Appl. Environ. Microbiol. doi:10.1128/AEM.02833-16

SybrGreen and Taqman-based *q*PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils

Seifeddine Ben Tekaya[1], Abirama Sundari Ganesan[1], Trina Guerra[1], Jeffrey O. Dawson[2], Michael R.J. Forstner[1], Dittmar Hahn[1/*]

[1]Texas State University, Department of Biology, 601 University Drive, San Marcos, TX 78666, USA

[2]University of Illinois at Urbana-Champaign, Department of Natural Resources and Environmental Sciences, 1201 South Dorner Drive, Urbana, IL 61801, USA

Journal: Applied and Environmental Microbiology

Pages: 20

Figures: 3

Tables: 1

*corresponding author Dittmar Hahn
Department of Biology
Texas State University
601 University Drive
San Marcos, TX 78666
(512) 245-3372 (voice)
(512) 245-8713 (fax)
dh49@txstate.edu

SybrGreen and Taqman-based *q*PCR approaches allow assessment of the abundance and relative distribution of *Frankia* clusters in soils

Seifeddine Ben Tekaya[1], Abirama Sundari Ganesan[1], Trina Guerra[1], Jeffrey O. Dawson[2], Michael R.J. Forstner[1], Dittmar Hahn[1/*]

[1]Texas State University, Department of Biology, 601 University Drive, San Marcos, TX 78666, USA

[2]University of Illinois at Urbana-Champaign, Department of Natural Resources and Environmental Sciences, 1201 South Dorner Drive, Urbana, IL 61801, USA

Abstract

The nodule-forming, actinobacterial genus *Frankia* can generally be divided into 4 taxonomic clusters, with cluster 1, 2 and 3 representing nitrogen-fixing strains of different host infection groups, and cluster 4 representing atypical, generally non-nitrogen-fixing strains. Recently, *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 *q*PCR method applied, with abundance estimates of 10 to 15 x 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 a subgroup of cluster 1b 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.

Importance

Root nodule formation by the actinobacterium *Frankia* is host plant-specific, and largely, but not exclusively correlates with assignments of strains to specific clusters within the genus. Due to the lack of adequate detection and quantification tools, studies on *Frankia* have been limited to clusters 1 and 3, and were generally excluding clusters 2 and 4. We have developed tools for the detection and quantification of clusters 2 and 4, which can now be used in combination with those developed for clusters 1 and 3 to retrieve information on the ecology of all clusters delineated within the genus *Frankia*. 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 some host and non-host trees.

Key words 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,*

77 *Retanilla, Kentrothamnus* and *Colletia* (5, 6). In addition to the typical nitrogen-fixing frankiae, atypical,
78 generally non-nitrogen-fixing and/or non-nodulating frankiae have been identified into cluster 4 (4).

79 Since the first report of an isolation of *Frankia* from root nodules in 1978 (7), a large number of
80 isolates has been obtained for clusters 1 and 3 (8-11), and a few for cluster 4 (12, 13). An isolate
81 representing cluster 2, however, has been obtained only recently (14). Consequently, most studies on
82 *Frankia* have been performed on those from clusters 1 and 3, and far less on frankiae of clusters 2 and 4.
83 Information on the ecology of cluster 2 and 4 frankiae is therefore quite limited. We have recently
84 developed SybrGreen-based *q*PCR methods that used *nif*H or 23S rRNA genes as target to quantify
85 uncultured *Frankia* populations in different soils (15-17). *nif*H as target only detected the combination of
86 members of cluster 1 and 3, but not those of cluster 2 and 4, while 23S rRNA genes as target covered all
87 frankiae on the genus level, i.e. clusters 1, 2, 3 and 4 together. Targeting the 23S rRNA gene also allowed
88 us to distinguish between cluster 1 and 3 frankiae, and subgroups within cluster 1 (i.e. clusters 1a, 1b, and
89 1c) (15). The sum of individual detections generally equaled those on the genus level with both *nif*H and
90 23S rRNA genes as target, indicating that members of cluster 2 and 4 were not present at all or not in
91 detectable numbers in the soils analyzed (15, 17, 18). However, this statement is highly speculative since
92 direct proof of presence or absence of cluster 2 and 4 frankiae in these soils has not been provided due to
93 the lack of adequate detection and quantification procedures.

94 In this study, we have amended our database of sequences of 23S rRNA gene fragments of *Frankia*
95 strains representing clusters 1 and 3 (15) to include sequences of frankiae assigned to clusters 2 and 4.
96 Sequence alignments were used to design primers for the specific detection and quantification of *Frankia*
97 clusters by SybrGreen-based *q*PCR, and subsequently for the design of probes differentiating members of
98 clusters 1, 2 and 3 (i.e. presumably all nitrogen-fixing frankiae) from those of cluster 4 (i.e. generally non-
99 nitrogen-fixing frankiae). These probes were then used in combination with the specific primer
100 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.

