Diversity of *Frankia* in root nodules of six *Morella sp.* from the Cape flora of South Africa

Dale A. Wilcox[*], Don A. Cowan[#]

[*]*Institute for Microbial Biotechnology and Metagenomics, Department of Biotechnology, University of the Western Cape, Cape Town 7535, South Africa*

[#]*Centre for Microbial Ecology and Genomics, University of Pretoria, Pretoria 0002, South Africa*

Corresponding Author:

Don A. Cowan,

e-mail: don.cowan@up.ac.za

tel: +27 (0)12 420 5873

Abstract

Africa hosts numerous endemic actinorhizal plants from the genus *Morella*, but the diversity of their *Frankia* endosymbionts has never been explored. The diversity of *Frankia* in root nodules collected from natural stands of six *Morella* species within the Cape flora of South Africa was investigated. The diversity of Frankia in root nodules collected from natural stands of six Morella species within the Cape flora of South Africa was investigated by comparative nitrogenase (nifH) gene sequence analysis. Gene sequences assigned nodular strains to both cluster I (*Alnus* host-infection group) and cluster III (*Elaeagnus* host-infection group), with sequences from both groups recovered from three hosts: *M. intergra*, *M. diversifolia* and *M. quercifolia*. Cluster I sequences were found in nodules from acidic soils exclusively. *Frankia* strains representing both groups were isolated and characterized by 16S rRNA and *nif*H analysis. This study is the first to investigate the diversity of *Frankia* associated with endemic African actinorhizal species in their natural habitats, and to report isolation of *Frankia* strains from African *Morella*.

Morella, Frankia, actinorhizal symbiosis, Cape flora, nitrogenase, nifH

Introduction

Frankia are geographically widespread gram-positive soil actinomycetes capable of entering into nitrogen fixing symbioses with more than 200 specific host plants from several genera distributed across eight families, in what are known as "actinorhizal" symbioses (Benson et al. 2004; Normand et al. 2014). Host plants, which occur on all continents except Antarctica, benefit from this association by gaining the ability to colonize marginal soils and are often pioneer species at newly formed or exposed sites (Benson et al 2004; Benson and Dawson 2007).

Because of their unusual ability to effectively nodulate with strains from both clusters I and III, alternatively referred to as the *Alnus-*, and *Elaeagnus-*Host Infection Groups (HIGs), when inoculated under laboratory conditions, the genus *Morella* (a member of the Myricaceae, along with actinorhizal *Myrica* and *Comptonia*) is considered promiscuous (Vandenbosch and Torrey 1983; Torrey 1990; Maggia and Bousquet 1994; Huguet et al. 2005). Although the symbiotic preferences of comparatively few *Morella* species have been studied under natural field conditions, host/endophyte associations have typically been found to be more specific than those found when hosts are greenhouse-grown, and it is thought that this may be a consequence of prevalent edaphic conditions. Whether this is due to the availability of infectious strains or the effect of soil conditions on either strain infectivity or host preference is unknown (Zimpfer et al. 1999; Clawson et al. 1999; Clawson and Benson 1999; Huguet et al. 2001; He et al. 2004; Huguet et al. 2005).

The African continent hosts an abundance of actinorhizal species, principally *Morella* (previously *Myrica*), of which there are 30 species (approximately half of the global total (Gtari et al. 2004; Huguet et al. 2005; The Plant List, Version 1.1 2014). Furthermore, *Morella* is the only actinorhizal genus indigenous to southern Africa. Despite this, within Africa endemic actinorhizal plants and their microsymbionts have been largely neglected, especially in the south of the continent (Gtari and Dawson 2011). The Cape Region of South Africa, an area of only 90000 square kilometers but with remarkable floral richness (Goldblatt 1978), is home to seven *Morella* species of which three are endemic (Killick et al. 1998; Goldblatt and Manning 2000; Gtari and Dawson 2011).

Morella from the Cape region vary widely in both their morphology and ecology, and most are easily identifiable in the field (Killick 1969; Killick et al. 1998; Herbert 2005). Species occupy a number of varied habitats (Goldblatt and Manning 2000); *Morella cordifolia* and *M. quercifolia* are lowland species, the former stabilizing coastal dunes in a role similar to that of the North American *M. cerifera* while the latter inhabits coastal flats, marshes and wetlands. Perennial stream and river banks may be colonized by *M. integra* or *M. serrata*, both of which are small trees. *M. kraussiana*

2

occurs on the mountains of the Western Cape Province, where it has access to moisture from perennial cloud cover during the region's drought prone summers. *M. diversifolia*, a possible hybrid of *M. kraussiana* and *M. quercifolia* (Bond 1971), occupies the same niche as *M. kraussiana*, but is restricted to the mountains of the Cape peninsula (Helme and Trinder-Smith 2006). Despite uncertainty over the classification of the Myricaceae elsewhere in the world (Herbert 2005), all of the Cape *Morella* are accepted species (The Plant List, Version 1.1 2015).

Early investigations found all southern African *Morella* to be nodule-bearing, demonstrated that the nodules were the site of nitrogen fixation and considerably expanded the number of plants from the Myricaceae known to be actinorhizal (Van Ryssen and Grobbelaar 1970; Bond 1971; Bond 1976). Beyond these investigations the subject of endophyte diversity in southern African actinorhizal hosts was not pursued, partly because *Frankia* strains had not yet been isolated (Callaham et al. 1978).

In this study we investigated the diversity of nodular *Frankia* associated with six African *Morella* species growing in their natural environments using partial *nif*H sequences amplified directly from root nodules. Additionally, endemic southern African *Frankia* strains were isolated from four hosts and classified using *nif*H and 16S rDNA *Frankia* sequences from nucleotide databases.

Methods

Nodule collection and processing

Root nodules were collected from six southern African M*orella* species (*M. cordifolia*, *M. diversifolia*, *M. quercifolia*, *M. integra*, *M. serrata* and *M. kraussiana*) from naturally occurring stands throughout the Western Cape province of South Africa, at sites indicated in Table 1 and Figure 1, between April of 2012 and December of 2013[1]. Of these species, *M. diversifolia*, *M. integra* and *M. kraussiana* are endemic to the Cape flora (Goldblatt and Manning 2000). Nodules were stored on ice or in habitat soil during transport to the laboratory, whereupon they were processed immediately. Individual nodules were cleaned of adherent soil particles by washing under a stream of water. Subsequent to this a single lobe was removed from each nodule, washed in sterile distilled water, carefully cleaned under a dissecting microscope and washed twice more. Lobes were then surface-sterilized in 30% H_2O_2

---

[1] Species occurrence data may be conveniently viewed using the South African National Biodiversity Institute Integrated Biodiversity Information System (SIBIS)

Table 1: Details of *Morella* species occurrence, with samples collection within 100 m of indicated coordinates for each site (Figure 1). The identity of the species found at site B2 is uncertain and may have been either *M. kraussiana* or *M. diversifolia*. For the purposes of this study we considered them the latter.

