populations of cluster 2 frankiae are lower in managed prairie soils, i.e., in restored and even more in cultivated prairie soils, compared to native prairie soils. Since native prairie harbors potential host plants for cluster 2 frankiae, such as *Ceanothus* species, it is tempting to assume that cluster 2 frankiae rely on the presence of host plants for growth. This situation would be similar to *Casuarina*-infective frankiae that, however, can persist long term in the absence of host plants. Long-term persistence in soil in the absence of host plants of cluster 2 frankiae would be in line with their detection in Tunisian soils lacking compatible host plants for more than 2 centuries, using bioassays with *Coriaria myrtifolia* as the capture plant (32).This assumption, however, needs to be assessed under controlled conditions, with different plant species as variables for population studies of cluster 2 frankiae.

Cluster 4 frankiae, i.e., presumably non-nitrogen-fixing frankiae, represented the most prominent *Frankia* population in all three prairie soils, while they were not detected

at all in soils under *A. glutinosa* or *B. nigra*. Studies on cluster 4 frankiae are scarce, even though they have been reported to form an important fraction of all frankiae in wet soils under *A. glutinosa* (33, 34), with natural resistance to infection exhibited by different progenies of *A. glutinosa* (12, 35). While these results were based on bioassays, we now also have the instruments to assess the importance of cluster 4 frankiae in different environments or to use different strains of the non-nitrogen-fixing frankiae in controlled inoculation studies to retrieve information on their ecology in soils. Our initial results indicate that vegetation affects the basic composition of frankiae in soils, with higher diversity in prairie soils compared to much more restricted diversity under host and non-host trees.

These results demonstrate the usefulness of the methodology developed, with the new forward primers and the probes providing more accurate coverage of the *Frankia* community in soils, even though members of cluster 2 and 4 might not be present in significant amounts in all soils. Future studies, however, should include next-generation sequencing analyses that assess the overall diversity of *Frankia* in a sequence of prairie soils, from native to highly managed, in order to determine environmental effects on diversity. In addition, our methodology now enables us to perform competition experiments that focus on the relationship between abundance and the nodule-forming capacity of cluster 1a/d and 1b frankiae, with frankiae of cluster 1a/d often found in nodules and cluster 1b frankiae dominant in soils.

MATERIALS AND METHODS

Cell sample preparation

Sequences of cluster 2 frankiae were obtained from uncultured endophytes of ethanol-preserved root nodules of *Datisca cannabina* collected in Rawalakot, Azad Kashmir, Pakistan (33.8472389, 73.7485194), *Coriaria nepalensis* was collected in Jhika Gali, Murree, District Rawalpindi, Pakistan (33.9112833, 73.4239306), *C. japonica* was from the Morton Arboretum in Lisle, IL, USA (41.8167861, –88.0679528), and *Ceanothus* sp. was from the Loda Cemetery Prairie Nature Preserve, IL, USA (40.5284721, –88.0717537). A single lobe was homogenized with a mortar and pestle in

1 ml of sterile water, and the homogenates were transferred to Eppendorf tubes and centrifuged at 14,000 x g for 1 min. The pellets were washed once with 0.1% (wt/vol) sodium pyrophosphate in water, followed by two washes with sterile distilled water. *Frankia* strains representing cluster 4 (CN3, CN4, CN6, CN7, CNm, CNm1, CNm3, CNm7, DC12, AgI5, AgW1.1, AgB1.5, AgB1.7, AgB1.9, and AgB1.10) (12, 13, 29, 36) were grown in P+N medium for 2 weeks (37), harvested by centrifugation at 14,000 x g for 5 min, and washed twice with sterile distilled water. Nodule pellets as well as pellets of pure cultures (approximately 50 mg) were resuspended in 95 µl of distilled water and lysed after addition of 5µl of proteinase K solution (30 U mg ^{1}; 10 mg ml ^{1} in water; Promega, Madison, WI) and incubation at 37°C for 20 min (23). Afterwards, 0.5 µl of 10% SDS solution was added and the mixtures incubated at 37°C for another 3 h, which was followed by a final incubation at 80°C for 30 min (23).

PCR amplification

From these lysates, 2 µl aliquots were used as the template in subsequent PCR-based analyses. 23S rRNA gene fragments (about 240 bp) were amplified using primers 23Fra1533f ($^{5'}$GTT GAT ATT CCC GTA CCG) and 23Fra1769r ($^{5'}$GGC TCG GCA TCA GGT CTC AG), targeting frankiae and some other actinobacteria. The PCR was carried out in a volume of 50 µl, containing 1 µl of a 10 mM deoxynucleoside triphosphate (dNTP) mix, 0.5 µl each primer (0.4 µM), 8.2 µl bovine serum albumin (BSA) (30 µg ml^{-1}), 5 µl of 10 x PCR buffer with 15 mM $MgCl_2$, 2 µl root nodule or pure culture lysate, and 0.2 µl Taq DNA polymerase (5 U µl^{-1}; GeneScript, Piscataway, NJ) that was added after an initial incubation at 96°C for 10 min. The addition of Taq polymerase was followed by 35 rounds of temperature cycling (96°C for 30 s, 60°C for 30 s, and 72°C for 45 s) and a final 7-min incubation at 72°C. Subsamples of the reaction mixtures (5 µl) were checked for amplification products by gel electrophoresis (1%, wt/vol, agarose in Tris-acetate-EDTA buffer) after staining with ethidium bromide (0.5 µg ml^{-1}) (38).

Sequence analyses

Amplified 23S rRNA gene fragments were cleaned using shrimp alkaline phosphatase and exonuclease I (Affymetrix, Santa Clara, CA) by following the manufacturer's protocols and then sequenced bidirectionally using BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA) with the same primers used for PCR. Sequences were analyzed on a 3500 genetic analyzer (Life Technologies, Carlsbad, CA) and deposited at GenBank under accession numbers LT576423 to LT576449.

Phylogenetic analyses

Sequences of amplified 23S rRNA gene fragments obtained from uncultured frankiae from root nodules of all plants analyzed and those of pure cultures of cluster 4 *Frankia* were trimmed to lengths between 141 and 152 bp to match those in our database (19, 39), assembled in Geneious 9.1.4 (Biomatters Ltd., Auckland, New Zealand), and checked in GenBank/EMBL databases using the BLAST algorithm (40). Representative sequences from confirmed *Frankia* strains of all 4 clusters were added from our and GenBank/EMBL databases and aligned by the Geneious alignment tool. The identity and relationship among the sequences amplified were evaluated using neighbor-joining (NJ) (41), maximum likelihood (ML) (42), and Bayesian analyses (43). All of these analyses were conducted from within Geneious 9.1.4. The neighbor-joining analyses utilized the HKY85 model to correct for substitution bias (49). Model parameters for maximum likelihood, which were estimated by the general timereversible model (GTR) with gamma (44), were used as input in an ML heuristic search using RAxML (45). Bootstrap values (46) were estimated from a heuristic search with random stepwise addition of sequence for 10,000 NJ and 1,000 ML iterations. MrBayes version 3.1.2 (43) was implemented for 10 million generations, saving every thousandth tree, with a burn-in of one million trees using the GTR with gamma substitution model.

Primer and probe design and evaluation

Aligned sequences were amended with sequences of other target and non-target organisms and used to manually check for and design forward primers specific for cluster 2 and 4 frankiae, i.e., primers 23Dat1578f and 23NNF1561, respectively, which could be used with reverse primer 23Fra1769r in SybrGreen-based *q*PCR (Table 1). In addition, two probes, one targeting all frankiae of clusters 1, 2, and 3 (and thus supposedly all

nitrogen-fixing frankiae), probe NF1715f ($^{5'}$-6carboxyfluorescein [FAM]-TGG TTG TCC TGG GGC AAG GGT GTA GG-6-carboxytetramethylrhodamine [TAMRA]), and a second targeting cluster 4 frankiae (and thus generally non-nitrogen-fixing frankiae), probe NNF1715f ($^{5'}$-6-FAM-CGG GGT AAG CGT GTA GG ACG ACG TGT A-TAMRA), were designed and subsequently evaluated in TaqMan-based qPCR using the genus- or subgroup-specific primer sets from SybrGreen-based applications for amplification. Selected primers and probes were checked for low potential of self- and heterodimer formation using OligoAnalyzer 3.1 (www.idtdna.com/calc/analyzer) and for target specificity using TestPrime 1.0 and TestProbe 3.0 (47) from the SILVA rRNA database project (www.arb-silva.de; accessed 29 September 2016) (48). Annealing temperatures for all primer combinations were tested in *q*PCRs with DNA of representative *Frankia* strains or PCR products from uncultured endophytes of clusters 1, 2, 3, and 4, respectively, and quantifications were compared between SybrGreen- and TaqMan-based analyses, both performed in an Eco real-time PCR system (Illumina, San Diego, CA).

Primer combinations for SybrGreen-based *q*PCR targeted *nif*H gene sequences (16) or 23S rRNA gene sequences (15) (Table 1). SybrGreen-based analyses were carried out in triplicate in a total volume of 10 µl containing 5 µl of SsoADV SybrGreen mix (Bio-Rad, Hercules, CA), 0.125 µl of forward and reverse primers (100 nM each), and 1 µl of DNA template using an initial denaturation at 95°C for 5 min and 40 cycles of denaturation at 95°C, annealing at 62, 64, or 66°C depending on the primer combination (Table 1), and extension at 72°C, each for 30 s (15, 17). The amplification was followed by a melting curve analysis.

Primer combinations targeting 23S rRNA sequences were also used for TaqMan-based quantification, though in combination with probe NF1715f or NNF1715f. Except for cluster 1b, all TaqMan-based analyses were carried out in triplicate in a volume of 10 µl containing 5 µl of Sso ADV probe mix (Bio-Rad), 0.2 µl of forward and reverse primers (100 nM each), 0.25 µl of probe (250 nM each), and 1 µl of DNA template. An initial denaturation at 95°C for 5 min was followed by 40 cycles of 60°C for 60 s. For cluster 1b, primer concentrations were 300 nM each, and cycles consisted of 58°C for 60 s followed by 72°C for 30 s.

Quantification was based on standard curves generated from purified PCR products of *nif*H or 23S rRNA genes of strains Ag45/Mut15, ArI3, CcI3, EAN1pec, and AgB1.9 or uncultured frankiae from root nodules of *Coriaria*, depending on the primer combination. Amplicons were generated using the genus-specific primers (Table 1), and concentrations were measured with a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA). Copy numbers were calculated from concentrations (http://cels.uri.edu/gsc/cndna.html) and normalized after *q*PCR quantification with the primer combination targeting all nitrogen-fixing frankiae. Copy numbers were divided by copy numbers of the 23S rRNA gene per genome (2 or 3 for pure cultures, 2.5 for unknown populations) to relate copy numbers to *Frankia* cell numbers (15).

Method evaluation

To assess probe specificity and effects of detection procedures on the quantification of frankiae, amplicons of 23S rRNA genes of strains Ag45/Mut15, ArI3, CcI3, EAN1pec, and AgB1.9 and uncultured frankiae from root nodules of *Coriaria* were generated using genus-specific primers. These amplicons were initially used as individual amplicons or in mixtures in Sybr Green-based *q*PCR to compare quantification with either genus- or cluster-specific primers. This approach was extended by the addition of probes in TaqMan-based *q*PCR.

Further method assessments used soil samples that were obtained from seven sites in Illinois, in close proximity to Urbana-Champaign. Soils included 4 previously analyzed sites, i.e., sites ABA (Arboretum at the University of Illinois; 40.093585, –88.218016) and BAHF (Horticulture Farm at the University of Illinois; 40.079306, –88.190558), planted with European alder (*Alnus glutinosa*), and sites LWRB (Lake of the Woods Park; 40.203501, –88.387924) and RBW (Illinois State Water Survey Campus; 40.083917, –88.242038), planted with river birch (*Betula nigra*)(17). Soils at sites ABA, BAHF, and RBW formed under tallgrass prairie on postglacial loess, while the soil at site LWRB formed under deciduous forest, all about 23,000 years before present. Additional soils were obtained from two sites at Loda Cemetery Prairie (40.5284721, –88.0717537), one representing native prairie dominated by *Sorghastrum nutans* (L.) (Indiangrass) and

Andropogon gerardii Vitman (Big bluestem) on black prairie soil, while the second was adjacent to the native prairie site but cultivated continuously with corn.

Table 1 Primer combinations for SybrGreen-based quantification of subgroups within the genus *Frankia*

Target group	Primer combination (5' → 3')	Anneal Temperature (°C)	Fragment size (bp)	Reference
Target gene: *nif*H				
Nitrogen-fixing *Frankia* strains of clusters 1 and 3				
	nifHf1 (5'GGC AAG TCC ACC ACC CAG C) nifHr158 (5'GAC GCA CTT GAT GCC CCA)	64	191	(29)
Target gene: 23S rRNA				
Genus *Frankia* (clusters 1, 2, 3 and 4)				
	23Fra1655f (5'CTG GTA GTA GGC AAG CGA TGG) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	64	133	(15)
Cluster 1 (*Alnus* and *Casuarina* host infection group)				
Subgroup 1a/d	23Ar1607f (5'GTG TCT TTT CGG AGA TGT GTC T) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	64	128	(16)
Subgroup 1b	23Mut1555f (5'TTG ATG CGT CCA TGC TGA GG) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	66	170	(15)
Subgroup 1c	23Cas1600f (5'GTG TCT CTT CGG AGG TGT GTT C) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	68	128	(15)
Cluster 2 (Rosaceae/Coriariaceae/Datiscaceae host infection group)				
	23Dat1578f (5'TGG TTC GTG CTA ACC GTC CGA) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	66	153	This study
Cluster 3 (Elaeagnaceae/Rhamnaceae host infection group)				
	23EAN1577f (5'GTT TGT GCT AAC CGT TCT GGT) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	64	146	(15)
Cluster 4 (Atypical, generally non-nitrogen-fixing and/or non-nodulating frankiae)				
	23NNF1561f (5'CCA ATG CTG AAT CTT CCT G) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	62	142	This study

The last soil was obtained from Meadow Brook Park (40.0789008, –83.7852567) and resembled restored prairie dominated by *S. nutans* (L.) and *A. gerardii* Vitman on black prairie soil. All soils were similar with respect to particle size distribution (silt loam), organic matter content (2.5 to 4.6%), and pH (6.0 to 7.1). At each site, soil samples of about 1 kg were taken from the upper 10 cm, with soils from sites with trees (ABA, BAHF, LWRB, and RBW) being sampled less than 1 m from the stem of one tree. Samples were obtained using a trowel that was cleaned with a wire brush and then rinsed in a bucket containing 50% ethanol between sample extractions. Soils were released from roots and homogenized by manipulating the entire sample in freezer bags and then stored at 4°C until further processing.

DNA was extracted from triplicate 250-mg (dry weight) soil samples using the SurePrep soil DNA isolation kit (Fisher Scientific, Houston, TX) with small modifications as described before (16). Ten fold dilutions were used as the template in both SybrGreen- and TaqMan-based *q*PCR analyses for members of the genus *Frankia* or subgroups within the genus, as described above. Results of all analyses were corrected for extraction efficiencies determined as the ratio of inoculated *Salmonella enterica* serovar Typhimurium (ATCC 14028) cells detected by *q*PCR-based quantification of a 268-bp *inv*A gene fragment before and after extraction as described previously (16).

Statistical analysis

One-way ANOVA and pairwise multiple-comparison procedures (Holm-Sidak method) were used in SigmaPlot 13.0 (Systat Software Inc., San Jose, CA) to assess the effects of different *q*PCR procedures on abundance estimates for frankiae, with a significance level at a P value of < 0.05. Accession number(s). Sequences determined here were deposited at GenBank under accession numbers LT576423 to LT576449.

ACKNOWLEDGMENTS

We are indebted to the Office of Sponsored Programs (Research Enhancement Program grant no. 9000000647), the Graduate College, and the Department of Biology at Texas State University for financial support.

REFERENCES

1. Schwintzer CR, Tjepkema JD. 1990. The biology of *Frankia* and actinorhizal plants. Academic Press, San Diego, CA.
2. Huss-Danell K. 1997. Actinorhizal symbioses and their N-2 fixation. New Phytol 136:375–405. https://doi.org/10.1046/j.1469-8137.1997.00755.x.
3. Benson DR, Silvester WB. 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol Rev 57:293–319.
4. Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L, Misra AK. 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Intl J System Bacteriol 46:1–9. https://doi.org/10.1099/00207713-46-1-1.
5. Benson DR, Dawson J. 2007. Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiol Plant 130: 318–330. https://doi.org/10.1111/j.1399-3054.2007.00934.x.
6. Dobritsa SV. 1998. Grouping of *Frankia* strains on the basis of susceptibility to antibiotics, pigment production and host specificity. Intl J System Bacteriol 48:1265–1275. https://doi.org/10.1099/00207713-48-4-1265.
7. Callaham D, Del Tredici P, Torrey JG. 1978. Isolation and cultivation in vitro of the actinomycete causing root nodulation in *Comptonia*. Science 199:899–902. https://doi.org/10.1126/science.199.4331.899.
8. Diem HG, Gauthier D, Dommergues YR. 1982. Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Can J Microbiol 28:526–530. https:// doi.org/10.1139/m82-079.
9. Weber A, Smolander A, Nurmiaho-Lassila EL, Sundman V. 1988. Isolation and characterization of *Frankia* strains from *Alnus incana* and *Alnus glutinosa* in Finland. Symbiosis 6:97–116.
10. Caru M, Becerra A, Sepulveda D, Cabello A. 2000. Isolation of infective and effective *Frankia* strains from root nodules of *Alnus acuminata* (Betulaceae). World J Microbiol Biotechnol 16:647–651. https://doi.org/ 10.1023/A:1008929520571.

11. Gtari M, Brusetti L, Skander G, Mora D, Boudabous A, Daffonchio D. 2004. Isolation of *Elaeagnus*-compatible *Frankia* from soils collected in Tunisia. FEMS Microbiol Lett 234:349–355. https://doi.org/10.1111/j.1574 -6968.2004.tb09554.x.
12. Hahn D, Starrenburg MJC, Akkermans ADL. 1988. Variable compatibility of cloned *Alnus glutinosa* ecotypes against ineffective *Frankia* strains. Plant Soil 107:233–243. https://doi.org/10.1007/BF02370552.
13. Mirza MS, Hahn D, Akkermans ADL. 1992. Isolation and characterization of *Frankia* strains from *Coriaria nepalensis*. Syst Appl Microbiol 15: 289–295. https://doi.org/10.1016/S0723-2020(11)80103-7.
14. Gtari M, Ghodhbane-Gtari F, Nouioui I, Ktari A, Hezbri K, Mimouni W, Sbissi I, Ayari A, 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.
15. Samant S, Amann RI, Hahn D. 2014. Evaluation of the 23S rRNA gene as target for qPCR based quantification of *Frankia* in soils. System Appl Microbiol 37:229–234. https://doi.org/10.1016/j.syapm.2013.11.001.
16. Samant S, Sha Q, Iyer A, Dhabekar P, Hahn D. 2012. Quantification of *Frankia* in soils using SYBR Green based *q*PCR. System Appl Microbiol 35:191–197. https://doi.org/10.1016/j.syapm.2011.12.004.
17. 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.
18. Samant SS, Dawson JO, Hahn D. 2015. Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. System Appl Microbiol 38:501–505. https://doi.org/10.1016/j.syapm.2015 .07.005.
19. Hönerlage W, Hahn D, Zepp K, Zeyer J, Normand P. 1994. A hypervariable 23S rRNA region provides a discriminating target for specific characterization of uncultured and cultured *Frankia*. System Appl Microbiol 17:433–443. https://doi.org/10.1016/S0723-2020(11)80061-5.

20. Nouioui I, Ghodhbane-Gtari F, Beauchemin NJ, Tisa LS, Gtari M. 2011. Phylogeny of members of the *Frankia* genus based on *gyr*B, *nif*H and *gln*II sequences. Antonie Van Leeuwenhoek 100:579–587. https:// doi.org/10.1007/s10482-011-9613-y.
21. Ghodhbane-Gtari F, Nouioui I, Chair M, Boudabous A, Gtari M. 2010. 16S-23S rRNA intergenic spacer region variability in the genus *Frankia*. Microb Ecol 60:487–495. https://doi.org/10.1007/s00248-010-9641-6.
22. Hahn D, Mirza B, Benagli C, Vogel G, Tonolla M. 2011. Typing of nitrogen fixing *Frankia* strains by matrix-assisted laser desorption ionization-time of-flight (MALDI-TOF) mass spectrometry. System Appl Microbiol 34:63–68. https://doi.org/10.1016/j.syapm.2010.11.009.
23. Mirza BS, Welsh A, Rasul G, Rieder JP, Paschke MW, Hahn D. 2009. Variation in *Frankia* populations of the *Elaeagnus* host infection group in nodules of six host plant species after inoculation with soil. Microb Ecol 58:384–393. https://doi.org/10.1007/s00248-009-9513-0.
24. Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. System Appl Microbiol 32:201–210. https://doi.org/10.1016/ j.syapm.2009.01.002.
25. Mirza BS, Welsh AK, Hahn D. 2009. Growth of *Frankia* strains in leaf litter-amended soil and the rhizosphere of a non-actinorhizal plant. FEMS Microbiol Ecol 70:132–141. https://doi.org/10.1111/j.1574-6941 .2009.00746.x.
26. Samant S, Dawson JO, Hahn D. 2016. Growth responses of introduced *Frankia* strains to edaphic factors. Plant Soil 400:123–132. https:// doi.org/10.1007/s11104-015-2720-1.
27. 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.
28. 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.

29. Mirza MS, Janse JD, Hahn D, Akkermans ADL. 1991. Identification of atypical *Frankia* strains by fatty acid analysis. FEMS Microbiol Lett 83: 91–97. https://doi.org/10.1111/j.1574-6968.1991.tb04395.x.
30. Lechevalier MP, Ruan JS. 1984. Physiology and chemical diversity of *Frankia* spp. isolated from nodules of *Comptonia peregrina* (L.) Coult. and *Ceanothus americanus* L. Plant Soil 78:15–22. https://doi.org/10.1007/ BF02277836.
31. Baker D, Newcomb W, Torrey JG. 1980. Characterization of an ineffective actinorhizal microsymbiont, *Frankia* sp EuI1 (Actinomycetales). Can J Microbiol 26:1072–1089. https://doi.org/10.1139/m80-180.
32. Nouioui I, Sbissi I, Ghodhbane-Gtari F, Benbrahim KF, Normand P, Gtari M. 2013. First report on the occurrence of the uncultivated cluster 2 *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.
33. Wolters DJ, Akkermans ADL, Van Dijk C. 1997. Ineffective *Frankia* strains in wet stands of *Alnus glutinosa* L Gaertn in the Netherlands. Soil Biol Biochem29:1707–1712.https://doi.org/10.1016/S0038-0717(97)00075-8.
34. Wolters DJ, Van Dijk C, Zoetendal EG, Akkermans AD. 1997. Phylogenetic characterization of ineffective *Frankia* in *Alnus glutinosa* (L.) Gaertn. nodules from wetland soil inoculants. Mol Ecol 6:971–981. https:// doi.org/10.1046/j.1365-294X.1997.00265.x.
35. Van Dijk C, Sluimer A. 1994. Resistance to an ineffective *Frankia* strain type in *Alnus glutinosa* (L.) Gaertn. New Phytol 128:497–504. https:// doi.org/10.1111/j.1469-8137.1994.tb02996.x.
36. Mirza MS, Akkermans WM, Akkermans ADL. 1994. PCR-amplified 16S rRNA sequence analysis to confirm nodulation of *Datisca cannabina* L.by the endophyte of *Coriaria nepalensis* Wall. Plant Soil 160:147–152. https://doi.org/10.1007/BF00150356.
37. Meesters TM, van Genesen ST, Akkermans ADL. 1985. Growth, acetylene reduction activity and localization of nitrogenase in relation to vesicle formation in *Frankia* strains Cc1.17 and Cp1.2. Arch Microbiol 143: 137–142. https://doi.org/10.1007/BF00411036.

38. Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
39. Maunuksela L, Zepp K, Koivula T, Zeyer J, Haahtela K, Hahn D. 1999. Analysis of *Frankia* populations in three soils devoid of actinorhizal plants. FEMS Microbiol Ecol 28:11–21. https://doi.org/10.1111/j.1574 -6941.1999.tb00556.x.
40. Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. Proc Natl Acad SciUSA85:2444–2448. https://doi.org/ 10.1073/pnas.85.8.2444.
41. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.
42. 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.
43. Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. https:// doi.org/10.1093/bioinformatics/btg180.
44. Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Am Math Soc Lect Math Life Sci 17:57–86.
45. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
46. Felsenstein J. 1985. Confidence limits of phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.2307/2408678.
47. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1. https://doi.org/10.1093/nar/gks808.
48. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner F-O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–D596. https://doi.org/10.1093/nar/gks1219.

49. Hasegawa M, Kishino H, Yano T. 1985. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.

CHAPTER III

FRANKIA COMMUNITY STRUCTURE IN NATIVE AND MANAGED TALLGRASS PRAIRIE SOILS

ABSTRACT

Long-term effects of crop rotation and different fertilization regimen on bacterial community structure with emphasis on members of the genus *Frankia* were assessed in soils from the Morrow Plots, the oldest agronomic experimental fields in the United States previously established on tallgrass prairie, and soils from the original prairie. In native and restored prairie soils, quantitative polymerase chain reaction (*q*PCR) detected members of all four *Frankia* clusters with clusters 1b, 2, 3, and 4, similar to prairie soils recently cultivated with corn that, however, were devoid of cluster 3 frankiae. Members of the poorly studied clusters 2 and 4 were generally present in high abundance. Soils of the Morrow Plots, with different crop rotation and fertilization regimen, harbored the same *Frankia* clusters as cultivated prairie soil, though with lower numbers of cluster 2 and 4 frankiae and generally higher numbers of cluster 1b frankiae. Illumina sequencing of 16S rRNA gene fragments detected Acidobacteria, Actinobacteria, Proteobacteria and Verrucomicrobia as major phyla in all soils, with no significant differences between sites, crop rotation or fertilization. Reads representing frankiae accounted for 0.1 to 1.0% of all reads, with generally higher percentages in fertilized soils. Reads represented frankiae of clusters 1a, 2, 3, and 4, but also a group of frankiae that could not reliably be assigned to a cultured relative. Additional studies using *nif*H gene targeted Illumina sequencing that excluded analyses of clusters 2 and 4, revealed frankiae of clusters 1a, 1b and 3 as the most prominent clusters while cluster 1d showed sporadic presence. The results provide evidence of long-term establishment of *Frankia* populations in soils under different management conditions. They also highlight methodological problems of our analyses since each of our methods targets different genes, with vastly different detection limits that impact both qualitative and quantitative analyses.

KEYWORDS: Frankiae, soil, qPCR, saprophytic, illumine

INTRODUCTION

New molecular techniques that focus on high throughput DNA sequencing such as 454 pyrosequencing and the MiSeq sequencing platforms revolutionized the field of microbial diversity studies (Michelsen *et al.*, 2014, Siles *et al.*, 2014, Embarcadero-Jimenez *et al.*, 2016, Fay *et al.*, 2016). To study phylogenetic diversity, ribosomal RNA (rRNA) genes are generally used as standard markers, and their sequences are widely used in high throughput DNA sequencing projects such as the Earth Microbiome Project (EMP), a collective attempt to establish microbial fingerprints in different environments of the planet (Gilbert *et al.*, 2014). Next generation sequencing (NGS)-based analyses of microbial communities generally focus on phylogenetic levels between phyla and families, and occasionally include assessments on the genus level (Li *et al.*, 2017, Mashiane *et al.*, 2017). Analyses on lower taxonomic levels such as species or subspecies levels are usually not attempted, since they are often impacted by the limited phylogenetic resolution of sequences of 16S rRNA gene fragments, the limited number of reads available at this level, as well as by the lack sufficient sequences for distinct species or subspecies in the databases.

Elaborate databases of 16S rRNA genes or gene fragments have been created for members of the genus *Frankia* that consists of diverse group of filamentous, gram-positive bacteria that are capable of fixing atmospheric di-nitrogen (N_2) (Benson, 1988). These soil actinobacteria are mostly detected as symbionts forming root nodules with members of eight plant families representing about 25 genera of woody, dicotyledonous, perennial angiosperms, collectively called “actinorhizal plants” (Benson & Dawson, 2007). These *Frankia* symbionts fix atmospheric N_2 and provide reduced nitrogen compounds to the plant that in exchange provides carbon sources to the symbiont. This symbiotic interaction enables actinorhizal plants to flourish even in nitrogen poor soils and in niches that have limited nitrogen availability (Dawson, 1986). Members of the genus *Frankia* can be assigned to four major clusters based on the N_2-fixing capability and host plant specificity. The first three clusters comprise nitrogen-fixing members, while the fourth cluster represents non-nitrogen fixing species (Normand *et al.*, 1996).

Quantitative polymerase chain reaction (*q*PCR) has demonstrated a high diversity of the genus *Frankia* in prairie soils under different management practices (native, restored and cultivated) (Ben Tekaya *et al*., 2017). Members of all *Frankia* clusters 1, 2, 3 and 4 were detected, with members of poorly studied clusters 2 and 4 generally present in high abundance (Ben Tekaya *et al.*, 2017). Tallgrass prairie soils in close proximity to our study sites have been used to establish the Morrow Plots, the oldest agronomic experimental fields in the United States. They were established in 1876 to evaluate the effects of different cropping systems and soil treatments on crop yield (Odell *et al.*, 1984, Khan *et al.*, 2007), and include the oldest continuous corn plots in the world (Velde & Peck, 2002). Major soil management steps were performed in 1904 and in 1955, with the start of MLP (Barnyard manure-**M**, Limestone-**L** and ground rock Phosphorus-**P**) and NPK (nitrogen-**N**, processed phosphorus-**P**, and potassium-**K**) fertilization regimen, respectively, to sub-plots to check whether addition of chemical fertilizer can improve soil productivity (Odell *et al.*, 1984).

Long-term effects of crop rotation and different fertilization regimen on bacterial community structure have been assessed recently for the Morrow Plots using next generation sequencing (NGS) of 16S rRNA gene fragments (Soman *et al.*, 2017). The goal of our study was to extend these studies and assess long-term effects of crop rotation and different fertilization regimen on the diversity of *Frankia* populations in the Morrow Plots, and adjacent tallgrass prairie. The studies included Illumina high-throughput sequencing of amplified 16S rRNA gene fragments as well as of amplified *nif*H gene fragments for comparative analyses of *Frankia* diversity among treatments.

MATERIALS AND METHODS

Soil samples

Composite samples from the upper 20 cm of soil were obtained September 15, 2014, from 3 sites: Loda Cemetery Prairie Nature Preserve, about 50 km north of Champaign (Illinois, USA) (40.5284721; -88.0717537), Meadow Brook Park (40.0789008; -83.7852567) and the Morrow plots, located near the center of the University of Illinois' Urbana campus (40.6331249; -89.3985283). Loda Cemetery Prairie

sites represented either original mesic black soil tallgrass prairie that had never been plowed or used for pasture, and soil cultivated continuously with corn, while Meadow Brook Park represented restored tallgrass prairie dominated by *Sorghastrum nutans* (L.) (Indiangrass) and *Andropogon gerardii* Vitman (Big bluestem) on black prairie soil. The Morrow plots represent the oldest agronomic experiment fields in the United States, established in 1876 on black soil tallgrass prairie. Nine sites were sampled, with different crop rotation and fertilization regimen (Table 1). All sites were classified as Aquic Arguidolls, that developed on Flanagan silt loam soil (6.4 - 9.0% sand, 67.1 - 66.8% silt, and 24.2 - 26.5% clay), under prairie vegetation with poor natural drainage and high available-moisture-holding capacity. Selected physiochemical characteristics of all soils and treatments (i.e. % organic material, pH, estimated nitrogen release, cation exchange capacity and macronutrient concentrations -P, K, Mg, Ca, NO_3^-, and NH_4^+-) were analyzed by Agricultural Soil Management, Inc (Champaign, IL) (Table 1).

DNA extraction

DNA was extracted from soils using the SurePrep™ Soil DNA Isolation Kit (Fisher Scientific, Houston, TX) with small modifications as described before (Samant *et al.*, 2012). Extractions of all samples were done in triplicate, and DNA concentrations measured with a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, USA). Concentrations ranged from 8 to 30 ng μl^{-1}, corresponding to soil DNA concentrations between 3.2 and 12.0 µg $(g\ soil)^{-1}$.

***q*PCR quantification**

Ten-fold dilutions were used as template in SybrGreen-based *q*PCR analyses for the genus *Frankia* or subgroups within the genus. Results of all analyses were corrected for extraction efficiencies determined as the ratio of inoculated *Salmonella* Typhimurium (ATCC14028) cells detected by *q*PCR-based quantification of a 268-bp *inv*A gene fragment before and after extraction as described previously (Samant *et al.*, 2012). SybrGreen-based analyses were carried out in triplicate in a total volume of 10 µl containing 5 µl of SsoADV SybrGreen Mix (BioRad, Hercules, CA), 0.125 µl of forward and reverse primers (100 nM each), and 1 µl of DNA template using an initial

denaturation at 95°C for 5 minutes, and 40 cycles of denaturation at 95°C, annealing at 62, 64 or 66°C depending on the primer combination (Table 2), and extension at 72°C, each for 30 seconds (Samant *et al.*, 2014, Samant *et al.*, 2016). The amplification was followed by a melting curve analysis.

Quantification was based on standard curves generated from purified PCR products of 23S rRNA genes of strains Ag45/Mut15, ArI3, CcI3, EAN1pec, and AgB1.9 or uncultured frankiae from root nodules of *Coriaria* depending on the primer combination. Amplicons were generated using the genus-specific primers (Table 1), and concentrations measured with a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, USA). Copy numbers were calculated from concentrations (http://www.uri.edu/gsc/cndna.html) and normalized after *q*PCR quantification with the primer combination targeting all frankiae. Copy numbers were divided by copy numbers of the 23S rRNA gene per genome (2 or 3 for pure cultures, 2.5 for unknown populations) to relate copy numbers to *Frankia* cell numbers (Samant *et al.*, 2014).

Illumina sequencing

For Illumina sequencing, 16S rRNA gene fragments were amplified from DNA extracts without addition of salmonellae using primer 515f and barcoded primers 806r, both of which included linker sequences following the instructions from the Earth Microbiome Project (www.earthmicrobiome.org/emp-standard-protocols/16s/) (Table 3) (Caporaso *et al.*, 2011, Caporaso *et al.*, 2012). PCR was carried out in a 100 µl volume with 1 x *Taq* Buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM primers, 2.5 µg $µl^{-1}$ BSA, 1U *Taq* polymerase (GenScript, Inc., Piscataway, NJ) and 1 µl DNA extract. PCR conditions followed the Earth Microbiome project with an initial denaturation at 94° for 3 min followed by 35 cycles of 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, with a final 72°C extension for 10 min. PCR products were cleaned using the UltraClean® 15 DNA Purification Kit (MoBio Laboratories, Inc., Carlsbad, CA), and then checked and quantified on a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA) using the Agilent DNA 7500 Kit. Samples were analyzed at the Genomic Sequencing and Analysis Facility at the University of Texas (Austin, TX), on the Illumina MiSeq v3 with paired

end 2 x 300 bp reads using the respective sequencing and index sequence primers (www.earthmicrobiome.org/emp-standard-protocols/16s/) (Table 2).

Table 2. Primers used for Illumina sequencing of a 263 bp region of the 16S rRNA gene of Bacteria

Primers	5'-Sequence-3'	
515F[1]	AATGATACGGCGACCACCGAGATCTACAC TATGGTAATT GT **GTG CCA GCM GCC GCG GTA A**	PCR
806R	CAAGCAGAAGACGGCATACGAGAT [GOLAY Barcode] [2] AGTCAGTCAG CC **GGA CTA CHV GGG TWT CTA AT**	
Read1 sequencing primer	TATGGTAATT GT **GTG CCA GCM GCC GCG GTA A**	Sequencing
Read2 sequencing primer	AGTCAGTCAG CC **GGA CTA CHV GGG TWT CTA AT**	
Indexing read primer	**AT TAG AWA CCC BDG TAG TCC** GG CTGACTGACT	

[1]Sequences of primers 515F and 806R are given in bold

[2]GOLAY barcode GACTTGGTATTC for sample ABA, barcode GCTGATGAGCTG for sample RBW

Table 3. Basic physicochemical conditions of soils from Loda Cemetery Prairie (40.5284721; -88.0717537), Meadow Brook Park (40.0789008; -83.7852567), and the Morrow Plots (40.6331249; -89.3985283)

Location	Organic	pH	ENR[3]	CAC[4]	Macronutrient concentration (ppm)					
Vegetation, Fertilization[1]	Material (%)		(ppm)	(meq/100 g)	P	K	Mg	Ca	NO_3^-	NH_4^+
Loda Cemetery Prairie										
Native Prairie	4.6	6.6	46.3	23.9	9	186	724	2997	11.4	6.1
Corn	3.7	6.0	36.6	20.9	17	194	646	2357	NS[5]	NS
Meadow Brook Park										
Restored Prairie	3.5	6.1	35.3	22.0	27	174	627	2796	4.4	3.9
Morrow Plots										
Corn-Corn, no fertilization	1.7	6.0	16.9	15.2	16	105	315	1894	4.6	2.5
Corn-Corn, MLP fertilization	1.9	6.6	19.1	17.0	30	163	360	2235	6.1	2.5
Corn-Corn, NPK fertilization	2.1 (0.1)	5.6 (0.1)	20.9 (0.9)	15.9 (0.3)	71 (9)	236 (30)	278 (9)	1943 (54)	4.7 (1.2)	3.8 (0.9)
Corn-Soy, no fertilization	2.0 (0.1)	5.9 (0.2)	19.6 (0.7)	16.5 (0.1)	11 (2)	86 (4)	347 (6)	1986 (47)	6.1 (0.7)	3.2 (0.4)
Corn-Soy, MLP fertilization	2.4	6.3	23.8	19.0	21	114	373	2639	10.9	3.7
Corn-Soy, NPK fertilization	2.0	5.9	20.1	16.4	57	177	320	2021	5.0	2.5
Corn-Oat-Alfalfa, no fertilization	2.9	5.7	28.6	19.3	12	142	386	2267	12.4	3.3
Corn-Oat-Alfalfa, MLP fertilization	2.6 (0.1)	6.0 (0.1)	25.4 (0.8)	18.2 (1.3)	36 (8)	226 (11)	395 (8)	2298 (26)	32.8 (8.7)	2.3 (0.4)
Corn-Oat-Alfalfa, NPK fertilization	2.3	5.7	22.7	17.7	72	170	388	2002	22.2	3.3

[1]Vegetation, Fertilization: Loda Cemetry Prairie (native prairie dominated by *Sorghastrum nutans* (L.) (Indiangrass) and *Andropogon gerardii* Vitman (Big bluestem) on black prairie soil; or cultivated continuously with corn); Meadow Brook Park (restored prairie dominated by *Sorghastrum nutans* (L.) (Indiangrass) and *Andropogon gerardii* Vitman (Big bluestem) on black prairie soil); and Morrow Plots (crop rotation on black prairie soil, with additional fertilization: MLP (manure+lime+P treatment implemented 1904) or NPK (lime+N+P+K treatment implemented 1955)

[2]ENR, Estimated Nitrogen Release, released over the season

[3]CAC, Cation Exchange Capacity

[4]NS, no samples

Bioinformatics

Fastq files containing the raw reads were received from the sequencing facility with barcodes, primers and adapter/linker sequences removed from them. The Dada 2 R package was used to process the demultiplexed paired reads as described by (Callahan *et al.*, 2016). Briefly, reads were checked for quality and trimmed accordingly using *plotqualityprofile(),* with a maximum of two expected errors per read (maxEE = 2). 16S rRNA gene reads were trimmed to 230 bp while *nif*H gene reads were trimmed to 200 bp. Reads were then dereplicated using *derepfastq().* In order to distinguish true sequences from errors, the *functiondada()* function was then applied. Inferred forward and reverse sequences were merged using *mergePairs()*. Sequences that did not perfectly overlap were considered errors and were therefore removed. Each sequence abundance was summarized in a table using the *makeSequenceTable()* function. Chimeras were then removed from the table using the *removeBimeraDenovo()* function. A new table with no chimeras was subsequently generated. Taxonomy is then assigned using *assignTaxonomy()* function. For 16S rRNA gene sequences, the SILVA references database file "silva_nr_v128_train_set.fa.gz" amended with *Frankia* sequences from our own database was used. For the analysis of *nif*H reads, a customized database was used to identify each sequence. Pooled sequences were then transferred to Geneious and compared to respective reference gene sequences.

RESULTS

Soil characteristics

Physicochemical characteristics were very similar in all soils or treatments, and present in sufficient amounts for agriculture. Prairie soils, independent of treatment (i.e. native, corn, restored), differed slightly from Morrow plot soils, with higher % organic material and seasonal nitrogen release estimates (Table 1). pH, cation exchange capacity as well as concentrations of different macronutrient were similar and within values for adequate plant growth. The only exception were concentrations for phosphorous (P), that were low in all treatments without fertilization, as well as in prairies soils (native, corn) (Table 1).

Table 4 Primer combinations for SybrGreen-based quantification of subgroups within the genus *Frankia*

Target group	Primer combination (5' → 3')	Anneal Temperature (°C)	Fragment size (bp)	Reference
Target gene: *nif*H				
Nitrogen-fixing *Frankia* strains (clusters 1, and 3)				
	nifHf1 (5'GGC AAG TCC ACC ACC CAG C) nifHr158 (5'GAC GCA CTT GAT GCC CCA)	64	191	Samant 2012
Target gene: 23S rRNA				
Genus *Frankia* (clusters 1, 2, 3 and 4)				
	23Fra1655f (5'CTG GTA GTA GGC AAG CGA TGG) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	64	133	Samant 2014
***Alnus* host infection group (cluster 1)**				
Cluster 1a (ArI3)	23Ar1607f (5'GTG TCT TTT CGG AGA TGT GTC T) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	64	128	Samant et al. 2015
Cluster 1b (Ag45/Mut15)	23Mut1555f (5'TTG ATG CGT CCA TGC TGA GG) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	66	170	Samant et al. 2014
Cluster 1c (CcI3)	23Cas1610f (5'GTG TCT CTT CGG AGG TGT GTT C) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	68	128	Samant et al. 2014
***Dryas* host infection group (cluster 2)**				
	23Dat1583f (5'CGT GCT AAC CGT CCG ATC TG) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	66	152	This study
***Elaeagnus* host infection group (cluster 3)**				
	23EAN1579f (5'GTT TGT GCT AAC CGT TCT GGT) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	64	146	Samant et al. 2014
Atypical, generally non-nitrogen-fixing frankiae (cluster 4)				
	23NNF1561f (5'CCA ATG CTG AAT CTT CCT G) 23Fra1769r (5'GGC TCG GCA TCA GGT CTC AG)	62	142	This study

Quantification of frankiae by *q*PCR

Members of the genus *Frankia* were detected in all soils by *q*PCR, with abundance estimates between 1 and 3 x 10^6 cells (g soil)$^{-1}$ (Table 4; Fig. 1). Numbers were statistically different between treatments ($F[11, 96]= 50.34$, $p<0.001$), with native and restored prairie showing similar values and significantly higher abundance of frankiae than non-fertilized soils with corn ($p=0.003$), corn-corn ($p<0.001$), or corn-soy ($p<0.001$), but not with corn-oat-alfalfa rotation ($p=0.054$). Soils with fertilization treatments had significantly increased *Frankia* abundance i.e corn-corn ($F[2,24]= 25.23$, $p<0.001$), and corn-soy rotations ($F[2,24] =97.95$, $p<0.001$) but not in corn-oat-alfalfa rotation ($F[2,24] =1.29$, $p=0.29$). Fertilization with organic matter, i.e. MLP fertilization compared to NPK fertilization, resulted in significantly higher *Frankia* abundance in corn-soy rotation ($p=0.02$), but not in corn-corn ($p=0.14$) and corn-oat-alfalfa rotation ($p=0.29$).

Values for quantification of frankiae by *q*PCR were affected by the choice of the target gene, i.e. the *nif*H or the 23S rRNA gene. Detection of *nif*H gene sequences that detect frankiae of clusters 1 and 3 only, generally resulted in lower values than 23S rRNA gene sequences that detect all clusters (i.e. clusters 1, 2, 3, and 4). These differences were most significant in all prairie soils, with only small differences obtained in soils from the Morrow plots (Fig. 1). Differences were mainly due to the detection of cluster 2 and especially cluster 4 frankiae in prairie soils. In native and restored prairie soils, members of all four clusters were detected, with cluster 1b and 4 being most prominent (Table 4; Fig. 1). Diversity was lower in managed systems of the Morrow Plots where cluster 3 frankiae were generally not detected, and members of clusters 2 and 4 were present in much lower numbers. In most soils of the Morrow Plots, the genus *Frankia* was largely represented by frankiae of cluster 1b (Table 4; Fig. 1).

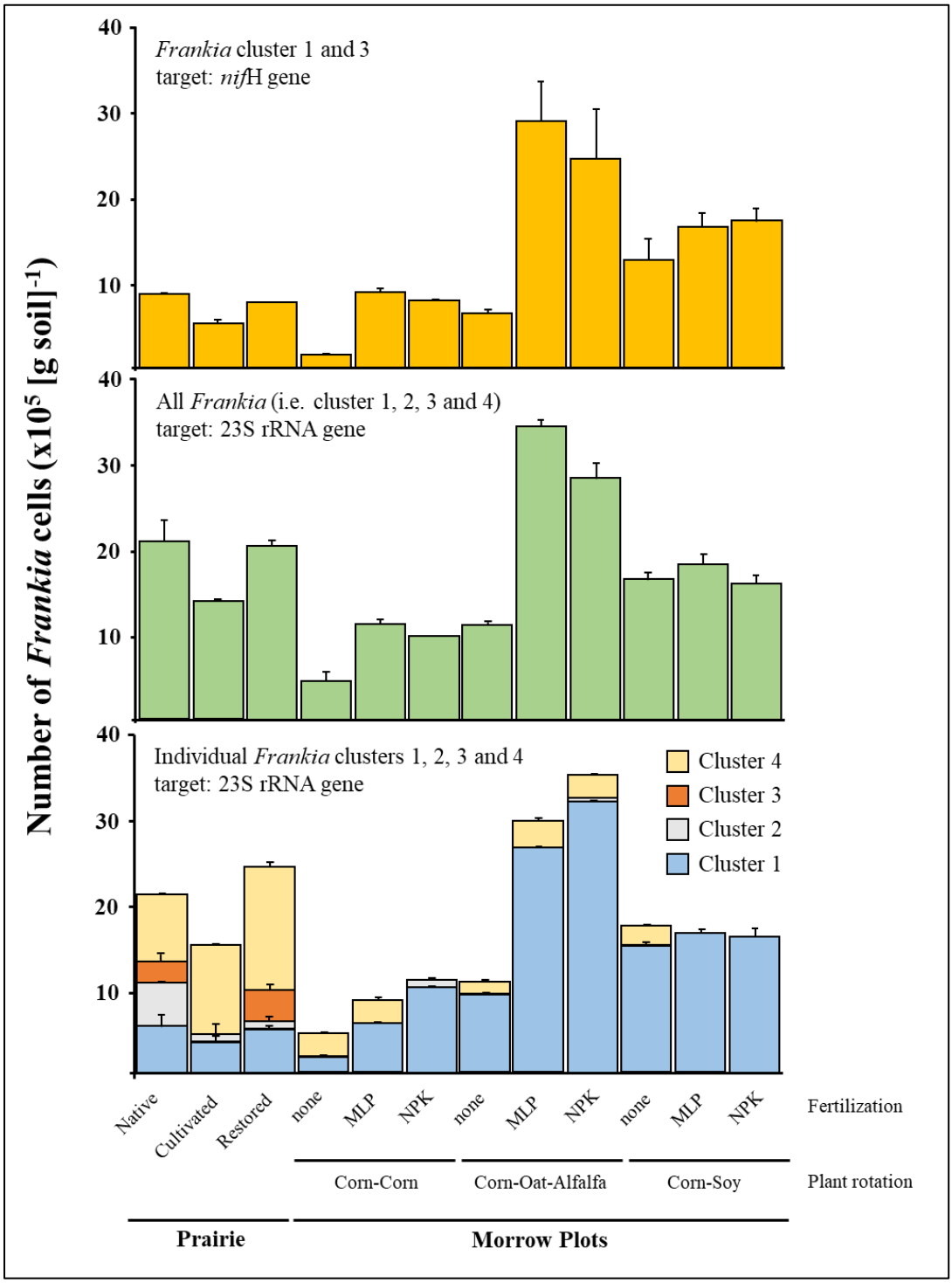

FIG 5: SybrGreen *q*PCR quantification of *Frankia* clusters 1, 2, 3 and 4 across soil treatments (controls labeled as none ; NPK fertilizers and MLP fertilizers). Quantification (from top to bottom) used *nif*H gene fragment as target detecting cluster 1 and 3 or 23s rRNA gene fragments generated with primer combinations that detecting the genus *Frankia* i.e. all clusters, or primer combinations specific for cluster 1, 2, 3 or 4, presented as the sum of individual clusters detected. Representatives of all clusters were detected across treatments.

