Diversity of *Frankia* associated with *Morella* species of the Cape Floristic Region of Southern Africa

UNIVERSITY *of the*
WESTERN CAPE

Thesis presented in partial fulfillment
of the requirements for the degree
of Doctor of Philosophy
at the
Institute for Microbial Biotechnology and Metagenomics
Department of Biotechnology
University of the Western Cape

by

Dale Adrian Wilcox

Thesis advisers/examiners:

Professor Don A. Cowan, Thesis adviser
Professor Ndiko Ludidi, Administrative supervisor

Cape Town, Republic of South Africa
July 7, 2016

Summary

Frankia is one of two partners in the globally distributed N_2-fixing actinorhizal symbiosis between this filamentous soil-dwelling actinomycete and almost 300 species of host plants from eight diverse angiosperm families. The actinorhizal symbiosis is a major contributor to nitrogen reservoirs in terrestrial ecosystems, and allows actinorhizal plants to perform the role of pioneers in newly formed and nitrogen-poor soils. *Frankia* are differentiated into four main host-infection groups (1: *Alnus/Comptonia/Myrica*-infective, 2: Rosaceae/*Datisca/Coriaria*-infective, 3: Elaeagnaceae/*Gymnostoma*-infective and 4: *Casuarina*-infective), and there is a large degree of phylogenetic clustering within these HIGs. Of these host lineages, species from the genus *Morella*, from the family Myricaceae, are notable as they have the ability to establish effective partnerships with *Frankia* from more than one host-infection group. Africa houses 16 of the world's 33 currently accepted *Morella* species, and *Morella* is the continents only genus containing endemic actinorhizal species. Despite this, the diversity of *Frankia* in symbiosis with African *Morella* has never been explored.

To address this lack of knowledge I investigated *Frankia* in root nodules of six *Morella* species from the Cape flora of Southern Africa, as well as in rhizosphere soils from selected hosts. Partial *nif*H gene fragments recovered from 202 root nodules yielded 26 unique sequences, which phylogenetic analysis assigned to *Frankia* Cluster I (the *Alnus* host infection group) and *Frankia* Cluster III (the *Elaeagnus* host infection group)[1]. Nineteen *nif*H sequences were assigned to three sub-clusters within *Frankia* Cluster III (CC-3, CC-4 and CC-5), and the remaining seven sequences to two sub-clusters within Cluster I (CC-1 and CC-2), one of which (CC-1) is novel to the current study. Identical sequences were recovered from

[1]Hereafter the terms "*Elaeagnus*-HIG" and "Cluster III" are used interchangeably, as are "Alnus-HIG and" "Cluster I".

nodules collected at geographically distant locations, suggesting a cosmopolitan distribution within the region for some subgroups from both clusters, but more localized distribution (or tighter host-specificity) for others. Soil pH correlated with strain presence in nodules, with Cluster I sequences being associated with hosts growing in acidic soils exclusively. Furthermore, three *Morella* species from the Cape flora of southern Africa are promiscuous in their natural habitats, with host infection group influenced by habitat edaphic conditions.

In order to explore the correlation between soil characteristics and *Frankia* presence in nodules, *nif*H soil libraries were created from selected host rhizospheres. While Cluster III sequences from these libraries corresponded closely to sequences found in nodules from the same sites, the dominant Cape Cluster I group (CC-1) was absent from all six libraries, even when present in nodules recovered from the same soils. Whether this was due to low abundance of -but strong selection for- these strains by hosts under particular conditions, or due to the absence in soil of hyphal forms of these strains could not be determined. Cluster III strains are known to be better able to persist saprophytically than their relatives from other host-infection groups. A second group of Cluster I strains, detected at only one sampling site, was present in that site's corresponding soil library. An *Alnus*-infective subgroup, cluster AI, which has been detected in soils collected on five continents, was also detected in the of the Cape soil libraries but never in nodules, raising questions as to this group's ability to persist in soil in the absence of known suitable hosts.

Ten *Frankia* strains representing all three of the numerically dominant subgroups (CC-1, CC-3 and CC-4, found in 186 of 202 root nodules) were isolated from four *Morella* species. These isolates represent six of the most abundant unique nodular *nif*H sequences found in the field survey, and display morphological and cultural characteristics typical of *Frankia*. Phylogenetic analysis confirmed their identity as *Frankia*, and multilocus analysis revealed that the isolates belong to three genospecies. Two of these genospecies fall into existing groups within the *Elaeagnus*-infective Cluster III, while the remaining genospecies is a novel addition to the otherwise well-described *Alnus*-infective Cluster I. Whole genome sequencing of a representative from each of the Cape genospecies allowed for basic annotation and genome descriptions, which agreed in each case with what has been previously found for strains from the *Elaeagnus* and *Alnus* host-infection groups, respectively.

