MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 3864–3877 doi: 10.1111/mec.12339

Phylogeny and assemblage composition of *Frankia* in *Alnus tenuifolia* nodules across a primary successional sere in interior Alaska

M. D. ANDERSON,*† D. L. TAYLOR† and R. W. RUESS†

*Department of Biology, Macalester College, 1600 Grand Ave, Saint Paul, MN 55105, USA, †Institute of Arctic Biology, University of Alaska, Fairbanks, AK 99775, USA

Abstract

In nitrogen (N) fixing symbioses, host-symbiont specificity, genetic variation in bacterial symbionts and environmental variation represent fundamental constraints on the ecology, evolution and practical uses of these interactions, but detailed information is lacking for many naturally occurring N-fixers. This study examined phylogenetic host specificity of *Frankia* in field-collected nodules of two *Alnus* species (*A. tenuifolia* and *A. viridis*) in interior Alaska and, for *A. tenuifolia*, distribution, diversity, spatial autocorrelation and correlation with specific soil factors of *Frankia* genotypes in nodules collected from replicated habitats representing endpoints of a primary sere. *Frankia* genotypes most commonly associated with each host belonged to different clades within the *Alnus*-infective *Frankia* clade, and for *A. tenuifolia*, were divergent from previously described *Frankia*. *A. tenuifolia* nodules from early and late succession habitats harboured distinct *Frankia* assemblages. In early succession, a single genotype inhabited 71% of nodules with no discernable autocorrelation at any scale, while late succession *Frankia* were more diverse, differed widely among plants within a site and were significantly autocorrelated within and among plants. Early succession *Frankia* genotype occurrence was strongly correlated with carbon/nitrogen ratio in the mineral soil fraction, while in late succession, the most common genotypes were correlated with different soil variables. Our results suggest that phylogenetic specificity is a significant factor in the *A. tenuifolia*-*Frankia* interaction and that significant habitat-based differentiation may exist among *A. tenuifolia*-infective genotypes. This is consistent with our hypothesis that *A. tenuifolia* selects specific *Frankia* genotypes from early succession soils and that this choice is attenuated in late succession.

Keywords: *Alnus*, *Frankia*, mutualism, nitrogen fixation, plant-microbe interactions

Received 7 January 2013; revision received 16 March 2013; accepted 19 March 2013

Introduction

Root nodule–based nitrogen (N)-fixing symbioses are ecologically and economically important interactions, providing large natural inputs to the global N cycle (Vitousek *et al.* 2002) and including many plants important in agriculture, forestry and bioremediation (Wheeler & Miller 1990; Graham & Vance 2003). These associations involve bacteria from α, β and possibly γ proteobacterial classes, actinobacteria and cyanobacteria and plants from at least ten angiosperm families and cycad gymnosperms (Vessey *et al.* 2005; Dawson 2008; Masson-Boivin *et al.* 2009). Across this broad group, specificity, in terms of phylogenetic breadth of the partner associated with a given plant or bacterial taxon, varies widely (Young & Johnston 1989; Swensen & Benson 2008), and diversity exists even within relatively specific associations (e.g. Spoerke *et al.* 1996; Simonet *et al.* 1999). Variation among bacterial symbionts can affect growth, N-fixation and reproduction of host plants (e.g. Parker 1995; Heath 2010), and conversely, variation

Correspondence: Mike Anderson, Fax: (651) 696 6443;
E-mail: andersonm@macalester.edu

among plant genotypes within a host species can affect bacterial reproduction (Heath 2010). Further, the benefits of the symbiosis for both partners can be affected by environmental variables such as nutrient concentration (Heath & Tiffin 2007; Heath *et al.* 2010) and herbivory (Heath & Lau 2011). Host specificity, genetic variation within suites of compatible bacteria and environmental variation thus represent important potential constraints on the co-evolution of N-fixing symbioses (Thompson 2005; Heath 2010), on colonization (Parker *et al.* 2006), growth and ecosystem functioning (Anderson 2011) of N-fixing plants, and on the practical applications of specific N-fixing systems (e.g. Wheeler & Miller 1990; Graham & Vance 2003). Understanding how these factors interact in specific N-fixing systems in natural habitats requires detailed field descriptions of such systems, yet such descriptions are either unavailable or incomplete for most N-fixing systems.

A large body of literature exists describing genetic and phenotypic variation in natural populations of symbiotic N-fixing bacteria. Such variation often correlates strongly with host distribution and/or environmental variation, but the relative importance of the two factors can vary widely among specific systems (Benson & Dawson 2007) and is unclear from many *in situ* studies, due to several factors. First, a large number of studies utilize 'trap plants' to examine genetic variation in soil bacteria across different environments and/or host species. While such studies can indicate important factors contributing to the structure of potentially compatible soil bacterial populations *in situ* (e.g. Valdés 2008), and potential fitness consequences to plants of such bacterial distributions (*e.g.* Thrall *et al.* 2007), without accompanying descriptions of nodule-dwelling bacteria, any effects of host–symbiont interactions *in situ* are unknown. Such interactions may include specificity among (Parker 2012) or within host species (Spoerke *et al.* 1996), and/or selection of bacterial genotypes based on symbiotic performance (i.e. 'host-choice') (*e.g.* Simms *et al.* 2006). Second, most studies of field-collected nodules are conducted in single locations (*e.g.* Wang *et al.* 2009), across large geographic areas (*e.g.* Simonet *et al.* 1999), or across assortments of unreplicated habitats (e.g. Navarro *et al.* 1999). Respectively, such studies fail to describe environmental variation, potentially confound geographic location and environmental variation as explanatory variables and have low power to discriminate explanatory signal from stochastic noise. Finally, very few field studies of nodule-dwelling populations examine bacterial variation within sampling sites. Descriptions of spatial patterns in such variation, partitioning of variation within and among host plants, correlations with specific environmental factors within sites and differences in all of these factors among different habitats can help constrain ecological hypotheses to testable sets (Bruns 1995) and are unavailable for most N-fixing symbioses.

Alders (*Alnus* spp., Betulaceae) are trees and shrubs, all species of which are 'actinorhizal', that is, form nodules with N-fixing actinobacteria of the genus *Frankia* (Dawson 2008). In Alaska, alders are ecologically important, occurring in the majority of ecoregions across the state, achieving extremely high abundance in some ecosystems, and making significant contributions to the N economy of areas in which they are present (Anderson 2011). In interior Alaska, symbiotic *Frankia* assemblages can differ widely between sympatric *Alnus* spp., and between successional habitats occupied by a given host species (Anderson *et al.* 2009). The present study aimed to characterize this variation in detail, focusing on five types of pattern: (i) phylogenetic relationships among *Frankia* in nodules of *A.incana* ssp. *tenuifolia* (hereafter *A. tenuifolia*) and *A. viridis* ssp. *fruticosa* (syn. *A. crispa*, hereafter *A. viridis*), and in *A. tenuifolia*: (ii) differences between early and late primary successional habitats, utilizing replicate sites representing each habitat, (iii) distribution of *Frankia* diversity within and among host plants in individual sites, and whether this differs between habitats, (iv) spatial patterns in *Frankia* distribution within sites, and (v) whether variation in *Frankia* is correlated with specific soil factors, and whether such correlations differ between habitats. This study was conducted at a restricted spatial scale (<15 km) to minimize the confounding effects of location and is part of a larger project examining the possible role of host-choice in the ecology of the *Alnus*-*Frankia* system in interior Alaska.