Table 5. Number of frankiae (x 10^5 [g soil]$^{-1}$) detected by quantitative PCR targeting 23S rRNA gene fragments (genus level detection), *nif*H gene fragments (detection of members of cluster 1, 2 and 3), as well as specific detection of sub-clusters 1a, 1b, 1c, cluster 2, cluster 3 and cluster 4 across prairie soils (native, cultivated and restored) and agricultural treatments (control, NPK fertilized and MLP fertilized)

Location	Genus	Cluster 1			Cluster 2	Cluster 3	Cluster 4
Vegetation, Fertilization[1]	*Frankia*	(a)	(b)	(c)			
Loda Cemetery Prairie							
Native Prairie	20.8 (2.5)	nd	5.8 (1.5)	nd	5.1 (0.0)	2.6 (1.0)	7.8 (0.0)
Corn	13.9 (0.1)	nd	3.9 (1.2)	nd	1.3 (0.9)	nd	10.1 (0.0)
Meadow Brook Park							
Restored Prairie	20.2 (0.7)	nd	5.4 (0.5)	nd	1.1 (0.6)	3.6 (0.9)	14.6 (0.5)
Morrow Plots							
Corn-Corn, no fertilization	4.5 (1.1)	nd	2.2 (0.2)	nd	nd	nd	2.8 (0.1)
Corn-Corn, MLP fertilization	11.2 (0.5)	nd	10.3 (0.3)	nd	0.7 (0.0)	nd	0.2 (0.1)
Corn-Corn, NPK fertilization	9.7 (0.1)	nd	6.2 (0.1)	nd	nd	nd	2.6 (0.4)
Corn-Soy, no fertilization	11.2 (0.3)	nd	9.4 (0.3)	nd	0.1 (0.0)	nd	1.5 (0.1)
Corn-Soy, MLP fertilization	34.1 (1.0)	nd	32.0 (0.2)	nd	0.7 (0.0)	nd	2.7 (0.2)
Corn-Soy, NPK fertilization	28.2 (1.8)	nd	26.9 (0.3)	nd	0.2 (0.0)	nd	3.0 (0.3)
Corn-Oat-Alfalfa, no fertilization	16.3 (0.9)	nd	15.4 (0.4)	nd	0.1 (0.0)	nd	2.0 (0.1)
Corn-Oat-Alfalfa, MLP fertilization	18.2 (1.3)	nd	16.4 (1.0)	nd	nd	nd	nd
Corn-Oat-Alfalfa, NPK fertilization	16.0 (0.9)	nd	16.7 (0.5)	nd	0.1 (0.0)	nd	nd

Illumina sequencing

Reads obtained for Illumina sequencing of *nif*H gene fragments differed significantly with extremes accounting for 9,753 reads (Morrow Plots, corn-oat-alfalfa, MLP) and 557,365 reads for cultivated prairie. Most soils, however, retrieved between 100,000 and 200,000 reads (data not shown). As expected, only cluster 1 and 3 frankiae were identified (Fig. 2), with large differences between treatments. Prairie soils were generally more diverse than soils from the Morrow Plots, with three clusters/subclusters detected compared to one or two clusters (Fig. 2). Clusters 1a, 1b and 3 were the most prominent clusters, with cluster 1d present occasionally in small percentages. Cluster 1a was present in all prairie soils, but not in soils from the Morrow Plots. Soils from the Morrow Plots were generally dominated by frankiae representing either cluster 1b or cluster 3, without a clear relationship to vegetation or fertilization (Fig. 2).

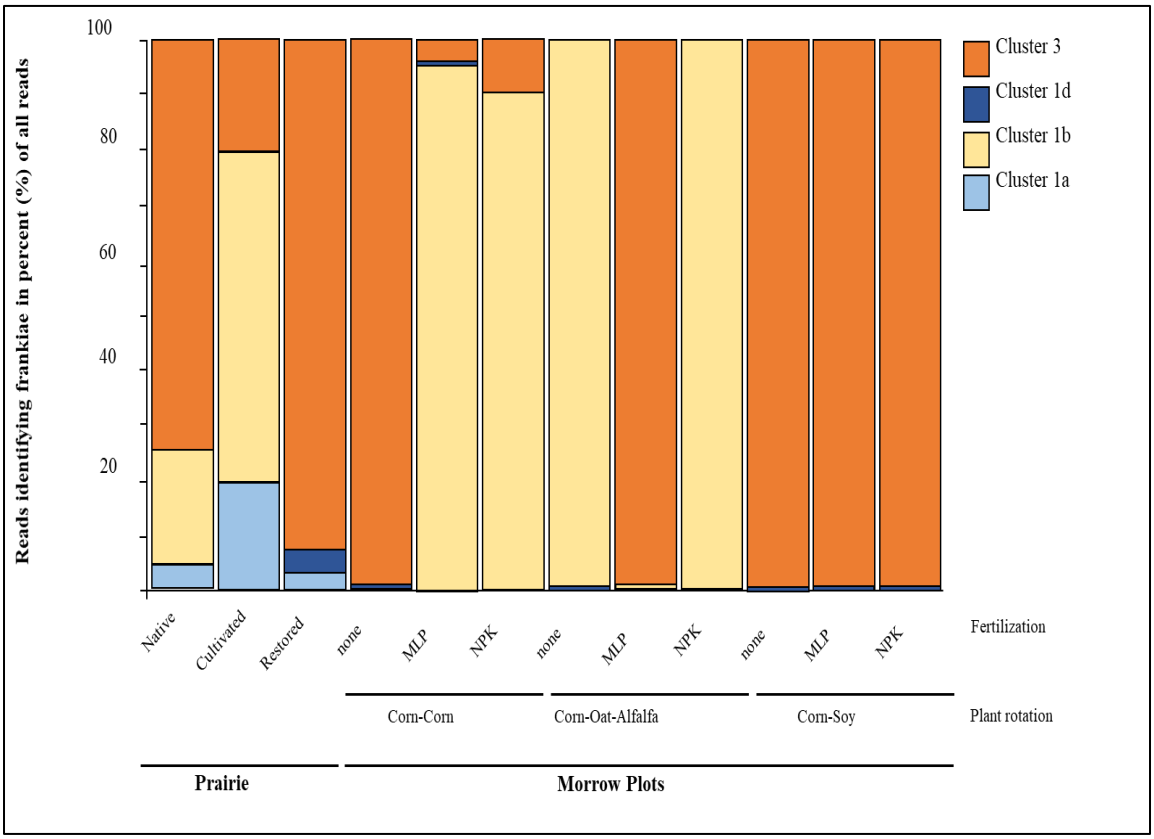

FIG 6 Percentage of *Frankia* reads across prairie soils (native prairie, cultivated prairie and restored prairie) and agricultural treatments (controls are labeled as none, NPK fertilized and MLP fertilized). Reads were generated using a *nif*H targeted Illumina sequencing approach. The figure shows two major dominant *Frankia* clusters i.e. cluster 3 and cluster 1b. Clusters 1a and 1d were represented at minor proportions.

Illumina-based 16S rRNA V3 amplicon sequencing resulted in a total of 26.47 M effective sequences from 44 samples, i.e. 12 soils with different vegetation and fertilization, and 3 to 6 replicates per soil, with 313,695 to 906,328 reads per sample. At a sequencing depth of 300,000 sequences for each sample, Acidobacteria (14.8-23.7%, average 17.9±2.8%), Actinobacteria (17.0-33.4%, average 19.5±4.4%), Proteobacteria (22.1-34.4%, average 28.9±3.1%) and Verrucomicrobia (4.4-16.6%, average 7.4±3.0%) were the most abundant bacterial phyla (67.6-81.7%, average 73.8±3.9%) present across all the soil samples, followed by the Bacteroidetes, Chloroflexi, Gemmatimonadetes, Nitrospirae, and Plancomycetes (0.5-5%) (Fig. 3). Among the Archaea, the Thaumarchaeota was the major phylum observed (Fig. 3). An average of 10.6±3.9% of the sequences could not be identified and/or belonged to phyla accounting for less than 1% of the reads. While the most abundant phyla Acidobacteria, Actinobacteria, Proteobacteria and Verrucomicrobia revealed differences between treatments, no trend was obtained for specific population developments as a function of vegetation or fertilization.

Reads identified as representing the genus *Frankia* were obtained from all soils, i.e. all prairie and Morrow Plot samples, and resemble between 0.1 and 1.0 % of all reads (Fig. 4). Specific identification to the cluster level revealed the presence of clusters 1a, 2, 3, and 4, but also a group of unknown sequences without a clear affiliation (i.e. >97% similarity) to a cultured relative. Prairie soils (native, restored) were the only soils harboring cluster 2 frankiae, while cultivated prairie as well as all Morrow Plot soils did not (Fig. 4). Cultivated prairie as well as all Morrow Plot soils were dominated by cluster 3 frankiae, with cluster 1a frankiae represented by much smaller numbers of reads in all soils (Fig. 4).

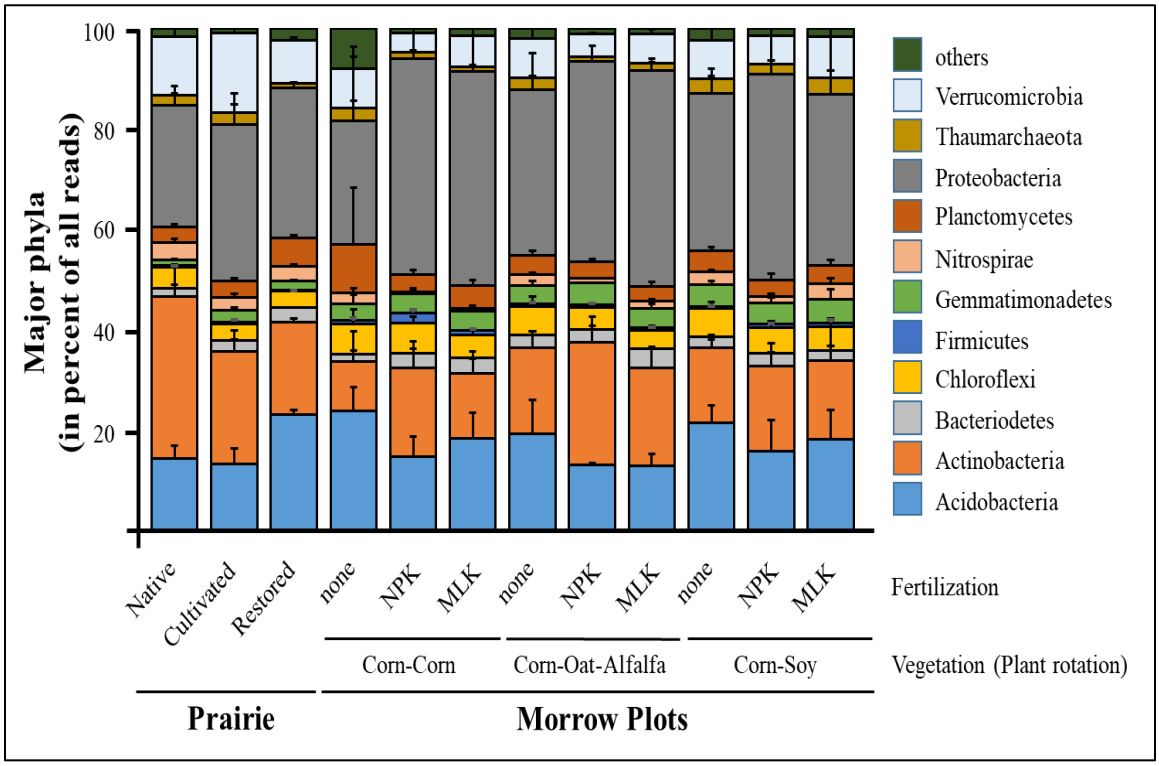

FIG 7 Major bacterial phyla abundance across prairie soils (native prairie, cultivated prairie and restored prairie) as well as agricultural treatments (controls labeled as none, NPK fertilized and MLP fertilized). Abundance is represented as percentage of read numbers of a phylum to the total numbers of read per sample. Reads were generated by 16S rRNA Illumina sequencing. In general, there was no treatment effect on microbial distribution.

DISCUSSION

In addition to their ability to form root nodules with specific actinorhizal plant species, frankiae are capable to live saprophytically in soils. Generally, the geographical distribution of symbiotic microbes is a reflection of the biogeography of their hosts (Higgins & Kennedy, 2012). However, soils supporting saprophytic growth include those deprived of potential host plants, with *Frankia* abundances of approx. $10^4 – 10^5$ cells (g soil)$^{-1}$. In native and restored prairie soils, frankiae were detected with abundances ranging between 1 x 10^5 and 5 x 10^5 cells (g soil)$^{-1}$ (Ben Tekaya *et al.*, 2017). Soils analyzed this study did not harbor any host plants, with exception of native and restored prairie soils that harbored *Ceanothus americana*, a host plant of cluster 2 frankiae. Members of cluster 2 frankiae have been suggested to rely on the presence of host plants

to grow, as suggested in a recent study (Battenberg *et al.*, 2017). Our data support a vegetation effect on the abundance of cluster 2 frankiae in our soils (Ben Tekaya *et al.*, 2017), and thus confirm the speculations outlined in the previous study (Battenberg *et al.*, 2017). The presence of cluster 2 frankiae in small numbers close to or at the detection

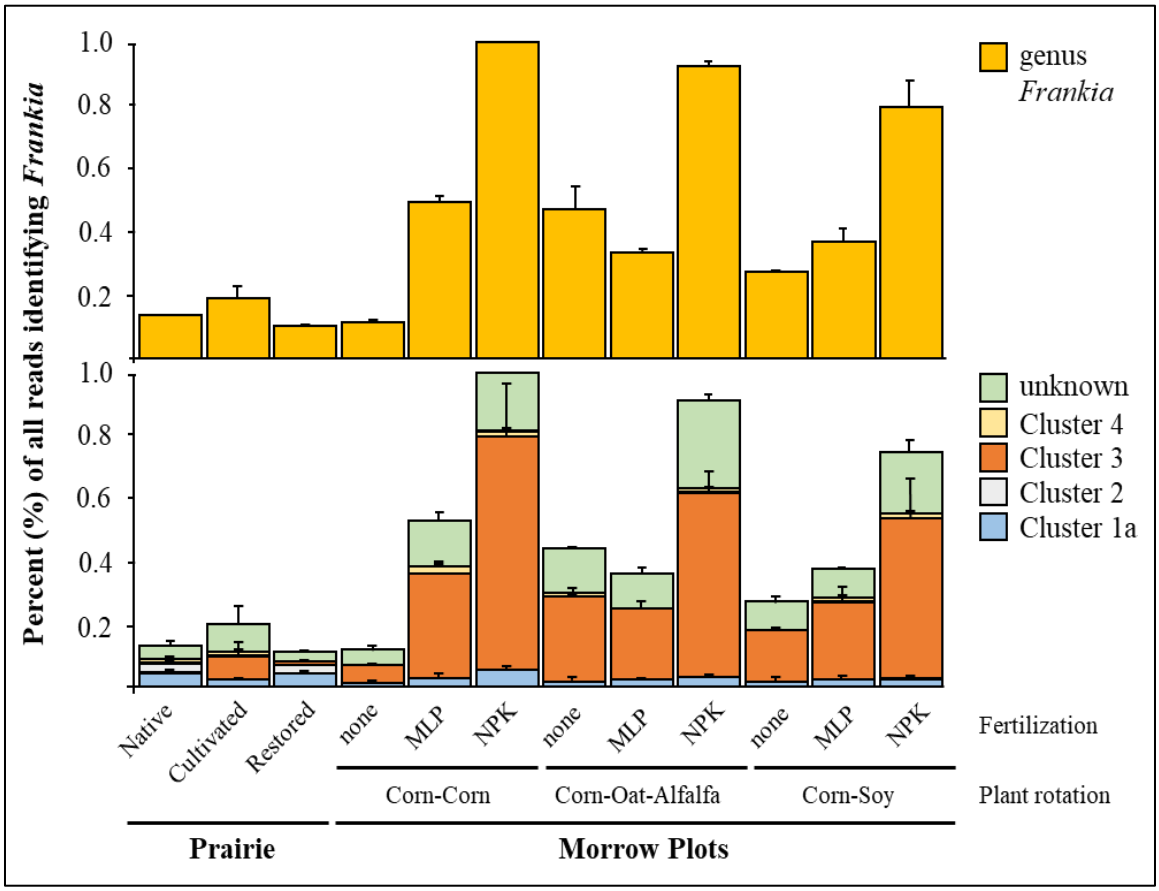

FIG 8 Percentage of *Frankia* reads to total bacteria per each site. Reads were generated by 16S rRNA gene targeted Illumina sequencing analyses. Upper barplots show percentage of genus *Frankia* i.e. all clusters, while bottom barplots show specific *Frankia* clusters assignments. The figure show a major presence of cluster 3 in all soil treatments except native and restored prairie. Cluster 1a, 2 and 4 were present in low numbers only. A number of reads did not match any known cultured frankiae and therefore, were classified as unknown

limit in soils from the Morrow plots support assumptions of long-term persistence of these frankiae in soils. These assumptions are in line with a research study showing the presence of cluster 2 frankiae in soils deprived of potential host for over two centuries (Nouioui *et al.*, 2013). Thus, the detection of low numbers of cluster 2 frankiae present in soils from the Morrow Plots was either the consequence of long-term persistence of these frankiae in soil long after removal of potential host plants, or the result of yet unknown dispersal mechanisms.

Soils from the Morrow Plots represented different treatments with respect to crop rotation and fertilization, established more than a century ago (control untreated, treated with NPK fertilizers and treated with MLP fertilizers). While physico-chemical characteristics were very similar between soils of different treatments, *q*PCR analysis detected a diverse *Frankia* distribution with numbers ranging from 4.5 to 34 x 10^5 cells (g soil)$^{-1}$. These numbers are comparable to those obtained in other studies (Samant *et al.*, 2016, Ben Tekaya *et al.*, 2017), though lower numbers have been observed as well (Mirza *et al.*, 2009). Analysis of specific sub-clusters revealed the consistent presence of members of cluster 1b across all soil treatments. These results corroborate those of previous studies that demonstrated the presence of this cluster in non-host rhizospheres (Samant *et al.*, 2016, Ben Tekaya *et al.*, 2017). Members of *Frankia* cluster 1b are generally dominant in soils with high organic matter content since they are capable of growing with leaf litter. However, no significant differences were observed between controls of crop rotation experiments and their respective fertilization with NPK or MLP, with exception of corn soy rotation where numbers were much higher in NPK and MLP treated soils. These data are in agreement with previous studies that showed advantage of cluster 1b members in soils rich in organic matter (Mirza *et al.*, 2009, Samant *et al.*, 2016).

Members of cluster 4 frankiae, referred to as atypical since they are generally either non nitrogen fixing, or non-infective frankiae, were prominent in most of the soils except for the NPK and MLP treated corn alfa oat rotation. Data on the occurrence of members of cluster 4 in soils are very scarce. However, they are meant to be generally absent in soils where host plants are predominant (Ben Tekaya *et al.*, 2017). Their presence in Morrow Plot soils deprived of host plants might infer a competition

disadvantage in rhizospheres where host specific strains are prominent.

*q*PCR analyses revealed the presence of members of cluster 3 frankiae both in native and restored prairie, respectively. These data were not confirmed by Illumina sequencing targeting *nif*H gene fragments. These data showed a prominent presence of members of cluster 3 frankiae across all treatments, with exception of corn oat alfa (unfertilized and NPK fertilized) rotations that only detected cluster 1b frankiae. Cluster 1d was observed in small proportions in the restored prairie and the corn corn rotation, while cluster 1a was observed in all prairie soils. Previous studies reported on the presence of members of cluster 3 and cluster 1b in soils harboring non-host plants (Rönkkö *et al.*, 1993, Mirza *et al.*, 2009). Our *nif*H gene targeted Illumina sequencing protocol retrieves only sequences of members of cluster 1 and cluster 3 (Rodriguez *et al.*, 2016). It is therefore not surprising that members of *Frankia* clusters 2 and 4 were not detected.

Illumina sequencing analysis of 16S rRNA fragments resulted in an abundance pattern dominated by Proteobacteria, Actinobacteria and Acidobacteria, respectively. These data are in agreement with a similar study on the same soils (Soman *et al.*, 2017). However, unlike the data presented in that study, our current results did not show any significant variation of phyla abundance across treatments. This might be due to several reasons, such as sampling strategies, seasonal variation or computational methods employed to analyze the data. Reads representing frankiae accounted for 0.1 to 0.9% of total reads, which represents the usual abundance of frankiae in the bacterial community encountered in soils (Hahn *et al.*, 1990, Samant *et al.*, 2016). Corn corn NPK treatment exhibited a relatively high abundance of reads representing frankiae accounting for nearly 1% of all reads. Clusters examination determined cluster 3 as the most prominent *Frankia* cluster in all treatments. They were also present in lower numbers within the cultivated and restored prairie soils but absent in native prairie soil. In agreement with the *q*PCR data, members of cluster 2 and cluster 4 were present.

All soil treatments exhibited large numbers of sequences related to *Frankia* sequences that could not be identified reliably. Unlike previous studies where novel reads matched *q*PCR-based quantification data (Ben Tekaya *et al.*, 2018), the current study did not show consistent abundance correlation between all three methods. It is therefore

tempting to conclude that results on *Frankia* diversity obtained by the 16S rRNA gene-based Illumina sequencing protocol are biased since that protocol is more suited to assess microbial community structure on a higher phylogenetic level than the species or genus level that we attempted for frankiae. An additional step might be to reassess sequences identified in the SILVA database to a more specific *Frankia* database to eliminate close relatives within the family Frankiaceae from misidentification. The accuracy of *q*PCR specificity and quantification has been demonstrated in earlier research (Samant *et al.*, 2014, Ben Tekaya *et al.*, 2017); therefore, differences with *nif*H based Illumina sequencing results might be due to differences in target genes specificity. Developing protocols such as *Frankia*-specific 23S rRNA gene targeted Illumina sequencing protocols or comparative analyses with digital PCR analyses might reduce such biases in future studies.