Species	Site	Latitude	Longitude	pH (CaCl$_2$)[1]	Nod[2]	HIG[3]
M. cordifolia	A 1.1	33° 11' 33.40" S	18° 4' 19.00" E	8.31 (0.07)	2	III
	A 1.2	33° 11' 51.70" S	18° 4' 41.80" E	8.43 (0.13)	4	III
	A 1.3	33° 8' 28.40" S	18° 6' 12.90" E	ND	1	III
	A 1.4	33° 14' 41.00" S	18° 11' 32.40" E	ND	2	III
	A 1.5	33° 13' 40.00" S	18° 9' 19.30" E	7.70 (0.24)	4	III
	A 2	33° 45' 14.49" S	18° 26' 31.70" E	ND	1	III
	A 4	34° 23' 20.00" S	21° 25' 29.14" E	8.16 (0.02)	8	III
	B 4	34° 4' 54.50" S	18° 27' 59.00" E	8.16 (0.02)	36	III
	B 7	34° 21' 27.70" S	18° 55' 28.00" E	7.06 (0.17)	10	III
M. quercifolia	A 1.5	33° 13' 40.00 "S	18° 9' 19.30" E	7.70 (0.24)	15	III
	B 1.1	34° 13' 59.90" S	18° 22' 51.10" E	ND	2	III
	B 3	34° 5' 31.30" S	18° 25' 30.80" E	ND	1	I
	B 5	33° 56' 11.00" S	18° 37' 7.90" E	7.45 (0.07)	21	III
M. diversifolia	B 2	33° 58' 16.90" S	18° 25' 15.20" E	3.45 (0.01)	18	III
	B 1.2	34° 13' 13.60" S	18° 22' 49.70" E	4.48 (0.03)	13	I/III
M. kraussiana	B 1.3	34° 13' 10.70" S	18° 22' 46.00" E	4.46 (0.02)	2	I
	B 3.1	34° 5' 33.80" S	18° 25' 35.60" E	ND	1	I
	B 3.2	34° 5' 34.60" S	18° 25' 59.60" E	4.48 (0.03)	3	I
M. serrata	B 6	33° 58' 23.20" S	18° 56' 13.40" E	5.21 (0.08)	6	I
M. integra	A 3	32° 6' 55.00" S	19° 3' 54.30" E	4.17 (0.06)	18	I/III
	A 3.1	32° 6' 6.80" S	19° 4' 3.10" E	3.83 (0.07)	3	III
	A 3.2	32° 6' 10.60" S	19° 3' 52.80" E	4.04 (0.02)	8	I
	A 3.3	32° 21' 41.25" S	19° 4' 19.52" E	ND	23	I
Total					202	

[1] Soil pH, mean of readings from three samples with standard deviation in parentheses.

[2] Number of nodules collected per locality for which *nif*H amplification was successful.

[3] *Frankia* host infection groups found in nodules at each site.
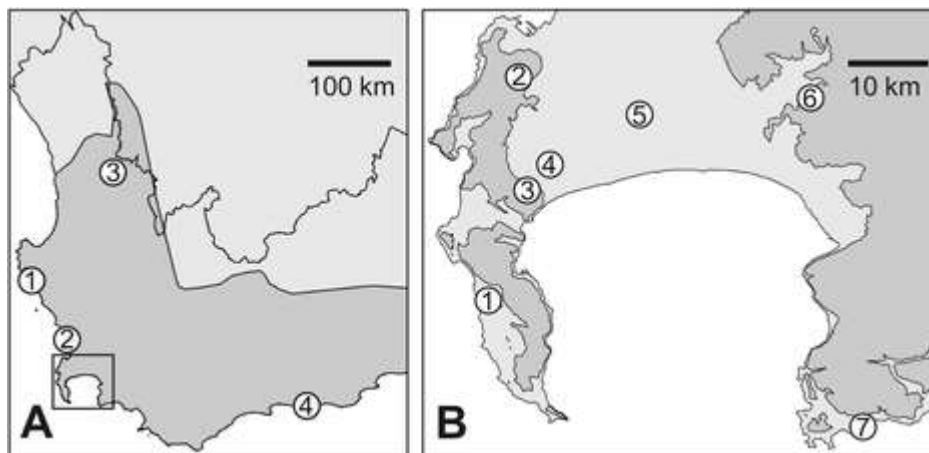
Figure 1: Map of *Morella* sampling locations within the Western Cape Province of South Africa, with the Cape Floristic Region indicated in dark grey (panel A). Site A1 is West Coast National Park, site A2 is the Blouberg Provincial Nature Reserve, site A3 is the Cedeberg wilderness area, site A4 is the Stilbaai Nature Reserve. The Cape peninsula and greater False Bay area are indicated by the boxed inset (panel B). Areas at altitudes above 100m are indicated in dark grey and correspond to the Table Mountain range on the Cape Peninsula in the West, and the Boland mountain range in the East.

Sites B1-B3 are within Table Mountain National Park and correspond to sampling sites on the Bonteberg, Table Mountain and the Steenberg respectively. Sites B4-B6 are the Zandvlei, Cape Flats and Jonkershoek nature reserves. Site B7 is a beach near the Kogelberg Nature Reserve. GPS coordinates and *Morella* root nodules collected at each site are detailed in Table 1.

for 30 min with periodic manual agitation, after which they were rinsed copiously with sterile distilled water.

For *Frankia* isolation two approaches were used: firstly, sterilized lobes were sectioned and incubated in 6 ml of DPM or FDM medium without a fixed nitrogen source and supplemented with cyclohexamide (50 μg mL^{-1}), nalidixic acid (10 μg mL^{-1}) and sodium azide (5 μg mL^{-1}) (Meesters et al. 1985; Lechevalier and Lechevalier 1990). For *Morella* species in which this first technique failed to produce isolates a second method was attempted. Lobe sections were incubated on similarly-formulated gellan gum-based solid media, both with and without ammonium chloride (Bass and Benson 2007). Plates and tubes were incubated in the dark at 28°C until colony outgrowth from nodule surfaces was observed. Filamentous outgrowths were excised under a dissecting microscope, homogenized in 1 mL of sterile dH2O and the suspension plated out on gellan gum-based *Frankia* medium containing pyruvate (1.2 g L^{-1}) and peptone (1.2 g L^{-1}). Plates were sealed with parafilm and incubated in the dark at 28°C for three weeks. Single colonies were then excised, processed as described above, and re-plated. This exercise was repeated a minimum of four times for each isolate.

Nucleic acids were extracted from nodule lobes, which were prepared in the same manner as those used for strain isolation, and *Frankia* isolates using a Mo-Bio PowerPlant Pro kit (MoBio, Carlsbad, CA) according to the manufacturer's instructions.

Soil pH determination

Soil pH was determined as previously described (Schofield and Wormald Taylor 1955). Measurements were taken in triplicate on individually prepared 1:5 suspensions of soil in 0.01 *N* CaCl2 and distilled water.

PCR conditions and molecular identification

Sequencing of *nif*H gene fragments from nodule extracts and *Frankia* isolates were performed in both directions using primers and nifHr (Welsh et al. 2009a). Each 25 μL reaction contained 2 μL dNTPs (2.5 mM each), 2.5 μL 10× PCR buffer, 1 μL of each primer (0.1 μM) 0.65 μL DMSO, 1.25 μL BSA (10 μg mL^{-1}), 0.1 μL ExTaq (5*U* μL^{-1}; TaKaRa) and 1 μL of template. PCR conditions were 95°C for 5 min; 35 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 45 s; and 72°C for 5 min.

Partial 16S rRNA genes from *Frankia* isolates were amplified using primers FGPS-1509'-153 and FGPS-281 bis (Normand et al. 1996). Each 50 μL reaction contained

4 μL dNTPs (2.5 mM each), 5 μL 10× PCR buffer, 2 μL of each primer (0.1 μM) 1.3 μL DMSO, 3 μL BSA (10 μg mL^{-1}), 0.2 μL ExTaq (5U μL^{-1}; TaKaRa) and 1 μL of template. PCR conditions were 95°C for 5 min; 35 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 2 min; and 72°C for 5 min. All reactions were carried out in an automated thermal cycler (T100, Bio-Rad) and amplification products visualized on 1% agarose gels. Amplified *nif*H and 16S rRNA gene fragments were purified with ExoSAP-IT PCR cleanup reagent (USB Corp., Cleveland, OH, USA) by combining 5 μL PCR product with 0.25 μL of reagent in a final reaction volume of 7 μL and incubating at 37°C for 45 min, followed by 80°C for 15 min. Sequencing was performed at the Stellenbosch University Central Analytical Facility using the primers used in their generation, and with the addition of internal sequencing primers F2 (5^t-ACT CCT ACG GGA GGC AGC AG) and R2 (5^t-GGA CTA CCI GGG TAT CTA ATC C) in the case of the 16S rRNA gene. Chromatograms were visually assessed, corrected by hand where necessary, and assembled in CLC Main Workbench (version 6.2.1). Representatives of each unique sequence were deposited in GenBank under accession numbers KP342075-KP342100 (nodule-derived *nif*H), KP342111-KP342120 (isolate-derived *nif*H) and KP342101-KP342110 (isolate 16S rRNA).