Materials and methods

Study system and field sites

Sites were located within the Bonanza Creek Experimental Forest (BNZ), part of the Long-Term Ecological Research (LTER) network and located approximately 30 km south-west of Fairbanks, Alaska, USA (64°48′N, 147°52′W). Detailed descriptions of the sites are available in the study by Anderson *et al.* (2009), and further details are on the BNZ-LTER website (http://www.lter.uaf.edu/). Briefly, sites represent early and late successional endpoints of a primary sere that initiates on alluvial deposits of the Tanana River. *A. tenuifolia* colonizes these deposits and forms a closed canopy ~5–10 years after initial substrate formation (Hollingsworth *et al.* 2010; Nossov *et al.* 2011). Succession proceeds through a stage dominated by balsam poplar (*Populus balsamifera* L.) beginning ~50 years postsubstrate, then white spruce (*Picea glauca* (Moench) Voss) after ~120 years (Chapin

et al. 2006). *A. tenuifolia* persists in the understory in these later stages, and *A. viridis*, which is generally absent in early stages, is relatively common in white spruce-dominated stands, where both species of *Alnus* often co-occur in close proximity.

Design and field methods

Three early (*A. tenuifolia*-dominated) and three late (white spruce-dominated) succession sites were sampled in the present study; these sites correspond to FPE1-3 and FPL1-3, respectively, in the study by Anderson *et al.* (2009). All sites were sampled between 21 and 28 July 2005. At each site, 20 *A. tenuifolia* plants were randomly selected by choosing the closest plant to a random point generated using three variables: (i) distance along (3–7 m), (ii) direction (right or left) and (iii) distance from (0–20 m) a 100-m transect line. In four of the six sites, this line was contiguous and oriented perpendicular to the trend of the river in order to capture the widest available diversity of habitat. At site FPE1, the narrowness of the alder stand made it necessary to orient the transect parallel to the river, and at site FPL1, the absence of *A. tenuifolia* from a large portion of the site made it necessary to construct the plot using two shorter (~50 m) transects perpendicular to the river but paralleling each other 250 m apart. Distance and compass bearing from each random point to the nearest plant were recorded, and bearing and GPS coordinates were recorded for the origin of each transect. To examine the diversity of *Frankia* occurring within individual plants, five of the twenty plants at each site were randomly chosen for 'intensive' sampling; ten nodule clusters were sampled from each of these plants and used for genetic characterization of symbionts. From each of the remaining 15 'extensive' plants, one nodule cluster was collected. For extensive plants, the first nodule cluster encountered for a plant was collected; for intensive plants, nodule clusters were collected haphazardly from all four quadrants in a 1 m radius circular plot surrounding the base of the plant. Where possible, nodule clusters from these plants were collected from different lateral root branches, but in late succession sites, low nodule numbers on plants made it necessary to collect multiple nodule clusters from individual lateral branches on most plants. For all collected nodules, lateral roots were traced to the plant stem prior to collection to ensure collection from the correct plant. At ten plants (every other plant along the transect), one soil core (5 cm diameter × 10 cm depth) was collected within 0.5 m of the base of the plant, separated into organic and mineral fractions by hand and transported to the laboratory in plastic zipper bags. Soil samples were dried at 65 °C for 72 h and stored at −20 °C until chemical analysis could be performed. Soil pH was measured on a 1:1 slurry of soil and deionized water. Total N and C were determined with a CN autoanalyzer (LECO Corporation, St. Joseph, MI, USA). Total phosphorus was determined colorimetrically using the molybdenum blue method with an amino-naphthol-sulfonic acid reducing agent (Jackson 1958). All soil variables were determined separately for organic and mineral horizons.

Frankia characterization

Nodule clusters were sterilized in the field immediately after collection by shaking in 10% bleach with Alconox for 3 min, followed by three 1-min rinses in DI water. Nodules were stored on dry ice until transport to the laboratory, then stored at −80 °C until DNA extraction could be performed. Prior to extraction, nodules were freeze-dried to constant weight and ground on a shaker mill (Retsch, Inc., Newtown, PA, USA). DNA was extracted from a subsample of each nodule cluster ($\leq$3 lobes) with the Plant DNeasy 96 kit (QIAGEN, Carlsbad, CA, USA), with the addition of 20 mg/mL lysozyme followed by a 30 min incubation at 37 °C, as recommended for Gram-positive bacteria. Polymerase chain reaction (PCR) (Mullis *et al.* 1986) was performed using dehydrated PCR 'beads' (Amersham Biosciences, Piscataway, NJ, USA), newly designed primers for the *nif*D-K intergenic spacer (IGS) region—forward, *nif* D1310frGC—CGCCAGATGCACTCCTGGGACTACT, reverse, *nif* KR331frGC—CGGGCGAAGTGGCTGCGG AA—and 35 cycles of 95 °C for 1 min, and 65 °C for 5 min. All successful PCRs were examined for sequence variation via restriction fragment length polymorphism (RFLP) based on separate digests (37 °C for 8 h) with *Cfo* I and *Hae* III (Promega, Madison, WI), as described in the study by Anderson *et al.* (2009). PCR and restriction digest products were measured via electrophoresis on 1.5% (SeaKem; Cambrex, Rockland, ME, USA) and 3% (1% SeaKem/2% NuSieve; Cambrex) agarose, respectively, and visualized via ethidium bromide/UV fluorescence. Digest fragments shorter than 50 bp were excluded from the analysis. Each unique combination of fragment patterns across both enzymes was given a numerical 'RF' designation.

To verify DNA sequence variation of *nif*D-K PCR product indicated by RFLP, and to examine each RF group for undetected variation, a subset of nodules from each RF group was sequenced which included multiple nodules from the most frequent RF groups. DNA extracts from nodules representing a subset of the *nif*D-K-based RF groups were also examined for variation in the 16S-23S intergenic spacer locus (rIGS), to verify diversity patterns observed for the *nif* locus, and

to aid in phylogenetic placement of the RF types. Primers used for rIGS were 16Sfr1—GTCACGAAAGTCG GTAACA and 23SRfr1—CCAGTGAGCTAATACGCAC, with a temperature programme consisting of 30 cycles of 94 °C for 1 min, 58 °C for 30 s and 72 °C for 4 min, with a 4 min 94 °C initial denaturing and 20 min 72 °C final extension step. For both loci, PCR product was cleaned using the QiaQuick PCR cleanup kit (QIAGEN) and quantified using a Nanodrop spectrophotometer (Nanodrop Products, Wilmington, DE, USA). Cycle sequencing reactions were performed with ABI BigDye (2.0 μL/rxn) (Applied Biosystems Inc., Carlsbad, CA, USA), using PCR primers and an annealing temperature of 58 °C. Capillary sequencing was performed on an ABI 3100 Genetic Analyzer (Applied Biosystems Inc.).

Data analysis

To determine whether most of the *Frankia* diversity in each sampling site had been detected, we generated sample-based rarefaction curves based on both interpolation-based [Mau Tau (Colwell *et al.* 2004)] and extrapolation-based [Chao1 and Chao2 (Chao 1984)] richness estimates. All estimates were calculated using Estimate S version 8.2 (Colwell 2009), with plant treated as sample. We concluded that saturation had been reached if the 95% confidence intervals for the interpolation- and extrapolation-based estimators overlapped in the final sample.

The contributions of among-plant (within-site) and among-site (within-stage) differences to variation in symbiotic *Frankia* assemblages were compared using analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) with Arlequin 3.1 (Excoffier *et al.* 2005). Separate analyses were run for intensive and extensive data sets, and for each successional stage. For the intensive data set, plants were included as populations and sites as groups; for the extensive data set, sites were included as populations and only within- and among-population variation was examined. Both sequence and RFLP data were examined in the AMOVA, and both gave qualitatively similar results.

Prior to spatial analysis, all distances and compass bearings for the location of plants and nodules in each site were converted to *x*-*y* coordinates. Global spatial autocorrelation for each site was assessed using Mantel tests (Mantel 1967) of correlation between a Euclidean distance matrix calculated for all nodules collected in a site (n = 51–61 nodules) and a binary connectivity matrix which coded nodules with shared RF haplotype as '1', and '0' otherwise. Significance of the correlation was determined using a permutation test that included 999 permutations. Local autocorrelation within each site was assessed using Mantel correlograms, with ten distance classes defined so that each class contained the same number of pairwise comparisons, but with the lowest distance class adjusted to $\leq$1 m to encompass all nodules collected within intensively sampled plants. Correlogram analyses were performed on both multivariate data (all RF groups included) and separately for each of the most common RF types in each site. Significance was assessed using Bonferroni-adjusted *P*-values from a permutation test that included 999 permutations. All spatial analyses were performed with the program PASSaGE2 (Rosenberg & Anderson 2011).