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* collected in Jhika Gali, Murree, District Rawalpindi, Pakistan (33.9112833, 73.4239306), *C. japonica* from the Morton Arboretum in Lisle, IL, USA (41.8167861, -88.0679528) and *Ceanothus* sp. from the Loda Cemetery Prairie Nature Preserve, IL, USA (40.5284721; -88.0717537). A single lobe was homogenized with a mortar and pestle in one ml of sterile water, and the homogenates transferred to Eppendorf tubes and centrifuged at 14,000 x g for one minute. The pellets were washed once with 0.1% sodium pyrophosphate in water (wt/vol), 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, 19, 20) were grown in P+N medium for 2 weeks (21), harvested by centrifugation at 14,000 x g for 5 minutes, and washed twice with sterile distilled water. Nodule pellets as well as pellets of pure cultures (approx. 50 mg) were re-suspended in 95 µl of distilled water, and lysed after addition of 5 µl of proteinase K solution (Promega, Madison, WI, 30 U mg^{-1}, 10 mg ml^{-1} in water) and incubation at 37ºC for 20 minutes (22). Afterwards, 0.5 µl of 10% SDS solution was added and the mixtures incubated at 37°C for another 3 hours which was followed by a final incubation at 80ºC for 30 minutes (22).

PCR amplification From these lysates, 2 µl were used as 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 dNTP mix, 0.5 µl each primer (0.4 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 or pure culture lysate, and 0.2 µl *Taq* DNA polymerase (5 U $µl^{-1}$; Gene Script, Piscataway,

126 NJ) that was added after an initial incubation at 96°C for 10 minutes. The addition of *Taq* polymerase was
127 followed by 35 rounds of temperature cycling (96°C for 30 seconds, 60°C for 30 seconds and 72°C for 45
128 second) and a final 7 minute incubation at 72°C. Sub-samples of the reactions (5 µl) were checked for
129 amplification products by gel electrophoresis (1% agarose in TAE buffer, wt/vol) after staining with
130 ethidium bromide (0.5 µg ml^{-1}) (23).

131 **Sequence analyses** Amplified 23S rRNA gene fragments were cleaned using Shrimp Alkaline
132 Phosphatase and Exonuclease I (Affymetrix, Santa Clara, CA) following the manufacturer's protocols,
133 and then sequenced bidirectionally using BigDye Terminator v3.1 (Applied Biosystems, Foster City,
134 CA), with the same primers used for PCR. Sequences were analyzed on a 3500 Genetic Analyzer (Life
135 Technologies, Carlsbad, CA), and deposited at Genbank under accession numbers LT576423 to
136 LT576449.

137 **Phylogenetic analyses** Sequences from amplified 23S rRNA gene fragments obtained from uncultured
138 frankiae from root nodules of all plants analyzed and those of pure cultures of cluster 4 *Frankia* were
139 trimmed to lengths between 141 and 152 bp to match those in our database (24, 25), assembled in
140 Geneious 9.1.4 (Biomatters Ltd, Auckland, New Zealand), and checked in GenBank/EMBL databases
141 using the BLAST algorithm (26). Representative sequences from confirmed *Frankia* strains of all 4
142 clusters were added from our and GenBank/EMBL databases, and aligned by the Geneious alignment
143 tool. The identity and relationship among the sequences amplified were evaluated using neighbor joining
144 (NJ) (27), maximum likelihood (ML) (28), and Bayesian analyses (29). All of these analyses were
145 conducted from within Geneious 9.1.4. The neighbor joining analyses utilized the HKY85 model to
146 correct for substitution bias (Hasegawa et al 1985). Model parameters for maximum likelihood, which
147 were estimated by the general time reversible model (GTR) with gamma (30), were used as input in a ML
148 heuristic search using RAxML (31). Bootstrap values (32) were estimated from a heuristic search with
149 random stepwise addition sequence for 10,000 NJ, and 1,000 ML iterations. MRBAYES version 3.1.2 (29)

150 was implemented for ten million generations, saving every thousandth tree, and with a burnin of one
151 million trees using the General Time Reversible with Gamma substitution model.