Phylogenetic analysis

Non-identical 606 bp partial *nif*H sequences (hereafter referred to as C̲ape F̲rankia N̲itrogenase, or CFN) were identified by nucleotide BLAST analysis and grouped into operational taxonomic units (OTUs) using the average-neighbor algorithm in MOTHUR (version 1.32.0). Additionally, sequences were grouped at ≥99%, ≥98%, ≥96%, ≥94%, ≥93% and ≥92% similarity levels. Non-identical *nif*H amplification fragments from Cape *Morella* nodules and isolates were aligned with *Frankia nif*H sequences downloaded from public databases, using CLC Main Workbench (version 6.2.1). The final data set comprised 26 sequences from Cape *Morella* nodules and 250 database sequences (75 from pure cultures and 175 obtained from nodules) representing previously defined clusters KL1-KL5 (Higgins and Kennedy 2012), KL6-KL8 (Kennedy et al. 2010) and clusters EI-EIV and AI-AV (Welsh et al. 2009a). The alignment was trimmed to 512 bp, corresponding to positions 227 to 738 of the *Frankia* ACN14a *nif*H gene (NC008278), sequences grouped into OTUs with MOTHUR (version 1.32.0) as described above, and identical sequences removed. Following this, the data set was further reduced by retaining only the most divergent sequences and sequences most similar to each CFN within each of the clusters recovered at 97% similarity.

The resultant 94 sequence data set was analyzed using neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods as previously described (Welsh et al 2009a). Neighbor joining was completed in PAUP

(version 4.0b) (Swofford 2003). A GTR+I+G model of sequence evolution and set values for the proportion of invariant sites (0.5250) and gamma shape parameter (0.4060) were estimated in jModeltest (version 2.1.4) (Guindon and Gascuel 2003; Darriba et al. 2012). Settings for jModeltest2 included 11 substitution schemes, 88 candidate models, rate variation I+G, nCat=4, an ML optimized base tree for likelihood calculation, and tree topology best of NNI and SPR. ML analysis was completed using the RAxML-HPC2 program on the CIPRES computer cluster (www.phylo.org) (Stamatakis 2006; Miller et al. 2010). Settings included GTR+CAT rate heterogeneity approximation, a proportion of invariant sites, empirical base frequencies and the number of bootstrap replicates required estimated during the run. MP analysis was completed with PAUP with 10000 random addition replicates, TBR and the multrees option set to "no" (Swofford 2003). Bootstrapping included 10000 replicates and a full heuristic search. Bayesian analysis was carried out using MRBAYES (version 3.2.2) on the CIPRES computer cluster (www.phylo.org) and included MCMC sampling, a GTR+I+G model estimated during the run and 5 million generations with sampling every 1000 trees. A 50% consensus tree was created with the first 25% of trees removed as burn-in. Support measures from each method were mapped onto a NJ tree using Dendropy (Sukumaran and Holder 2010) and displayed in Figtree version 1.4 (Rambaut 2009).

16S rRNA amplification products from Cape *Frankia* isolates were aligned with those of 34 strains retrieved from public databases and trimmed to 1380 bp. Sequences were checked for the presence of chimeras using DECIFER (Wright et al. 2012). BLAST analysis was performed to confirm the identity of the isolates, and a NJ tree constructed to determine their positions within previously defined *Frankia* clusters (Normand et al. 1996). Bootstrap values are expressed as percentages and were determined from 1000 replicates (Felsenstein 1985).

Results

Using the *Frankia* nitrogenase-specific primer pair nifHf1/nifHr, PCR products of the expected size (641 bp) were obtained from 202 of 210 (96%) *Morella* nodule lobe extracts, collected from natural stands of all six species of *Morella* examined (Table 2). Nodules in which *nif*H sequences could not be detected were collected from *M. kraussiana* on the northern plateau of Table Mountain exclusively (not indicated in Figure 1). These nodules were subsequently tested for the presence of cluster II *Frankia nif*H sequences, which were also found to be absent (results not shown). Following this, nodules from this site were not further investigated.

Sequence analysis revealed the presence of twenty-six non-identical 606 bp *nif*H sequences (CFN 1 to 26). Average similarity between non-identical Cape frankiae *nif*H sequences CFN1-CFN26 was 95.25% (SD 0.021) and cluster analysis in

Table 2: Occurrence of 606 bp *nif*H sequences in nodules of Cape *Morella* species. Sequences are listed by Cape Cluster genotype, defined at >97% similarity, then by prevalence. In-silico translation products are indicated, as are *Frankia* isolates corresponding to nodule derived *nif*H sequences. CC-1 and CC-2 correspond to cluster I *Frankia* (*Alnus*-HIG). CC-3 to CC-7 correspond to cluster III *Frankia* (*Elaeagnus*-HIG).

| Genotype	Unique *nif*H	*M. cordifolia*	*M. diversifolia*	*M. quercifolia*	*M. integra*	*M. kraussiana*	*M. serrata*	Total	Translations[1]	Isolates
	CFN1				42		6	48	1	FMi1, FMi2
	CFN3		7	1	2	6		16	1	FMk1, FMq1
CC-1	CFN20				1			1	1	
	CFN22				1			1	1	
	CFN23				1			1	1	
CC-2	CFN10		6					6	2	
	CFN16		2					2	2	
	CFN5	12		1				13	3a	FMc1, FMc2, FMc3
	CFN6	11	2	6				19	3b	
	CFN7	6		2	4			12	3b	
	CFN8		9					9	3b	
CC-3	CFN9	8						8	3b	FMc4
	CFN11			4				4	3b	
	CFN13	4						4	3b	
	CFN14			2				2	3b	FMc5
	CFN17	2						2	3b	
	CFN26	1						1	3b	
	CFN2	20						20	4	FMc6, FMc7
	CFN4			20				20	4	
CC-4	CFN18	2						2	4	
	CFN19			1				1	4	
	CFN21					1		1	4	
	CFN25	1						1	4	
CC-5	CFN12		4					4	5	
CC-6	CFN15		1	2				3	6	
CC-7	CFN24	1						1	7	
Total		68	31	39	52	6	6	202		

[1] In-silico translation of 606-bp fragments numbered to indicate identical translation products.

MOTHUR assigned all CFNs to one cluster at ≥92% identity. At 97% sequence similarity, CFNs grouped into a total of seven *Frankia* genotypes, which are hereafter referred to as Cape Clades (CC) 1 to 7. In silico translations of the twenty-six non-identical 606 bp CFNs revealed identical peptide sequences within each CC, with the sole exception of CFN5 whose translation product differed from the rest of CC-3 at one position (Table 2).