Correlations between abundance of RF types and soil variables were investigated using multiple nominal logistic regression in JMP 9.0 with RF type as the dependent variable. Separate analyses were performed for each successional stage. Variables included were pH, N, C and P concentration, as well as C/N, C/P and N/P ratios for both organic and mineral horizons. Initial model construction utilized a backward stepwise procedure which began with all variables in the model, then removed the variable with the smallest chi-square at each step until $P \leq 0.05$ for all independent variables. Previously removed variables were retested in the resulting model until the whole model chi-square and log likelihood difference between the full and reduced models were maximized, and $P \leq 0.05$ for all included variables. Selection of variables to add during the forward stepwise phase was informed by bivariate logistic regressions performed between RF type and each soil variable prior to model construction, which aided in selection of promising variables during the forward procedure. Independent variables found to be significant for all RF types were tested in separate models constructed for each of the most common RF types in each stage.

Phylogenetic analysis

Separate phylogenetic analyses were performed for *nif*D-K and rIGS loci. Comparison sequences included several derived from nodules of *Alnus viridis* ssp. *fruticosa* that co-occurred with *A. tenuifolia* in our FPL sites (sampling methods described in Anderson *et al.* 2009), as well as all sequences of comparable length that were available on the NCBI database and found via BLASTn searches (accession numbers provided in tree figures). Alignments were generated using ClustalX (Thompson *et al.* 1997) and corrected by eye and trimmed to common length using BioEdit (Hall 1999). Substitution models were selected for each locus using jModeltest (Posada 2008), with all 88 available models included, the base tree ML optimized and Akaike's corrected

information criterion (AIC_C) as the selection criterion. The best-fit model for the rIGS data was the General Time-Reversible model (Tavaré 1986) with gamma-distributed rate heterogeneity (GTR+G); for the *nif*D-K data, it was Kimura's three-parameter (Kimura 1981) with unequal base frequencies and gamma-distributed rate heterogeneity (TPM3uf+G).

Three phylogenetic analysis programs were utilized in order to assess the robustness of our results to different tree selection criteria and search algorithms. All programs were run on the CIPRES network (Miller *et al.* 2010). Maximum-likelihood (ML) analyses were performed using both RAXML (Stamatakis 2006) and GARLI (Zwickl 2006), which utilize different search algorithms. RAXML was run using the GTR+G substitution model for both loci, as the program settings do not offer subsets of the GTR model. GTR+G was used for both final tree selection and bootstrap analysis, which was run using the rapid option and 100 replicates. GARLI was run with default population and termination settings, and the model recommended by jModeltest for each locus. Bayesian analysis was performed using MrBayes with the most general model settings offered (GTR+G+I) and flat Dirichlet priors. Markov chain Monte Carlo simulations were set to 5 million generations, sampled every 1000 generations with the first 25% of samples discarded as burn-in.

Because the spacer portion of the *nif*D-K locus was highly divergent and difficult to align, we also assessed the robustness of our phylogenetic analysis to inclusion of different portions of the alignment. We used the program Gblocks (Castresana 2000) to remove portions of the alignment deemed unreliable according to the most stringent, the least stringent and an intermediate set of parameters. The resulting alignments were then analysed using RAxML and MrBayes, and the best tree qualitatively compared with the best tree based on the entire alignment.

Results

Frankia characterization and phylogenetic host specificity patterns

PCR/RFLP yielded interpretable data for 342 of 390 nodules (88%) collected from *A. tenuifolia*. The majority of RF types observed were identical to those previously reported from *Alnus* species in these sites (Anderson *et al.* 2009), with the exception of RF11-RF16. *Nif*D-K sequence data were obtained from nodules representing all RF types except RF12 and RF14, and for multiple representatives of RF1-RF4 and RF7. All sequences derived from a single RF group were identical. Ribosomal IGS sequences were obtained from a subset of the *nif*D-K-based RF types, namely RF1-4, RF7, RF9 and RF16.

Phylogenetic analysis of the rIGS locus using both ML and Bayesian approaches placed all sequences from both host species we examined in the 'Clade 2' group of *Frankia*, which includes strains infective on *Alnus*, *Myrica* and *Casuarina* (Clawson *et al.* 2004) (Fig. 1). RF9, which is dominant in *A. viridis* but rare in *A. tenuifolia* nodules, and RF4, which infects both *A. viridis* and *A. tenuifolia* at intermediate abundance, are similar to sequences derived from other alder host species. The majority of RF groups observed in *A. tenuifolia* nodules form a well-supported clade in all analyses that is comparatively distant from other Clade 2 *Frankia* (hereafter referred to as the 'AT clade').

For the *nif*D-K locus, ML and Bayesian analyses placed RF8-10 from *A. viridis* and RF4 in distinct clades within the Clade 2 *Frankia*. Similar to the results from the rIGS locus, all three tree construction algorithms also yielded a well-supported 'AT clade' consisting of sequences representing the dominant RF groups found on *A. tenuifolia* (RFs 1–3, 5–7 and 11–16) that is relatively distant from other *Alnus*-infective strains (Fig. 2). Placement of this clade differed among analytical methods: both ML programs placed it on a relatively long branch within the Clade 2 *Frankia*, while in the Bayesian analysis, it formed a sister group to Clade 2. Placement of this clade by ML was sensitive to the characters included in the analysis: with the least stringent (348 of 753 characters) and intermediate (291 characters) settings, both RAxML and GARLI still placed the group within Clade 2, but with the most stringent settings (156 characters), both programs placed it as a sister to known Clade 2 sequences.

Variation in Frankia assemblage structure

Composition of symbiotic *Frankia* assemblages observed in the present study is similar to that previously reported for *A. tenuifolia* in these sites (Anderson *et al.* 2009), which suggests assemblage composition may be temporally stable, at least at the scale of time between the two samples (3 years). As in the earlier study, early succession sites were dominated by RF7, and late succession sites supported a much higher overall diversity (Table 1), with differences in *Frankia* structure paralleled by differences in soil chemistry between habitats (Table 2). The largest difference with the earlier study is the appearance of RF16 as a dominant component of late succession assemblages. This may be related to either the larger sampling area and/or sample size per site, or the much higher success rate of the PCR-RFLP procedure in the current study. Given the scale at which this haplotype is clumped in most sites (see