152 **Primer and probe design and evaluation** Aligned sequences were amended with sequences of other
153 target and non-target organisms and used to manually check for and design forward primers specific for
154 cluster 2 and 4 frankiae, i.e. primers 23Dat1578f and 23NNF1561, respectively, which could be used with
155 reverse primer 23Fra1769r in SybrGreen-based *q*PCR (Table 1). In addition, two probes, one targeting all
156 frankiae of clusters 1, 2, and 3 (and thus supposedly all nitrogen-fixing frankiae), probe NF1715f ($^{5'}$[6-
157 FAM]-TGG TTG TCC TGG GGC AAG GGT GTA GG-[TAMRA]), and a second targeting cluster 4
158 frankiae (and thus generally non-nitrogen-fixing frankiae), probe NNF1715f ($^{5'}$[6-FAM]-CGG GGT AAG
159 CGT GTA GG ACG ACG TGT A-[TAMRA]) were designed, and subsequently evaluated in Taqman-
160 based *q*PCR using the genus or subgroup-specific primer sets from SybrGreen-based applications for
161 amplification. Selected primers and probes were checked for low potential of self- and hetero-dimer
162 formation using OligoAnalyzer 3.1 (www.idtdna.com/calc/analyzer) and for target specificity using
163 TestPrime 1.0 and TestProbe 3.0 (33) from the SILVA rRNA database project (www.arb-silva.de,
164 accessed 09/29/2016) (34). Annealing temperatures for all primer combinations were tested in *q*PCRs
165 with DNA of representative *Frankia* strains or PCR products from uncultured endophytes of clusters 1, 2,
166 3, and 4, respectively, and quantifications compared between SybrGreen and Taqman-based analyses,
167 both performed in an Eco Real-time PCR system (Illumina, San Diego, CA).

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

175 Primer combinations targeting 23S rRNA sequences were also used for Taqman-based quantification,
176 though in combination with probes NF1715f or NNF1715f. Except for cluster 1b, all Taqman-based
177 analyses were carried out in triplicate in a volume of 10 µl containing 5 µl of Sso ADV Probe Mix
178 (BioRad), 0.2 µl of forward and reverse primers (100 nM each), 0.25 µl of probe (250 nM each) and 1 µl
179 of DNA template. An initial denaturation at 95°C for 5 minutes was followed by 40 cycles of 60°C for 60
180 seconds. For cluster 1b, primer concentrations were 300 nM each, and cycles consisted of 58°C for 60
181 seconds, followed by 72°C for 30 seconds.

182 Quantification was based on standard curves generated from purified PCR products of *nif*H or 23S
183 rRNA genes of strains Ag45/Mut15, ArI3, CcI3, EAN1pec, and AgB1.9 or uncultured frankiae from root
184 nodules of *Coriaria* depending on the primer combination. Amplicons were generated using the genus-
185 specific primers (Table 1), and concentrations measured with a Qubit® 2.0 Fluorometer (Life
186 Technologies, Carlsbad, USA). Copy numbers were calculated from concentrations
187 (http://www.uri.edu/gsc/cndna.html) and normalized after *q*PCR quantification with the primer
188 combination targeting all nitrogen-fixing frankiae. Copy numbers were divided by copy numbers of the
189 23S rRNA gene per genome (2 or 3 for pure cultures, 2.5 for unknown populations) to relate copy
190 numbers to *Frankia* cell numbers (15).

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

197 Further method assessments used soil samples that were obtained from seven sites in Illinois, in close
198 proximity to Urbana-Champaign, IL, USA. Soils included 4 previously analyzed sites, i.e. sites ABA
199 (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 post-glacial 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. 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 soil samples (dry wt.) using the SurePrep™ Soil DNA Isolation Kit (Fisher Scientific, Houston, TX) with small modifications as described before (16). Ten-fold dilutions were used as 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* Typhimurium (ATCC14028) cells detected by *q*PCR-based quantification of a 268-bp *inv*A gene fragment before and after extraction as described previously (16).

223 **Statistical Analysis** One way ANOVA and pairwise multiple comparison procedures (Holm-Sidak
224 method) were used in SigmaPlot 13.0 (Systat Software Inc., San Jose, USA) to assess the effects of
225 different *q*PCR procedures on abundance estimates for frankiae, with a significance level $P<0.05$.

226 **Results**

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

236 **Primer and probe design and evaluation** The amended database of partial 23S rRNA gene sequences of
237 *Frankia* strains was used to design two forward primers, 23Dat1578f targeting cluster 2 frankiae, and
238 23NNF1561f targeting cluster 4 strains. Primer 23Dat1578f was specific with no mismatches to
239 sequences of all cluster 2 frankiae, and 3-7 mismatches to those of other *Frankia* strains in our database.
240 TestPrime 1.0 and TestProbe 3.0 analyses using the SILVA database revealed that primer 23Dat1578f (*E.*
241 *coli* position 1415) was specific for the two target organisms in the database, i.e. the *Frankia* endophyte
242 from *Datisca glomerata* (Dg1, CP002801) and strain BMG5.1 (JWIO01000013), and displayed at least 3
243 mismatches to sequences of non-target frankiae (i.e. 3 mismatches for representatives of cluster 1c, CcI3,
244 BMG5.23 and Thr; and 4 mismatches for cluster 1a and 3 *Frankia* strains), while other organisms
245 (*Prochlorococcus* sp.) displayed at least 5 mismatches. Together with primer 23Fra1769r, specific
246 detection was achieved for the target organisms, i.e. cluster 2 frankiae represented by the endophyte from
247 *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, 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, 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, with 62-68°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 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.