NCBI nucleotide BLAST identified all twenty-six unique 606 bp amplicons as *Frankia* nitrogenase reductases (*nif*H). Phylogenetic analysis of these sequences, trimmed to 512 bp, clearly assigned *nif*H sequences recovered in this study to either

cluster I (the *Alnus*-HIG) or cluster III (the *Elaeagnus*-HIG), with Neighbor joining, Maximum likelihood, Maximum parsimony and Bayesian analysis of the 512 bp *nif*H alignment producing similar tree topologies (Figures 2 & 3).
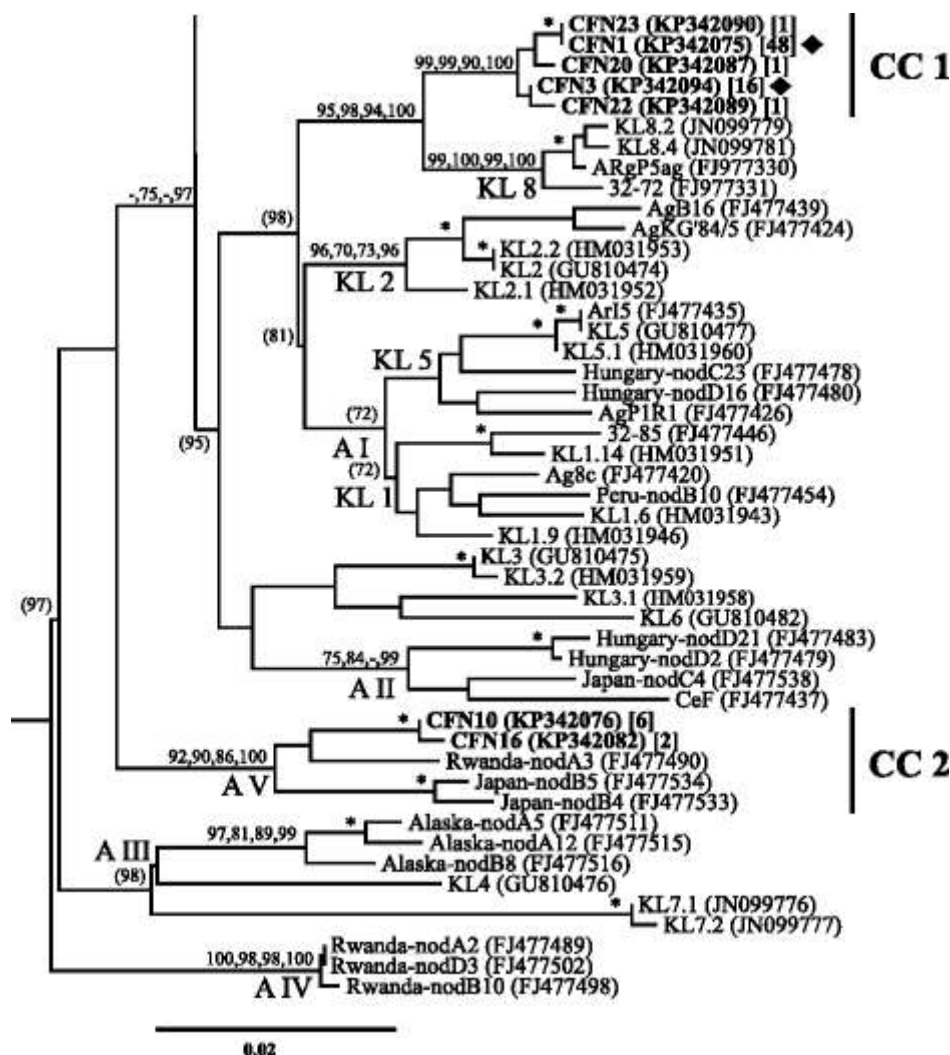
Figure 2: Phylogenetic tree generated using 512 bp *nif*H gene sequences from Frankia isolates and nodule-derived sequences, cut to display sequences from cluster I (the *Alnus* host infection group) only. Sequences are labelled with country of origin, genotype name or strain designation and GenBank accession number. Sequences encountered in this study are designated in bold, with GenBank accession numbers indicated in brackets followed by the number of nodules in which each sequence was detected. Diamonds indicate Cape *Frankia* isolates corresponding to nodule-derived *nif*H sequences. Numbers at nodes reflect bootstrap (BS) measures from neighbour joining, maximum likelihood and maximum parsimony analyses, and posterior probabilities (PP) from Bayesian analysis, respectively. Only values above 70% are shown. Where nodes were supported by Bayesian analysis only, PP values are indicated in parentheses. Asterisks indicate terminal nodes with support from at least 3 of the 4 phylogenetic methods. *Frankia Alnus* sub-groups AI-AV (Welsh et al. 2009a), KL1-5 and KL6-8 (Kennedy et al. 2010; Higgins and Kennedy 2012) are labelled where appropriate, and Cape clusters (CC) are indicated. An uncultured *Frankia* strain from *Datisca* nodules (X76398) was included as the outgroup (not shown).

9
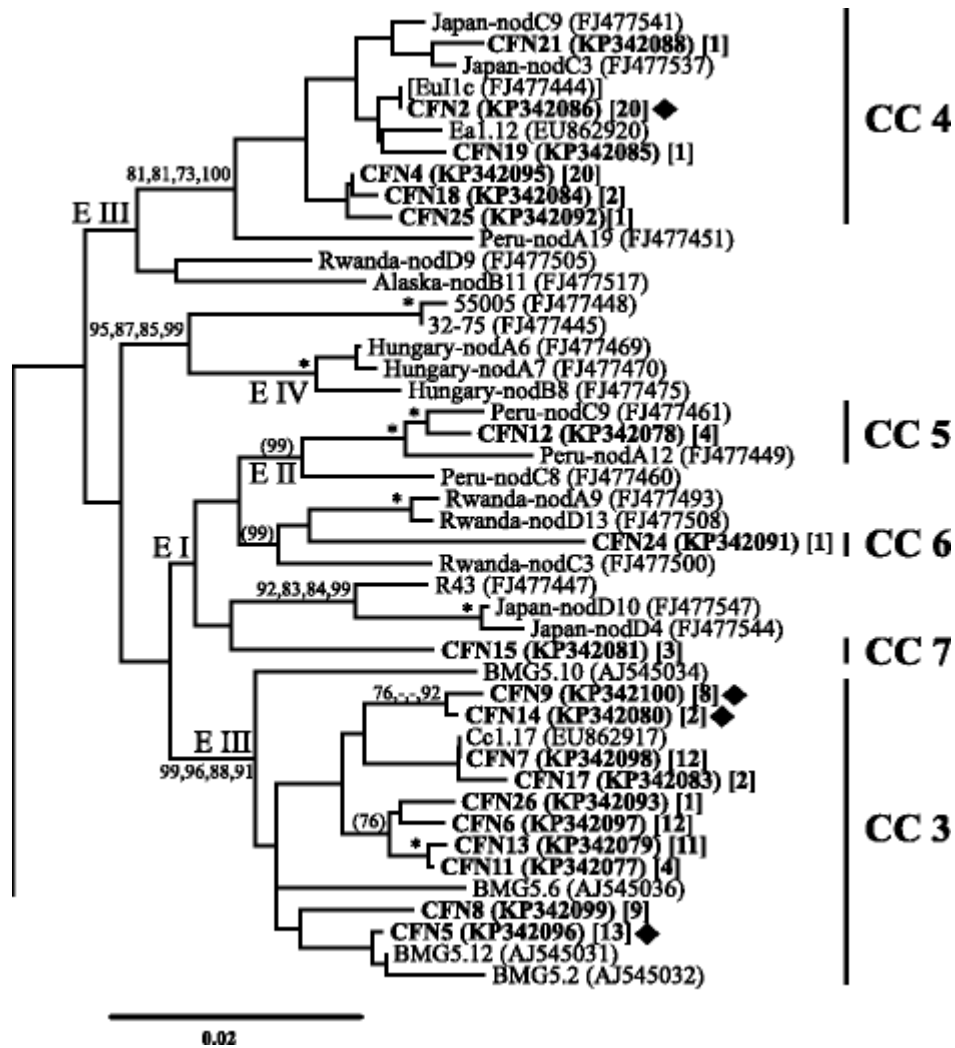
Figure 3: Phylogenetic tree generated using 512 bp *nif*H gene sequences from Frankia isolates and nodule-derived sequences, cut to display sequences from cluster III (the *Elaeagnus* host infection group) only. Sequences are labelled with country of origin, strain designation and GenBank accession number. Sequences encountered in this study are designated in bold, with Genbank accession numbers indicated in brackets followed by the number of nodules in which each sequence was detected. Diamonds indicate nodule-derived *nif*H sequences corresponding to Cape *Frankia* isolates. Numbers at nodes are bootstrap (BS) measures from neighbour joining, maximum likelihood and maximum parsimony analysis, and posterior probabilities (PP) from Bayesian analysis, respectively. Only values above 70% are shown. Where nodes were supported by Bayesian analysis only, PP values are indicated in parentheses. Asterisks indicate terminal nodes with support from at least 3 of the 4 phylogenetic methods. *Elaeagnus* clusters EI-EIV (Welsh et al. 2009a) are labelled, and Cape clusters (CC) are indicated on the right. EUI1c's *nif*H sequence (bracketed) is aberrant as the gene is now known to be absent from this strain's genome (NC_014666). An uncultured *Frankia* strain from *Datisca* nodules (X76398) was included as the outgroup (not shown).