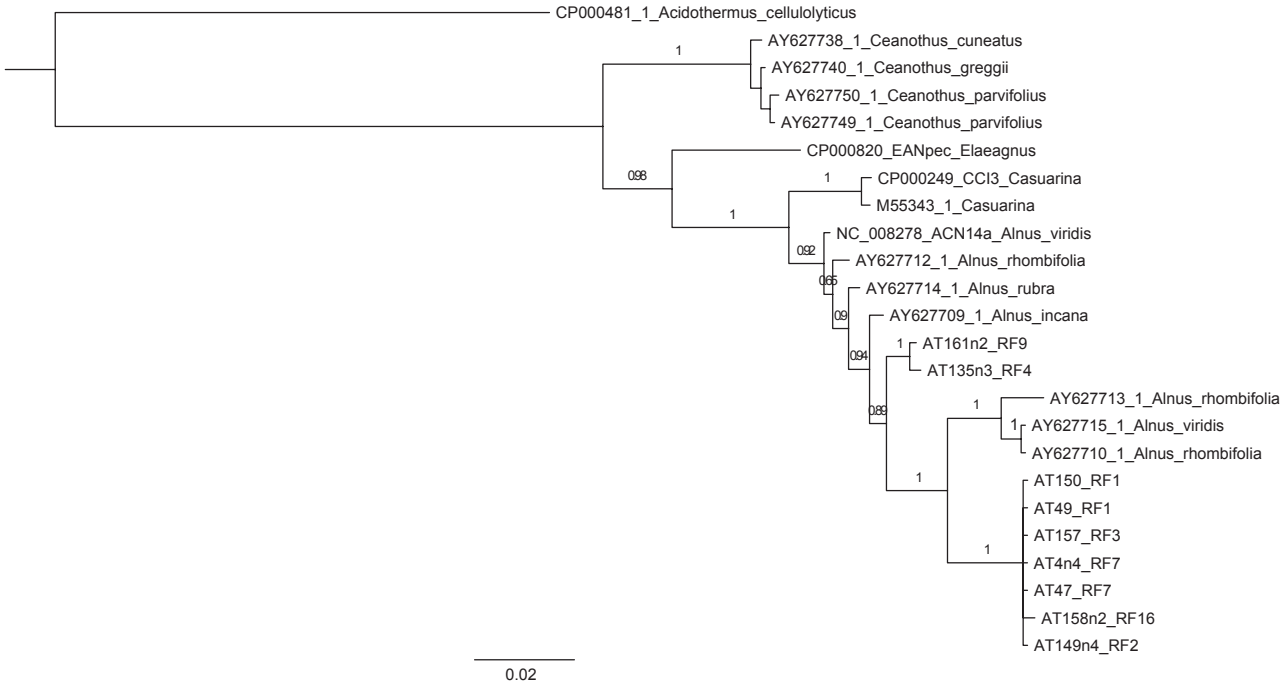

Fig. 1 *Frankia* phylogeny based on the rIGS locus. Bayesian consensus tree based on DNA sequences (1143 characters) of the ribosomal intergenic spacer (rIGS) locus derived from *Alnus tenuifolia* ('AT__') and *A. viridis* ('AC__') nodules collected in early and late succession habitats in the Bonanza Creek Experimental Forest, interior Alaska. 'RF' designations refer to restriction fragment patterns for each sample based on PCR-RFLP of the *nif*D-K locus. Host species and accession numbers are given in the tree for comparison sequences downloaded from GenBank. Branch labels are Bayesian posterior probabilities.

'spatial structure of *Frankia* assemblages' subsection), the latter explanation seems more likely.

Accumulation curves for RF richness indicated that saturation had been reached for all six sites, based on overlapping 95% confidence intervals between Mau Tau (interpolation-based) and Chao1 and Chao2 (extrapolation-based) indices. Mau Tau richness estimates were also always within one 'species' of all other estimators calculated by Estimate S, which included ACE, ICE, bootstrap and jackknife-based estimates. Thus, it appears that this study captured the available RF diversity in each site.

AMOVA indicated differences between habitats in both total molecular variation and the partitioning of variation between among-plant within-site and among-site components. In the intensive data set, a nearly equal proportion of variation (PV) occurred within plants in both early (79%) and late (77%) succession habitats, but total variation was much higher in late succession (sum of squares (SS) = 47.6, $n = 128$) than in early (SS = 22.05, $n = 129$). In early succession, the remaining ~20% was split nearly equally between among-plant within-site (PV = 10.1%; $P = 0.009$) and among-site (PV = 10.9%; $P = 0.016$) components, while in late succession, nearly all occurred among plants within sites (PV = 23%; $P < 0.0001$). The extensive data set indicated small but significant variation among late succession sites (PV=9.3%: $P = 0.02$), but no significant among-site variation in early succession.

Spatial structure of Frankia assemblages

Among early succession sites, global Mantel tests were significant for site FPE2 ($P < 0.02$), but not for the other two sites. Correlograms detected no significant local autocorrelation in any early succession site, including FPE2 (Table 1).

All late succession sites yielded significant global autocorrelation when all RF types were included in the analysis ($P = 0.002$ for FPL1, 0.006 for FPL2, and 0.03 for FPL3). For all three sites, multivariate correlograms produced the largest autocorrelation values in the first distance class (<1 m), which corresponds to within-plant distances (Table 3), although for site FPL2, this peak was not significant. For two sites, autocorrelation values dropped to insignificance in the second distance class, which corresponded closely with first nearest-neighbour distances among plants (Table 2), suggesting a patch size for all RF types that is approximately equivalent with individual plants (Legendre & Fortin 1989). In the third site, significant autocorrelation was only observed in the third distance class (13–24 m), sug-

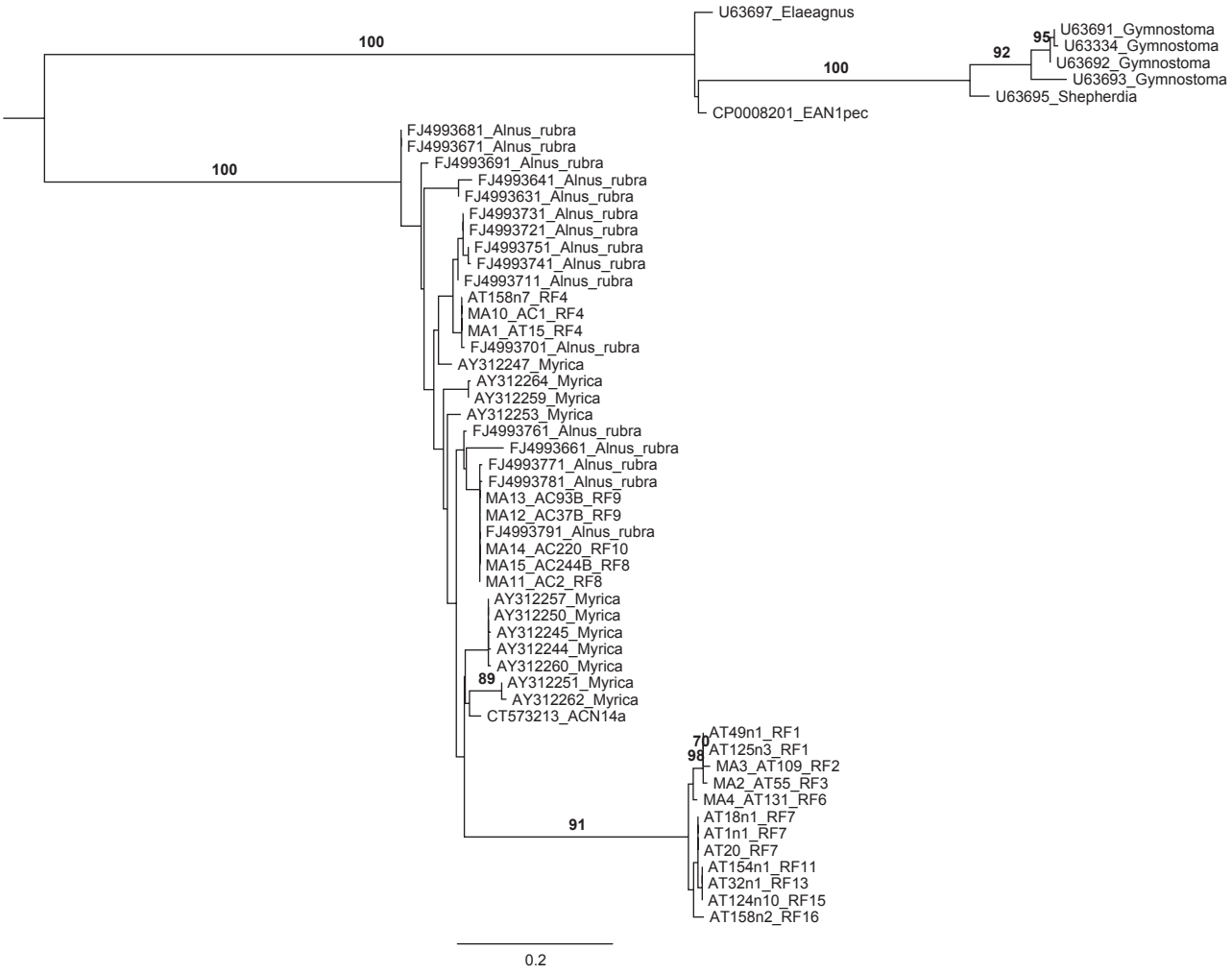

Fig. 2 *Frankia* phylogeny based on the *nif*D-K locus. Maximum-likelihood tree based on DNA sequences (753 characters) of the *nif*D-K spacer locus derived from *Alnus tenuifolia* ('AT__') and *A. viridis* ('AC__') nodules collected in early and late succession habitats in the Bonanza Creek Experimental Forest, interior Alaska. 'RF' designations refer to restriction fragment patterns for each sample based on PCR-RFLP of the *nif*D-K locus. Host species and accession numbers are given in the tree for comparison sequences downloaded from GenBank. Significant bootstrap values are given as branch labels.