Similarly, the organization of nitrogenase gene clusters in each of the sequenced strains mirrors that found in other strains from their respective host-infection groups, indicating that this gene cluster is highly conserved in different *Frankia* lineages.

For the first time the diversity of *Frankia* nodulating endemic African *Morella*, and present in root-associated soils of these species, has been explored. This is also the first study to report isolation and description of *Frankia* strains from actinorhizal plants endemic to Africa.

Structure of the dissertation

A general introduction to the actinorhizal symbiosis (Chapter 1) provides background, introducing the topic and describing actinorhizal plants, their *Frankia* microsymbionts and discussing their symbiotic associations. The first experimental chapter reports on *Frankia* diversity in the root nodules of six *Morella* species from the Cape flora within their natural habitats (Chapter 2). As *Frankia* are recruited from soil, a study was conducted to determine the diversity of frankiae in rhizosphere soils at six sampling locations (Chapter 3). *Frankia* was isolated from four host species, and genetic and phenotypic characterizations performed (Chapter 4). Chapter 5 reports on the sequencing of three isolates and the structure of their nitrogenase gene clusters. Finally, a summary of the work and its limitations along with future research directions is provided (Chapter 6).

Figure 1: *Morella cordifolia* root nodules with South African one Rand coin for scale (diameter 23 mm).

Declaration

I declare that "Diversity of *Frankia* associated with *Morella* species of the Cape Floristic Region of Southern Africa" is my own work, that it was done wholly while in candidature for a research degree at this University and that it has not previously been submitted for a degree or any other qualification. Sources I have consulted are in each instance clearly attributed. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Cape Town,
July 7, 2016

Dale A. Wilcox

Acknowledgments

After a period of just over five years, today I am writing this note of thanks as the finishing touch on my doctoral thesis. It represents the last effort in what has been a prolonged ordeal in the pursuit of a personal goal, which I am now satisfied I have accomplished. I would like to acknowledge those who helped and supported me during my study.

I would first like to thank the Lord Jesus Christ, the Messiah. He's real, He's good, and He's true to His Word. Every time I've been in real trouble (for which I was sometimes myself responsible) He's provided a way out.

Johan Victor, Theoniel Potgieter and Anne Pope. Without your selfless legal assistance and advice I would *never* have been able to complete my degree. You invested your time and effort in order to help someone you didn't know, and you have my profound, sincere and eternal gratitude.

My family, and especially my brother Gareth, whose support and encouragement were invaluable. You have aided me greatly, and I will always be grateful for what you've done.

My supervisor, Don Cowan, for the opportunity to conduct my research in what was, at the time, your laboratory. Also for the freedom and support to independently pursue my own research interests.

The IMBM laboratory manager, Heide Goodman. In addition to maintaining an orderly and well managed laboratory environment (no small task), you were always willing to help with the many small tasks that needed doing. Thank you for this.

My friends Tyrone Genade, Andrew de Groot, Jean Louis de Klerk and Michael J. Vismer. Your assistance, especially with the field work, is greatly valued. Tyrone was particularly helpful, providing intellectual support and encouragement, and editing my thesis (any remaining errors are my own).

Richard Knight, who helped me to find my first *Morella* specimen in the field. This was a crucial first step in pursuing my thesis topic, you have my thanks.
My laboratory co-inhabitants, but especially Gerda Theron, who was always helpful, pleasant and hard-working.
The staff of SANParks[2], Cape Nature, and the University of the Western Cape and Zandvlei nature reserves. Every person I interacted with from each of these organizations was both professional and helpful, to a fault. I greatly appreciate the assistance they gave me in securing my sampling permits, and in providing data on species occurrence.
Finally, this study would not have been possible without financial support from the South African National Research Foundation, the University of Pretoria Genomics Research Institute, and the private contributions of the Wilcox family. Thank you all!

Cape Town, July 7, 2016 Dale A. Wilcox

[2]My apologies to the park rangers of Table Mountain National Park. Had I not on one particular day neglected to inform park authorities that I'd be sampling, an unfortunate incident could have been avoided.