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 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 ($p = 0.1$).

SybrGreen-based quantification using the *nif*H gene as target to quantify frankiae of clusters 1 and 3 resulted in abundance estimates of 10 to 15 x 10^5 cells [g soil]$^{-1}$ depending on the soil (Fig. 3). Except for soil BAHF ($p = 0.002$), these estimates were not significantly different from 23S rRNA gene 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. 3). 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).

297 For prairie soils, SybrGreen-based detection using *nif*H as target resulted in significantly lower
298 abundance estimates (i.e. 5 to 10 x 10^5 cells [g soil]$^{-1}$) than SybrGreen-based detection of all clusters in
299 the genus (i.e. 15 to 25 x 10^5 cells [g soil]$^{-1}$) ($p < 0.001$) (Fig. 4). Frankiae of clusters 1, 2, 3 and 4 were
300 detected in two of the three soils, with their sum of their abundance estimates matching the estimates on
301 the genus level (Fig. 4). Taqman-based *q*PCR resulted in similar composition of the subgroups, and
302 detection of nitrogen fixing frankiae equal to the sum of subgroups in clusters 1, 2, and 3 (Fig. 4).
303 Subgroup 1a represented by strain ArI3, and subgroup 1c represented by strain CcI3, were not detected at
304 all and thus not found to be abundant in any of the soils tested. Despite incomplete coverage of cluster 4
305 *Frankia* strains by both forward primer 23NNF1561f and probe NNF1715f, comparable abundance
306 estimates for cluster 4 frankiae in these soils were obtained by both SybrGreen- and Taqman-based *q*PCR
307 (p values between 0.05 and 0.5) (Fig. 4).

308 **Discussion**

309 Despite the short length of the sequences and thus low or no bootstrap support for major clades, the
310 topology of our phylogenetic analyses confirmed cluster assignments of *Frankia* strains and uncultured
311 endophytes from root nodules derived from comparative analyses of 16S rRNA gene sequences (4).
312 Cluster assignments were also consistent with previous analyses using comparative sequence analyses of
313 the insertion in Domain III of the 23S rRNA gene (24), partial gyrase B (*gyr*B), nitrogenase reductase
314 (*nif*H) or glutamine synthetase II (*gln*II) sequence analyses (35), sequence analysis of the 16S-23S rRNA
315 internally transcribed spacer (ITS) (36), or protein mass fingerprints from whole cells (37). While our
316 analyses of uncultured endophytes in root nodules of *Datisca cannabina*, *Coriaria nepalensis*, *C.*
317 *japonica* and *Ceanothus* sp. revealed a limited sequence diversity and thus resulted in the assembly of a
318 concise cluster 2, sequences of cluster 4 frankiae were more diverse, with significant sequence diversity
319 establishing 3 subgroups.

320 Subgroups were previously described for clusters 1 and 3 (22, 38), with those of cluster 1 depicting
321 remarkable physiological differences between each other (39). 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 nutrient resource (39), and subgroup 1c, the *Casuarina*-infective strains that require the presence of host plants for growth (40). 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 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 USA 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 (41, 42), 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 frankiae 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*,

347 *Datisca cannabina*, and *Ceanothus americanus* (CN4, AgB1.9, AgW1.1, EuI1c, DC12 and CaI1) (12, 13,
348 20, 43, 44). The reduction in annealing temperature from 66 to 62°C allowed us to increase coverage to
349 two of the three subgroups detected, without losing specificity for other *Frankia* strains. Taqman-based
350 quantification using genus-specific primers and a probe targeting cluster 4 frankiae resulted in the
351 detection and reliable quantification of representative strains of these subgroups, and thus might provide
352 an adequate alternative to detection using specific forward primers. The new forward primers 23Dat1578f
353 and 23NNF1561f, as well as the design of probes targeting nitrogen-fixing frankiae (clusters 1, 2 and 3)
354 and cluster 4 frankiae (atypical, generally non-nitrogen fixing strains), now allows us to expand *q*PCR
355 analyses of frankiae in soils from clusters 1 and 3 and subgroups within cluster 1 to include cluster 2 and
356 4, and distinguish between nitrogen-fixing and non-nitrogen-fixing populations (Table 1).