Table 1 Occurrence of *Frankia* haplotypes in *A. tenuifolia* nodules. Number of nodules collected from *Alnus tenuifolia* plants that were found to contain each haplotype ('RF'= *nif*D-K RFLP pattern) of *Frankia* in replicate sites representing early and late successional habitats in the Bonanza Creek Experimental Forest, interior Alaska

Stage	Site	RF1	RF2	RF3	RF4	RF6	RF7	RF8	RF9	RF11	RF12	RF13	RF14	RF15	RF16
Early	FPE1	6	5	7	0	1	35	0	0	0	0	0	0	0	0
	FPE2	4	0	0	1	3	47	0	0	0	0	1	0	0	1
	FPE3	2	1	0	0	0	58	0	0	0	0	0	0	0	0
Late	FPL1	7	2	3	9	3	7	2	3	3	1	0	4	1	11
	FPL2	13	12	3	1	6	1	0	0	5	0	0	0	0	13
	FPL3	12	8	2	9	3	0	0	1	1	0	0	0	0	24

gesting patchiness at a scale that includes multiple plants. In the former sites, additional significant peaks occurred following the drop to insignificance beyond the smallest distance classes, suggesting the occurrence of multiple patches among all RF types (Legendre & Fortin 1989). Autocorrelation patterns for the most

Table 2 Study site characteristics. *Alnus tenuifolia* density (stems/ha) and means (±1 standard error) of soil variables and first nearest-neighbour distances between sampled *A. tenuifolia* plants in replicate sites representing early and late successional habitats in the Bonanza Creek Experimental Forest, interior Alaska. Total C, N and P values represent percentage by mass

		Total C		Total N		Total P		pH	
Stage	Site	Min	Org	Min	Org	Min	Org	Min	Org
Early	FPE1	1.4 ± 0.05	12.3 ± 0.7	0.10 ± 0.003	0.71 ± 0.04	0.08 ± 0.001	0.09 ± 0.002	8.3 ± 0.03	6.9 ± 0.06
	FPE2	1.4 ± 0.05	8.8 ± 0.3	0.09 ± 0.004	0.56 ± 0.02	0.07 ± 0.001	0.08 ± 0.001	8.2 ± 0.02	7.5 ± 0.06
	FPE3	1.2 ± 0.03	16.1 ± 0.6	0.07 ± 0.002	0.98 ± 0.03	0.08 ± 0.001	0.09 ± 0.001	8.2 ± 0.04	7.9 ± 0.02
Late	FPL1	1.5 ± 0.14	23.1 ± 2.8	0.09 ± 0.006	0.91 ± 0.10	0.07 ± 0.001	0.08 ± 0.001	7.4 ± 0.23	5.7 ± 0.05
	FPL2	5.5 ± 0.47	30.7 ± 1.1	0.23 ± 0.014	1.08 ± 0.03	0.07 ± 0.001	0.08 ± 0.003	5.7 ± 0.16	5.6 ± 0.08
	FPL3	2.1 ± 0.22	21.6±	0.11 ± 0.010	0.83 ± 0.03	0.06 ± 0.001	0.07 ± 0.002	6.1 ± 0.18	5.2 ± 0.04

C/N		N/P		C/P			
Min	Org	Min	Org	Min	Org	Plant density (stems/ha)	1st NN
14.6 ± 0.2	17.4 ± 0.3	1.3 ± 0.04	8.2 ± 0.3	18.6 ± 0.6	142 ± 6	2827	8.8 ± 3.2
16.6 ± 0.2	15.7 ± 0.1	1.2 ± 0.05	6.8 ± 0.2	20.4 ± 0.8	107 ± 4	4853	9.9 ± 2.7
18.2 ± 0.2	16.4 ± 0.2	0.8 ± 0.04	10.4 ± 0.4	14.8 ± 0.5	167 ± 7	2015	8.4 ± 3.9
17.1 ± 0.5	25.3 ± 0.3	1.2 ± 0.09	11.8 ± 1.4	20.8 ± 2.1	299 ± 37	132	12.0 ± 3.5
23.7 ± 0.8	28.4 ± 0.5	3.0 ± 0.19	13.0 ± 0.7	73.9 ± 6.2	351 ± 19	537	9.9 ± 3.5
18.1 ± 0.4	26.1 ± 0.3	1.8 ± 0.18	11.9 ± 0.5	34.2 ± 4.0	308 ± 12	56	11.3 ± 10.6

frequent RF types in each site generally paralleled the multivariate patterns (Table 3). Taken together, these results indicate clumping of RFs both within and among plants in late succession sites, and variable clump size among RFs and sites.

Correlation with soil variables

For early succession soils, the best logistic regression model indicated RF occurrence was significantly correlated with N and C concentration, pH and C/N ratio in the mineral horizon (Table 4A). The strongest correlation was with mineral C/N ratio. This relationship was particularly strong in the dominant RF type, RF7, which was positively related to C/N ratio, while the other two RF types tested only occurred at low C/N ratios.

For late succession soils, logistic regression with all RF types yielded two nearly identical best models, each with three explanatory variables (Table 4B). Both models included N and C concentration in the organic horizon, but one also included organic C/N ratio, while in the other pH in the mineral horizon was the third significant variable. Analysis of individual RF types indicated correlations with specific soil variables that differed among RF types (Table 4). The strongest correlations were with pH in RF6 (positive correlation) and RF16 (negative correlation). RF7 retained a positive relationship with C/N ratio in late succession, although this relationship was much weaker than in early succession.

Discussion

This study had two major objectives: (i) to characterize phylogenetic affinities of *Frankia* associated with *A. tenuifolia* and *A. viridis* in our interior Alaskan field sites and (ii) to provide a detailed description of small-scale habitat-based differences in symbiotic *Frankia* assemblages for the former host which included consistency of structure among replicate sites, and scale, spatial pattern and edaphic correlations within sites. The following are our most significant findings: (i) *A. tenuifolia*-dominant *Frankia* appear to belong to a single clade that is relatively distinct from other described *Frankia*, while *A. viridis* symbionts appear to be closely related to the majority of *Alnus*-infective *Frankia*, and (ii) large habitat-based differences in *A. tenuifolia* symbiont assemblage structure were observed which included differences in the scale and spatial pattern of *Frankia* assemblages, as well as correlations with specific soil variables. These differences were largely consistent among replicate sites, and overall structure was largely independent of sampling year and sampling intensity.