Acknowledgements

The authors are indebted to the Graduate College (Doctoral Research Support Fellowship to S. Ben Tekaya), and the Department of Biology at Texas State University for financial support.

REFERENCES

Battenberg K, Wren JA, Hillman J, Edwards J, Huang LJ & 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 Microb* **83**: e02661-16.

Ben Tekaya S, Guerra T, Rodriguez D, Dawson JO & Hahn D (2018) *Frankia* diversity in host-plant root nodules is independent of abundance or relative diversity of *Frankia* in corresponding rhizosphere soils. *Appl Environ Microbiol*. **84**:e02248-17

Ben Tekaya S, Ganesan AS, Guerra T, Dawson JO, Forstner MRJ & Hahn D (2017) SybrGreen- and TaqMan-based quantitative PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. *Appl Environ Microb* **83**:e02833-16.

Benson DR (1988) The genus *Frankia*: actinomycete symbionts of plants. *Microbiol Sci* **5**: 9-12.

Benson DR & Dawson J (2007) Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. *Physiol Plant* **130**: 318-330.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA & Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581-583.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N & Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* **108**: 4516-4522.

Caporaso JG, Lauber CL, Walters WA, *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621-1624.

Dawson JO (1986) Actinorhizal plants: their use in forestry and agriculture. *Outlook on Agriculture* **15**: 202-208.

Embarcadero-Jimenez S, Rivera-Orduna FN & Wang ET (2016) Bacterial communities estimated by pyrosequencing in the soils of chinampa, a traditional sustainable agro-ecosystem in Mexico. *J Soil Sediment* **16**: 1001-1011.

Fay BJ, Corrigan A & Murphy RA (2016) Short-term effects of mechanical drainage on fungal and bacterial community structure in a managed grassland soil. *Appl Soil Ecol* **101**: 93-100.

Gilbert JA, Jansson JK & Knight R (2014) The Earth Microbiome project: successes and aspirations. *BMC Biol* **12**: 69.

Hahn D, Kester R, Starrenburg MJC & Akkermans ADL (1990) Extraction of ribosomal RNA from soil for detection of *Frankia* with oligonucleotide probes. *Arch Microbiol* **154**: 329-335.

Higgins LM & Kennedy PG (2012) Symbiotic *Frankia* bacteria in *Alnus* forests in Mexico and the United States of America: is geographic location a good predictor of assemblage structure? *Botany* **90**: 423-431.

Khan SA, Mulvaney RL, Ellsworth TR & Boast CW (2007) The myth of nitrogen fertilization for soil carbon sequestration. *J Environ Qual* **36**: 1821-1832.

Li HZ, Liang TT, Chu QP, Xu FL, Li Y, Fu LL & Zhou B (2017) Effects of several in-feed antibiotic combinations on the abundance and diversity of fecal microbes in weaned pigs. *Can J Microbiol* **63**: 402-410.

Mashiane RA, Ezeokoli OT, Adeleke RA & Bezuidenhout CC (2017) Metagenomic analyses of bacterial endophytes associated with the phyllosphere of a Bt maize cultivar and its isogenic parental line from South Africa. *World J Microb Biot* **33**: 1-12.

Michelsen CF, Pedas P, Glaring MA, Schjoerring JK & Stougaard P (2014) Bacterial diversity in Greenlandic soils as affected by potato cropping and inorganic versus organic fertilization. *Polar Biol* **37**: 61-71.

Mirza BS, Welsh AK & Hahn D (2009) Growth of *Frankia* strains in leaf litter-amended soil and the rhizosphere of a non-actinorhizal plant. *FEMS Microbiol Ecol* **70**: 132-141.

Mirza BS, Welsh AK, Rieder JP, Paschke MW & Hahn D (2009) Diversity of frankiae in soils from five continents. *System Appl Microbiol* **32**: 558-570.

Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L & Misra AK (1996) Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. *Int J System Bacteriol* **46**: 1-9.

Nouioui I, Sbissi I, Ghodhbane-Gtari F, Benbrahim KF, Normand P & Gtari M (2013) First report on the occurrence of the uncultivated cluster 2 *Frankia* microsymbionts in soil outside the native actinorhizal host range area. *J Biosci* **38**: 695-698.

Odell RT, Melsted SW & Walker WM (1984) Changes in organic carbon and nitrogen of morrow plot soils under different treatments. 1904-1973. *Soil Sci* **137**: 160-171.

Rodriguez D, Guerra TM, Forstner MRJ & Hahn D (2016) Diversity of *Frankia* in soil assessed by Illumina sequencing of *nif*H gene fragments. *System Appl Microbiol* **39**: 391-397.

Rönkkö R, Smolander A, Nurmiaho-Lassila EL & Haahtela K (1993) *Frankia* in the rhizosphere of nonhost plants: A comparison with root-associated nitrogen-fixing *Enterobacter*, *Klebsiella* and *Pseudomonas*. *Plant Soil* **153**: 85-95.

Samant S, Amann RI & Hahn D (2014) Evaluation of the 23S rRNA gene as target for *q*PCR based quantification of *Frankia* in soils. *System Appl Microbiol* **37**: 229-234.

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.

Samant S, Sha Q, Iyer A, Dhabekar P & Hahn D (2012) Quantification of *Frankia* in soils using SYBR Green based *q*PCR. *System Appl Microbiol* **35**: 191-197.

Siles JA, Rachid CTCC, Sampedro I, Garcia-Romera I & Tiedje JM (2014) Microbial diversity of a mediterranean soil and its changes after biotransformed dry olive residue amendment. *PLoS One* **9**: e103035.

Soman C, Li DF, Wander MM & Kent AD (2017) Long-term fertilizer and crop-rotation treatments differentially affect soil bacterial community structure. *Plant Soil* **413**: 145-159.

Velde B & Peck T (2002) Clay mineral changes in the Morrow Experimental Plots, University of Illinois. *Clay Clay Miner* **50**: 364-370.

Chapter IV

FRANKIA DIVERSITY IN HOST-PLANT ROOT NODULES IS INDEPENDENT OF ABUNDANCE OR RELATIVE DIVERSITY OF *FRANKIA* IN CORRESPONDING RHIZOSPHERE SOILS

published as

Ben Tekaya, S., Guerra, T., Rodriguez, D., Dawson, J.O., Hahn, D. 2018. *Frankia* diversity in host-plant root nodules is independent of abundance or relative diversity of *Frankia* in corresponding rhizosphere soils. Appl. Environ. Microbiol. 84:e02248-17. https://doi.org/10.1128/AEM.02248-17.

ABSTRACT

Actinorhizal plants form nitrogen-fixing root nodules in symbiosis with soil-dwelling actinobacteria within the genus *Frankia*, and specific *Frankia* taxonomic clusters nodulating plants in corresponding host infection groups. In same soil microcosms, we observed some host species were nodulated (*Alnus glutinosa*, *A. cordata*, *Shepherdia argentea*, *Casuarina equisetifolia*) while others were not (*Alnus viridis*, *Hippophaë rhamnoides*). Nodule populations were represented by eight different sequences of *nif*H gene fragments. two of these sequences characterized frankiae in *S. argentea* nodules, and three frankiae in *A. glutinosa* nodules. frankiae in *A. cordata* nodules were represented by five sequences, one of which was also found in nodules *A. glutinosa* and *C. equistifolia*, while another one was detected in nodules from *A. glutinosa*. Quantitative PCR assays showed vegetation generally increased the abundance of frankiae in soil, independent of the target gene (i.e. the *nif*H or the 23S rRNA gene). Targeted Illumina sequencing of *Frankia*-specific *nif*H gene fragments detected 24 unique sequences from all rhizosphere soils four of which were also found in nodules, while the remaining four sequences in nodules were not found in soils. Seven of the 24 sequences from soils represented more than 90% of the reads obtained in most samples, with the two most abundant sequences not found in root nodules and only two of these sequences detected in nodules. These results demonstrate large differences between detectable *Frankia* populations in soil and those in root nodules, therefore suggesting that

root nodule formation is not a function of abundance or relative diversity of specific *Frankia* populations in soils.

KEYWORDS: abundance - actinorhiza - *Frankia* – Illumina – *nif*H - *q*PCR - quantification - soil

INTRODUCTION

The genus *Frankia* resembles nitrogen- and non-nitrogen-fixing actinobacteria that live in soils (1) and can form root nodules in symbiosis with a variety of woody plants (2, 3). Root nodule formation requires interactions between members of specific *Frankia* taxonomic clusters and plants of specific host infection groups. Frankiae of cluster 1 form nodules on plants of the genera *Alnus*, *Morella* and *Comptonia*, with a subgroup specifically nodulating the genera *Casuarina* and *Allocasuarina.* Frankiae of cluster 2 form nodules with *Ceanothus*, *Cercocarpus*, *Chamaebatia*, *Coriaria*, *Datisca*, *Dryas*, and *Purshia*, while those of cluster 3 form nodules on members of the genera *Elaeagnus*, *Hippophaë*, *Shepherdia*, *Myrica*, *Morella* and *Colletia* (2, 4). The last cluster, cluster 4, represents atypical, generally non-nitrogen-fixing frankiae from a variety of different host plants including *Alnus*, *Ceanothus*, *Coriaria*, *Datisca*, *Purshia* and *Elaeagnus* (5-8).

The compatibility of frankiae to plants of specific host infection groups is very well documented (9-11), as is intrageneric variation in host plant compatibility with specific *Frankia* populations (12-15). Host plant species have been shown to determine the selection of *Frankia* strains from soil for potential nodule formation (14, 15), with nodules of different plant species from the same genus harboring unique *Frankia* populations, and none of them capturing the entire diversity of nodule-forming frankiae (15). While large differences in nodule-forming frankiae were observed for different soils on the same host plant (14), diversity in nodule-forming frankiae from individual soils was found to be low with generally one or two populations dominating in nodules formed (14, 16). Comparative analyses of *Frankia*-specific *nif*H gene clone libraries from soil with root nodule populations often displayed large differences between sequences

retrieved from clone libraries and those obtained from nodules, with matching sequences only rarely encountered (17) or not detected at all (14).

Many of these data have been retrieved from plant bioassays in which nodule-forming frankiae were identified on roots of a specific host plant after inoculation with a soil slurry (16, 18-21). Since root nodules represent a natural locale of enrichment of usually one *Frankia* population, molecular tools such as polymerase chain reaction (PCR)-assisted sequence analyses can be used to retrieve information on *Frankia* populations in root nodules. The case is different for highly complex and diverse environments such as soil with *Frankia* populations present in small numbers in a large and complex microbial community (22) in which even molecular tools such as gene clone library analyses provide limited information (14, 17). The development of *Frankia*-specific quantitative polymerase chain reaction (*q*PCR) (23-25) or targeted Illumina sequencing (26) analyses now provides increasingly sophisticated molecular tools for the analyses of *Frankia* populations in soil.

The aim of this study was to assess *Frankia* diversity in root nodules of different host plant species growing on the same soil and to relate this diversity to the abundance and relative distribution of indigenous frankiae in rhizosphere soils. Different actinorhizal plant species, including *Alnus glutinosa*, *A. cordata*, *A. viridis*, *Casuarina equisetifolia*, *Shepherdia argentea*, and *Hippophaë rhamnoides* as well as the non-actinorhizal plant species *Betula pendula* were grown in soil microcosms which, together with a non-vegetated control, were analyzed by *Frankia*-specific *q*PCR and Illumina sequencing to assess the abundance and relative distribution of indigenous frankiae in rhizosphere soils. Results were related to *Frankia* diversity in root nodules analyzed by PCR-assisted sequence analyses.

RESULTS

Qualitative Analyses of *Frankia* in Root Nodules by Sequencing

Healthy, surviving plants were growing with heights between 15 and 35 cm, depending on the species, and roots generally filling the entire microcosm. Root nodules were obtained on four of the seven plant species used for analyses, i.e. *Alnus glutinosa*

(30 ± 3 nodules per microcosm, n=3 microcosms), *Alnus cordata* (57 nodules, n=1 microcosm), *Casuarina equisetifolia* (1 nodule, n=3 microcosms), and *Shepherdia argentea* (23 ± 3 nodules, n=2 microcosms), while *Alnus viridis* (n=3 microcosms), *Hippophaë rhamnoides* (n=3 microcosms) and the non-host plant *Betula pendula* (n=3 microcosms) did not produce nodules (Figure 1). Sequences were obtained from 141 lobes (51 *A. glutinosa*, 46 *A. cordata*, 1 *C. equisetifolia*, and 43 *S. argentea*), with a total of eight different sequences. Two distinct sequences characterized frankiae in nodules from *S. argentea*, three sequences characterized frankiae in nodules from *A. glutinosa*, and five sequences characterized frankiae in nodules from *A. cordata* of which two were shared with *A. glutinosa*, and one sequence identical to that obtained from *C. equisetifolia* (Figure 1).

Maximum likelihood analyses revealed that all sequences obtained from *S. argentea* represented frankiae of cluster 3, with high similarity values of a minor percentage of sequences (7%) to *Frankia elaeagni* BMG5.12 (99.6 % similarity) and the majority of sequences (93%) to strain EUN1f (98.3% similarity) (Figure 1). Frankiae of cluster 3, related to strain EAN1pec (99.4% similarity), were also found in one nodule from *A. glutinosa* corresponding to 2% of the sequences in nodules of *A. glutinosa*. Sequences present in nodules from both *A. glutinosa* and *A. cordata* were either closely related to *Frankia* strain ARgP5 (99.0% similarity) representing cluster 1d (62% of the sequences from *A. glutinosa*, and 22% of the sequences from *A. cordata*), or formed a separate branch within cluster 1 characterized by sequences of uncultured frankiae without cultured relative in the database, here tentatively named cluster 1e (36% of the sequences from *A. glutinosa*, and 62% of the sequences from *A. cordata*) (Figure 1). Two other distinct sequences from endophytes in nodules from *A. cordata* (7% of the sequences from *A. cordata* each) as well as the one from *C. equisetifolia* were assigned to this cluster as well.

Quantitative Analyses of *Frankia* in Soil by *q*PCR

Microcosms with different vegetation harbored different numbers of *Frankia* cells, with a general increase in numbers with any vegetation compared to a non-vegetated control ($p < 0.001$, except for *A. viridis* ($p = 0.003$), and a non-significant value for *H. rhamnoides* ($p = 0.1$)). Numbers about 4-5 times higher than those 10^6 cells [g

soil]$^{-1}$ in the non-vegetated control soil were obtained for soils with *A. glutinosa*, *A. cordata* and *C. equisetifolia*, while those for the remaining plant species ranged between 2 and 3 x 10^6 cells [g soil]$^{-1}$ (Fig. 2).

Within treatments, results were identical independent of the target gene, i.e. the *nif*H gene allowing the retrieval of information on cluster 1 and 3 together only, and the 23S rRNA gene that provides information on all four clusters representing the genus *Frankia* (Fig. 2). These results suggest the absence of clusters 2 and 4 which was confirmed by specific analyses that only detected frankiae of clusters 1b, the newly assigned cluster 1e, and cluster 3, while clusters 1a/d, 1c, 2, and 4 were not detected (Fig. 2). All three clusters were present only in the non-vegetated control soil, and soil vegetated with *A. glutinosa*, while soils with the remaining plant species generally harbored cluster 3 and a smaller population of cluster 1e frankiae. Post treatment microcosm soils with *C. equisetifolia* were an exception because they harbored cluster 1b instead of cluster 3 frankiae, and again a smaller population of cluster 1e frankiae (Fig. 2).

Qualitative Analyses of *Frankia* in Soil by Illumina Sequencing

Numbers of reads between 273,068 (± 14,956) (*A. glutinosa*) and 425,752 (*A. cordata*) were identified as representing the genus *Frankia* (Fig. 3) with a total of 24 unique sequences (LT934508 – LT934531) retrieved from all treatments together (Table 1). Seven of these sequences were present in the majority of samples, with these reads combined representing generally more than 90% of all reads obtained [except for *A. cordata* (78.1%) and *S. argentea* (85.4%)]. The majority of reads were represented by read 147378 (generally between 15 to 30%) and read 170193 (generally 45 to 60%), both of which identified cluster 3 frankiae as closest relatives but were not found in root nodules of our test plant species (Table 1). Only two of the seven sequences were identical to those found in nodules, with read 101236 representing cluster 3 frankiae (generally 3 to 5 % of the reads, except for *S. argentea*) and read 148038 representing cluster 1e frankiae (generally 4 to 9% of the reads) (Table 1). Read 101236 was present in soil and nodules of *S. argentea*, both in high percentages (20% of reads from soil, 93% of the nodules) (Figs. 1, 3), while read 148038 was detected in soil and nodules of *A. glutinosa* and *C. equisetifolia*, but not in soils with *A. cordata* even though it was found

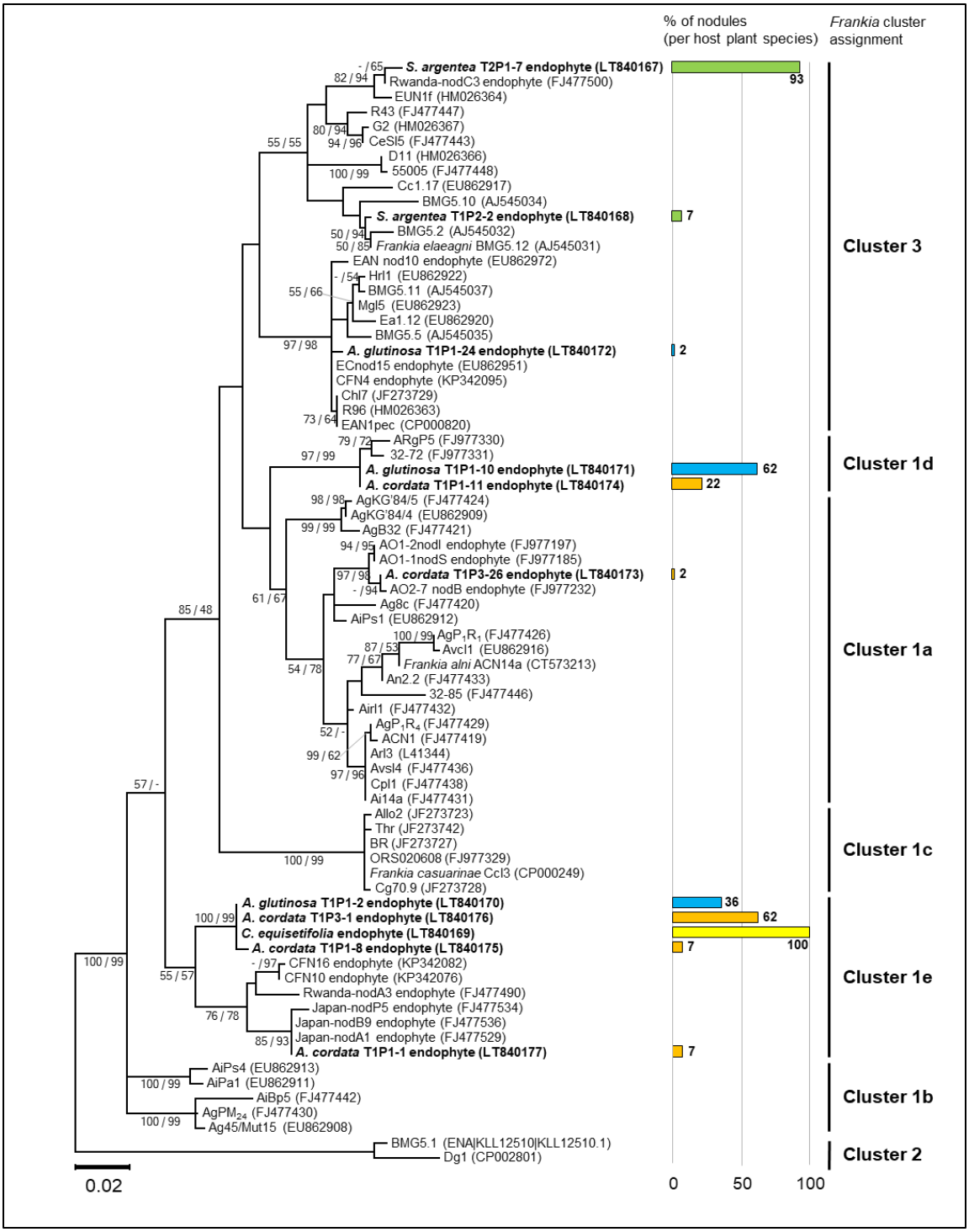

FIG 9 Neighbor-joining topology showing sequence relationships for selected *Frankia* strains and uncultured endophytes in root nodules to demonstrate *Frankia* cluster assignments. Numbers above the branches represent the bootstrap values from maximum likelihood (1,000 replicates) /neighbor-joining (10,000 replicates) bootstrap analyses for clades with >50% bootstrap support. Support values less than 50% are indicated by a dash. These two phylogenetic criteria resolved generally similar topologies. The bar chart shows the percentage of nodules assigned to known *Frankia* strains per host plant (green, *S. argentea*; blue, *A. glutinosa*; orange, *A. cordata*; yellow, *C. equisetifolia*).

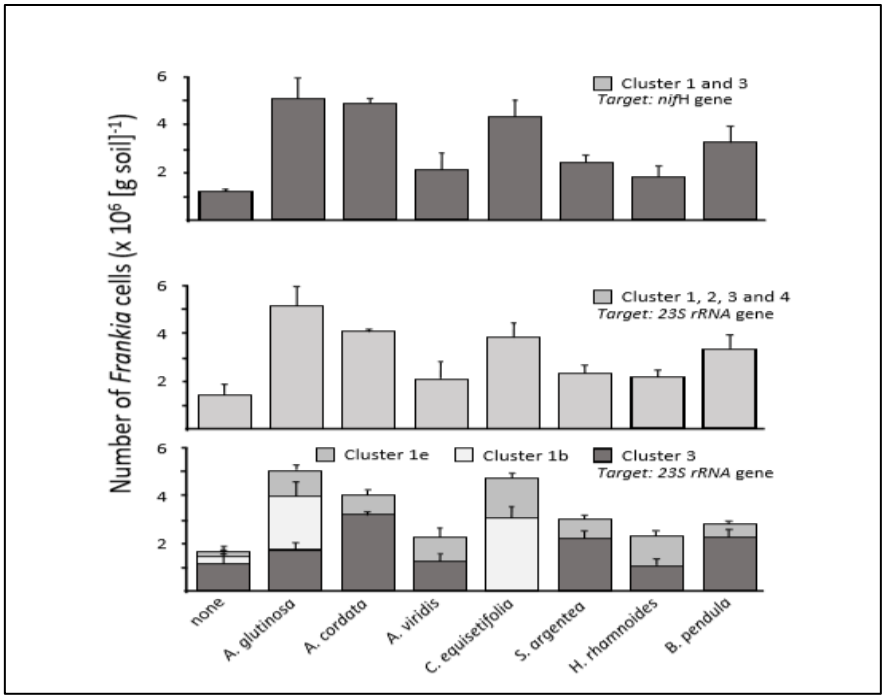

FIG 10 SybrGreen based *q*PCR quantification of specific *Frankia* clusters in soils vegetated with host trees (*Alnus glutinosa*, *Alnus cordata*, *Alnus viridis*, *Casuarina equisetifolia*, *Shepherdia argentea* and *Hippophaë rhamnoides*), or non-host trees (*Betula pendula*). A control was kept non-vegetated. Quantification (from top to bottom) used *nif*H gene fragments as a target detecting clusters 1 and 3, or 23S rRNA gene fragments generated with primer combinations detecting the genus *Frankia*, i.e. all clusters 1, 2, 3 and 4, or primer combinations specific for clusters 1a/d, 1b, 1c, 1e, 2, 3, and 4 (presented as the sum of the individual clusters and subgroups detected). Only frankiae of clusters 1b, 1e and 3 were detected in these soils, while the remaining clusters and subgroups remained undetected

in 62% of the nodules on *A. cordata* (Figs. 1, 3). Of the reads present in minor abundance (i.e. just above the threshold of 1%), two were found to be identical to those in root nodules, one representing a cluster 3 *Frankia* population in a nodule on *A. glutinosa*, and another one representing cluster 1e in nodules on *A. cordata* but again not present in soil vegetated with *A. cordata* (Table 1). The remaining sequences were generally most closely related to sequences representing cluster 3 frankiae, with 2 exceptions that represented cluster 1a and 1b frankiae.