357 Previous analyses of soils ABA, BAHF, LWRB and RBW revealed cell densities of about 10^6 cells (g
358 soil)$^{-1}$, with cluster 1b representing the most prominent *Frankia* population, while cluster 3 frankiae were
359 present in small numbers, and cluster 1a, 1d and 1c usually absent (17). These results were largely
360 confirmed in our current analyses, where specific analyses retrieved frankiae of cluster 1b only (soils
361 ABA and BAHF) while soils LWRB and RBW harbored small numbers of cluster 3 frankiae as well (Fig.
362 3). Frankiae of clusters 2 and 4 remained undetected indicating that these populations are either absent or
363 present in numbers below the detection limit. Cluster 1b frankiae have been detected as major populations
364 in several studies, with absolute numbers depending on the sampling depth, physicochemical conditions
365 and the vegetation (15, 17, 18). These results demonstrated differential effects of environmental
366 conditions including plant species, carbon resources, and matric potentials on the fate of specific *Frankia*
367 strains in soil. These factors could potentially affect subpopulations of indigenous frankiae of clusters 2
368 and 4 that were both detected in all three prairie soils, but not in soils under *A. glutinosa* or *B. nigra*. Their
369 presence in prairie soils but not in soils under host plant *A. glutinosa* and non-host plant *B. nigra* suggests
370 that vegetation is affecting the abundance of frankiae of these clusters. Populations of cluster 2 frankiae
371 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 two centuries, using bioassays with *Coriaria myrtifolia* as capture plant (45). This assumption, however, needs to be assessed under controlled conditions, with different plant species as variable 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 prairies 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* (46, 47), with natural resistance to infection exhibited by different progenies of *A. glutinosa* (12, 48). 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 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 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.

Acknowledgements

The authors 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.
2. **Huss-Danell K.** 1997. Actinorhizal symbioses and their N-2 fixation. New Phytol **136:**375-405.
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.
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.
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.
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.
8. **Diem HG, Gauthier D, Dommergues YR.** 1982. Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Can J Microbiol **28:**526-530.

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.

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.

12. **Hahn D, Starrenburg MJC, Akkermans ADL.** 1988. Variable compatibility of cloned *Alnus glutinosa* ecotypes against ineffective *Frankia* strains. Plant Soil **107:**233-243.

13. **Mirza MS, Hahn D, Akkermans ADL.** 1992. Isolation and characterization of *Frankia* strains from *Coriaria nepalensis*. System Appl Microbiol **15:**289-295.

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. Scientific Reports **5:**13112; doi: 13110.11038/srep13112.

15. **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.

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.

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.

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.

19. **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.

20. **Mirza MS, Janse JD, Hahn D, Akkermans ADL.** 1991. Identification of atypical *Frankia* strains by fatty acid analysis. FEMS Microbiol Lett **83:**91-97.

21. **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.

22. **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.

23. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

24. **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.

25. **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.

26. **Pearson WR, Lipman DJ.** 1988. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA **85:**2444-2448.

27. **Saitou N, Nei M.** 1987. The Neighbor-joining Method: A new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406-425.

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

29. **Ronquist F, Huelsenbeck JP.** 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572-1574.

30. **Tavaré S.** 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. American Mathematical Society: Lect Math Life Sci **17:**57-86.

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

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

33. **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. doi:10.1093/nar/gks1808.

34. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner F-O.** 2013. Nucleic Acids Research. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools**.** Nucleic Acids Res **41:**D590-D596. doi: 10.1093/nar/gks1219.

35. **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. Anton Leeuw **100:**579-587.

36. **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.

37. **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.

38. **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.

39. **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.

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

41. **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.

42. **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.

43. **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.

44. **Baker D, Newcomb W, Torrey JG.** 1980. Characterization of an ineffective actinorhizal microsymbiont, *Frankia* sp. EuI1 (Actinomycetales). Can J Microbiol **26:**1072-1089.

45. **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.

46. **Wolters DJ, Akkermans ADL, Van Dijk C.** 1997. Ineffective *Frankia* strains in wet stands of *Alnus glutinosa* L. Gaertn. in the Netherlands. Soil Biol Biochem **29:**1707-1712.

47. **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.

48. **Van Dijk C, Sluimer A.** 1994. Resistance to an ineffective *Frankia* strain type in *Alnus glutinosa* (L.) Gaertn. New Phytol **128:**497-504.