Our phylogenetic analyses, together with distribution of RF types on both hosts in this and a prior study (Anderson *et al.* 2009), suggest a relatively high degree of specificity between *Alnus* species and *Frankia* genotypes in our sites. *A. tenuifolia* and *A. viridis* harbour distinct symbiont assemblages, with the dominant RF types for each host occurring rarely on the other host

Table 3 Spatial structure of *Frankia* genotypes. Results of Mantel correlogram analysis of the distribution of *Frankia nif*D-K restriction fragment (RF) patterns in *Alnus tenuifolia* nodules collected in three late succession sites in the Bonanza Creek Experimental Forest in interior Alaska. Separate analyses were performed for all RF types and for the most frequent RF types observed in each site. Numbers are correlation coefficients from Mantel tests for each distance class between a geographic distance matrix and a binary distance matrix indicating shared RF types. All distance classes were designated to contain an equal number of pairwise distances except the first, which was set to 0–1 m to encompass nodules collected within the same plant. Negative coefficients indicate positive autocorrelation. Asterisks indicate Bonferroni-adjusted *P*-values from permutation tests (999 permutations) for each Mantel test as follows: $^{**}P < 0.01$, $^{*}P < 0.05$.

	Site												
	FPL1						FPL2						
Distance class	Range (m)	ALL RFs	RF1	RF4	RF7	RF16	Range (m)	ALL RFs	RF1	RF2	RF6	RF11	RF16
1	0–1	−0.25**	−0.18**	−0.17*	−0.12*		0–1						−0.17*
2	1–15						1–13						
3	15–19						13–24	−0.09**					−0.14*
4	19–30						24–31				−0.09*		
5	30–50	−0.06*				−0.12*	31–39						
6							39–52						
7							52–58						
8							58–75						
9							75–84						
10							84–94						

Site FPL3 Range (m)	ALL RFs	RF1	RF2	RF4	RF16
0–1	−0.22**		−0.17*		
1–19					
19–27					
27–46					
46–56					
56–63	−0.07*				−0.09**
63–72					
72–77					
77–91					
91–93					

even in sympatry (Anderson *et al.* 2009). In the present study, this pattern was found to contain a significant phylogenetic component for both *nif* and rIGS loci for RF types dominant on *A. tenuifolia*. Phylogenetic host specificity thus appears to be a significant constraint on the structure of *Frankia* assemblages in *Alnus* nodules in our study sites. This result is somewhat at odds with both a long history of cross-inoculation studies of *Alnus* and *Frankia*, which collectively suggest little specificity at the level of nodule formation (*reviewed in* Anderson 2011), and recent phylogenetic studies of *Alnus*-infective *Frankia*, in which host-specific clades appear to be the exception rather than the rule (*e.g.* Hahn 2008; Welsh *et al.* 2009a; Kennedy *et al.* 2010; Pokharel *et al.* 2010). It is possible that the specificity observed in the present study is exaggerated by sampling one endpoint of a genetic cline: interior Alaska is the northern extremity of the range of *A. tenuifolia*, which extends south nearly to Mexico (eFloras 2008). A more complete geographic sample will be necessary to examine this possibility.

We also observed strong patterns in distribution of RF types belonging to the *A. tenuifolia*-infective clade with respect to successional stage. Variation in *Frankia* symbiotic with *Alnus* across different environments has been reported in several studies (e.g. Khan *et al.* 2007; Welsh *et al.* 2009b; Kennedy *et al.* 2010). To our knowledge, however, this is the first field study of any N-fixing system to simultaneously characterize variation in terms of within- vs. among-plant variance components, spatial structure, consistency across multiple sites with similar environmental conditions, and with respect to specific edaphic factors. The large, consistent

Table 4 (A) Correlations between *Frankia* genotypes and soil variables. Summary of multiple logistic regressions examining the response of *Frankia* genotype (*nif*D-K restriction fragment (RF) pattern) in *Alnus tenuifolia* nodules to variation in several measured soil variables in mineral and organic horizons from early (A) and late (B) succession habitats (n = 3 sites each). Separate models were examined for each habitat and for all RF types as well as the most common individual RF types in each habitat. *P*-values for individual RF types are unadjusted, but significance after Bonferroni adjustment is indicated with an asterisk. 'Dir' indicates the direction of the observed correlation between individual RF type frequency and the indicated soil variable ('pos' = positive, 'neg' = negative). Late succession data yielded two equally good models

	All RF types					RF1		RF3		RF7	
(A) Early succession	−logL	r^2	χ^2	df	*P*	*P*	Dir	*P*	Dir	*P*	Dir
Whole model	31.5	0.52	62.9	20	<0.0001	0.014*		0.0003*		<0.0001*	
Mineral N			13.1	5	0.022	ns		0.023	pos	ns	
Mineral C			13.7	5	0.018	ns		0.022	pos	ns	
Mineral pH			11.7	5	0.038	ns		0.003*	pos	ns	
Mineral C/N			15.0	5	0.010	0.014*	neg	0.013*	neg	<0.0001*	pos

	All RF types					RF1		RF2		RF3		RF4		RF6		RF7		RF16	
(B) Late succession	−logL	r^2	χ^2	df	*P*	*P*	Dir	*P*	Dir	*P*	Dir	*P*	Dir	*P*	Dir	*P*	Dir	*P*	Dir
Model 1																			
Whole model	35.8	0.38	71.6	27	<0.0001	ns		0.043		0.012		0.0059*		0.0009*		0.05		0.001*	
Organic N			22.3	9	0.008	ns		0.043	pos	ns		ns				0.007*	neg	ns	
Organic C			18.1	9	0.034	ns		ns		0.021	pos	ns				0.008*	neg	ns	
Mineral pH			21.4	9	0.011	ns		0.026	pos	ns		ns						0.047	neg
Organic pH			ns					ns		ns		ns		0.001*	pos			0.007*	neg
Model 2																			
Whole model	35.8	0.38	71.6	27	<0.0001														
Organic N			19.9	9	0.019														
Organic C			19.9	9	0.019														
Organic C/N			21.4	9	0.011	ns		ns		0.031	neg	0.0059*	neg	0.003*	pos	0.01	pos		

differences observed between early and late successional habitats in this and an earlier study (Anderson *et al.* 2009) suggest a strong relationship between successional factors (environmental variation and/or time) and symbiont structure in *A. tenuifolia* nodules. These differences are paralleled in the present study by patterns in within- vs. among-plant structure, spatial structure and correlations with edaphic factors. Considering the geographic scale of the study and the highly outcrossing habit of *Alnus* (Bousquet & Lalonde 1990), genetic structure in hosts, similar to that reported in the *Amphicarpaea-Bradyrhizobium* system (Spoerke *et al.* 1996), is unlikely to explain these results. More likely causal mechanisms at this scale are direct structuring of soil *Frankia* assemblages by successional factors and/or nonrandom selection of symbiont genotypes by host plants. We have previously hypothesized that this difference between early and late succession is primarily due to selection of optimal *Frankia* genotypes by *A. tenuifolia* (*i.e.* 'host-choice') in early succession sites (Anderson *et al.* 2009; Anderson 2011), in which low soil N availability and high plant N demand place a high premium on fixed N. The ability of hosts to choose symbionts based on performance is predicted by numerous theoretical studies of evolutionary stability in mutualisms (*e.g.* Simms & Taylor 2002; Kiers & Denison 2008; but see Marco *et al.* 2009) and has experimental support in a number of rhizobial systems, based on apparent honest signalling (Heath & Tiffin 2009; Gubry-Rangin *et al.* 2010), 'sanctioning' of inferior mutualists (Kiers *et al.* 2003, 2006; Oono *et al.* 2011) and selective allocation of nodule biomass to superior mutualists (Simms *et al.* 2006). In our early succession sites, such choice should result in low diversity in nodules if only one or a few genotypes meet host criteria. In late succession, lower N demand due to both lower light and higher N/P ratio in soils should lower allocation to nodules overall (Wall 2000) and may also result in less stringent or nonexistent host-choice (i.e. random sampling of soil *Frankia* by alder), as has been observed in some species of *Mimosa* (Elliot *et al.* 2009). In the present study, within-site distribution of RF types and correlations with edaphic factors are consistent with both direct and plant-mediated mechanisms in early succession, but more strongly suggest direct mechanisms in late succession, consistent with the second part of our host-choice hypothesis.