Two of the seven CC genotypes were assigned to *Frankia* cluster I, and the remaining five to cluster III (Table 2, Figures 2 & 3). The dominant genotypes (CC-1, CC-3 and CC-4) were found in 186 of 202 nodules (92%) and totaled 67, 74 and 45 nodules per clade, respectively. Three of the cluster III genotypes (CC-5, CC-6, CC-7) were represented by single non-identical sequences only.

Within the *Alnus* HIG, genotype CC-2 belonged to subgroup AV, as defined by Welsh et al. (2009a). Genotype CC-1 is novel to our study and is most closely related

to KL8, described by Kennedy et al. (2010). Within the *Elaeagnus* HIG the majority of sequences (119/127) fell within EIII, previously described by (Welsh et al. 2009a). The remaining 8 sequences comprised genotypes clustering within EI (CC-6, CC-7) and EII (CC-5).

Ten clonally pure *Frankia* strains, representing each of the numerically dominant CC genotypes found in the nodules (CC-1, CC-3 and CC-4), were successfully isolated from nodules of *M. cordifolia* (FMc1-FMc6), *M. integra* (FMi1, FMi2), *M. kraussiana* (FMk1) and *M. quercifolia* (FMq1) (Table 2, Figure 4). In each case *nif*H sequences from the isolates were identical to those found in the nodules of the host species from which they were isolated (Table 2). In each case 16S rRNA BLAST analysis confirmed the identity of the isolates as *Frankia*, and phylogenetic analysis assigned them to the same HIGs indicated by *nif*H analysis (Figure 4).
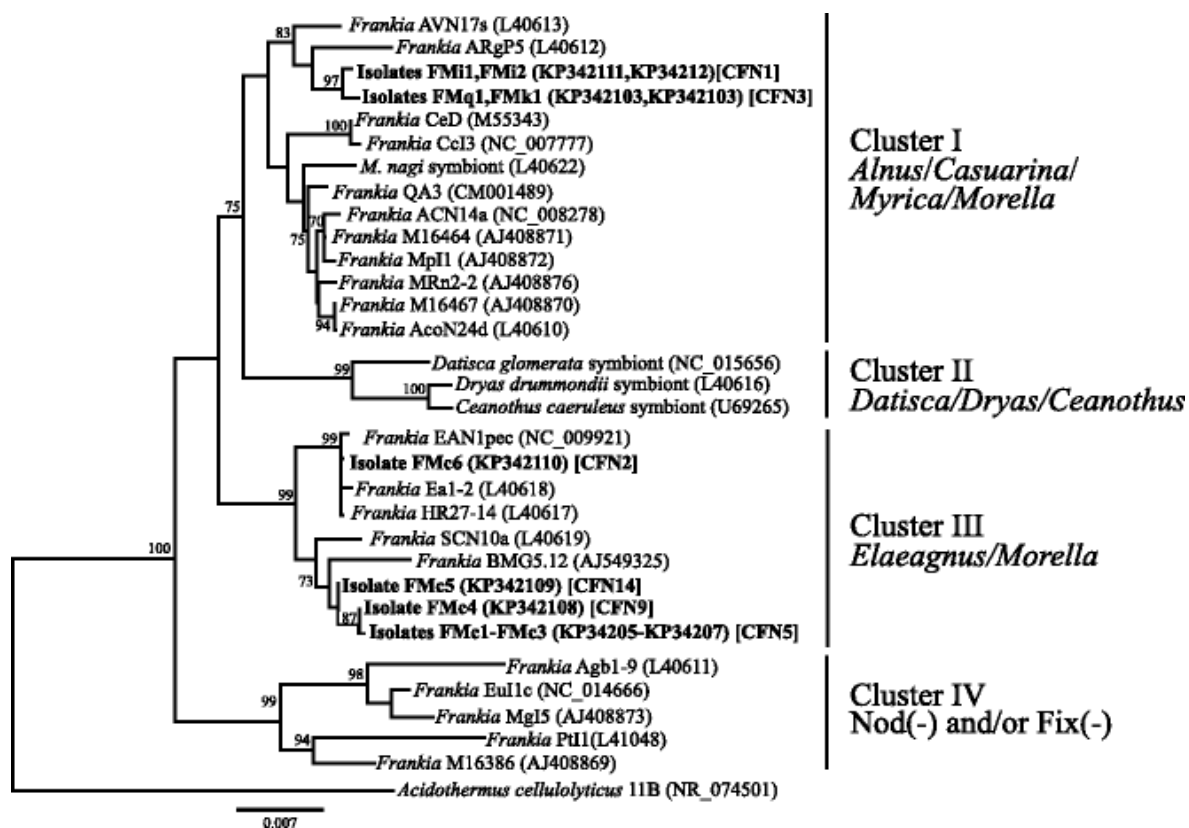
Figure 4: Neighbour-joining phylogenetic tree of 1428 bp partial 16S ribosomal RNA gene sequences showing positions of *Frankia* strains within clusters according to Normand et al. (1996). Isolates are labeled with trivial designations and GenBank accession number. Cape isolates are indicated in bold, with trivial strain designation, GenBank accession number and their corresponding *nif*H sequence. Uncultivated strains are indicated by the name of the host plant species. Accession numbers are shown in parentheses; *Frankia* clusters are indicated in the margin. *Acidothermus cellulolyticus* 11B was included as outgroup.

Discussion

Frankiae *nif*H sequences were recovered from the majority of nodules collected in

this study. From 202 successful PCR reactions, a total of 26 unique 606 bp *nif*H sequences were obtained, all of which were identified as *Frankia* nitrogenase reductases by BLAST analysis (Table 2). A ≥97% similarity threshold is commonly used to describe *Frankia* diversity as it allows for differentiation between cluster I and cluster III strains, with reliable assignment to sub-clusters within these groups (Higgins and Kennedy 2012; Welsh et ala 2009a; Mirza et al 2009; Lipus and Kennedy 2011). When Cape *Frankia nif*H sequences were clustered at ≥97% similarity a total of seven genotypes were found, namely (CC) 1 to 7 (Table 2).

Two of these genotypes were assigned to cluster I and the remaining five CC's to cluster III (Table 2). Both major *Frankia* host infection groups are therefore associated with Cape *Morella* in their natural habitats. The overall number of detected strains (26) was comparatively low, which is in keeping with what has been found in other environments (Huguet et al. 2001; Welsh et al. 2009a; Vanden Heuvel et al. 2004; Pokharel et al. 2011).

In order to determine the position of Cape nodular *nif*H sequences in relation to previously described *Frankia*, CFNs were aligned with *Frankia nif*H sequences from public databases. As many database sequences are incomplete, the alignment was trimmed to 512 bp to allow the incorporation of the greatest number of sequences possible. Following trimming the number of unique Cape *Morella* derived sequences was reduced from 26 to 25, as CFN1 and CFN23 are identical across the remaining alignment positions. CFN cluster assignments at a ≥97% identity threshold for the 512 bp truncations were identical to those found for the 606 bp sequences.