Table 6 Similarity of *Frankia*-specific *nif*H reads from soil microcosms to sequences of uncultured frankiae in root nodules of different plant species grown in these soils, and to sequences of closest *Frankia* isolates or clones

Plant species	*Shepherdia argentea*		*Casuarina equisetifolia*	*Alnus glutinosa*			*Alnus cordata*					Closest *Frankia* sequence	Cluster assignment
Sample	T_2P_1-7	T_1P_2-2	T_1P_2-1	T_1P_1-2	T_1P_1-10	T_1P_1-24	T_1P_3-26	T_1P_3-1	T_1P_1-11	T_1P_1-8	T_1P_1-1		
Reads present in the majority of samples, with abundance of combined reads up to 96% of all reads													
5797		97.7%										98.1% (Cc1.17)	3
34873		95.8%										96.7% (BMG5.15)	3
101236	**100%**											96.7% (EUN1f)	3
147378		94.9%										95.8% (G2)	3
170193	95.8%	95.8%				95.8%	95.8%					96.8% (D11)	3
10509			95.3%	95.3%				95.3%		95.3%		94.9% (EUN1f)	3
148038			**100%**	**100%**				**100%**				96.4% (CFN10)	1e
Reads present in few of the samples, generally at low abundance (less than 1% of all reads for each sample)													
139714						**100%**						100% (EAN1pec)	3
137910	97.2%											96.3% (Chl7)	3
203721		98.6%										99.1% (Cc1.17)	3
115049		95.3%										96.3% (G2)	3
115039		96.7%				96.7%						97.2% (Cc1.17)	3
114592		96.7%				94.9%						97.2% (Cc1.17)	3
106934					95.8%	95.8%			95.8%			95.8%, EAN1pec	3
221741	97.7%											97.7% (G2)	3
78840					96.7%	96.7%	96.7%					96.8% (D11)	3
75193	97.7%											97.2% (Cc1.17)	3
57086	96.7%											100% (EUN1f)	3
43921		96.3										96.3% (BMG5.12)	3
39637		99.1										99.1% (BMG5.12)	3
16382	95.8%		95.8%	95.8%				95.8%		95.8%		96.7% (EUN1f)	3
69195							97.2%					100% (ArI3)	1a
192360			96.3%	96.3%				96.3%				96.3% (Ai7a)	1b
255602											**100%**	100% (Japan-nodA1)	1e

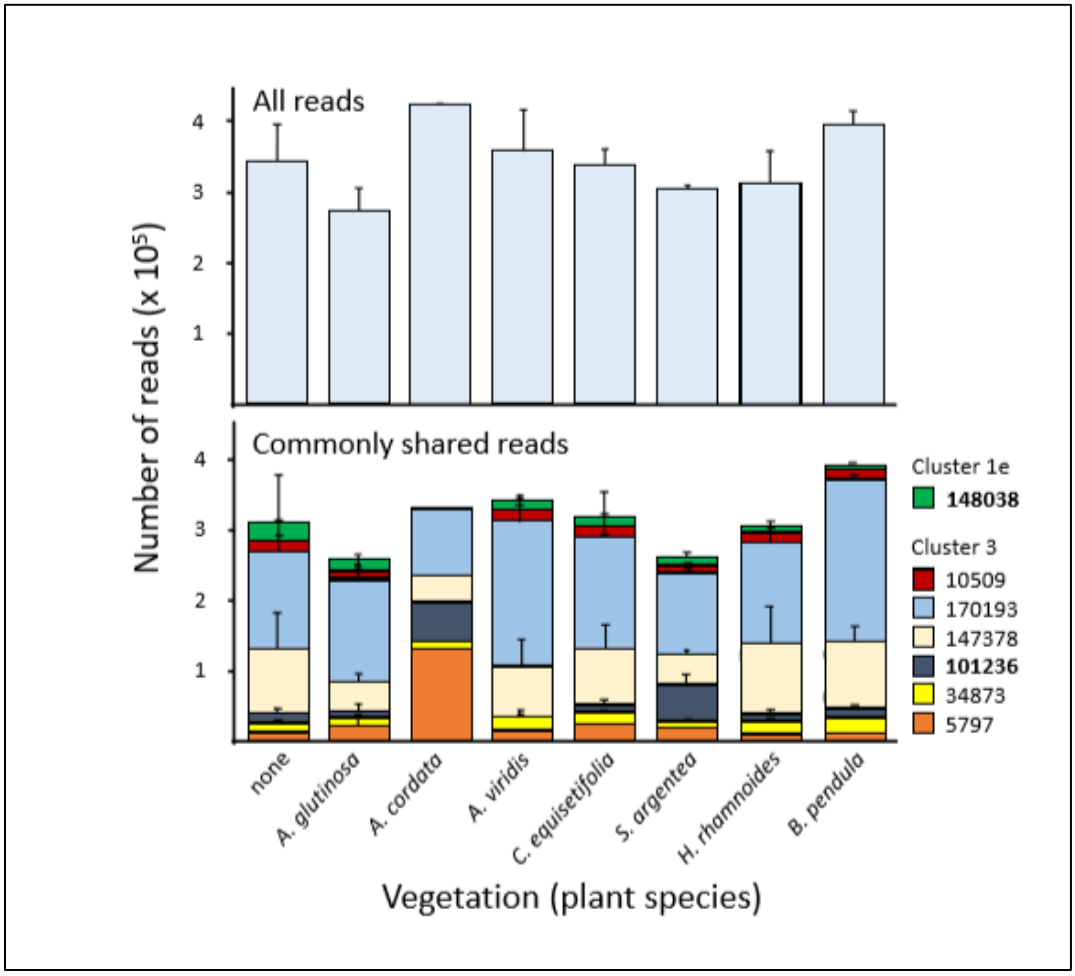

FIG 11 Number of reads representing *Frankia*-specific *nif*H gene fragments obtained by Illumina sequencing for DNA extracts obtained from soils vegetated with host trees (*Alnus glutinosa*, *Alnus cordata*, *Alnus viridis*, *Casuarina equisetifolia*, *Shepherdia argentea* and *Hippophaë rhamnoides*), or non-host trees (*Betula pendula*). A control was kept non-vegetated. Quantification of major specific reads (lower panel) that were present in the majority of samples, with these reads combined representing generally more than 90% of all reads obtained [except for *A. cordata* (78.1%) and *S. argentea* (85.4%)] revealed that most reads represented *Frankia* populations of cluster 3, and only few those of cluster 1e, while reads representing other clusters were absent or present in numbers below 1%.

DISCUSSION

Host plant species as well as *Frankia* populations seem to affect nodulation capacity and subsequently plant growth performance as demonstrated in inoculation studies where spore-producing nodule homogenates produced significantly more nodules than non-spore-producing nodule homogenates on different alder species (27, 28). Spore-producing nodule homogenates produced twice as many nodules with *A. glutinosa* than

with *A. incana* (27), while *A. crispa* produced more than twice as many nodules as *A. rubra* and six times as many nodules as *A. rugosa* (28). Differential root nodule formation of different alder species has been reported before, with *A. rubra* producing more nodules than *A. incana* which ultimately produced more nodules than *A. glutinosa* (12). While the opposite sequence was found for different soils with more nodules formed on *A. glutinosa* than on *A. incana* (13), none of these studies reported the lack of nodulation of some species (*A. viridis* and *H. rhamnoides*) while others of the same host infection group produced abundant nodules (*A. glutinosa*, *A. cordata*, *S. argentea*) as in our study. Above- and below-ground biomass of both *A. viridis* and *H. rhamnoides* was much smaller than that of the other plant species (data not shown), and *Frankia* populations were only marginally or not significantly different from those in soils that were kept non-vegetated [*A. viridis* ($p = 0.03$) and *H. rhamnoides* ($p = 0.102$)] while *Frankia* populations in soils vegetated by all other plant species significantly increased in abundance (all $p < 0.001$). These data indicate potential experimental effects that reduced plant growth performance of both *A. viridis* and *H. rhamnoides*, which subsequently might have affected *Frankia* growth in soil and rhizosphere and finally root nodule formation.

Although our study was not designed to assess quantitative differences in nodule formation on different host plant species, it clearly established differences in diversity and abundance of specific *Frankia* populations in nodules and in soil. Frankiae in root nodules on each plant species were generally dominated by one or two populations with additional populations present in a few nodules only. Two nodules were found with *Frankia* populations not typically detected on the respective plant host species. One nodule of *A. glutinosa* harbored a cluster 3 frankiae instead of the usual cluster 1 populations, confirming previous reports of occasional detections of cluster 3 frankiae in nodules of the Betulaceae (2). A single nodule was found on *C. equisetifolia*, with a *Frankia* population representing cluster 1e. *Casuarina* species usually produce nodules with *Frankia* of cluster 1c (25) meant to represent a group of highly-specialized frankiae that depend on co-introduction with their exotic host plant species outside their native range (29, 30). Consequently, *Casuarina*-infective *Frankia* were not expected to form nodules in soils that have never supported *Casuarina* spp. (22, 24, 31). Cluster 1e

frankiae, that can be distinguished from other cluster 1 frankiae by highly distinctive signature sequences on a 23S rRNA gene insertion as demonstrated in this study, have been observed in previous studies in nodules of different *Morella* species growing on acidic soils in South Africa (32), as well as in nodules from *Morella pensylvanica* used as capture plants in bioassays to assess *Frankia* diversity in soils from different countries throughout the world (e.g. from Japan and Rwanda) (17).

Soil *Frankia* populations in vegetated microcosms showed higher abundance values than in non-vegetated soil indicating growth enhancing effects of vegetation on frankiae as demonstrated previously (22, 31, 33). This result confirms previous results documenting that all *Frankia* strains tested so far grow in the rhizosphere of their host plants (34) but also potentially in that of other plants (34-36). Diversity of *Frankia* populations, however, was relatively low with frankiae of cluster 1b, cluster 1e and cluster 3 present in non-vegetated controls, and generally cluster 1e and cluster 3 frankiae in vegetated microcosms only. Previous studies reported on low diversity of *Frankia* populations in different soil environments, e.g. the presence of cluster 1b only in wet or even water-logged soils in natural stands of *A. glutinosa* (24), clusters 1b and 3 in forest soils with *A. glutinosa* or non-host plants including *B. nigra* (22, 25), clusters 1a, 1b and 3 in forest soils with *A. glutinosa* (22), clusters 1a and 3 in microcosms with *A. glutinosa* or *C. equisetifolia* (31), or clusters 1b, 2, 3, and 4 in prairie soils with *Ceanothus* as potential host plant species (25). These studies suggest soil-specific environmental effects on *Frankia* populations in soils, with frankiae of clusters 1b and 3 being quite ubiquitous and independent of the presence of host plant species. Cluster 1b frankiae had been shown to grow with leaf litter of host and/or non-host plants (34, 37), and thus could be adapted to carbon resources provided by the decomposition of plant material. Similarly, cluster 3 frankiae grew in bulk soil, the rhizosphere and with leaf litter, independent of matric potential and plant species, and thus represent a group with broad physiological adaptations (31). In contrast, cluster 2 frankiae were not detected. Prior to European settlement of the study region in the early nineteenth century, the landscape encompassing the site was tallgrass prairie and oak savanna having the native actinorhizal shrub *Ceanothus americanus* L. as a common component. Existing native prairie remnants contain *C. americanus* and the corresponding, detectable soil frankiae specific

to cluster 2 (25). It has been speculated that cluster 2 frankiae depend on their host plant for growth and establishment (25, 38), and thus the conversion of native vegetation almost entirely to agricultural and urban uses, primarily cultivation of maize, legume hay crops and, in recent decades, soybeans may have resulted in the loss of cluster 2 frankiae at this site and other cultivated sites in this region.

*q*PCR data for soil *Frankia* populations were not exactly matched by targeted Illumina sequencing that retrieved sequences representing clusters 1a, 1b, 1e and 3, with all samples dominated by cluster 3 sequences, a small numbers of cluster 1e reads and very similar composition of specific reads (Fig. 3). Root nodule formation does not seem to be a function of abundance of specific *Frankia* populations in soils. In nodules of *A. cordata*, more than 70% of the nodules harbored frankiae of cluster 1e detected by *q*PCR, but not by Illumina sequencing. The remaining 30% of nodules harbored cluster 1a frankiae that were not detected by either method. Similar results were obtained for nodules of *A. glutinosa* that harbored cluster 1e frankiae as expected from detection in soils by both methods. However, most nodules contained cluster 1a frankiae that were not detected in soils by any of the two methods, while cluster 1b frankiae detected as a major population by *q*PCR did not occur in nodules at all. On a more specific scale than cluster assignments, eight distinct sequences were found in root nodules, but only 4 were found back in soil by Illumina sequencing. Although specific reads from soil samples were detected in nodules, read abundance in soil samples did not reflect their presence in root nodules. These results corroborate previous studies where clone libraries of *Frankia*-specific *nif*H gene fragments from soils revealed large differences in cluster assignments to sequences obtained from nodules, with assignments to the same cluster only rarely encountered for individual soils (17).

These results demonstrate large differences between detectable *Frankia* populations in soil and those in root nodules indicating the inadequacy of bioassays for the detection of specific nodulating frankiae in soil and the role of plants in the selection of frankiae from soil for root nodule formation. The data also highlight the necessity to use a combination of different assessment tools to adequately address methodological constrains that could provide contradictory datasets, or quantify apparently small, threshold levels of infectious soil frankiae for specific hosts. Future studies could

therefore use the availability of pure culture representatives of different *Frankia* clusters to assess the reliability of individual assessment tools including *q*PCR analyses and Illumina sequencing approaches, for both qualitative and quantitative analyses of *Frankia* populations in soils and nodules.

MATERIALS AND METHODS

Soil microcosms

Soil microcosms were established each in 50-ml falcon tubes using 70 g of soil (dry wt.) collected from a dense stand of tall (4-5 m) *Elaeagnus angustifolia* plants established on a Drummer silty clay loam (taxonomic classification = fine-silty, mixed, superactive, mesic Typic Endoaquoll) in Illinois, USA. The climate is continental with mean annual precipitation of 940 mm, and mean annual air temperature of 11°C. The fertile soil formed under tallgrass prairie vegetation at an elevation of 214 m, is slightly acidic (pH = 6.9) and rich in organic matter (circa 3 percent). The site is adjacent to a temperate deciduous forest and has a history of row cropping, but not in the last 25 years (40.207797, -88.366524). Seeds of all plant species were surface sterilized in 3% hydrogen peroxide for 10 min., washed twice in sterile water for 10 min. each and then germinated on a water agar. Three microcosms were set up for each plant species with three seedlings per microcosm, and three microcosms without plants were used for comparison. All microcosms were kept at 23°C with a photoperiod of 16/8 h (day/night, respectively), and a matric potential of -0.005 MPa using round-bottom ceramic suction tubes (8 cm long with an outer diameter of 6 mm [#0652X02-B01M1], Soilmoisture Equipment Corp., Santa Barbara, CA) and 50 cm water columns (18).

Microcosms were harvested after 7 months, and root nodules as well as rhizosphere soil -defined as the soil adhering to plant roots when the plants were removed from the remaining bulk soil- were collected. Root nodules were kept in 100% ethanol at -20°C, while soil was stored at -20°C until further analysis.

Qualitative Analyses of *Frankia* in Root Nodules by Sequencing

Root nodules generally consisted of single lobes which were homogenized in 1 ml of sterile water with a mortar and a pestle. The homogenate was centrifuged at 14,000 x g for 3 min., and the pellet washed once in 1 ml of 1% sodium pyrophosphate. The pellet was then dissolved in 95 µl of sterile water, to which 5 µl of Proteinase K solution (10 mg ml^{-1}) was added (15, 25). This solution was transferred to a 0.5 ml PCR tube and incubated in a thermocycler at 37°C for 20 min., after which 0.5 µl of 10% SDS solution was added. After incubation at 37°C for 30 min., the Proteinase K was inactivated by incubation at 80°C for 20 min.

Primers nifHf1 ($^{5'}$GGC AAG TCC ACC ACC CAG C) and nifHr ($^{5'}$CTC GAT GAC CGT CAT CCG GC) (39) were used to amplify 606 bp *nif*H gene fragments in a reaction volume of 50 µl, containing 1 µl of a 10 mM dNTP mix, 1 µl each primer (0.2 µM), 8.2 µl BSA (30 µg ml^{-1}), 5 µl of 10 x PCR buffer with 15 mM $MgCl_2$, 2 µl root nodule lysate, and 0.2 µl *Taq* DNA polymerase (5 U μl^{-1}; GeneScript, Piscataway, NJ). Amplification conditions included an initial incubation at 96°C for 10 minutes, followed by 35 rounds of temperature cycling (96°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds) and a final 7 minute incubation at 72°C. Amplified *nif*H gene fragments were cleaned using Shrimp Alkaline Phosphatase and Exonuclease I (Affymetrix, Santa Clara, CA) following the manufacturer's protocols, and then sequenced bidirectionally using BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA), using primers nifH1 and nifHr. Sequences were analyzed on a 3500 Genetic Analyzer (Life Technologies, Carlsbad, CA), and representative sequences deposited at GenBank under accession numbers LT840167 to LT840177.

Sequences were assembled in Geneious 9.1.4 (Biomatters Ltd, Auckland, New Zealand), and checked in GenBank/EMBL databases using the BLAST algorithm (40). Representative sequences from confirmed *Frankia* strains of all clusters representing nitrogen-fixing frankiae were added from GenBank/EMBL databases, aligned by the Geneious alignment tool and trimmed to 516 bp long to match those in the database (15). Geneious was also used to build a neighbor-joining topology based on a distance matrix generated using HKY85 correction from the resulting alignment. Clade confidence values were estimated using 10,000 bootstrap iterations. Maximum likelihood (ML) analyses (41) to evaluate the relationship among the sequences were carried out through the

Cyber-Infrastructure for Phylogenetic Research project portal (www.phylo.org) (42). Model parameters were estimated via GTRCAT with 25 rate categories (43) and used as input in a ML heuristic search using RAxML version 8.2.10 (44). Rapid bootstrapping was conducted (45) for 1,000 ML iterations while searching for the best-scoring ML tree.

Quantitative Analyses of *Frankia* in Soil by *q*PCR

DNA was extracted from rhizosphere samples in triplicate 250 µg soil samples using the SurePrep™ Soil DNA Isolation Kit (Fisher Scientific, Houston, TX) (23). Soils had been amended with *Salmonella* Typhimurium (ATCC14028) cells to account for extraction efficiencies as described previously (23). DNA was used in SybrGreen-based qPCR reactions to quantify clusters or subclusters of *Frankia* with either *nif*H gene fragments (cluster 1 and 3 together) or 23S rRNA gene fragments (cluster 1, 2, 3 and 4 representing the entire genus, and clusters 1a/d, 1b, 1c, 2, 3 and 4 separately) as targets (25). Amplicons of 23S rRNA gene fragments [about 240 bp generated using primers 23Fra1533f ($^{5'}$GTT GAT ATT CCC GTA CCG) and 23Fra1769r ($^{5'}$GGC TCG GCA TCA GGT CTC AG)] (25) were also used to reanalyze endophytes in nodules that had been assigned to a phylogenetically distinct group of yet uncultured frankiae without cultured relative (referred to as cluster 1e) in our nodule analyses (Fig. 1). PCR and sequence analyses were carried out as described above, and selected sequences deposited at Genbank under accession numbers LT840178 to LT840183. Sequences were aligned in Geneious 9.1.4, and the alignment amended with sequences of target and non-target organisms from Genbank databases. This alignment was used to design a forward primer specific for 23S rRNA gene sequences of cluster 1e, i.e. primer 113f ($^{5'}$GGA TGT GTG TGT GAG GTC GGG A), respectively, which could be used with reverse primer 23Fra1769r in SybrGreen-based *q*PCR (24, 25). Primer 113f had 3 mismatches to *Frankia* strain ARgP5 (previously assigned to cluster 1d), and 5 or more mismatches to sequences of all other previously tested strains and uncultured endophytes in root nodules (24, 25). The potential of self- and hetero-dimer formation, target specificity and *q*PCR conditions were checked as described before (25). Briefly, low potential of self- and hetero-dimer formation was checked in OligoAnalyzer 3.1 (www.idtdna.com/calc/analyzer), while target specificity was checked using TestPrime

1.0 and TestProbe 3.0 (46) from the SILVA rRNA database project (www.arb-silva.de, accessed 07/29/2017) (47). TestProbe revealed high specificity of primer 113f, with at least 4 or 5 mismatches to non-target organisms (*Sulfolobus* sp. and *Haemophilus influenza*, respectively), while TestPrime on the combination of primers 113f and 23Fra1769r generally retrieved organisms showing at least 9 mismatches (5 and 4) to primers 113f and 23Fra1769r, respectively. Sequences of target organisms for cluster 1 e were not available in the database, and neither TestPrimer nor TestProbe retrieved *Frankia* sequences based on the search limit of 5 mismatches.

Sybr Green-based analyses were carried out in triplicate in a total volume of 10 µl containing 5 µl of SsoADV SybrGreen Mix (BioRad, Hercules, CA), 0.125 µl of forward and reverse primers (100 nM each), and 1 µl of DNA template using an initial denaturation at 95°C for 5 minutes, and 40 cycles of denaturation at 95°C, annealing at 62, 64 or 66°C depending on the primer combination, and extension at 72°C, each for 30 seconds (22, 24). Annealing temperature for primer 113f was 64°C. The amplifications were followed by melting curve analyses. Results of all analyses were corrected for extraction efficiencies determined as the ratio of inoculated *Salmonella* cells detected by *q*PCR-based quantification of a 268-bp *inv*A gene fragment before and after extraction (23).

Qualitative Analyses of *Frankia* in Soil by Illumina Sequencing

A second set of triplicate rhizosphere samples was used to extract DNA without addition of *Salmonella* cells. DNA was used in a nested PCR using the *nif*H gene of *Frankia* as target for amplification as described before (26). Briefly, 606-bp fragments were retrieved using primers nifH1f and nifHr as described above for root nodule endophytes. These were cleaned using the UltraClean® 15 DNA Purification Kit (Mo Bio Laboratories, Inc., Carlsbad, CA) and quantified on a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA) using the Agilent DNA 7500 Kit. Dilutions of these amplicons were used as templates in nested PCR reactions using modified primers nifHf1 and nifH269 to create a 263 bp product (48). In the nested reactions, primer nifH269 was barcoded, and both primers modified with Illumina adapters, primer pads, and linkers (26, 49). Samples were analyzed at the Genomic Sequencing and Analysis Facility

(GSAF) at the University of Texas, Austin, on an Illumina MiSeq with paired end 250 bp reads using the respective sequencing and index sequence primers (26). Illumina sequence data were accessioned into the NCBI Sequence Read Archive (SRP119405).

Illumina sequence data were analyzed as described previously (26). Briefly, paired-end reads were aligned in PEAR 0.9.6 with default parameters and then parsed and filtered in QIIME 1.8.0 (50) using *split_library.py* accepting reads with a Phred quality score of 26 or better (-q 25). The *nif*H gene sequence from *Frankia* strain EuIK1 (U53362) was used as a blast reference in addition to the seq.fna file from *split_library.py* output to run *exclude_seqs_by_blast.py* to exclude any chimera or truncated reads due to PCR errors (26). The output file (matching.fna) served to get unique reads by running *pick.denovo_otus.py* with a similarity threshold set as 1.0. Sequences representing less than 1% of the total numbers of reads per sample were removed using the resulting biom table and a customized R script (26). The collected unique sequences were aligned in Geneious and compared to *nif*H gene sequences of confirmed *Frankia* strains or uncultured endophytes from nodules or soils obtained from GenBank.