Variation in *Frankia* structure in the present study was very low in early succession and was similar at all scales examined – within plants, among plants within sites and among sites. This similarity suggests that factors determining *Frankia* structure act similarly within and among plants, and among replicate sites. Factors known to affect *Frankia* assemblages in soils, such as plant community structure, vertical soil development, dispersal vectors such as water and soil chemistry (Valdés 2008), are homogeneous across the same range of scales in these sites, so this pattern is consistent with both direct and plant-mediated effects. In contrast to this homogeneity, the clearest edaphic factor correlated with *Frankia* structure in this study occurred in early succession. This was C/N ratio in the mineral horizon, high values of which favoured RF7 at the expense of RF1 and RF3. C/N ratio is a well-known indicator of substrate quality for saprotrophs such as *Frankia* and also strongly affects N availability to plants (e.g. Yamakura & Sahunalu 1990). Thus, this correlation is consistent with both direct and plant-mediated effects. However, two observations from additional studies provide support for host-choice in early succession: (i) RF7 appears to have a lower respiratory unit cost of N-fixation than the other two most common RF types in early succession (Ruess *et al. in press*), suggesting a basis for its selection, and (ii) preliminary data suggest that RF7 is less frequent in *A. tenuifolia* rhizosphere soil than in nodules, suggesting at least a degree of selectivity on the part of alder (Anderson, Ruess and Taylor, *unpublished data*). Nevertheless, a rigorous test of the host-choice hypothesis will require a controlled inoculation study with isolates of RF types observed in the field.

Frankia structure was much more heterogeneous in late succession: RF diversity (both richness and evenness) was higher in all sites, and strong patterns were observed in the scale and spatial distribution of variation. A much greater proportion of *Frankia* variation occurred among plants within a site than among sites, and distribution of RF types had a strong spatial component both within and among plants. These results suggest that factors affecting symbiont structure in these sites act both within plants (strong spatial structure) and among plants within a site (spatial structure and among-plant variance) and that these factors are largely consistent among replicate late succession sites (no among-site variance). This pattern parallels the pattern of environmental variation in these sites, in which individual plants often occupy distinct microhabitats within a site—for example, riverbanks at the forest edge, canopy gaps, patches of deciduous overstory, wet depressions from old sloughs and patches underlain with permafrost—but different sites are composed of the same microhabitat types. Thus, among-plant patterns in these sites are also consistent with both direct and plant-mediated mechanisms. However, within-site distribution patterns suggest a more important role for direct effects. Within-plant spatial structure and correlations with different edaphic factors among RF types suggest either direct structuring of *Frankia* assemblages

by patchy soil resources or resource-neutral historical factors (e.g. small-scale dispersal, accumulation of taxa over successional time (Valdés 2008), colonization history and competitive exclusion (Bruns 1995)) that spatially covary with resources. It is conceivable that symbiont structure is still largely mediated by plant selection, that is, that plants in specific microhabitats select symbionts optimal for each microhabitat (among-plant spatial structure) from among spatially structured soil *Frankia* (within-plant spatial structure). However, in addition to the nonparsimonious nature of this scenario, which posits both spatial structure in soil *Frankia* and selection by plants, this seems unlikely for several reasons: (1) high within-plant diversity of RF types is counter to the idea of one or a few optimal genotypes for a given microhabitat, (2) wide variability of RF structure among closely occurring plants suggests a lack of optimal symbiont(s) in a given microhabitat, and (3) high soil N and low alder leaf N compared with early succession sites (Anderson *et al*. 2009) suggest that late succession plants rely less on fixed N, which is also supported by the lower abundance and smaller size of nodules on late succession alder roots (Anderson & Ruess, personal observations). Point (3) would undermine both the need for and the effect of any plant selection in these sites.

While the results of this study could not differentiate between host-related indirect and soil-related direct effects in early succession, our data suggest direct effects may be most important in late succession, which is consistent with the second part of our host-choice hypothesis. Further work is needed to investigate (i) whether *A. tenuifolia* is capable of selecting optimal symbionts under controlled laboratory conditions, and whether selection varies with environmental conditions, particularly those suggested to be important in field studies, and (ii) whether nodule assemblages in early and late succession sites are random or nonrandom samples of soil assemblages with respect to *Frankia* genetic structure.

References

Anderson MD (2011) Two in the far north: the alder-*Frankia* symbiosis, with an Alaskan case study. In: *Ecological Aspects of Nitrogen Metabolism in Plants* (eds Polacco JC, Todd CD), pp. 138–166. Wiley & Sons, West Sussex, UK.

Anderson MD, Ruess RW, Myrold DD, Taylor DL (2009) Host species and habitat affect nodulation by specific *Frankia* genotypes in two species of *Alnus* in interior Alaska. *Oecologia*, **160**, 619–630.

Benson DR, Dawson JO (2007) Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. *Physiologia Plantarum*, **130**, 318–330.

Bousquet J, Lalonde M (1990) The genetics of actinorhizal Betulaceae. In: *The Biology of Frankia and Actinorhizal Plants* (eds Schwintzer CR, Tjepkema JD), pp. 239–261. Academic Press, Inc., San Diego, CA.

Bruns TD (1995) Thoughts on the processes that maintain local species diversity of ectomycorrhizal fungi. *Plant and Soil*, **170**, 63–73.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Chao A (1984) Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265–270.

Chapin FS III, Viereck LA, Adams PC, *et al*. (2006) Successional processes in the Alaskan boreal forest. In: *Alaska's Changing Boreal Forest* (eds Chapin FS III, Oswood MW, Van Cleve K, Viereck LA & Verbyla DL). Oxford University Press, New York.

Clawson ML, Bourret A, Benson DR (2004) Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences. *Molecular Phylogenetics and Evolution*, **31**, 131–138.

Colwell RK (2009) EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2. http://viceroy.eeb.uconn.edu/estimates. [accessed 21 April 2013].

Colwell RK, Mao CX, Chang J (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, **85**, 2717–2727.

Dawson JO (2008) Ecology of actinorhizal plants. In: *Nitrogen Fixation: Origin, Applications, and Research Progress, Vol. 6 Nitrogen-Fixing Actinorhizal Symbioses* (eds Pawlowski K, Newton WE), pp. 199–234. Springer, Dordrecht, The Netherlands.

Elliot GN, Chou J-H, Chen W-M, *et al*. (2009) *Burkholderia* spp. are the most competitive symbionts of *Mimosa*, particularly under N-limited conditions. *Environmental Microbiology*, **11**, 762–778.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

eFloras (2008) Published on the Internet http://www.efloras.org [accessed 23 July 2012] Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA.

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. *Plant Physiology*, **131**, 872–877.

Gubry-Rangin C, Garcia M, Béna G (2010) Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proceedings of the Royal Society B*, **277**, 1947–1951.