520 **Legends**

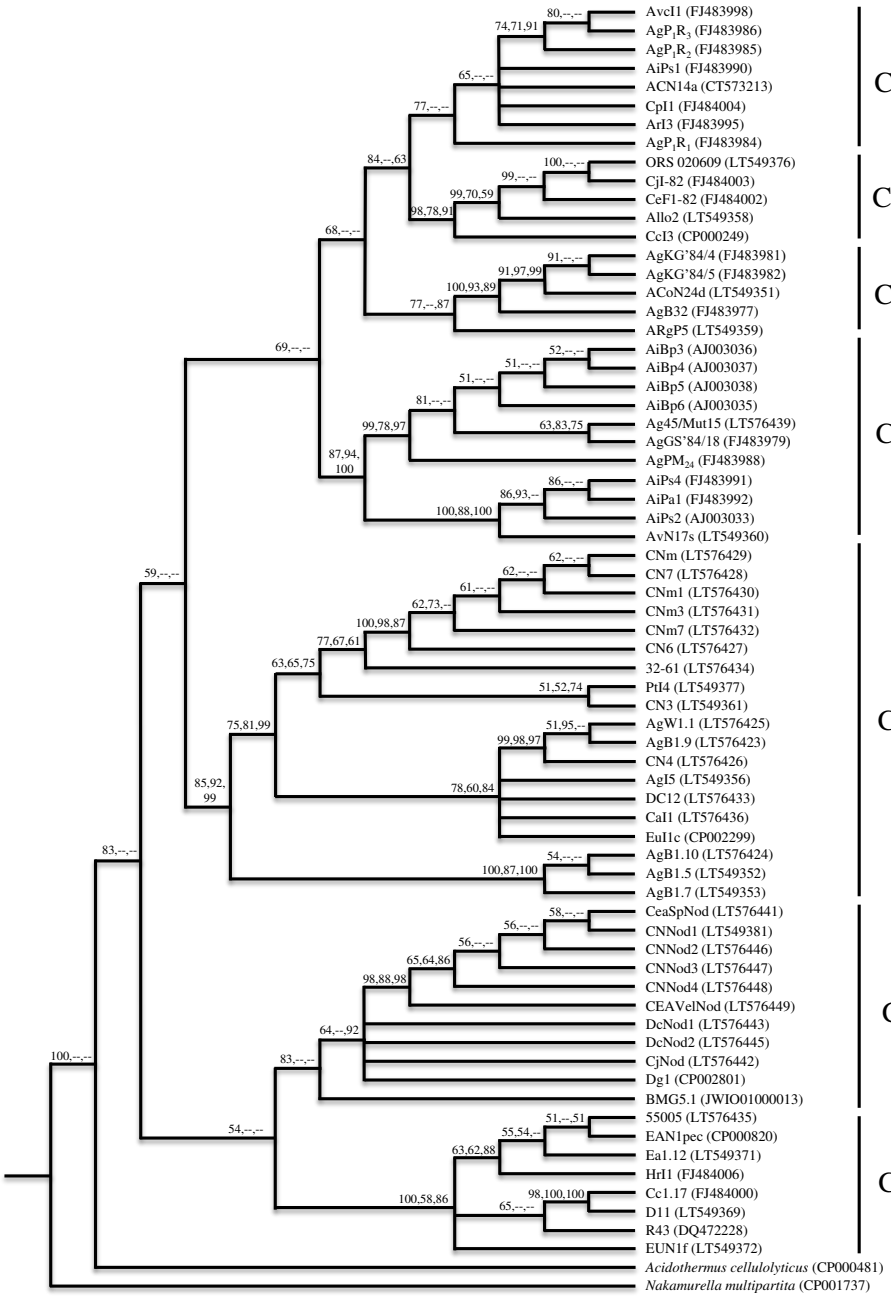
521 **Figure 1** Sequence relationships for selected *Frankia* strains provided to show probe utility. Numbers
522 above the branches represent the bootstrap values from a neighbor-joining (NJ) bootstrap analysis (10,000
523 replicates) using the HKY85 correction, followed by maximum likelihood bootstrap (1,000 replicates)
524 and Bayesian analyses values, respectively, noted for clades with greater than 50% bootstrap support. All
525 results are plotted on the NJ bootstrap topology. Both the ML and Bayes analyses resolved generally
526 similar topologies, although there are fewer supported tip nodes in the latter two analyses and the larger
527 clades' basal relationships are either not resolved (ML) or have an alternative arrangement of the main
528 clades (Bayes basal arrangement not depicted).

529 **Figure 2** SybrGreen and Taqman based *q*PCR quantification of amplicons from representative pure
530 cultures or endophytes of *Frankia* clusters 1, 2, 3 and 4, and of subgroups a, b and c within cluster 1.
531 Individual amplicons or mixtures of amplicons were quantified using either genus- or (sub)cluster-
532 specific primer combinations in SybrGreen-based analyses, or additional probes targeting either cluster 1,
533 2 and 3 frankiae or cluster 4 frankiae.