Contents

Summary

Declaration **iv**

Acknowledgments **v**

List of figures **ix**

List of tables **xii**

List of Abbreviations **xiv**

1 The Actinorhizal symbiosis **1**

Abstract 1

1.1 *Frankia* and the actinorhizal symbiosis 1

1.2 Actinorhizal plant phylogeny and origin of symbiosis 5

1.3 Diversity of the genus *Frankia* 9

1.3.1 *Frankia* host-infection groups 10

1.3.2 *Frankia* phylogenetics 11

1.3.3 *Frankia* genospecies 13

1.4 Global distribution, taxonomy and symbiotic interactions within the Myricaceae 15

1.4.1 Symbiotic associations within the Myricaceae 17

1.4.2 South African *Morella* 20

1.4.3 *Morella* and the Cape Flora of Southern Africa 23

1.4.4 Studies of South African *Morella* prior to 2011 26

1.5 *Frankia* isolation and culture 28

1.5.1 Isolation procedures 30

1.5.2 *Frankia* morphology in culture . . . 32
1.6 Problem statement and methods used . . . 33
1.7 Conclusion . . . 33
References . . . 35

2 Natural diversity of *Frankia* in field collected nodules from Cape *Morella* species 58
Abstract . . . 58
2.1 Introduction . . . 59
2.2 Materials and Methods . . . 62
2.2.1 Identification of Cape *Morella* species in the field . . . 62
2.2.2 Nodule collection and processing . . . 62
2.2.3 Soil pH determination . . . 65
2.2.4 PCR conditions . . . 65
2.2.5 Sequencing and sequence analysis . . . 66
2.2.6 Phylogenetic analysis . . . 66
2.3 Results . . . 67
2.4 Discussion . . . 74
2.5 Conclusion . . . 80
References . . . 81

3 Diversity of frankiae in root-associated soils from Cape *Morella* 89
Abstract . . . 89
3.1 Introduction . . . 90
3.2 Materials and Methods . . . 91
3.2.1 Soil selection and sample preparation . . . 91
3.2.2 PCR conditions . . . 92
3.2.3 Phylogenetic analysis . . . 93
3.2.4 Clone coverage . . . 94
3.2.5 Intra-library diversity assessment . . . 95
3.2.6 Inter-library diversity comparisons . . . 95
3.3 Results . . . 96
3.4 Discussion . . . 109
3.5 Conclusion . . . 114

References . . . 116

4 Isolation and characterization of *Frankia* 122

Abstract . . . 122

4.1 Introduction . . . 123

4.2 Materials and Methods . . . 125

4.2.1 Isolation and cultivation of Cape *Frankia* . . . 125

4.2.2 Light and electron microscopy . . . 127

4.2.3 DNA extraction . . . 128

4.2.4 PCR conditions . . . 128

4.2.5 Sequence analysis . . . 129

4.2.6 Phylogenetic analysis . . . 130

4.3 Results . . . 132

4.3.1 *Frankia* isolation and morphological characteristics . . . 132

4.3.2 Sole carbon substrate utilization . . . 134

4.3.3 PCR amplification, phylogenetic and multilocus analysis . . . 137

4.4 Discussion . . . 147

4.5 Conclusion . . . 153

References . . . 154

5 Genome sequencing of *Frankia* strains FMc2, FMc6 and FMi1 163

Abstract . . . 163

5.1 Introduction . . . 164

5.2 Materials and Methods . . . 165

5.3 Results . . . 166

5.4 Discussion . . . 170

5.5 Conclusion . . . 171

References . . . 172

6 Summary and future prospects 178

A Supplementary material 185

List of Figures

1.1 Plant orders relationship to *Frankia* host infection groups 11
1.2 Global distribution of actinorhizal genera from the Myricaceae . . . 15
1.3 Illustrations of South African *Morella* leaf shapes. 21
1.4 Phytogeographic regions within the Western Cape 25
1.5 Pommer's illustration of his *Frankia* isolate, circa 1959 29

2.1 Typical Cape *Morella* leaf morphologies for six species 63
2.2 Unidentified *Morella* species 64
2.3 Map of study sites 71
2.4 *Frankia* Cluster I *nif*H phylogenetic tree 72
2.5 *Frankia* Cluster III *nif*H phylogenetic tree 73