Hahn D (2008) Polyphasic taxonomy of the genus *Frankia*. In: *Nitrogen Fixation: Origins, Applications, and Research Progress, Vol. 6 Nitrogen-fixing Actinorhizal Symbioses* (eds Pawlowski K, Newton WE), pp. 25–47. Springer, Dordrecht, The Netherlands.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Heath KD (2010) Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution*, **64**, 1446–1458.
Heath KD, Lau JA (2011) Herbivores alter the fitness benefits of a plant-rhizobium mutualism. *Acta Oecologica*, **37**, 87–92.
Heath KD, Tiffin P (2007) Context dependence in the coevolution of plant and rhizobial mutualists. *Proceedings of the Royal Society B*, **274**, 1905–1912.
Heath KD, Tiffin P (2009) Stabilizing mechanisms in a legume-rhizobium mutualism. *Evolution*, **63**, 652–662.
Heath KD, Stock AJ, Stinchcombe JR (2010) Mutualism variation in the nodulation response to nitrate. *Journal of Evolutionary Biology*, **23**, 2494–2500.
Hollingsworth TN, Lloyd AH, Nossov D, *et al.* (2010) Twenty-five years of vegetation change along a putative successional chronosequence on the Tanana River, Alaska. *Canadian Journal of Forest Research*, **40**, 1273–1287.
Jackson ML (1958) *Soil Chemical Analysis*. Prentice-Hall, Englewood Cliffs, NJ.
Kennedy PG, Weber MG, Bluhm AA (2010) *Frankia* bacteria in *Alnus rubra* forests: genetic diversity and determinants of assemblage structure. *Plant and Soil*, **335**, 479–492.
Khan A, Myrold DD, Misra AK (2007) Distribution of *Frankia* genotypes occupying *Alnus nepalensis* nodules with respect to altitude and soil characteristics in the Sikkim Himalayas. *Physiologia Plantarum*, **130**, 364–371.
Kiers ET, Denison RF (2008) Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 215–236.
Kiers ET, Rousseau RA, West SA, Denison RF (2003) Host sanctions and the legume-rhizobium mutualism. *Nature*, **425**, 78–81.
Kiers ET, Rousseau RA, Denison RF (2006) Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evolutionary Ecology Research*, **8**, 1077–1086.
Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences USA*, **78**, 454–458.
Legendre P, Fortin M-J (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.
Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
Marco DE, Carbajal JP, Cannas S, *et al.* (2009) An experimental and modelling exploration of the host-sanctions hypothesis in legume-rhizobia mutualism. *Journal of Theoretical Biology*, **259**, 423–433.
Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium species? *Trends in Ecology and Evolution*, **17**, 458–466.
Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 Nov. 2010, New Orleans, LA, pp. 1–8.
Mullis K, Faloona F, Scharf S, *et al.* (1986) Specific enzyme amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium in Quantitative Biology*, **51**, 263–273.
Navarro E, Jaffre T, Gauthier D, *et al.* (1999) Distribution of *Gymnostoma* spp. microsymbiotic *Frankia* strains in New Caledonia is related to soil type and to host-plant species. *Molecular Ecology*, **8**, 1781–1788.
Nossov DR, Hollingsworth TN, Ruess RW, Kielland K (2011) Development of *Alnus tenuifolia* stands on an Alaskan floodplain: patterns of recruitment, disease and succession. *Journal of Ecology*, **99**, 621–633.
Oono R, Anderson CG, Denison RF (2011) Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings of the Royal Society B*, **278**, 2698–2703.
Parker MA (1995) Plant fitness variation caused by different mutualist genotypes. *Ecology*, **76**, 1525–1535.
Parker MA (2012) Legumes select symbiosis island sequence variants in *Bradyrhizobium*. *Molecular Ecology*, **21**, 1769–1778.
Parker MA, Malek W, Parker IM (2006) Growth of an invasive legume is symbiont limited in newly occupied habitats. *Diversity and Distributions*, **12**, 563–571.
Pokharel A, Mirza BS, Dawson JO, Hahn D (2010) *Frankia* populations in soil and root nodules of sympatrically grown *Alnus* taxa. *Microbial Ecology*, **61**, 92–100.
Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
Rosenberg MS, Anderson CD (2011) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution*, **2**, 229–232.
Ruess RW, Anderson MD, McFarland JM, Kielland K, Olson K, Taylor DL (in press) Ecosystem-level consequences of symbiont partnerships in a N_2-fixing shrub from interior Alaskan floodplains. *Ecological Monographs*.
Simms EL, Taylor DL (2002) Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integrative and Comparative Biology*, **42**, 369–380.
Simms EL, Taylor DL, Povich J, *et al.* (2006) An empirical test of partner-choice mechanisms in a wild legume-rhizobium interaction. *Proceedings of the Royal Society B*, **273**, 77–81.
Simonet P, Navarro E, Rouvier C, *et al.* (1999) Coevolution between *Frankia* population and host plants in the family Casuarinaceae and consequent patterns of global dispersal. *Environmental Microbiology*, **1**, 525–533.
Spoerke JM, Wilkinson HH, Parker MA (1996) Nonrandom genotypic associations in a legume-*Bradyrhizobium* mutualism. *Evolution*, **50**, 146–154.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
Swensen SM, Benson DR (2008) Evolution of actinorhizal host plants and *Frankia* endosymbionts. In: *Nitrogen Fixation: Origin, Applications, and Research Progress, Vol. 6 Nitrogen-Fixing Actinorhizal Symbioses* (eds Pawlowski K, Newton WE), pp. 73–104. Springer, Dordrecht, The Netherlands.
Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: *Some Mathematical Questions in Biology – DNA Sequence Analysis* (ed. Miura M), pp. 57–86. American Mathematical Society, Providence, RI.
Thompson JN (2005) *The Geographic Mosaic of Coevolution*. University of Chicago Press, Chicago Il.
Thompson JD, Gibson TJ, Plewniak F, *et al.* (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.
Thrall PH, Slattery JF, Broadhurst LM, Bickford S (2007) Geographic patterns of symbiont abundance and adaptation

in native Australian *Acacia*-rhizobia interactions. *Journal of Ecology*, **95**, 1110–1122.

Valdés M (2008) *Frankia* ecology. In: *Nitrogen Fixation: Origin, Applications, and Research Progress, Vol. 6 Nitrogen-Fixing Actinorhizal Symbioses* (eds Pawlowski K, Newton WE), pp. 49–71. Springer, Dordrecht, The Netherlands.

Vessey JK, Pawlowski K, Bergman B (2005) Root-based N_2-fixing symbioses: legumes, actinorhizal plants, *Parasponia* sp. and cycads. *Plant and Soil*, **274**, 51–78.

Vitousek PM, Cassman K, Cleveland C, *et al.* (2002) Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry*, **58**, 1–45.

Wall LG (2000) The actinorhizal symbiosis. *Journal of Plant Growth Regulation*, **19**, 167–182.

Wang H, Man CX, Wang ET, Chen WX (2009) Diversity of rhizobia and interactions among the host legumes and rhizobial genotypes in an agricultural-forestry ecosystem. *Plant and Soil*, **314**, 169–182.

Welsh A, Mirza BS, Reider JP, *et al.* (2009a) Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. *Systematic and Applied Microbiology*, **32**, 201–210.

Welsh AK, Dawson JO, Gottfried GJ, Hahn D (2009b) Diversity of *Frankia* populations in root nodules of geographically isolated Arizona alder trees in central Arizona (United States). *Applied and Environmental Microbiology*, **75**, 6913–6918.

Wheeler CT, Miller IM (1990) Current and potential uses of actinorhizal plants in Europe. In: *The Biology of Frankia and Actinorhizal Plants* (eds Schwintzer CR & Tjepkema JD). Academic Press, San Diego, CA.

Yamakura T, Sahunalu P (1990) Soil carbon/nitrogen ratio as a site quality index for some southeast Asian forests. *Journal of Tropical Ecology*, **6**, 371–377.

Young JPW, Johnston AWB (1989) The evolution of specificity in the legume-rhizobium symbiosis. *Trends in Ecology and Evolution*, **4**, 341–349.

Zwickl DJ (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD Dissertation. University of Texas, Austin, TX.

M.D.A. contributed to experimental design, performed sample collection in the field, performed PCR, RFLP and DNA sequencing, performed phylogenetic and statistical analyses, and drafted and revised the manuscript. D.L.T. contributed to experimental design, developed new primers and optimized PCR conditions, assisted with generation of DNA sequence data, assisted with phylogenetic analysis, and provided edits and revisions to the manuscript. R.W.R. contributed to experimental design, assisted with field collection, generated soil chemistry data, assisted with statistical analysis, and provided edits and revisions to the manuscript.

Data accessibility

DNA sequence data are available from GenBank: accession numbers KC822369–KC822377 for rIGS sequences, and KC814788–KC814807 for *nif*D-K spacer sequences.

Raw data files available on the Bonanza Creek Long-Term Ecological Research website: URL: http://www.lter.uaf.edu/.

Datafile: 532_MDA_05_Frankia_RF_and_soildat.txt.