534 **Figure 3** SybrGreen and Taqman based *q*PCR quantification of *Frankia* clusters 1, 2, 3 and 4 in soils
535 vegetated with host trees (*Alnus glutinosa*, soils ABA and BAHF) or non-host trees (*Betula nigra*, soils
536 LWRB and RBW). Quantification (from left to right) used *nif*H gene fragments as a target detecting
537 clusters 1 and 3, or 23S rRNA gene fragments generated with primer combinations detecting the genus
538 *Frankia*, i.e. all clusters 1, 2, 3 and 4, or primer combinations specific for clusters 1a/d, 1b, 1c, 2, 3, and 4
539 (presented as the sum of the individual clusters and subgroups detected). Only frankiae of subgroups 1b
540 and cluster 3 were detected in these soils, while the remaining clusters and subgroups remained
541 undetected (all SybrGreen). Taqman-based detection of 23S rRNA fragments resulted in similar
542 composition of frankiae with respect to clusters and subgroups as SybrGreen-based detection, with their
543 sum representing similar quantification values as 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 (*).

Figure 4 SybrGreen and Taqman based *q*PCR quantification of *Frankia* clusters 1, 2, 3 and 4 in prairie soils (native, restored and cultivated). 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). Frankiae of subgroups 1b and clusters 2, 3 and 4 were generally detected in these soils, while subgroups 1a/d and 1c remained undetected (all SybrGreen). Statistically significantly different values between treatments of the same sample are highlighted with one or two asterisk (*, **).