3.1 Soil library pairwise similarity: shared *nif*H sequences. 101
3.2 Soil library structure comparisons: unique *nif*H sequences. 101
3.3 Soil library pairwise similarity: shared genotypes (97% similarity) . 102
3.4 Soil library structure comparisons: shared genotypes (97% similarity) 102
3.5 Bonteberg soil library *nif*H phylogenetic tree (BB) 103
3.6 Cederberg soil library *nif*H phylogenetic tree (CW) 104
3.7 Steenberg soil library *nif*H phylogenetic tree (SB) 105
3.8 Table Mountain soil library *nif*H phylogenetic tree (TM) 106
3.9 West Coast National Park soil library *nif*H phylogenetic tree (WC) 107
3.10 Zandvlei Wildlife Sanctuary soil library *nif*H phylogenetic tree (ZV) 108

4.1 Light microscopy of characteristic *Frankia* sporangia 133
4.2 Scanning electron microsocopy of (SEM) showing typical *Frankia* cellular morphology 134
4.3 Light microscopy of *Frankia* hyphae and vesicles 134

4.4 Substrate-dependent variation in pigment production by FMc2 . . . 135
4.5 Colony formation on solid media containing single carbon sources . 136
4.6 *Frankia nif*H phylogenetic tree: Cape isolates in relation to sequences from databases and Cape nodules 139
4.7 *Frankia nif*H phylogenetic tree: Cape isolates in relation to globally sourced isolates 140
4.8 *Frankia* 16S phylogenetic tree 141
4.9 *Frankia* 23S rRNA actinomycete-specific insertion phylogenetic tree 142
4.10 MLST maximum-likelihood tree of *Frankia* strains 145
4.11 MLST neighbornet graph of *Frankia* strains 146

A.2 *Frankia* subgroup VI 23S IS signature sequence alignment 192
A.3 *Frankia* subgroup IIIb 23S IS signature sequence alignment 193
A.4 Nitrogenase gene-cluster organization in from *Frankia* Cluster I strains 194
A.5 Nitrogenase gene-cluster organization in *Frankia* Cluster II strains . 194
A.6 Nitrogenase gene-cluster organization in *Frankia* Cluster III strains 195

List of Tables

1.1 List of actinorhizal host plants. 8
1.2 Distribution of ten indigenous *Morella* species within the provinces of South Africa. 22
1.3 Cross-inoculation studies with South African *Morella* 28

2.1 Collection sites for nodules 69
2.2 CFN sequences detected in individual *Morella* species 70

3.1 Collection sites for soils 97
3.2 Soil library metrics 98
3.3 Genotypes common to nodules and soil from the same site 99
3.4 Genotypes unique to either nodules or soil from the same site . . . 99
3.5 Soil library *nif*H sequences identical to sequences recovered from nodules 100

4.1 PCR primers used to characterize *Frankia* isolates 132
4.2 Morphological characteristics of *Frankia* isolates 133
4.3 *Frankia* growth on single carbon substrates 136
4.4 Pairwise distances between Cape *Frankia* isolates according to a concatenated 3149 bp multi-gene alignment 143
4.5 *Frankia* genomospecies according to DDH, AFLP and MLSA 143

5.1 Genome assembly statistics 168
5.2 Genome characteristics of isolates FMc2, FMc6 and FMi1 168
5.3 Number of genes associated with subsystem functional categories in six *Frankia* strains as estimated by RAST 169

A.1 Nodules per species at each individual sampling location 186

A.2 Soil clone library *nif*H BLAST results 188
A.3 Similarity of individual nucleotide sequences from genes within the nitrogenase gene cluster from isolate FMc2 to those from other *Frankia* strains 196
A.4 Similarity of individual peptide sequences from genes within the nitrogenase gene cluster from isolate FMc2 to those from other *Frankia* strains 197
A.5 Similarity of individual nucleotide sequences from genes within the nitrogenase gene cluster from isolate FMc6 to those from other *Frankia* strains 198
A.6 Similarity of individual peptide sequences from genes within the nitrogenase gene cluster from isolate FMc6 to those from other *Frankia* strains 199
A.7 Similarity of individual nucleotide sequences from genes within the nitrogenase gene cluster from isolate FMi1 to those of other *Frankia* strains 200
A.8 Similarity of individual peptide sequences from genes within the nitrogenase gene cluster from isolate FMi1 to those from other *Frankia* strains 201