When different phylogenetic treeing methods have previously been applied to *Frankia nif*H sequences, similar tree topologies have been found irrespective of the methodology used (Welsh et al 2009a). Sub-clusters within these trees were assigned primarily based on tree topology and node support measures rather than a strict sequence similarity threshold. Consequently, clusters outlined by (Welsh et al. 2009a) are defined at a range of similarity values from 93% to 99%. It is known that defined similarity values (or OTU assignments) for both *nif*H and 16S rRNA gene sequences, as well as other commonly used markers, do not necessarily agree with *Frankia* genospecies as defined in DNA/DNA hybridization and AFLP studies (Fernandez et al. 1989; Welsh et al. 2009a; Bautista et al. 2011). In line with this, the common practice of assigning microbial genetic marker sequences (usually 16S rDNA) to operational taxonomic units at defined thresholds has been criticized, principally because of the lack of support from an underlying theory of microbial speciation (Gevers et al. 2005)[2].

Nevertheless, cut-offs of 97% similarity for *nif*H sequences are still commonly

---

[2] In one study *nif*H OTUs defined at 99.2 to 99.4% similarity were found to be ecologically meaningful in the grouping of *Alnus*-associated *Frankia* on a global scale (Sergei et al. 2014).

used in investigations of *Frankia* diversity (Higgins and Kennedy 2012), and we found that in-silico translation of CFN sequences resulted in identical peptide sequences for nucleotide sequences within CC genotypes. Furthermore, peptide sequences were unique for each of the seven CC genotypes (Table 2). This finding supports the use of a 3% nucleotide dissimilarity threshold in cluster assignment for investigating *Frankia* diversity within the Cape region.

Identical *nif*H sequences were found in nodules from geographically distant sites, with eight of the thirteen general localities sharing identical CFNs with between one and three other locations (data not shown). Of the 26 non-identical *nif*H sequences, five (consisting of totals of 64 individual sequences from CC-1, and 44 from CC-3) were found at multiple sampling sites. In terms **of** host specificity, *Frankia* from both clusters I and III were represented in nodules from three of the six species examined (*M. integra*, *M. diversifolia* and *M. quercifolia*).

Cluster I sequences displayed low local diversity, a finding often encountered in studies of this host infection group (Huguet et al. 2001; Welsh et al. 2009b; Kennedy et al. 2010). Cluster I strains are typically associated with specific hosts, as demonstrated in sympatric *Alnus* stands, and where *Alnus* and *Myrica gale* grow together (Huguet et al. 2001; Pokharel et al. 2011). Furthermore, strains specifically infective on *Myrica gale* in Europe and North America display evidence of divergence, supporting the hypothesis that some degree of co-evolution exists between *Frankia* from this cluster and their hosts (Huguet et al. 2001; Huguet et al. 2004). While cluster I strains are cosmopolitan, local dominance by sub-groups within the cluster found in other studies and the presence of only two specific groups (CC-1 and CC-2) within the Cape region is evidence for greater selectivity between strains from this cluster and their hosts than is found with cluster III strains.

Within cluster I sequences from genotype CC-1 were numerically dominant, with five unique sequences recovered from a total of 67 nodules, derived from five *Morella* species at geographically dispersed sites. While soil conditions were not determined for all study sites, cluster I sequences were found in nodules from acidic soils exclusively (Table 1), and were dominant in *M. integra* collected at riparian sites in the Cederberg mountains (Figure 1, site A2). Sequence CFN1 was also detected in six nodules from *M. serrata* (Figure 1, site 6B) where this species inhabits a similar riparian habitat. Sequence CFN3 was detected in nodules from riparian *M. integra*, as well as nodules from *M. diversifolia*, *M. quercifolia* and *M. kraussiana*, all from montane habitats on the Table Mountain range.

According to both cluster and phylogenetic treeing analyses, CC-1 is a sister group to genotype KL8, recently described from four *Alnus* nodules collected in Mexico (Higgins and Kennedy 2012). Nodules from a *M. serrata* stand on the banks of the Eerste river (site B6) harbored sequence CFN1, identical to those dominant in *M. integra* nodules collected at riparian sites in the Cederberg mountains (Figure 1, site A3), exclusively. Similarly, sequence CFR2 was found in *M. kraussiana*, *M.*

13

diversifolia and *M. quercifolia* nodules from acidic soils on Table Mountain, as well as *M. integra* nodules collected in the Cederberg (Figure 1, site A3). *Myrica gale* is known to be nodulated by cluster I strains exclusively in its natural habitats, where soils are typically acidic and water-logged (Clawson and Benson 1999), similar to the soil in which *M. integra* grew.

Cluster III strains were more diverse, with 18 unique sequences from five Cape clusters distributed amongst three of the four clusters identified by (Welsh et al. 2009a). Within cluster III sequences CFN5 and CFN6, the most abundant sequences in this cluster, were found at sites with the highest degree of geographical separation, indicating cosmopolitan distribution for strains from this cluster within the region. Similarly, cluster I sequences CFN1 and CFN3 were found in all montane areas surveyed, ranging from the Boland and Table Mountain ranges in the south to the Cederberg range in the north. As with the cluster I sequences, unique genotypes were associated with *M. diversifolia* at site B1 (Figure 1) exclusively, and were found in numerous nodules. In our analysis sub-cluster EIII was resolved into two separate groups (CC-3 and CC-4), with CFN sequences present in both. Genotypes CC-2, CC-5 and CC-6 were detected in *M. diversifolia* nodules only.

This last species is restricted to the mountains of the Cape peninsula (Goldblatt and Manning 2000; Helme and Trinder-Smith 2006) and it is possible that these genotypes are associated uniquely with it. More intensive sampling would be necessary to establish whether specific relationships between these *Frankia* genotypes and *M. diversifolia* exist. Furthermore, cluster III sequence CFN8 was found in this species which, while belonging to CC-3, was not detected in the nodules of any other host species. Assuming host/symbiont co-evolution *Alnus*-cluster *nif*H sequences from CC-2 may therefore originate from *Frankia* evolving towards a preference for this host, and further investigation is warranted. No isolates from genotype AV are known to exist (Welsh et al. 2009a) and despite intense effort we were unable to isolate *Frankia* strains representing genotype CC-2.

Soil pH and other edaphic factors are known to influence both strain presence in nodules and the degree of nodulation (Smolander 1990; Zitzer and Dawson 1992; Dawson and Klemp 1987). For example, soil pH had clear effects on nodulation of *Alnus glutinosa* and *Elaeagnus angustifolia* species in greenhouse trials, with the former more heavily nodulated at lower, and the latter at higher, pH's (Zitzer and Dawson 1992). In our study *M. cordifolia* was found in neutral to alkaline coastal soils (Table 1, Figure 1), was the most widely and intensively sampled host species with 67 nodules recovered from 5 geographically widely-dispersed coastal sites, accounting for 11 of 26 unique CFNs (Table 2), and was nodulated by cluster III strains exclusively. Similarly, *M. quercifolia* was nodulated by cluster III strains exclusively when sampled at sites with alkaline soils. A single *M. quercifolia* nodule recovered from soil at site B3 (Figure 1) yielded a cluster I sequence (CFN3), indicating that this species too is promiscuous in the field. This sample was recovered from similar

soil to that of *M. kraussiana* at the same site, for which soil pH was found to be acidic (Table 1). This supports suggestions that soil pH may play a determining role in *Frankia* strain selection (Huguet et al. 2001).

Welsh et al. (2009a) found local dominance in soils collected from 5 globally distributed sites, with *Frankia* clusters present in any one soil rarely detected in others. Our findings mirror this, as sequences from cluster EIII from the *Elaeagnus* HIG were dominant in Cape *Morella* nodules, with sequences assigned to clusters EI and EII only rarely detected. Similarly, Cape sequences assigned to the *Alnus* HIG were limited to two sub-clusters only, namely Welsh's cluster AV and a new cluster: CC-1.