Acknowledgements

The authors are indebted to the Graduate College (Doctoral Research Support Fellowship to S. Ben Tekaya), and the Department of Biology at Texas State University for financial support.

REFERENCES

1. Chaia EE, Wall LG, Huss-Danell K. 2010. Life in soil by the actinorhizal root nodule endophyte *Frankia*. A review. Symbiosis 51:201-226.
2. Benson DR, Dawson J. 2007. Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiol Plant 130:318-330.
3. Dawson JO. 1986. Actinorhizal plants: their use in forestry and agriculture. Outlook on Agriculture 15:202-208.

4. Dobritsa SV. 1998. Grouping of *Frankia* strains on the basis of susceptibility to antibiotics, pigment production and host specificity. Intl J System Bacteriol 48:1265-1275.
5. Hahn D, Starrenburg MJC, Akkermans ADL. 1988. Variable compatibility of cloned *Alnus glutinosa* ecotypes against ineffective *Frankia* strains. Plant Soil 107:233-243.
6. Mirza MS, Hahn D, Akkermans ADL. 1992. Isolation and characterization of *Frankia* strains from *Coriaria nepalensis*. System Appl Microbiol 15:289-295.
7. Mirza MS, Hahn D, Dobritsa SV, Akkermans AD. 1994. Phylogenetic studies on uncultured *Frankia* populations in nodules of *Datisca cannabina*. Can J Microbiol 40:313-8.
8. Baker D, Newcomb W, Torrey JG. 1980. Characterization of an ineffective actinorhizal microsymbiont, *Frankia* sp. EuI1 (Actinomycetales). Can J Microbiol 26:1072-1089.
9. Baker DD. 1987. Relationships among pure cultured strains of *Frankia* based on host specificity. Physiol Plant 70:245-248.
10. Dobritsa SV, Novik SN, Stupar OS. 1990. Infectivity and host specificity of strains of *Frankia*. Microbiology (New York) 59:210-214.
11. Huang JB, Zhao ZY, Chen GX, Liu HC. 1985. Host range of *Frankia* endophytes. Plant Soil 87:61-65.
12. Huss-Danell K, Myrold DD. 1994. Intragenic variation in nodulation of *Alnus*: Consequences for quantifying *Frankia* nodulation units in soil. Soil Biol Biochem 26:525-531.
13. Maunuksela L, Hahn D, Haahtela K. 2000. Effect of freezing of soils on nodulation capacities of total and specific *Frankia* populations. Symbiosis 29:107-119.
14. Pokharel A, Mirza BS, Dawson JO, Hahn D. 2011. *Frankia* populations in soil and root nodules of sympatrically grown *Alnus* taxa. Microb Ecol 61:92-100.
15. Mirza BS, Welsh A, Rasul G, Rieder JP, Paschke MW, Hahn D. 2009. Variation in *Frankia* populations of the *Elaeagnus* host infection group in nodules of six host plant species after inoculation with soil. Microb Ecol 58:384-393.

16. Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. System Appl Microbiol 32:201-210.
17. Mirza BS, Welsh AK, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in soils from five continents. System Appl Microbiol 32:558-570.
18. Nickel A, Hahn D, Zepp K, Zeyer J. 1999. *In situ* analysis of introduced *Frankia* populations in root nodules of *Alnus glutinosa* grown under different water availability. Can J Bot 77:1231-1238.
19. Zepp K, Hahn D, Zeyer J. 1997. *In-situ* analysis of introduced and indigenous *Frankia* populations in soil and root nodules obtained on *Alnus glutinosa*. Soil Biol Biochem 29:1595-1600.
20. Maunuksela L, Hahn D, Haahtela K. 2000. Effect of freezing of soils on nodulation capacities of total and specific *Frankia* populations. Symbiosis (Rehovot) 29:107-120.
21. Maunuksela L, Zepp K, Koivula T, Zeyer J, Haahtela K, Hahn D. 1999. Analysis of *Frankia* populations in three soils devoid of actinorhizal plants. FEMS Microbiol Ecol 28:11-21.
22. 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.
23. Samant S, Sha Q, Iyer A, Dhabekar P, Hahn D. 2012. Quantification of *Frankia* in soils using SYBR Green based *q*PCR. System Appl Microbiol 35:191-197.
24. Samant S, Amann RI, Hahn D. 2014. Evaluation of the 23S rRNA gene as target for *q*PCR based quantification of *Frankia* in soils. System Appl Microbiol 37:229-234.
25. Ben Tekaya S, Ganesan AS, Guerra T, Dawson JO, Forstner MRJ, Hahn D. 2017. Sybr Green- and TaqMan-based quantitative PCR approaches allow assessment of the abundance and relative ristribution of *Frankia* clusters in soils. Appl Environ Microbiol 83: e02833-16.

26. Rodriguez D, Guerra TM, Forstner MRJ, Hahn D. 2016. Diversity of *Frankia* in soil assessed by Illumina sequencing of *nif*H gene fragments. System Appl Microbiol 39:391-397.
27. Cotin-Galvan L, Pozzi AC, Schwob G, Fournier P, Fernandez MP, Herrera-Belaroussi A. 2016. In-planta sporulation capacity enhances infectivity and rhizospheric competitiveness of *Frankia* strains. Microb Environ 31:11-18.
28. Markham JH. 2009. Variability of nitrogen-fixing *Frankia* on *Alnus* species. Botany 87:1011-1011.
29. Zimpfer JF, Smyth CA, Dawson JO. 1997. The capacity of Jamaican mine spoils, agricultural and forest soils to nodulate *Myrica cerifera*, *Leucaena leucocephala* and *Casuarina cunninghamiana*. Physiol Plant 99:664-672.
30. Diem HG, Dommergues YR. 1990. Current and potential uses and management of Casuarinaceae in the tropics and subtropics, p 365-385. *In* Schwintzer CR, Tjepkema JD (eds), The biology of *Frankia* and actinorhizal plants. Academic Press, San Diego, CA.
31. Samant SS, Dawson JO, Hahn D. 2015. Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. System Appl Microbiol 38:501-505.
32. Wilcox DA, Cowan DA. 2016. Diversity of *Frankia* in root nodules of six *Morella* sp. from the Cape flora of South Africa. Plant Soil 406:375-388.
33. Samant S, Dawson JO, Hahn D. 2016. Growth responses of introduced *Frankia* strains to edaphic factors. Plant Soil 400:123-132.
34. Mirza BS, Welsh AK, Hahn D. 2009. Growth of *Frankia* strains in leaf litter-amended soil and the rhizosphere of a non-actinorhizal plant. FEMS Microbiol Ecol 70:132-141.
35. Rönkkö R, Pennanen T, Smolander A, Kitunen V, Kortemaa H, Haahtela K. 1994. Quantification of *Frankia* strains and other root-associated bacteria in pure cultures and in the rhizosphere of axenic seedlings by high-performance liquid chromatography-based muramic acid assay. Appl Environ Microbiol 60:3672-3678.

36. Rönkkö R, Smolander A, Nurmiaho-Lassila EL, Haahtela K. 1993. *Frankia* in the rhizosphere of nonhost plants: A comparison with root-associated nitrogen-fixing *Enterobacter, Klebsiella* and *Pseudomonas*. Plant Soil 153:85-95.
37. Mirza BS, Welsh A, Hahn D. 2007. Saprophytic growth of inoculated *Frankia* sp. in soil microcosms. FEMS Microbiol Ecol 62:280-289.
38. Battenberg K, Wren JA, Hillman J, Edwards J, Huang LJ, 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.
39. Normand P, Simonet P, Bardin R. 1988. Conservation of *nif* sequences in *Frankia*. Mol Gen Genet 213:238-246.
40. Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444-2448.
41. Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.
42. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p 1-8. *In* Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA
43. Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. American Mathematical Society: Lectures on Mathematics in the Life Sciences 17:57-86.
44. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
45. Felsenstein J. 1985. Confidence limits of phylogenies: an approach using the bootstrap. Evolution 39:783-791.
46. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner F-O. 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucl Acids Res:doi:10.1093/nar/gks1808.

47. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner F-O. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools Nucl Acids Res:doi: 10.1093/nar/gks1219.
48. Mirza BS, Welsh AK, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in soils from five continents. System Appl Microbiol 32:558-570.
49. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME Journal 6:1621-1624.
50. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335-336.

CHAPTER V

META-ANALYSIS OF INDIGENOUS *FRANKIA* POPULATIONS IN SOILS FROM FIVE CONTINENTS

ABSTRACT

Diversity and relative abundance of the actinorhizal symbiont *Frankia* were assessed in soils from countries from five different continents (i.e. Rwanda, Hungary, Alaska, Peru and Japan) using an Illumina sequencing approach that targeted *nif*H gene fragments. Results were compared to results of previous studies on indigenous *Frankia* populations using bioassays and subsequent sequence analyses of *Frankia*-specific *nif*H gene fragments in nodules as well as *Frankia*-specific *nif*H gene clone libraries from soils. Additional studies included quantitative analyses by *q*PCR targeting 23S rRNA gene fragments that represented specific clusters and sub-clusters within the genus *Frankia*. Illumina sequencing resulted in a generally low diversity of *Frankia* populations, with only 18 distinct reads obtained from all soils, and with few sequences identical or closely related to those of cultured relatives. *Frankia* populations in individual soils were generally represented by only one or two abundant reads, with additional reads often very similar. *q*PCR analysis detected representatives of all clusters in soils from Rwanda, Hungary and Japan, while that from Peru harbored cluster 1a, 2 and 4 frankiae and Alaskan soil cluster 1b frankiae only. Meta-analyses including results from bioassays and clone libraries revealed large quantitative, but also qualitative differences, suggesting the presence of methodological biases such as selective nodulation, PCR amplification artifacts, or short reads length biases that could affect taxonomical assignments.

KEYWORDS: clone libraries - *Frankia* - *nif*H - nitrogenase – root nodules

INTRODUCTION

Members of the genus *Frankia* are soil-dwelling actinobacteria that are generally characterized as nitrogen-fixing symbionts forming root nodules on more than 260 actinorhizal plant species belonging to 8 families of angiosperms (1). Typical nitrogen-fixing *Frankia* strains are assigned to clusters 1, 2 and 3 within the genus (2), with *Frankia alni* strain ACN14a (3), *Frankia casuarinae* strain CcI3 (3), *Frankia coriariae* strain BMG5.1 (4) and *Frankia elaeagni* strain BMG5.12 (3) representing frankiae of cluster 1a, 1c, 2 and 3, respectively. In addition to typical *Frankia* strains, atypical, generally non-nitrogen-fixing and/or non-nodulating frankiae such as *Frankia inefficax* strain EuI1c (5) have been identified within the genus *Frankia* representing cluster 4 (2).

Host plant species have large effects in the selection of *Frankia* strains from soil for potential nodule formation, which is well documented for frankiae belonging to different host infection groups (6-8), but has also been observed for *Frankia* populations in nodules from plants belonging to the same host infection group (9-12). Studies on abundance and diversity of frankiae in soils using bioassays in which nodule populations of *Frankia* are analyzed on specific capture plants, are therefore likely underestimating both since individual host plant species do not capture the entire diversity of nodule-forming frankiae in soil (11). In order to circumvent bias introduced by host plant specificity of bioassays, molecular techniques such as clone libraries of nitrogenase reductase (*nif*H) gene fragments amplified from DNA extracts from soils have been used to assess diversity. However, molecular studies displayed large differences in sequences and cluster assignments to those obtained from nodules, with identical sequences in libraries and nodules and assignments to the same cluster only rarely encountered for individual soils (13). These studies thus did not only highlight the inadequacy of bioassays for the analysis of frankiae in soil, but also revealed some limitations of gene clone libraries for the analyses of *Frankia* populations in soil.

Recently, targeted Illumina sequencing of *nif*H gene fragments and analyses of pair-end reads through a modified QIIME pipeline were used to assess the diversity of *Frankia* populations in soils (14). This approach was affected by the low abundance of *Frankia* in soils that did not allow us to amplify fragments for Illumina sequencing

directly from soil DNA extracts, and thus a nested PCR was used as a basis for analyses. In addition, our approach provided sufficient coverage only for analyses of frankiae from cluster 1 and 3, while frankiae from clusters 2 and 4 were not detected, and thus analyses of the entire genus *Frankia* in soils might be limited (14). Analyses of reads using 1% abundance and 97% similarity cutoffs generally retrieved a low diversity of *Frankia* in individual soils, with few sequences identical or closely related to those of a cultured relative. Nevertheless, targeted Illumina sequencing provides an efficient and economical method for assessing haplotype diversity of ecofunctional genes such as *nif*H in microorganisms such as *Frankia.*

Previous studies on indigenous *Frankia* populations using bioassays and subsequent sequence analyses of *Frankia*-specific *nif*H gene fragments in nodules (11, 15) as well as *Frankia*-specific *nif*H gene clone libraries from soils (13) resulted in the retrieval of *nif*H sequences representing frankiae of clusters 1 and 3 only. Thus, despite the coverage limitations of our targeted Illumina sequencing protocol, a comparative analyses of *nif*H gene fragments from the previous studies using bioassay or clone library information (11, 13, 15) with reads obtained from the same soils by Illumina sequencing could provide a more comprehensive and accurate view on the diversity of cluster 1 and 3 frankiae in these soils. The original soil samples had been collected in October 2006, with subsamples that were kept frozen at -20°C used in this study. The soil samples had been obtained from sites in 5 continents, i.e. Africa (Rwanda, -1.94604/30.05372), Europe (Hungary, 46.87347/19.38878), Asia (Japan, 38.71913/139.85805), South America (Peru, -6.95694/-78.38), and North America (Alaska, USA, 61.16866/-149.76095) (11, 13, 15).

MATERIALS AND METHODS

DNA extraction and PCR analysis

DNA was extracted from triplicate 250 mg soil samples (dry wt.) using the SurePrep^{TM} Soil DNA Isolation Kit (Fisher Scientific, Houston, TX) (16), and used in a nested PCR targeting the *nif*H gene of cluster 1 and 3 *Frankia* (14). Briefly, 606-bp fragments were amplified using primers nifH1f ($^{5'}$GGC AAG TCC ACC ACC CAG C) and nifHr ($^{5'}$CTC GAT GAC CGT CAT CCG GC) (17) in a reaction volume of 50 µl,

containing 1 µl of a 10 mM dNTP mix, 1 µl each primer (0.2 µM), 8.2 µl BSA (30 µg ml^{-1}), 5 µl of 10 x PCR buffer with 15 mM $MgCl_2$, 2 µl DNA extract, and 0.2 µl *Taq* DNA polymerase (5 U μl^{-1}; Gene Script, Piscataway, NJ). An initial incubation at 96°C for 10 minutes was followed by 40 rounds of temperature cycling (96°C for 30 seconds, 60°C for 30 seconds and 72°C for 30 seconds) and a final 7-minute incubation at 72°C. Amplification products were cleaned using the UltraClean® 15 DNA Purification Kit (Mo Bio Laboratories, Inc., Carlsbad, CA) and quantified on a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA) using the Agilent DNA 7500 Kit.

Diluted amplicons (1 µl) were used as templates in nested PCR reactions using modified primers nifHf1 and nifH269 ($^{5'}$CCG GCC TCC TCC AGG TA) and reaction conditions described above to create a 263 bp product (18). Primer nifH269 was barcoded, and both primers modified with Illumina adapters, primer pads, and linkers (14, 19). Samples were analyzed at the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas, Austin, on an Illumina MiSeq with paired-end 250 bp reads using the respective sequencing and index sequence primers (14). Illumina sequence data were accessioned into the NCBI Sequence Read Archive.

Bioinformatics

Paired-end reads were aligned in PEAR 0.9.6 with default parameters and then parsed and filtered in QIIME 1.8.0 (20) using *split_library.py* accepting reads with a Phred quality score of 26 or better (-q 25) as described previously (14). To exclude any chimera or truncated reads due to PCR errors the *nif*H gene sequence from *Frankia* strain EuIK1 (U53362) was used as a blast reference in addition to the seq.fna file from *split_library.py* output to run *exclude_seqs_by_blast.py* (14). Unique reads were retrieved from the output file (matching.fna) by running *pick.denovo_otus.py* with a similarity threshold set as 1.0. Sequences representing less than 1% of the total numbers of reads per sample were removed using the resulting biom table and a customized R script (14). The collected unique sequences were aligned in Geneious 9.1.4 (Biomatters Ltd, Auckland, New Zealand), and compared to *nif*H gene sequences of confirmed *Frankia* strains or uncultured endophytes from nodules or soils obtained from GenBank/EMBL

databases using the BLAST algorithm (21). Representative reads were deposited in GenBank under accession numbers LT964921 to LT964938.

The identity and relationship among the sequences amplified were evaluated using neighbor joining (NJ) (22), and maximum likelihood (ML) (23) analyses, conducted from within Geneious 9.1.4. The neighbor joining analyses utilized the HKY85 model to correct for substitution bias (24). Model parameters for maximum likelihood were estimated by the general time reversible model (GTR) with gamma (25), and used as input in a ML heuristic search using RAxML (26). Bootstrap values (27) were estimated from a heuristic search with random stepwise addition sequence for 10,000 NJ, and 1,000 ML iterations.

Meta-analyses

Previously published sequences of *nif*H gene fragments from root nodules obtained in bioassays using the same soils (11, 15) as well as of *Frankia*-specific *nif*H gene clone libraries from these soils (13) were retrieved from GenBank/EMBL databases with accession numbers FJ477419 to FJ477548 and GQ141268 to GQ141513, respectively, and used for comparative analyses on *Frankia* populations in soils assessed by bioassays, gene clone libraries and Illumina sequencing.

RESULTS

Analyses of Illumina sequence data for soils from sites in Rwanda, Hungary, Alaska, Peru and Japan retrieved a total of 170,227, 100,620, 164,667, 130,261, and 74,110 reads for *Frankia*, respectively. The analyses revealed 18 distinct sequences with an abundance threshold larger than 1% of the total number of reads in all soils. Frankiae in soil from Rwanda were represented by seven distinct sequences, while those in soils from Hungary, Alaska, Peru and Japan were represented by two, three, three, and six sequences, respectively (Table 1, Fig. 1). Only one sequence, read 58286, was retrieved from more than one soil, while all other sequences were unique for their respective soil. Read 58286 was detected in soils from Rwanda, Hungary, Alaska, and Peru, but not in soil from Japan (Table 1; Fig. 2). In soils from Rwanda, Hungary, and Alaska, read 58286 represented 62%, 93%, and 64% of all reads, respectively, while it was retrieved in

much smaller percentage in soil from Peru (2%) (Table 1). In soil from Peru, the vast majority of frankiae were represented by read 43329 (96%), while those in soil from Japan were mainly represented by read 269251 (79%). Additional reads for individual soils often had similar sequences to the most abundant read for the soil (e.g. for Rwanda, Alaska, Peru, Japan) (Fig. 1), and thus represented the same *Frankia* cluster (Fig. 2) Reads from soils from Rwanda and Alaska were all assigned to represent cluster 3 frankiae, while those from soil from Hungary were prominently cluster 3 frankiae (93%), with cluster 1b frankiae accounting for the remaining 7% of the reads (Fig. 2, 3). Frankiae of cluster 1a were dominant in soil from Peru (98% of the reads), but a minor component in soil from Japan (4%). Frankiae in soil from Japan were generally represented by reads that could not reliably be assigned to specific clusters since sequence similarities close or below the threshold of 97% to sequences of cultured *Frankia* strains provided ambiguous results, i.e. similar values to both strain D11 representing cluster 3 frankiae, and strain An2.2 representing cluster 1a frankiae (Table 1). Overall, only two reads were identical to sequences obtained from confirmed *Frankia* strains, i.e. the most prominent read 58286 present in four soils to the sequence of strain Cc1.17 (cluster 3), and a minor component of soils from Japan (read 128047) identical to the sequence of strain ArI3 (cluster 1a).

*q*PCR analysis of DNA extracted from the five soils resulted in various diversity patterns for *Frankia* populations, that were generally different from the Illumina read profiles. *Frankia* populations in the Alaskan soil consisted of *Frankia* cluster 1b only, while the other four soils harbored frankiae of all 4 clusters, except for the Peru soil where cluster 3 frankiae were absent. Members of cluster 1a were present in Rwanda, Peru and Japan soils, while members of cluster 1b, generally prominent in soil rich in organic materials, were detected in soils from Rwanda, Hungary and Japan at an abundance rate of 15%, 50% and 65%, respectively. Members of cluster 2 were prominent frankiae in all soils except Alaska, with an abundance ranging from 10% of all frankiae in Japan to 32% of all frankiae in Hungary. Members of cluster 4 frankiae were also prominent in all soils except Alaska, with an abundance ranging from 12% of all frankiae in Hungary to 45% of all frankiae in Rwanda.

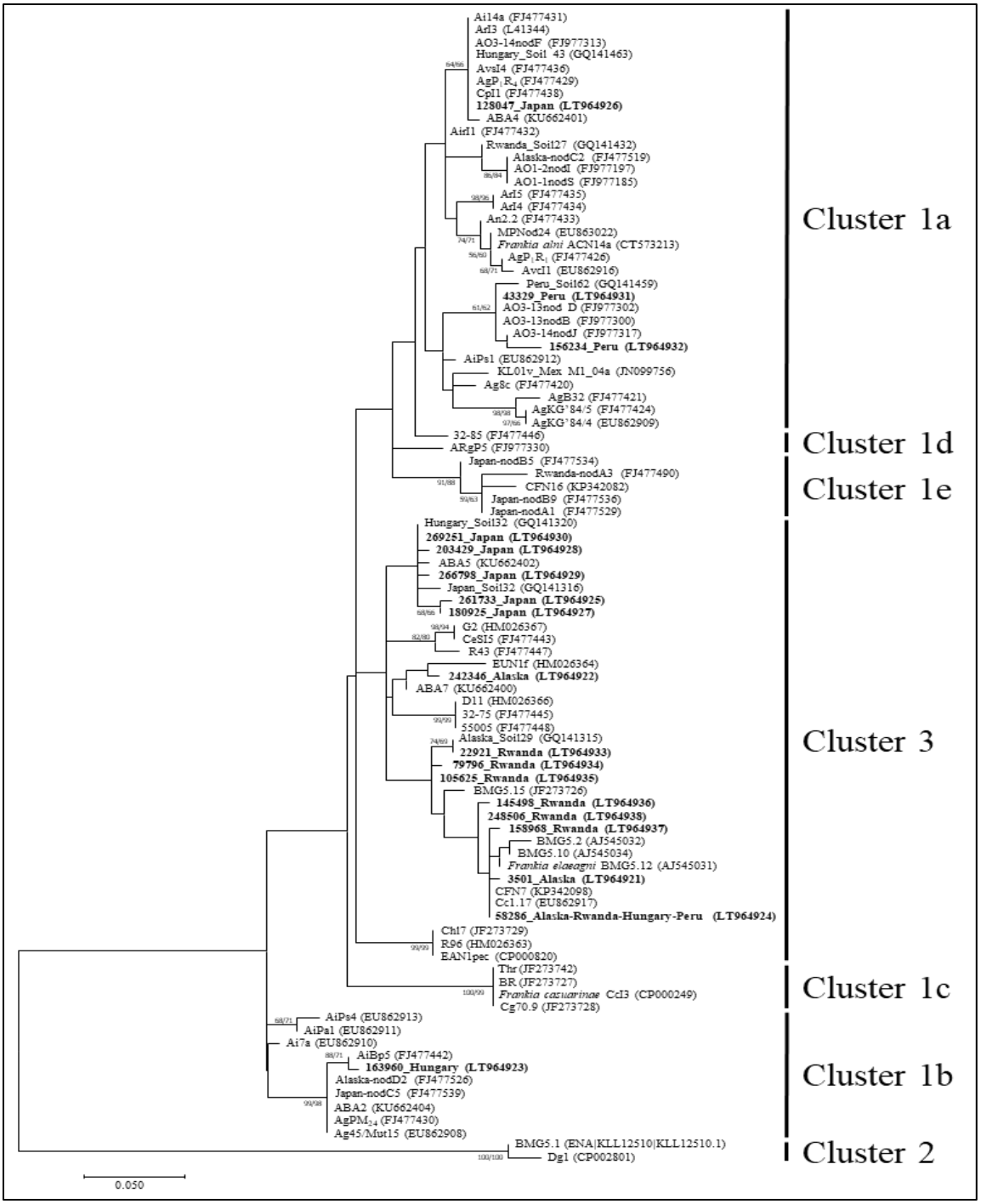

FIG 12 A Neighbor-joining topology showing sequence relationships for selected *Frankia* strains, uncultured endophytes in root nodules and gene clone libraries, and reads obtained from soil DNA samples from sites in 5 continents, i.e. Africa (Rwanda), Europe (Hungary), North America (Alaska, USA), South America (Peru), and Asia (Japan). Numbers above the branches represent the bootstrap values from a Maximum-likelihood (1,000 replicates) and Neighbor-joining (10,000 replicates) bootstrap analysis (ML/NJ) for clades with greater than 50% bootstrap support. Support values less than 50% are denoted with a dash (-). Both phylogenetic criteria resolved generally similar topologies.