List of Abbreviations

AFLP . . . . Amplified fragment length polymorphism

***atp*D** . . . . . ATP synthase subunit beta gene

bp . . . . . Base pair

CFN . . . . . Cape *Frankia* nitrogenase

CFR . . . . . Cape floristic region

DDH . . . . . DNA-DNA hybridization

DNA . . . . . Deoxyribonucleic acid

***dna*A** . . . . . Chromosomal replicator initiator protein DnaA

DPM . . . . . Defined propionate medium

FDM . . . . . *Frankia* defined minimal medium

***fts*Z** . . . . . Prokaryote cell division protein FtsZ gene

***gln*A** . . . . . Glutamine synthetase A gene

***gln*II** . . . . . Glutamine synthetase II gene

gu . . . . . Genomic units

***gyr*B** . . . . . DNA gyrase subunit B gene

HIG . . . . . Host-infection group

IGS . . . . . Intergenic spacer region

IS . . . . . Insertion sequence

Mbp . . . . . Million base pairs

MLSA . . . . Multilocus sequence analysis

MYA . . . . . Million years ago

nif	Nitrogen fixation
***nif*D**	nitrogenase molybdenum-iron protein subunit alpha gene
***nif*H**	Nitrogenase reductase gene
***nif*K**	nitrogenase molybdenum-iron protein subunit beta gene
nu	Nodulation units
ORf	Open reading frame
PCR	Polymerase chain reaction
pgk	Phosphoglycerate kinase gene
rDNA	Ribosomal deoxyribonucleic acid
***rpo*B**	Ribosomal polymerase subunit B gene
***rpo*D**	Ribosomal polymerase subunit D gene
rRNA	Ribosomal ribonucleic acid
RNA	Ribonucleic acid
Sp+/Sp-	spore-positive/spore-negative

1 The Actinorhizal symbiosis

Abstract

This chapter provides a brief introduction to the actinorhizal symbiosis, focusing primarily on interactions specific to the Myricaceae. The phylogeny of actinorhizal hosts and current taxonomic status of the actinomycete *Frankia* is presented. A history of research into actinorhiza either in South Africa itself, or involving South African species of *Morella*, is provided. Host-associations between *Frankia* lineages and those of actinorhizal plants, focusing on the Myricaceae, are outlined. Due to the acknowledged difficulty associated with isolation and culture of *Frankia*, a brief description of isolation procedures and characteristics in culture is provided. Finally, the problem addressed by this study is stated and a description and justification of the methods employed is provided.

1.1 *Frankia* and the actinorhizal symbiosis

Frankia are geographically widespread filamentous, gram-positive actinomycetes present in soil in low abundance, and which are capable of entering into nitrogen-fixing symbioses with a wide range of angiosperms by forming nodules on their roots [24]. This is referred to as the "actinorhizal" symbioses; host plants benefit from this association by gaining the ability to grow in nitrogen-poor habitats and are thus often pioneer species in new soils, such as those formed by vulcanism or

exposed by glacial retreat [26]. Microsymbionts gain nutrients from their hosts, and protection in soils where conditions are unfavorable for free-living *Frankia* [26, 136]. Actinorhizal plant species are found globally and, along with legumes, are major contributors of fixed nitrogen compounds to terrestrial ecosystems.

The first recorded description of actinorhizal nodules was by Meynen in 1829, who considered them to be parasitic plants [134]. It was only in 1866 that Woronin demonstrated that the organism was a microbe, although he incorrectly concluded that it was fungal in nature [204]. The view that the organism was parasitic persisted until the first demonstration of nitrogen fixation in nodules of greenhouse-grown *Alnus glutinosa*[1], whereafter this activity was ascribed to the microorganism [94]. The name *Frankia subtilis* was first given to the microsymbiont in root nodules of non-leguminous plants by Brunchorst, in honor of his professor B. Frank and after he had decided, based on what he considered to be hyphae and sporangia, that the organism was possibly an actinomycete [38]. Following years of unsuccessful isolation attempts, J. H. Becking considered the organism an obligate symbiont and in 1970 he proposed the family Frankiaceae, adopting the name "*Frankia*" for the single type genus it contained [18]. This is the name in use today [116, 201].