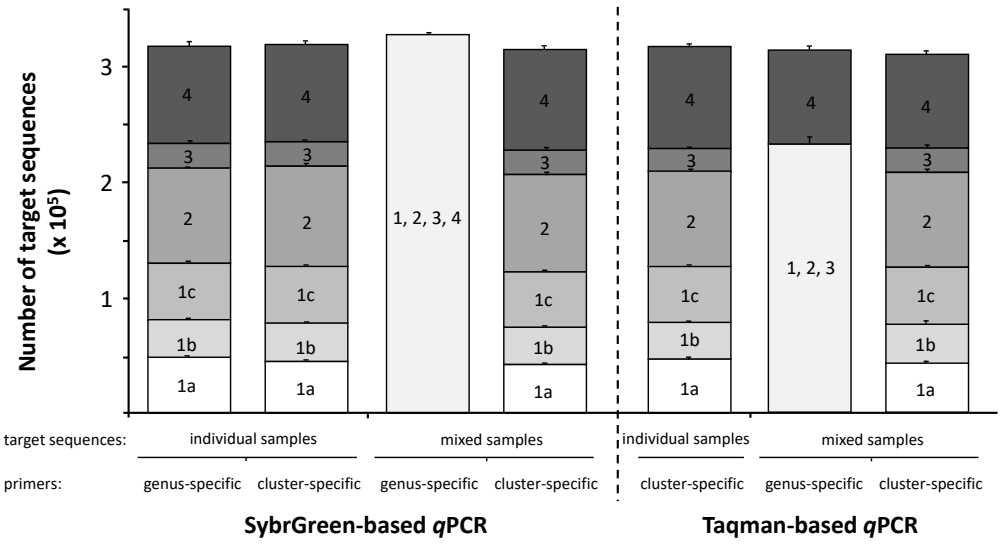


Fig. 2

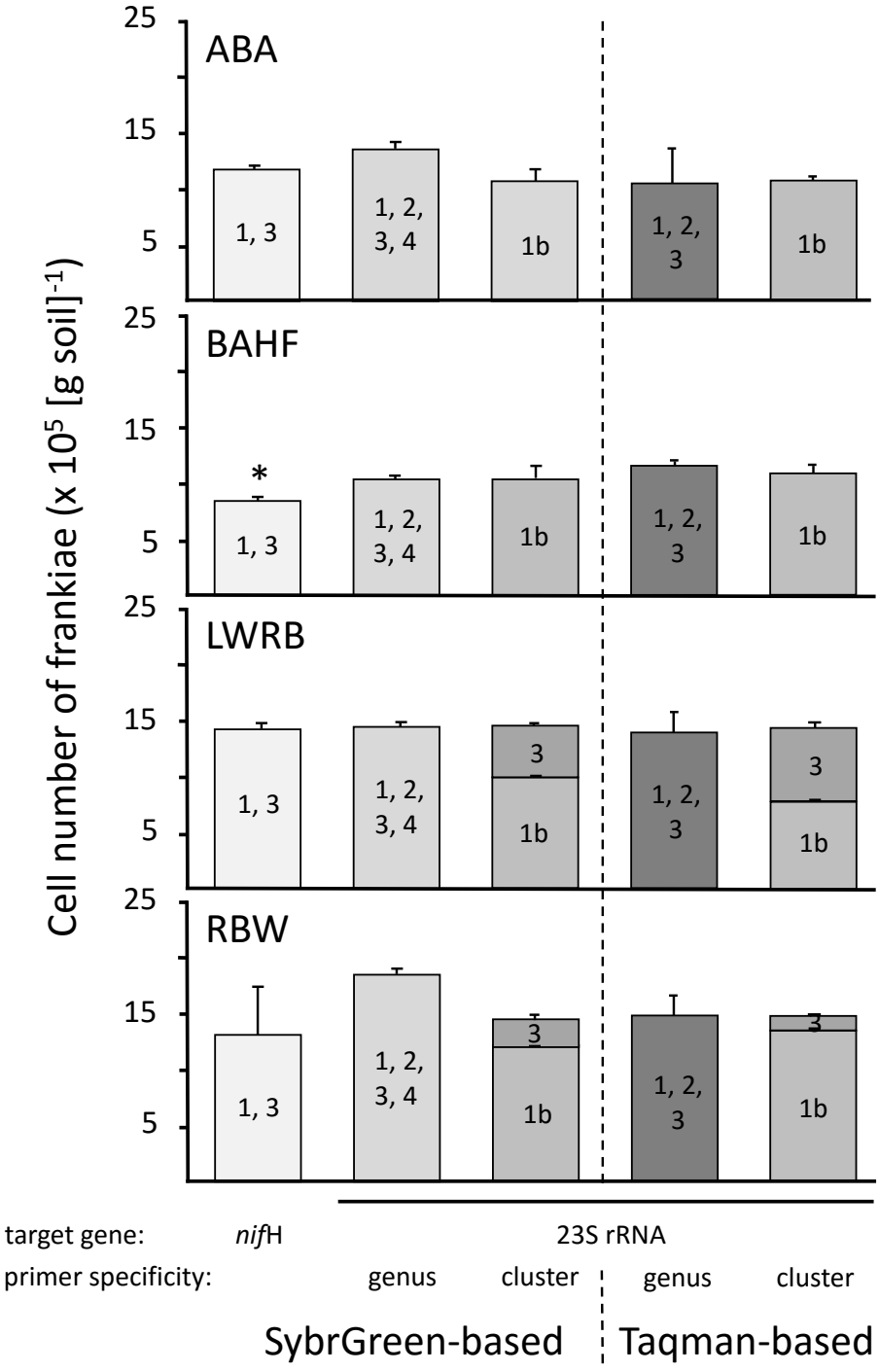


Fig. 3

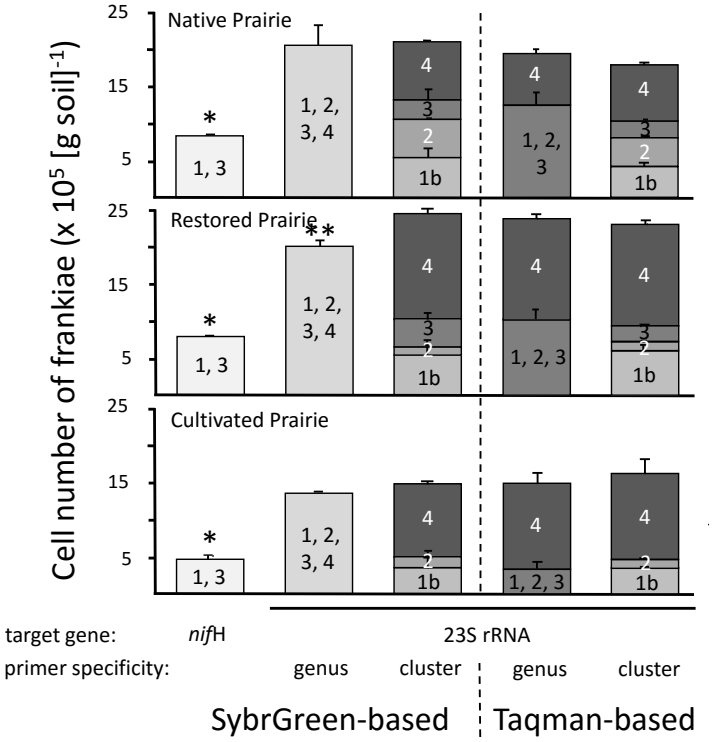


Fig. 4

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	(16)
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	(17)
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	(17)
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