Higgins and Kennedy (2012) reported a new *Frankia nif*H lineage (KL8) in their survey of *Alnus* nodules across North America. In our analysis *Frankia* isolates ARgP5ag and 32-72 (which were excluded from their analysis) clustered within this genotype and separately from CC-1, at 97% sequence identity and with good support from all four treeing methods (Figure 2). ARgP5ag is currently the sole member of *Frankia* genospecies G3 (Fernandez et al 1989; Bautista et al. 2011). With a average *nif*H similarity of only 96.8% to KL8, the previously described *Frankia* cluster most similar to it, Cape *Frankia* genotype CC-1 may comprise a new group of *Frankia* within the otherwise well-described cluster I.

Notably, we did not detect strains clustering with *Casuarina* genotypes (*nif*H subgroup AII) in our survey. *Morella* is the only extant actinorhizal genus native to southern Africa, as endemic *Casuarina* became extinct in the middle Miocene (Coetzee and Praglowski 1984; Linder 2003). This study thus provides additional support for the claim that *Morella* do not serve as a reservoir for typical *Casuarina* strains in the absence of their normal hosts (Simonet et al. 1999).

While *Frankia* strains have been previously isolated from indigenous African actinorhizal hosts, they have never been obtained from species endemic to the continent (Gtari et al. 2004; Gomaa et al. 2008). Of the six *Morella* investigated in this study, we isolated a total of ten strains from four host species (Table 2). Near-complete 16S and partial *nif*H gene fragments were obtained from ten *Frankia* isolates, and all isolate-derived *nif*H sequences corresponded to sequences obtained from nodules (Figure 4). These strains displayed typical *Frankia* phenotypes, including slow growth rates, diazovesicle formation and multilocular sporangia (Whitman et al. 2012). Isolates from *M. quercifolia* and *M. kraussiana* were identical in terms of their *nif*H and 16S rRNA gene sequences. The nodules from which they were isolated were collected at the same site (Figure 1, site B3) and it is thus possible that they are the same strain. *Frankia* strains FMc1, FMc2 and FMc3 had identical 16S and *nif*H sequences, but displayed significant phenotypic variation, including distinct differences in pigment production, substrate utilization and sporulation intensity (Wilcox, unpublished data). Previous studies have reported that phenotypically distinct isolates may have identical marker sequences (Bern`eche-D'Amours et al. 2011).

Root nodules have previously been found to be formed on actinorhizal hosts by atypical cluster IV *Frankia*, which would not be detected with the described methodology, and other actinomycetes found in actinorhizal nodules may be involved in their induction (Valdes et al. 2005; Gtari et al. 2012; Carro et al. 2013). In addition to this, nitrogen-fixing actinomycetes other than *Frankia* have previously been isolated from actinorhizal nodules (Valdes et al. 2005) and a *Frankia nif*H sequence has recently been found in a non-*Frankia* actinomycete (BMG5.6) isolated from *Elaeagnus* nodules from Tunisia (Gtari et al. 2004). For this reason, nodule derived *nif*H sequences by themselves do not guarantee the identity of nodular microsymbionts as *Frankia*, just as root nodules on actinorhizal plant roots do not guarantee the presence of *Frankia*. Our *Frankia* isolates have identical 606 bp *nif*H sequences to the numerically dominant nodular sequences in genotypes CC-1, CC-3 and CC-4, as well as two less abundant sequences from CC-3 (Table 2). These represent a majority of nodular sequences recovered in the study (111 of 202). The isolation of nodule endophytes, their characteristic cell morphologies, and confirmation of their identity by 16S rRNA and *nif*H sequencing, strongly supports the conclusion that Cape *Morella* microsymbionts detected in the current survey of nodular *nif*H sequences are indeed *Frankia.*

Cape cluster I *Frankia* have, following the local extinction of *Casuarina*, nodulated *Morella* exclusively and it will be interesting to deter- mine whether they are infective/effective on "traditional" cluster I hosts from the Betulaceae and Casuarinaceae. As representatives of this cluster are included among our isolates, future genomic comparison and cross-inoculation studies with *Alnus*, *Casuarina* and hosts from the Myricaceae will provide a deeper understanding of *Frankia* host specificity, genome evolution and the molecular mechanisms of infection for these strains.

The influence of geographic location on *Frankia* occurrence was not investigated in depth as it could not be detached from various potentially confounding factors, which included the limited number of populations sampled for most host species (particularly *M. serrata* and *M. kraussiana*), and varying soil conditions between sites.

Conclusions

Southern African actinorhizal hosts of the genus *Morella* are nodulated by *Frankia* from both cluster I (*Alnus*-infective) and cluster III (*Elaeagnus* -infective) under natural field conditions. Within Africa the diversity of *Frankia* infective on endemic *Morella* species had previously not been investigated. *Morella* of the Cape flora are promiscuous in the field, with sequences from cluster I and cluster III recovered from nodules of three of the six species sampled. Cluster I strains were represented by seven unique sequences in two genotypes, assigned to AV and a new sub-cluster closely related to KL8. Cluster III strains were more diverse, with a total of 18 unique sequence distributed across five

genotypes within previously described groups. Three *Morella* species (*M. integra*, *M. diversifolia* and *M. quercifolia*) were found to be promiscuous in the field. Two (*M. serrata*, *M. kraussiana*) were nodulated by cluster I strains while *M. cordifolia* was nodulated by cluster III strains exclusively. *Morella* displayed an apparent preference for cluster I strains under acidic soils conditions, with these strains completely absent in nodules recovered from neutral or alkaline soils. Ten *Frankia* strains representing six non-identical *nif*H sequences, two from a single cluster I genotype (CC-1), four from two cluster III genotypes (CC-3 and CC-4). This work constitutes the first investigation into the diversity of *Frankia* infective on endemic actinorhizal plant species of southern Africa, and reports the first isolation of *Frankia* strains from African *Morella*.

Acknowledgements

List of references

Bassi CA, Benson DR (2007) Growth characteristics of the slow-growing actinobacterium *Frankia* sp. strain cci3 on solid media. Physiol Plant 130:391–399

Bautista GHH, Cruz HA, Nesme X, Valdes M, Mendoza HA, Fernandez MP (2011) Genomospecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of *Frankia* spp. Syst Appl Microbiol 34:200–206

Benson DR, Dawson JO (2007) Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiol Plant 130(3):318–330

Benson DR, Vanden Heuvel BD, Potter D (2004) Actinorhizal symbioses: diversity and biogeography. In: Gillings M (ed) Plant Microbiology, BIOS Scientific Publishers Ltd., Oxford

Bern`eche-D'Amours A, Ghinet MG, Beaudin J, Brzezinski R, Roy S (2011) Sequence analysis of rpoB and rpoD gene fragments reveals the phylogenetic diversity of actinobacteria of genus *Frankia*. Can J Microbiol 57:244–249

Bond G (1971) Fixation of nitrogen in two African species of *Myrica*. New Phytol 70:1–8

Bond G (1976) The results of the IBP survey of non-legume root nodules formation in non-leguminous angiosperms. In: Nutman PS (ed) Symbiotic Nitrogen Fixation in Plants, Cambridge Univ. Press, pp 443–474

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

Carro L, Pujic P, Trujillo ME, Normand P (2013) *Micromonospora* is a normal occupant of actinorhizal nodules. J Biosci 38(4):685–693

Clawson ML, Benson DR (1999) Natural diversity of *Frankia* strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Appl Environ Microbiol 65(10):4521–4527

Clawson ML, Gawronski J, Benson DR (1999) Dominance of *Frankia* strains in stands of *Alnus incana* subsp. *rugosa* and *Myrica pensylvanica*. Can J Botany 77:1203–1207, 11111

Coetzee JA, Praglowski J (1984) Pollen evidence for the occurrence of *Casuarina* and *Myrica* in the Tertiary of South Africa. Grana 23(1):23–41