The study of *Frankia* has always been challenging. Despite the symbiosis being actively investigated since the late 19th century (reviewed by Wheeler, Akkermans and Berry [198]), the bacterial partner was not available in pure culture until 1978 [40]. Because of this, when Becking proposed the family Frankiaceae the ten *Frankia* species he described were based principally on the host plant's genus and species, the assumption that *Frankia* species had defined host specificity, and the microorganism's morphology in symbiosis (primarily that of its vesicles) [18]. Shortly thereafter Callaham isolated *Frankia* CpI1 from *Comptonia peregrina* and, after this initial success, other isolation methods and means of distributing strains were quickly developed [40, 66, 166]. As pure cultures became widely available it became apparent that Becking's species designations were untenable, as cross-inoculation of host genera with pure cultures showed that individual *Frankia* strains were able to infect a broader range of hosts than suggested by cross-inoculations

[1]George Bond later showed that nitrogen fixation in *Alnus* occurred under normal field conditions [32].

conducted using root-nodule homogenates [13, 24, 36, 108, 112, 184]. It also became clear that *in planta* morphology, the primary basis for Becking's species descriptions, was under the control of the host and not *Frankia* [108]. Becking's species were thus rejected, and to date the only species remaining in the genus is *Frankia alni* [201].

The isolation of CpI1 was a watershed[2], and caused much excitement in the plant research community [185]. Prior to 1979 all nitrogen-fixing symbioses outside of those interactions between microorganisms and plants of the Fabaceae were grouped under the collective term "non-leguminous plant symbioses". A more positive and descriptive name, specific to *Frankia* symbiosis, was now sought and at the first international meeting on "Symbiotic Nitrogen Fixation in Actinomycete-nodulated Plants" in 1978 the terms "actinomycetorhizal", "actinomycorrhizal" and "actinorhizal" were proposed [185]. The last term (actinorhizal) was ultimately accepted, and is used to describe *Frankia*-plant symbioses today [198][3]. Further details of work conducted between 1829 and 1978 are excluded here, but are covered in a comprehensive review by Anton Quispel, to which the reader is referred [156]. An exception is made for George Bond's contribution to the International Biological Programme, as his survey considerably expanded the number of known actinorhizal species, including for the first time all of the South African *Morella*[4] [34], and for the work of South African researchers or studies involving South African *Morella* species from 1966 to 1976.

Frankia is a free-living nitrogen-fixing soil actinomycete which enters into symbioses with members of 25 genera of actinorhizal plants from eight families of dicotyledons [26, 146, 178]. Despite a century of research into "non-legume" symbioses, the determination that root nodules were the site of nitrogen fixation and the discovery that the microsymbiont within these nodules was an actinomycete, the exact identity of nodulating organism remained uncertain [19, 35]. Principally, this was because *Frankia* had proven difficult to isolate and was thus considered an

[2]Pommer had reported isolation from *Alnus glutinosa* in 1959 [155], but this report went largely ignored and his strains were subsequently lost. Later examination of his laboratory records revealed, from the morphological characteristics of his isolates, that he had doubtlessly succeeded in isolating *Frankia* (Figure 1.5) [24, 40, 116, 156]

[3]A further change in nomenclature was proposed in 1988 at the seventh "International Meeting on *Frankia* and Actinorhizal Plants", where it was decided to abandon the use of the term "endophyte" in favor of "microsymbiont", as the former implied that *Frankia* was a plant [14].

[4]At the time these were classified as "*Myrica*".

obligate symbiont. Pommer reported isolation from *Alnus glutinosa* in 1959, but this report went largely ignored and his strains were subsequently lost [155, 198]. Later examination of his laboratory records revealed, from the morphological characteristics of his isolates, that he had doubtlessly succeeded in isolating *Frankia* (Figure 1.5) [24, 40, 116, 198].

It was not until 1978 that successful isolation was once again reported [24, 40, 156]. The isolate, CpI1 from *Comptonia peregrina*, was able to both re-infect actinorhizal hosts and induce nitrogen fixation in root nodules, thus partially fulfilling Koch's postulate [40, 107, 108, 183]. Following this initial success isolation methods were quickly developed [10, 21, 82, 116].

Economic importance of actinorhizal plants

Actinorhizal plants have diverse uses, both directly as timber, fuel, fruit and in the remediation of degraded lands, as well as indirectly due to their stimulatory effect on associated plant species in plantations [20, 54, 55].