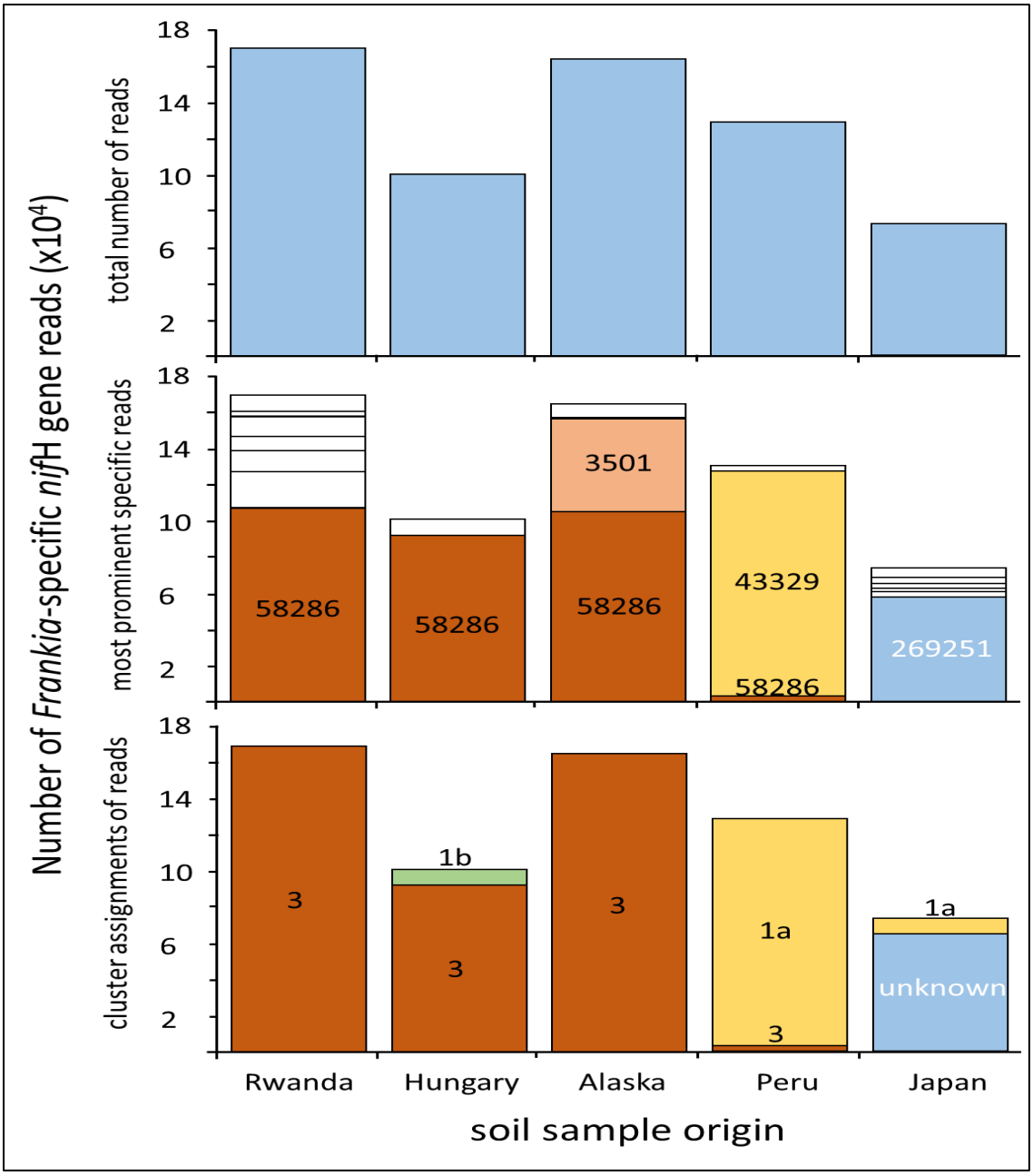

FIG 13 Number of *Frankia*-specific *nif*H gene fragment reads retrieved from soil DNA samples from sites in 5 continents, i.e. Africa (Rwanda), Europe (Hungary), North America (Alaska, USA), South America (Peru), and Asia (Japan). The upper panel represents the number of all *Frankia*-specific reads obtained for each sample, while the middle panel represents their distribution into individual *Frankia*-specific reads, with the most abundant read(s) highlighted. The lower panel depicts the assignment of the individual reads to *Frankia*-specific clusters (i.e. 1a, 1b, 3, and unknown).

Table 7 Abundance of *Frankia* specific *nif*H reads in sample locations and their sequence similarity to *Frankia* strains and previously obtained sequences from nodules and soils from the same locations

Read	Number of specific reads from sample location					Closest relative (>97% similarity) to specific read		Assigned
	Rwanda	Hungary	Alaska	Peru	Japan	*Frankia* strain	*Frankia* previously obtained in samples from study locations	Cluster
58286	107,155	93,335	105,860	2,470	0	Cc1.17 (100%)	Alaska_Soil29 (97.6%); 97.6% Hungary-nodA7 (97.6%)	3
43329	0	0	0	125,406	0	AiPs1 (97.6%)	Peru_Soil62 (99.0%)	1a
269251	0	0	0	0	58,421	D11 (96.6%)	Hungary_Soil32 (100%); Japan_Soil32 (99.0%); Rwanda-nodD9 (99.0%); Alaska-nodB11 (98.5%); Peru-nodC10 (97.6%)	Unknown
163960	0	7,285	0	0	0	AiBp5 (99.5%)	Alaska-nodD2 (98.5%); Hungary_Soil30 (98.1%)	1b
3501	0	0	56,408	0	0	Cc1.17 (99.5%)	Alaska_Soil29 (97.1); Hungary-nodA7 (97.1%)	3
242346	0	0	2,399	0	0	BMG5.6 (98.5%)	Japan_Soil32 (97.6%); Peru-nodC1 (97.6%); Hungary_Soil32 (97.1%); Rwanda-nodD9 (97.1%)	3
156234	0	0	0	2,385	0	ArI5 (96.6%)	Peru_Soil62 (97.1%)	unknown
248506	20,687	0	0	0	0	Cc1.17 (99.5%)	Alaska_Soil29 (98.1%); Hungary-nodA7 (98.1%)	3
22921	11,270	0	0	0	0	Cc1.17 (97.6%)	Alaska_Soil29 (100%); Hungary-nodA7 (100%); Peru-nodC10 (97.1%)	3
105625	7,983	0	0	0	0	Cc1.17 (98.5%)	Alaska_Soil29 (99.0%); Hungary-nodA7 (99.0%); Japan_Soil28 (97.1%); Peru-nod10 (97.1%)	3
158968	11,441	0	0	0	0	Cc1.17 (99.5%)	Alaska_Soil29 (98.1%); Hungary-nodA7 (98.1%)	3
145498	2,796	0	0	0	0	Cc1.17 (99.0%)	Alaska_Soil29 (98.5%); Hungary-nodB8 (97.1%)	3
79796	8,895	0	0	0	0	Cc1.17 (98.1%)	99.5% Alaska_Soil29 (99.5%); Hungary-nodA7 (99.5%)	3
180925	0	0	0	0	1,897	An2.2 (96.6%)	Hungary_Soil32 (99.0%); Rwanda-nodD9 (98.1%); Japan_Soil28 (97.6%); Alaska-nodB11 (97.6%)	Unknown
128047	0	0	0	0	2,883	ArI3 (100%)	Hungary_Soil43 (100%); Alaska_Soil41 (99.5%); Peru_Soil55 (98.5%); Rwanda_Soil27 (98.5%)	1a
203429	0	0	0	0	2,446	D11 (97.1%)	Hungary_Soil32 (99.5%); Japan_Soil32 (98.5%); Rwanda-nodD9 (98.5%); Alaska-nodB11 (98.1%); Peru-nodC1 (97.1%)	Unknown
266798	0	0	0	0	5,669	D11 (97.1%)	Hungary_Soil32 (99.5%); Japan_Soil32 (98.5%); Rwanda-nodD9 (98.5%); Alaska-nodB11 (98.1%); Peru-nodC1 (97.1%)	Unknown
261733	0	0	0	0	2,794	An2.2 (97.1%)	Hungary_Soil32 (98.5%); Japan_Soil32 (97.6%); Rwanda-nodD9 (97.6%); Alaska-nodB11 (97.1%)	Unknown

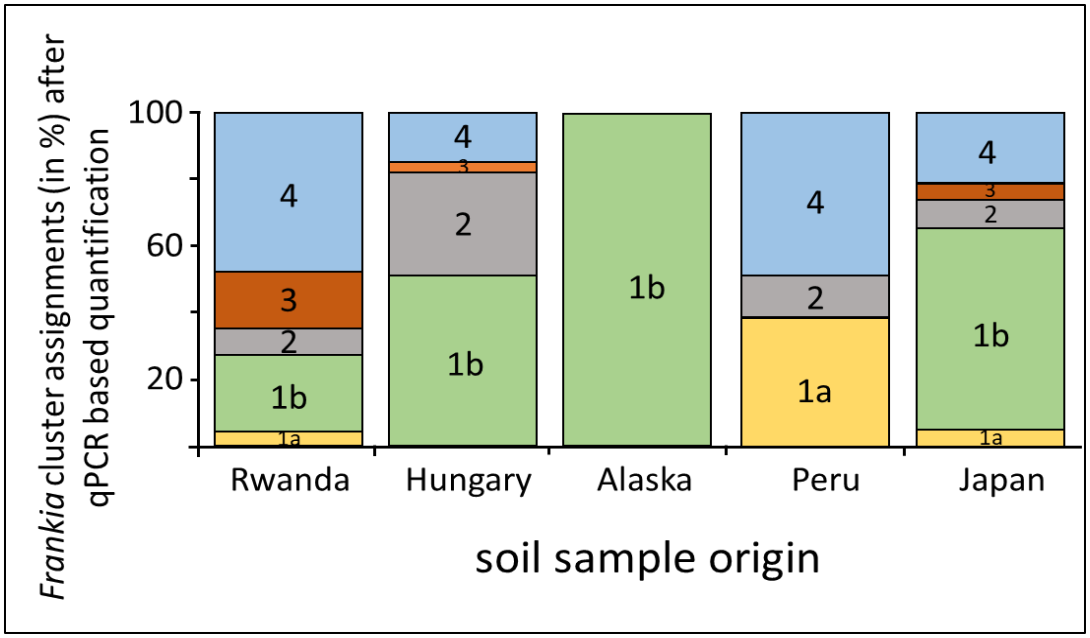

FIG 14 SybrGreen *q*PCR quantification of *Frankia* clusters 1, 2, 3 and 4 in the five soil samples. Quantification used 23S rRNA gene fragments generated with primer combinations specific for cluster 1a, 1b, 1c, 1d, 1e, 2, 3 and 4, presented as the sum of individual clusters per soil sample. Frankiae of subgroup 1a were detected in soil from Rwanda, Peru and Japan. Cluster 1b was observed in soils from Rwanda, Hungary, Alaska and Japan. Members of cluster 3 were detected in Rwanda, Hungary and Japan while members of cluster 2 and 4 were prominent in all soils except that from Alaska.

Comparative analyses of Illumina reads and sequences obtained from gene clone libraries from the same soils found only little overlap, with only three of the 18 distinct Illumina reads found in clone libraries from one of these soils (Table 1). Detection of reads in soils was generally not matching results in clone libraries. For example, read 128047, a unique component in soil from Japan and thus absent in the other four soils, was not found at all in gene clone libraries from soil from Japan, but in clone libraries from Hungary (100% similarity) and Alaska (99.5% similarity). Similar results were obtained for read 269251, even though this sequence was found at lower similarity value in clone libraries from soil from Japan (99.0%) as well (Table 1). And finally, read 22921 present in soil from Rwanda only, was not found in clone libraries from soil from Rwanda, but only in that from Alaskan soil.

Comparative analyses of Illumina reads and sequences obtained from nodules resulted in even fewer matches, with only read 22921 unique for Rwandan soil detected in root nodules. These nodules, however, were formed on plants growing in Hungarian soil, and not those on Rwandan soil (Table 1). Reduction of sequence diversity by assignment of sequences to *Frankia* clusters results in the identification of *Frankia*

populations in soils that are heavily impacted by the analyses methodology, resulting in a different composition of *Frankia* populations for each soil as a function of methodology (Fig. 3). As an example, assignment of Illumina reads from Rwandan soil to clusters identified all sequences as representing cluster 3 frankiae, while sequence analyses of gene clone libraries identified only a minor component as cluster 3 frankiae, with the major number of sequences representing cluster 1a frankiae and fewer sequence cluster 1b frankiae (Fig. 3). Root nodules were dominated by cluster 3 frankiae, but 30% of the sequences represented frankiae with unknown assignment. Similar differences were obtained for the other soils, with all Illumina reads for Alaskan soil, for example, representing cluster 3 frankiae, while gene clone libraries also detected clusters 1a and 1b in about 20% of the clones, and sequences in root nodules were dominated by cluster 1b frankiae, with only few harboring cluster 1a and cluster 3 frankiae (Fig. 3).

DISCUSSION

Targeted Illumina sequencing of *nif*H gene fragments has successfully been used to assess the diversity of *Frankia* populations in soils (14, 28). Using a conservative approach with 1% abundance and 97% similarity cutoffs for the analyses of reads, individual soils generally harbored a low diversity of *Frankia* populations, with few sequences identical or closely related to those of a cultured relative (28). Our studies corroborate these results with only 18 distinct reads obtained from 5 soils. *Frankia* populations in individual soils were generally represented by only one or two abundant reads, with additional reads often very similar. Since the low abundance of *Frankia* in soils required a nested PCR as basis for Illumina sequencing, target enrichment using PCR amplification could have introduced artifacts through polymerase errors generating sequence changes not present in the original samples and thus may have resulted in sequence variants present at low abundance in final sequence reads (29). Consequently, all soils in this study might only harbor one or two major *Frankia* populations.

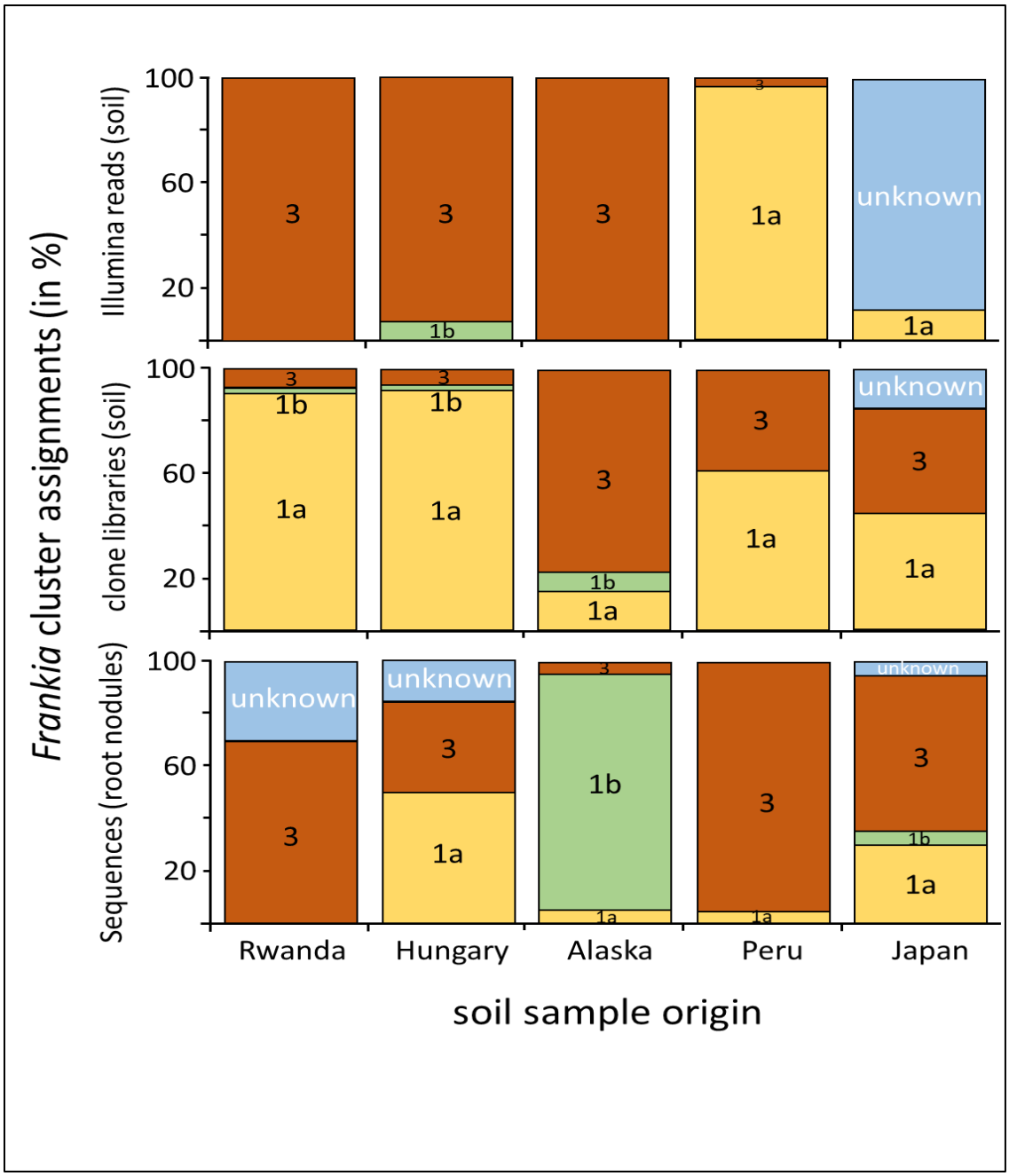

FIG 15 Assignments of sequences of *nif*H gene fragments to *Frankia* clusters (in %) retrieved from soil DNA samples from sites in 5 continents, i.e. Africa (Rwanda), Europe (Hungary), North America (Alaska, USA), South America (Peru), and Asia (Japan) after comparative analyses with sequences from GenBank/EMBL databases. The upper panel represents data based on reads retrieved by *Frankia*-specific *nif*H Illumina sequencing in this study, while the middle panel is based on data analyses of previously published *Frankia*-specific *nif*H gene clone libraries from the same soils (13) and the lower panel based on sequence analyses of previously published *nif*H gene fragments of nodule populations obtained in bioassays (11, 15).

Assignments of reads to *Frankia* clusters corroborate results highlighting a generally limited diversity of frankiae in soils. In Rwandan soil, only cluster 3 frankiae were detected, represented by one major read identical to the sequence from strain Cc1.17. The same read represented the most prominent *Frankia* population in soil from Hungary, with a minor population belonging to cluster 1b, while Alaskan soil harbored two distinct cluster 3 populations, one of which was again represented by the read identical to the sequence from strain Cc1.17. This read also represented a small population in Peruvian soil, in which the remaining *Frankia* population was identified as cluster 1a *Frankia* characterized by one read only. Cluster 1a frankiae were present in soil from Japan as well, however, the majority of reads could not reliably be assigned to a specific *Frankia* cluster at the 97% sequence similarity level. The identical sequences representing cluster 3 frankiae in four of the five soils from countries in different continents suggest a ubiquitous distribution of these frankiae. Cluster 3 frankiae have been shown to represent a group with broad physiological adaptations since they have been shown to grow in bulk soil, the rhizosphere and with leaf litter, independent of matric potential and plant species (30, 31).

Cluster assignments of reads and the detection of limited diversity in our soils are in agreement with analyses by different methods. Quantitative polymerase chain reaction *(q*PCR) analyses targeting *Frankia* cluster-specific sequences on the 23S rRNA gene, for example, detected frankiae of cluster 1a and 3 as major populations in soils (30), although cluster 1b instead of cluster 1a was more prominent in others (32). Cluster 1b was dominating *Frankia* diversity in soils under *Alnus glutinosa* and *Betula nigra*, with only cluster 1 and 3 frankiae detected (33). In microcosm soil planted with different actinorhizal hosts, cluster 1b and 3 frankiae were detected as major *Frankia* populations, however, a newly assigned cluster 1e was present as well (28). In prairie soils diversity was higher with frankiae of all 4 clusters detected and those of cluster 4 dominating (33).

More specific comparison of *q*PCR-based data with those obtained by other methods on our soils, however, depicted highly variable outcomes, independent of the soil. As an example, *Frankia* populations in soil from Rwanda analyzed by *q*PCR were represented by clusters 1a, 1b, 2, 3, and 4, while root nodules from bioassays only harbored cluster 3 and unknown frankiae, clone libraries cluster 1a, 1b and 3 frankiae and

Illumina sequencing only cluster 3 frankiae. These data need to be compared considering the obvious differences with respect to targets, i.e. *q*PCR targeting 23S rRNA gene sequences of all subgroups, while the remaining methods, i.e. bioassays, clone libraries and Illumina sequencing target *nif*H genes of frankiae of clusters 1 and 3 only, thus excluding cluster 2 and 4 frankiae (16, 34). Unknown frankiae are represented by sequences that have less than 97% similarity values to those of pure cultures in the database, but with similar values to both cluster 1a and 3 frankiae (data not shown). Thus, some of the qualitative differences in results could be easily explained by differences in methodology.

All methods have additional drawbacks that might have affected the qualitative and quantitative composition of the *Frankia* populations detected in these five soils. Bioassays rely on nodule formation by specific host plant species, and thus are limited to those frankiae infective on the respective host plant. Although *Morella pensylvanica* that was used in bioassays on these soils has been characterized as promiscuous host plant for frankiae, nodulation is restricted to frankiae representing cluster 1 and 3 (15). Even within clusters, *Frankia* strains have been shown to select for different host plants, even those closely related (i.e. on the species level) resulting in both qualitative (12) and quantitative (9) biases. Thus, despite accurate identification of frankiae in root nodules by comparative sequence analyses (e.g. of large *nif*H gene fragments) (15, 35), bioassays do not allow researchers to recover the entire diversity of *Frankia* and consequently analyses trying to assess diversity and quantitative detection of frankiae in soils by bioassays is extremely biased and unreliable.