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772

Dawson JO, Klemp MT (1987) Proceedings Central Hardwoods Forest Conference VI. In: Variation in the capacity of Black Alder to nodulate in central Illinois soils, Dept. Forestry, Univ. Tenn., Knoxville, TN, pp 255–260

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791

Fernandez MP, Meugnier H, Grimont PAD, Bardin R (1989) Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Intl J Syst Evol Micr 39(4):424–429

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van De Peer Y, Vandamme P, Thompson FL, Swings J (2005) Re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739

Goldblatt P (1978) An analysis of the flora of southern Africa: its characteristics, relationships, and origins. Ann Missouri Bot 65(2):369–436

Goldblatt P, Manning JC (2000) Cape Plants. A conspectus of the Cape flora of South Africa. National Botanical Institute of South Africa. pp. 547

Gomaa AM, Abo-Aba SEM, Awad NS (2008) Isolation, characterization and genetic differentiation of *Frankia* sp. isolated from ecologically different Egyptian locations. Res J Cell Mol Biol 2(1):6–17

Gtari M, Dawson JO (2011) An overview of actinorhizal plants in Africa. Func Plant Biol 38:653–661

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

Gtari M, Ghodhbane-Gtari F, Nouioui I, Beauchemin NJ, Tisa LS (2012) Phylogenetic perspectives of nitrogen-fixing actinobacteria. Arch Microbiol 194:3–11

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst Biology 52:696–704

He XH, Chen LG, Hu XQ, Asghar S (2004) Natural diversity of nodular microsymbionts of *Myrica rubra*. Plant Soil 262:229–239

Helme N, Trinder-Smith T (2006) The endemic flora of the Cape Peninsula, South Africa. S Afr J Bot 72:205–210

Herbert J (2005) Systematics and biogeography of Myricaceae. Dissertation, University of St. Andrews

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

Huguet V, McCray Batzli J, Zimpfer JF, Normand P, Dawson JO, Fernandez MP (2001) Diversity and specificity of *Frankia* strains in nodules of sympatric *Myrica gale*, *Alnus incana*, and *Sheperdia canadensis* determined by *rrs* gene polymorphism. Appl Environ Microbiol 67(5):2116–2122

Huguet V, Mergeay M, Cervantes E, Fernandez MP (2004) Diversity of *Frankia* strains associated to *Myrica gale* in western Europe: impact of host plant (*Myrica* vs. *Alnus*) and of edaphic factors. Environ Microbiol 6(10):1032–1041

Huguet V, Ojeda Land E, Garcia Casanova J, Zimpfer JF, Fernandez MP (2005) Microbial diversity of microsymbionts from the relict species *Myrica rivas-myrica faya* (Ait.) and *Martinezii* (S.) in Canary Islands and Hawaii. Microbial Ecol 49(4):617–625

Kennedy PG, Weber MG, Bluhm AA (2010) *Frankia* bacteria in *Alnus rubra* forests: genetic diversity and determinants of assemblage structure. Plant Soil 335:479–492

Lechevalier MP, Lechevalier HA (1990) Systematics, Isolation and Culture of *Frankia*. In: Schwintzer CR, Tjepkema JD (eds) The Biology of *Frankia* and Actinorhizal Plants, Academic Press, Inc., New York, N.Y., pp 35–60

Linder HP (2003) The radiation of the Cape flora, southern Africa. Biol Rev 78:597–638

Lipus A, Kennedy PG (2011) *Frankia* assemblages associated with *Alnus rubra* and *Alnus viridis* are strongly influenced by host species identity. Int J Plant Sci 172(3):403–410

Maggia L, Bousquet J (1994) Molecular phylogeny of the actinorhizal Hamamelidae and relationships with host promiscuity towards *Frankia*. Mol Ecol 3:459–467

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

Miller MA, Pfeiffer W, Schwartz T (14 Nov. 2010) Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA pp 1–8

Mirza BS, Welsh A, Rieder JP, Paschke MW, Hahn D (2009) Diversity of frankiae in soils from five continents. Syst Appl Microbiol 32:558–570

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

Normand P, Benson DR, Berry AM, Tisa LS (2014) The Family Frankiaceae. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds) The Prokaryotes, Springer-Verlag, Berlin Heidelberg, pp 339–356

Pokharel A, Mirza BS, Dawson JO, Hahn D (2011) *Frankia* populations in soil and root nodules of sympatrically grown *Alnus* taxa. Microb Ecol 61:92–100

Rambaut A (2009) Figtree v1.2.3. Institute of evolutionary biology, university of Edinburgh. http://tree.bio.ed.ac.uk/software/figtree

Schofield RK, Wormald Taylor A (1955) The measurement of soil pH. Soil Sci Soc Am J 19(2):164–167

Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P (2014) The phylogeny of actinobacteria revisited in the light of complete genomes, the orders Frankiales and Micrococcales should be split into coherent entities. Proposal of *Frankiales* ord. nov., *Geodermatophilales* ord. nov., *Acidothermales* ord. nov. and *Nakamurellales* ord. nov. Int J Syst Evol Microbiol 4(11):3821–3832

Sergei P, Bahram M, Urmas K, Tendersoo L (2014) Global biogeography of *Alnus*-associated *Frankia* actinobacteria. New Phytol 204(4):979–988

Simonet P, Navarro E, Rouvier C, Reddell P, Zimpfer J, Dommergues Y, Bardin R, Combarro P, Hamelin J, Domenach aM, Gourbi`ere F, Prin Y, Dawson JO, Normand P (1999) Co-evolution between *Frankia* populations and host plants in the family Casuarinaceae and consequent patterns of global dispersal. Environ Microbiol 1(6):525–533

Smolander A (1990) *Frankia* populations in soils under different tree species with special emphasis on soils under *Betula pendula*. Plant Soil 121:1–10

South African National Biodiversity Institute Integrated Biodiversity Information System (SIBIS) (2014) http://sibis.sanbi.org/

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688–2690

Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571

Swofford DL (2003) PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4, Sinauer Associates, Sunderland, Massachusetts

The Plant List, Version 11 (2014) www.theplantlist.org. Accessed 12 November 2015

Torrey JG (1990) Cross-inoculation groups within *Frankia* and host-endosymbiont associations. In: Schwintzer CR, Tjepkema JD (eds) The Biology of *Frankia* and Actinorhizal Plants, Academic Press, Inc., New York, N.Y., pp 83–106

Valdes M, Perez N, Estrada-de los Santos P, Caballero-Mellado J, Pena-Cabriales JJ, Normand P, Hirsch PR (2005) Non-*Frankia* actinomycetes isolated from surface-sterilized roots of *Casuarina equisetifolia* fix nitrogen. Appl Environ Microbiol 71(1):460–466

Van Ryssen FWJ, Grobbelaar N (1970) The nodulating and nitrogen fixing ability of South African *Myrica*. S Afr JSci 66(1):22–25

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

Vandenbosch KA, Torrey JG (1983) Host-endophyte interactions in effective and ineffective nodules induced by the endophyte of *Myrica gale*. Can J Bot 61:2898–2909

Welsh A, Mirza BS, Rieder JP, Paschke MW, Hahn D (2009a) Diversity of Frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Syst Appl Microbiol 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). Appl Environ Microbiol 75(21):6913–6918

Whitman WB, Parte A, Goodfellow M, Kampfer P, Busse HJ, Trujillo ME, Ludwig W, Suzuki K (2012) Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria. v. 5, pts. 1-2, Springer

Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl Environ Microbiol 78(3):717–725

Zimpfer JF, Kennedy GJ, Smyth CA, Hamelin J, Navarro E, JDawson JO (1999) Localization of *Casuarina*-infective *Frankia* near *Casuarina cunninghamiana* trees in Jamaica. Can J Botany 77(9):1248–1256

Zitzer SF, Dawson JO (1992) Soil properties and actinorhizal vegetation influence nodulation of *Alnus glutinosa* and *Elaeagnus angustifolia* by *Frankia*. Plant Soil 140:197–204