Natural populations of *Alnus incana* are always nodulated, and when the species is included in European pine plantations it enhances growth of surrounding trees [8]. In North American timber plantations *Alnus rubra* plays a role in controlling root parasites like *Poria werii* [179]. *Casuarina* has been used in land-reclamation in West Africa and Asia [56], as have members of the Elaeagnaceae [79]. In South Africa *Casuarina* are used as a windbreak and to stabilize sand dunes, and is commonly planted on the edges of fields on farms in the Western Cape (personal observations). *Casuarina* have a high calorific value and are considered the best firewood in the world, burning easily even when green. This, combined with its rapid growth, makes it an important genus especially in emerging economies [69]. An emerging application for actinorhizal trees is in remediation

Actinorhizal plants also serve as food crops. Whereas many, such as Sea buckthorn (*Hippophae rhamnoides*), produce edible berries with limited economic importance [79], *Morella rubra* is a major fruit crop in China and on the Indian subcontinent [90]. Despite reports that *Morella faya* was used in wine making

by Portuguese sailors [92], interviews with natives and descendants of Portuguese settlers in the Azores revealed no local knowledge of such use [124]. One account of the berries being used to flavor brandy was given, but *M. faya* berries were reported to be too small to collect, and to contain too little juice for wine making [124]. In the Azores it is used for forage and as firewood; the only beneficial use of *M. faya* in Hawaii was in erosion control [124].

Specific uses for the Myricaceae are recorded, in brief, by Jane Herbert [92]. *Myrica gale* has insect-repellent qualities and work is being undertaken to make commercial use in Europe of both this and its anti-bacterial qualities [92]. In North America *M. cerifera* has similar uses [64]. In Africa various species have applications both as food sources and in traditional medicine [81]. The fatty coatings of *M. cordifolia* berries were eaten by the Khoisan peoples of southern Africa, and are used in small-scale production of candles and soaps (personal observations). In the 19th century a considerable export market existed for *M. cordifolia* wax, which fetched the same price as tallow [168]. In South Africa *M. quercifolia* and *M. diversifolia* possibly have similar applications to *M. gale* and *M. cerifera* as both of these species are fragrant, often powerfully so[5], and their leaves were never found to be attacked by insects in the field (personal observation). In contrast to these two species, and even when it was found growing in close proximity to them, the odorless *M. kraussiana* invariably possessed leaves damaged by insects. Gordon and Dawson (1979) suggest that the South African *Morella pilulifera* could find use as a provider of nitrogen in tropical conifer plantations, in a role similar to that played by *Alnus incana* in Europe [79].

1.2 Actinorhizal plant phylogeny and origin of symbiosis

Since the discovery of actinorhiza at least 276 host species nodulated by *Frankia* have been identified (Table 1.1). According to the morphology-based taxonomic system of Cronquist, these species fall into eight distinct and highly diverse families

[5]In the current study I found several populations of *M. quercifolia* merely by being downwind of them.

within four of six major angiosperm subclasses, all of which, with the exception of *Datisca*, are woody trees or shrubs (Table 1.1) [52, 53]. While they are widely distributed and occur on all major landmasses except Antarctica, most species are temperate with only a minority found in tropical and circumpolar environments (Table 1.1). In terms of ecology actinorhizal plants are usually pioneers on nitrogen-poor soils, and frequently don't persist past primary succession [26, 54, 163].

Despite their apparent diversity [52], the first molecular systematic study based on the ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO or *rbc*L) indicated that plants capable of entering root nodule N_2-fixing symbioses were more closely related than their morphology-based taxonomies suggested, assigning all of them to the Eurosid I clade [7, 45]. This was later confirmed using three loci (*rbcL*, 18S rDNA and chloroplast-encoded *atpB* genes)[6] [7, 170–172]. This lineage included actinorhizal species (eight families within the well-supported orders Rosales, Curcurbitales and Fagales, Figure 1.1), as well as legumes (Fabaceae) and *Rhizobium*-nodulated *Parasponia* from the family Ulmaceae. It was subsequently proposed that these families descended from a common ancestor which possessed a predisposition for nodulation, approximately 100 MYA [61, 171], and that genes involved in nodulation and nitrogen-fixing symbioses were recruited from older arbuscular mycorhizal symbioses-related genes after a whole genome multiplication event in this common ancestor [61, 95, 169].