Problems have also been shown to occur in gene clone library analyses that consequently often do not accurately reflect the abundance and diversity of microorganisms (13, 33). Clone libraries are generally using PCR amplicons cloned into a vector that is subsequently transformed into an *E. coli* strain. Transformed *E. coli* colonies are picked, cultured and amplicons sequenced. The results are affected at many stages, starting with the specificity of the primers selected for the initial amplification as well as by basic PCR errors during the amplification as a consequence of the quality of the polymerase but also the purity of the DNA extract or the amount and complexity. Sequence artifacts introduced by these PCR biases might be accompanied by cloning bias

resulting in a distorted distribution of sequences within libraries (36). The selection of a limited number of colonies, e.g. 40 colonies in our previous study (13), might further affect the reliable detection and quantification of *Frankia* populations in soil. Considering the potential problems for bioassays and clone libraries, it is not surprising that quantitative data on *Frankia* populations obtained by these methods do not match, and that even qualitative analyses are difficult to interpret.

Similar to bioassays and clone libraries, Illumina sequencing has methodological drawbacks that might impact both qualitative and quantitative analyses. As outlined before, *nif*H gene targeted Illumina sequencing provided sufficient coverage only for frankiae from cluster 1 and 3, while frankiae from clusters 2 and 4 were not detected (14). The analyses were also impacted by potential PCR bias since the low abundance of *Frankia* in soils required a nested PCR for the analyses. Illumina sequencing specific effects include problems with the MySeq driven clusters amplification, as well as with the computational analysis of resulting data (37). High cluster densities tend to suppress reads with high G+C DNA content, and read data retrieval might further be affected by the sequencing protocol, run to run variability or even variability between flow-cell lanes (38). Short read lengths, together with a limited taxonomic coverage of our primers of cluster 1 and 3, but not 2 and 4, as well as limited availability of sequences in the EMBL/GenBank databases might impact accurate identification of reads and thus assignments of reads retrieved from soils to *Frankia* clusters (14). Conservative read filtering with a threshold of 1% instead of the usually proposed 0.005% did not change the diversity pattern (14, 39).

The current study shows that different methodological approaches to analyze *Frankia* diversity in soils qualitatively and quantitatively might retrieve considerably different diversity patterns, impacted by biases and limitations of each of the approaches used. In comparative approaches, outcomes might therefore not be compatible. Currently, *q*PCR seems to be the most reliable method, compared to the other methods, since it allows the analyses of all 4 *Frankia* clusters and provides data similar to those of previous studies (32). However, this statement needs to be confirmed by additional studies that avoid the amplification step and allows absolute quantification, e.g. by dPCR based approaches.

Acknowledgements

The authors are indebted to the Graduate College (Doctoral Research Support Fellowship to S. Ben Tekaya), and the Department of Biology at Texas State University for financial support.

REFERENCES

1. Benson DR, Dawson J. 2007. Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiologia Plantarum 130:318-330.

2. Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L, Misra AK. 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. International Journal of Systematic Bacteriology 46:1-9.

3. Nouioui I, Ghodhbane-Gtari F, Montero-Calasanz MD, Goker M, Meier-Kolthoff JP, Schumann P, Rohde M, Goodfellow M, Fernandez MP, Norrnand P, Tisa LS, Klenk HP, Gtari M. 2016. Proposal of a type strain for *Frankia alni* (Woronin 1866) Von Tubeuf 1895, emended description of *Frankia alni*, and recognition of *Frankia casuarinae* sp nov and Frankia elaeagni sp nov. International Journal of Systematic and Evolutionary Microbiology 66:5201-5210.

4. Nouioui I, Ghodhbane-Gtari F, Rohde M, Klenk HP, Gtari M. 2017. *Frankia coriariae* sp nov., an infective and effective microsymbiont isolated from *Coriaria japonica*. International Journal of Systematic and Evolutionary Microbiology 67:1266-1270.

5. Nouioui I, Ghodhbane-Gtari F, Montero-Calasanz MD, Rohde M, Tisa LS, Gtari M, Klenk HP. 2017. *Frankia inefficax* sp nov., an actinobacterial endophyte inducing ineffective, non nitrogen-fixing, root nodules on its actinorhizal host plants. Antonie Van Leeuwenhoek 110:313-320.

6. Baker DD. 1987. Relationships among pure cultured strains of *Frankia* based on host specificity. Physiologia Plantarum 70:245-248.

7. Dobritsa SV, Novik SN, Stupar OS. 1990. Infectivity and host specificity of strains of *Frankia*. Microbiology (New York) 59:210-214.

8. Huang JB, Zhao ZY, Chen GX, Liu HC. 1985. Host range of *Frankia* endophytes. Plant and Soil 87:61-65.

9. Huss-Danell K, Myrold DD. 1994. Intragenic variation in nodulation of *Alnus*: Consequences for quantifying *Frankia* nodulation units in soil. Soil Biology & Biochemistry 26:525-531.

10. Maunuksela L, Hahn D, Haahtela K. 2000. Effect of freezing of soils on nodulation capacities of total and specific *Frankia* populations. Symbiosis (Rehovot) 29:107-120.

11. Mirza BS, Welsh A, Rasul G, Rieder JP, Paschke MW, Hahn D. 2009. Variation in *Frankia* populations of the *Elaeagnus* host infection group in nodules of six host plant species after inoculation with soil. Microbial Ecology 58:384-393.

12. Pokharel A, Mirza BS, Dawson JO, Hahn D. 2011. *Frankia* populations in soil and root nodules of sympatrically grown *Alnus* taxa. Microbial Ecology 61:92-100.

13. Mirza BS, Welsh AK, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in soils from five continents. System Appl Microbiol 32:558-570.

14. Rodriguez D, Guerra TM, Forstner MRJ, Hahn D. 2016. Diversity of *Frankia* in soil assessed by Illumina sequencing of *nif*H gene fragments. Systematic and Applied Microbiology 39:391-397.

15. Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Systematic and Applied Microbiology 32:201-210.

16. Samant S, Sha Q, Iyer A, Dhabekar P, Hahn D. 2012. Quantification of *Frankia* in soils using SYBR Green based qPCR. Systematic and Applied Microbiology 35:191-197.

17. Normand P, Simonet P, Bardin R. 1988. Conservation of nif sequences in *Frankia*. Molecular and General Genetics 213:238-246.

18. Mirza BS, Welsh AK, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in soils from five continents. Systematic and Applied Microbiology 32:558-570.

19. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME Journal 6:1621-1624.

20. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335-336.

21. Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences of the United States of America 85:2444-2448.

22. Saitou N, Nei M. 1987. The Neighbor-joining Method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

23. Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

24. Hasegawa M, Kishino H, Yano T. 1985. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160-174.

25. Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. American Mathematical Society: Lectures on Mathematics in the Life Sciences 17:57-86.

26. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

27. Felsenstein J. 1985. Confidence limits of phylogenies: an approach using the bootstrap. Evolution 39:783-791.

28. Ben Tekaya S, Guerra T, Rodriguez D, Dawson JO, Hahn D. 2018. *Frankia* diversity in host-plant root nodules is independent of abundance or relative diversity of *Frankia* in corresponding rhizosphere soils. Appl Environ Microbiol. 84: e02248-17.

29. Peng Q, Satya RV, Lewis M, Randad P, Wang YX. 2015. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. BMC Genomics 16:589. doi.org/10.1186/s12864-015-1806-8.

30. Samant SS, Dawson JO, Hahn D. 2015. Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. Systematic and Applied Microbiology 38:501-505.

31. Samant S, Dawson JO, Hahn D. 2016. Growth responses of introduced *Frankia* strains to edaphic factors. Plant and Soil 400:123-132.

32. 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. Microbial Ecology 71:473-481.

33. Ben Tekaya S, Ganesan AS, Guerra T, Dawson JO, Forstner MRJ, Hahn D. 2017. Sybr Green- and TaqMan-based quantitative PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. Applied and Environmental Microbiology 83: e02833-16.

34. Samant S, Amann RI, Hahn D. 2014. Evaluation of the 23S rRNA gene as target for *q*PCR based quantification of *Frankia* in soils. Systematic and Applied Microbiology 37:229-234.

35. Welsh AK, Dawson JO, Gottfried GJ, Hahn D. 2009. Diversity of *Frankia* in root nodules of geographically isolated Arizona alders in central Arizona (USA). Applied and Environmental Microbiology 75:6913-6918.

36. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. 2005. PCR-induced sequence artifacts and bias: Insights from comaprison of two 16S rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology 71:8966-8969.

37. Chiang DY, Getz G, Jaffe DB, O'Kelly MJT, Zhao XJ, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES. 2009. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nature Methods 6:99-103.

38. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology 12:R18; doi 10.1186/gb-2011-12-2-r18.

39. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10:57-59.

Chapter VI

SUMMARY

The current research was aimed at understanding population dynamics of members of the genus *Frankia* in soils under a panoply of conditions, mainly involving the presence or absence of host plants. Slow growth and cultivation constraints led to the development of molecular tools targeting *Frankia* DNA extracted from soils to bypass these limitations. Over the past few years, techniques evolved rapidly from clone library development to quantitative PCR targeting specific *Frankia* groups in soil. Previous studies succeeded at developing *q*PCR tools for detection of frankiae that belonged to cluster 1 and cluster 3 by targeting *nif*H gene sequences. In addition, a *q*PCR protocol targeting 23S rRNA gene sequences was developed to allow detection of frankiae at the genus level, but also at subgroup levels, specifically for clusters 1 and 3. Analyses of frankiae were then established using a combination of both targets to assess population dynamics of frankiae in their natural habitats (2–5). Furthermore, next-generation sequencing was used to target the *nif*H gene in order to assess the diversity of clusters 1 and 3 in soils (1). The work described in this dissertation was expanding coverage of these methods to different *Frankia* populations by 1) developing a SybrGreen-based *q*PCR protocol for the detection of clusters 2 and 4 in soil, combined with method modifications that included the development of Taqman-based detection protocols for *Frankia* in soil; and 2) using a combination of molecular tools to study *Frankia* dynamics in soils in the presence or absence of potential host plants.

In order to accomplish the first objective, specific primers were designed for detection of cluster 2 frankiae, i.e. 23DAT1578F and 23NNF1561F, which can be combined with the reverse primer, 1769R, in a SybrGreen-based *q*PCR protocol, resulting in specific detection of this group. In addition, two specific probes that could be used for Taqman-based detection were designed: NF1715f that detected all nitrogen-fixing frankiae, i.e. clusters 1, 2 and 3, and NNF1715f that allowed the detection of cluster 4 frankiae. Cluster 2 frankiae have been the last of the four clusters to be isolated with, to date, only one reported successful isolation (6). Members of cluster 4 have been

obtained in pure culture more often, however, seem to be highly diverse with respect to host plant compatibility, root nodule formation and nitrogen-fixing capacity. While data on the ecology of frankiae of these two clusters are scarce, phylogenetic analysis confirm their relationship to *Frankia* as separate clusters 2 and 4. Sequences of uncultured endophytes representing cluster 2, retrieved from root nodules of *Datisca*, *Coriaria* or *Ceanothus* located in the United States and Pakistan, fell within one concise, non-diverse clade. Pure cultures representing cluster 4, on the other hand, showed more phylogenetic diversity resulting in three separate subgroups. To evaluate the newly developed methods, two sets of soils were analyzed: soils sampled from *Alnus glutinosa* stands, a host tree of cluster 1 frankiae, and from *Betula nigra* stands, a non-host tree taxonomically close to the *Alnus* genus. Another set of samples was collected from prairie soils, where *Ceanothus*, a host plant of cluster 2, was the only actinorhizal host plant on the site. Surprisingly, both *Frankia* abundance and diversity were higher in prairie soils than in soils vegetated with *Alnus glutinosa* or *Betula nigra*. Diversity of frankiae found in soils with *Alnus glutinosa* and *Betula nigra* was restricted to *Frankia* cluster 1b, and 3 that are known to be ubiquitous. *Frankia* representatives of all four clusters were detected in prairie soils, with cluster 4 frankiae being more abundant than frankiae of any other cluster. Cluster 2 was also prominent in smaller numbers, leading to speculations about a correlation to their host presence. In line with our findings, a similar study using Next Generation Sequencing targeting *gln*A and 16S rRNA genes concluded an intimate relation between host plant presence and cluster 2 abundance in soils (7).

In the third chapter, we attempted to use previously described techniques to assess *Frankia* diversity in the Morrow Plot agricultural experiment. The Morrow Plots, located at the University of Illinois at Urbana Champaign, consist of a crop rotation experiment that has been ongoing for more than a century. SybrGreen based *q*PCR analysis of the different soils resulted in a diverse trend. The control samples represented by native, cultivated or restored prairie soils was the most diverse with all *Frankia* clusters detected. This was consistent with a previous study on the same prairie soils confirming presence of all clusters using both SybrGreen- and Taqman-based *q*PCR protocols (15). Clusters 4 and 2 were prominent across all treatments, with cluster 4 present generally in higher numbers, and cluster 2 present in low numbers outside of the control soils. This suggests

host interference on the abundance of specific symbionts in these soils. The presence of *Ceanothus* in the prairie soils might explain the difference in abundance compared to the crop rotation treatments. The low numbers within the crop rotation soils could be due to the persistence of frankiae after removal of host plants, or recent dispersal from the adjacent prairie soils. This is supported by similar studies of nodulated *Coriaria* plants from soils that were deprived of host plants for two centuries (17). Illumina analysis targeting the *nif*H gene resulted in a different pattern, which showed higher numbers of cluster 1b and cluster 3 frankiae. Prominence of these two clusters in nutrient-rich soils has been previously reported, confirming their saprophytic potential for thriving independently of the presence of plant hosts (5). Differences in the diversity might be due to methodological issues. Different targeted genes might be one of the major reasons. Analysis using universal 16S rRNA gene targeted Illumina sequencing resulted in a microbial pattern dominated by Proteobacteria, Actinobacteria and Acidobacteria, respectively. This correlates with a similar study on the same site (18). However, the mentioned study observed actual shifts in microbial communities across treatments, while our study did not. Many reasons could explain these divergent results, including sampling strategies, seasonal variations or computational biases on data analysis. The percentage of frankiae accounted for 0.01% to 0.1% of total microbiota. The highest percentage was observed within the corn NPK treatment. These proportions are in line with the usual abundance of *Frankia* encountered in soils (13, 19). Sequences aligned using BLAST adjusted to target only Frankia resulted in a diverse pattern dominated by cluster 3. Similar to *q*PCR, cluster 4 was detected. Cluster 2 was observed only within the native and restored prairie. Furthermore, there was no clear pattern change across treatments. These three methods showed evidence of *Frankia* diversity in soils devoid of host plants, confirming previous reports on the saprophytic properties of frankiae (4, 5). However, there was a clear bias on these techniques leading to differences in data delivery.

In the fourth chapter of this dissertation, we tried to associate previous and newly developed *q*PCR methods, with a recently published *nif*H targeted Illumina approach (1). The study was designed to assess the impact of vegetation on a given indigenous *Frankia* population in soil, and then assess *Frankia* diversity within nodules formed on host plants growing on these soils. Six host plants, *Alnus glutinosa*, *Alnus cordata*, *Shepherdia*

argentea, *Casuarina equisetifolia*, *Alnus viridis* and *Hippophaë rhamnoides*, and a non-host plant, *Betula pendula*, were subjected to a seven month microcosm experiment. A non-vegetated soil was used as a control. The study demonstrated that bioassay experiments used to relate *Frankia* diversity in soils to nodulation by specific strains are biased (8–10). Four out of the six host plants did nodulate while *Alnus viridis* and *Hippophaë rhamnoides* did not (data not shown). *Frankia* abundance and diversity in the rhizosphere of these two plants was comparable to the non-vegetated control, while a significant increase in abundance was observed within other host plant rhizospheres. This suggests a direct host mediated effect on *Frankia* rhizosphere populations. Potential experimental effects might explain the reduced growth performance of *Alnus viridis* and *Hippophaë rhamnoides*, which compromised *Frankia* growth in the soil leading to absence of nodulation. Analysis of nodule populations resulted generally in one dominant *Frankia* species per unit, with occasional presence of additional populations. For instance, a sequence belonging to cluster 3 was observed in nodules of *Alnus glutinosa*, confirming previous reports of occasional infection of Betulaceae by these *Frankia* groups (11). In addition, Cluster 1e sequences, previously observed in soils from Japan, Rwanda and South Africa (12, 13), were isolated from *Alnus glutinosa*, *Alnus cordata* and *Casuarina equistifolia*. The latter one was rather surprising, since the genus *Casuarina* is known to form root nodules with members of cluster 1c only. Illumina analysis revealed a higher sequence diversity in soils than found in the root nodules, with 24 major sequences, only four of which were encountered in the root nodules also. Analysis of population abundance resulted in higher values within vegetated soils, supporting previous reports on plant enhancing effects on *Frankia* populations in soils (4, 5, 14). Diversity within non-vegetated soils was dominated by cluster 1e, cluster 1b and cluster 3, while in vegetated soils, diversity was limited to clusters 3 and 1e. These data did not perfectly match *q*PCR results, which demonstrated dominance of cluster 3, cluster 1b and smaller fractions of cluster 1e. The ubiquity of cluster 1b and cluster 3 in soils independent of specific host plants has previously been reported (3, 5, 15). In general, there was a clear distinction between populations in the rhizosphere and those within root nodules, i.e. cluster 1a sequences were found in nodules of *Alnus glutinosa* but never detected in the rhizosphere by either method. Frankiae of cluster 1e, which accounted for

70% of *Alnus cordata* nodule sequences, were found in their rhizospheres by Illumina sequencing but never by *q*PCR. In addition, cluster 1b was prominent in soils but absent in nodules. Previous studies targeting *nif*H sequences have shown similar differences in diversity from nodules to respective rhizospheres (13, 16).

A previous study done in our lab shed light on *Frankia* populations from distant geographical locations. Two different research articles were published, contrasting *Frankia* nodulation capacities when using *Morella pensylvanica* as a trapping plant (16) and assessing *Frankia* diversity in these soils through a *nif*H based clone library approach (13). Generally, community composition was different from soils to nodules, supporting our recent findings reported in chapter 2. In this chapter, we used the same soil samples, collected from five locations from different continents, Alaska, Japan, Hungary, Peru and Rwanda, to re-assess *Frankia* populations with up to date protocols such as *nif*H targeted Illumina sequencing and *q*PCR targeting 23S rRNA. Illumina analysis revealed cluster 3 dominant in Hungary, Alaska and Rwanda, cluster 1b in Peru and cluster 1a in Japan. Data for the Alaska soil, which was sampled from under an alder stand, matched the clone library data with an abundance rate of 100% and 75% respectively. The four other soils showed dissimilar patterns with Illumina sequencing, revealing one major dominant cluster (ranging from 96% to 100%), while clone library analysis resulted in more diverse patterns. For example, the Hungary soil showed an abundance of cluster 3 accounting for 93% of the sequences obtained by Illumina sequencing compared to only 7% by clone library analyses. Other clades detected within the same soil by the clone library were cluster 1b representing 55% of the sequences and cluster 1a representing 37%. The high resolution of the Illumina sequencing approach, resulting in an average of 150,000 *Frankia* reads, was not comparable to the much smaller numbers, i.e. about 300 sequences, generated by the clone library. The *q*PCR analysis of the five soils resulted in a more diverse pattern highlighting the presence of cluster 4 and cluster 2 *Frankia* in four of the five soils excluding Alaska, which was dominated by cluster 1b only (confirming the pattern seen in the Illumina sequencing and the clone library for this soil sample). The *q*PCR data confirms results from chapters 1 and 3 that evidenced more diverse *Frankia* populations in soils devoid of host plants compared to the limited diversity in soils where host trees are present (5, 15).

This study made it possible to investigate *Frankia* dynamics in soils using different methodological tools, from SybrGreen and Taqman-based *q*PCR that target all clades to next generation sequencing, that has been proven successful in assessing nitrogen-fixing *Frankia* diversity in soils. In this dissertation, different molecular methods were used to show their usefulness on actual comprehensive ecological studies. However, in order to draw accurate conclusions on *Frankia* occurrence outside the ranges of their host plants, more studies need to be done on a broader range of samples. For example, other agricultural rotation experiments could be targeted and contrasted in order to see if the actual observed results are generally applicable. In addition, this research tried to compare *q*PCR to next generation sequencing, which proved to be a challenge. For future purposes, developing Illumina sequencing methods that target specific rRNA *Frankia* fragments, such as 23S rRNA gene fragments, would offer a more comparable tool to the actual *nif*H gene based protocol. In addition, as digital PCR is emerging as a state of the art quantitative tool, it could offer a more suitable method for *Frankia* quantification generating absolute data rather than the relative numbers delivered by *q*PCR. Furthermore, in order to better understand *Frankia* dynamics in soil habitats, factors such as the nature of organic material and its concentration in soil still need to be monitored in a controlled experiment in order to better understand the saprophytic and host associated behavior of *Frankia*.

REFERENCES

1. Rodriguez D, Guerra TM, Forstner MRJ, Hahn D. 2016. Diversity of *Frankia* in soil assessed by Illumina sequencing of *nif*H gene fragments. Syst Appl Microbiol 39:391–397.
2. Samant S, Sha Q, Iyer A, Dhabekar P, Hahn D. 2012. Quantification of *Frankia* in soils using SYBR Green based *q*PCR. Syst Appl Microbiol 35:191–197.
3. Samant S, Amann RI, Hahn D. 2014. Evaluation of the 23S rRNA gene as target for *q*PCR based quantification of *Frankia* in soils. Syst Appl Microbiol 37:229–234.

4. Samant SS, Dawson JO, Hahn D. 2015. Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. Syst Appl Microbiol 38:501–505.
5. 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.
6. Gtari M, Ghodhbane-Gtari F, Nouioui I, Ktari A, Hezbri K, Mimouni W, Sbissi I, Ayari A, Yamanaka T, Normand P, Tisa LS, Boudabous A. 2015. Cultivating the uncultured: Growing the recalcitrant cluster-2 *Frankia* strains. Sci Rep 5:1–9.
7. 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 83:1–18.
8. Smolander A. 1990. *Frankia* populations in soils under different tree species - with special emphasis on soils under *Betula pendula*. Plant Soil 121:1–10.
9. Smolander A, Sundman V. 1987. *Frankia* in acid soils of forests devoid of actinorhizal plants. Physiol Plant 70:297–303.
10. Paschke MW, Dawson JO. 1992. *Frankia* abundance in soils beneath *Betula-nigra* and other non-actinorhizal woody plants. Acta Oecol 13:407–415.
11. Benson DR, Dawson JO. 2007. Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiol Plant 130:318–330.
12. Wilcox DA, Cowan DA. 2016. Diversity of *Frankia* in root nodules of six *Morella* sp. from the Cape flora of South Africa. Plant Soil 406:375–388.
13. Mirza BS, Welsh A, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in soils from five continents. Syst Appl Microbiol 32:558–570.
14. Samant S, Dawson JO, Hahn D. 2016. Growth responses of introduced *Frankia* strains to edaphic factors. Plant Soil 400:123–132.
15. Tekaya S Ben, Ganesan AS, Guerra T, Dawson JO, Forstner MRJ, Hahn D. 2017. SybrGreen- and TaqMan-based quantitative PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils. Appl Environ Microbiol 83:1–12.

16. Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D. 2009. Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Syst Appl Microbiol 32:201–210.
17. Nouioui I, Sbissi I, Ghodhbane-Gtari F, Benbrahim KF, Normand P, Gtari M. 2013. First report on the occurrence of the uncultivated cluster 2 *Frankia* microsymbionts in soil outside the native actinorhizal host range area. J Biosci 38:695–698.
18. Soman C, Li D, Wander MM, Kent AD. 2017. Long-term fertilizer and crop-rotation treatments differentially affect soil bacterial community structure. Plant Soil 413:145–159.
19. Hahn D, Kester R, Starrenburg MJC, Akkermans ADL. 1990. Extraction of ribosomal RNA from soil for detection of *Frankia* with oligonucleotide probes. Arch Microbiol 154:329–335.