Notably, not all families within the Eurosid I clade are capable of entering into N_2-fixing symbioses, and there are marked differences in nodule morphology, phylogenetic position and the identity of the symbiotic partner between actinorhizal and leguminous plants [154, 171, 178]. Together, these observations suggest that the symbiosis has arisen independently several times [103, 104, 177, 178]. Furthermore, actinorhizal plants are often interspersed with non-actinorhizal species within the same family (Table 1.1). In some families all members bear nodules (Casuarinaceae, Coriariaceae, Datiscaceae and Elaeagnaceae), while in others only a proportion are nodulated (Betulaceae, Myricaceae, Rhamnaceae and the most extreme case:

[6] *Gunnera* from the family Gunneraceae hosts the cyanobacteria *Nostoc* in leaf glands rather than root nodules. It does not belong to the Eurosid I clade, and represents an independently evolved N_2-fixing symbiosis. Similarly, nitrogen-fixing plant/cyanobacterial symbioses exist in liverworts, cycads and aquatic vasularized ferns such as *Azolla* [29, 133, 171]

Rosaceae), which suggests that loss of symbiosis may also have occurred more than once [54, 178]. This is perhaps not surprising, as the energetic cost of maintaining symbiosis would exert negative evolutionary pressure in environments where fixed nitrogen is freely available. Under such conditions, loss of symbiosis could easily result from even a single gene deletion. The result would be an immediate competitive advantage for the plant no longer burdened by the cost of maintaining its symbiont [178].

Table 1.1: List of actinorhizal host plants.

Subclass[a]	Order[b]	Family	Genus	Species[c,d]	Geographical distribution[d]	Habitat[d]	*Frankia* cluster[e]
Hamamelidae	Fagales	Betulaceae (1/6)	*Alnus*	31	Circumpolar, montane South America	River banks, forests	I
		Causarinaceae (4/4)	*Allocasuarina*	57	Oceania	Sandy soil, beaches	I
			Casuarina	14	Oceania, tropics, global	Sandy soil, beaches	I
			Ceauthostoma	2	Oceania	Sandy soil, beaches	?
			Gymnostoma	13	New Caledonia	Sandy soil, beaches	III
		Myricaceae (3/4)	*Comptonia*	1	North America	Sandy soil, forest	I
			Morella	**33**[f]	**Global**	**Montane, riparian, sandy soil, beaches**	I & III
			Myrica	3	North America, Europe	Acidic bogs lake shores	I
Rosidae	Rosales	Elaeagnaceae (3/3)	*Elaeagnus*	13	Circumpolar	Sandy soil	III
			Hippophae	7	Asia, Europe	Sandy beaches, mountain slopes	III
			Shepherdia	2	North America	Sandy soil	III
		Rhamnaceae (5/55)	*Ceanothus*	59	North America	Dry chaparral	II
			Colletia	5	South America	Dry chaparral	III
			Discaria	8	South America	Dry chaparral	III
			Kentrothamnus	1	South America	Dry chaparral	III
			Trevoa	1	South America	Dry chaparral	III
		Rosaceae (4/100)	*Cercocarpus*	3	North America	Rocky substrate	II
			Chamaebatia	1	North America	Pine forest understory	II
			Dryas	1	North America	Glacial moraines	II
			Purshia	2	North America	Forest soil	II
Magnoliidae Dilleniidae	Cucurbitale	Coriariaceae (1/1)	*Coriaria*	12	Disjunct[a]	Dry chaparral	II
		Datiscaceae (1/1)	*Datisca*	2	California, Europe, Asia	Mountain slopes	II
				271			

[a] According to Cronquist [53].
[b] According to the Angiosperm Phylogeny Group [7].
[c] Number of nodulated species according to Gtari et al. [83]
[d] Normand et al. [147], Gtari and Dawson [81] and Gtari et al. [83]
[e] Natural *Frankia* association in the field.
[f] Accepted species according to www.theplantlist.org (accessed 27 October 2015).

1.3 Diversity of the genus *Frankia*

It is clear that *Frankia* are highly diverse, and several reviews deal specifically with the taxonomy and relationships of the genus [24, 83, 113, 146]. Despite this diversity, and following the rejection of Becking's species [112], only *Frankia alni* is currently recognized within the genus [201]. Once isolates became available numerous attempts were made to describe species using classical physiological testing [24, 27, 116]. These were almost universally unsuccessful, not only because of the difficulties associated with *Frankia*'s slow growth, but also because test results often varied depending on how long the strains had been in culture [24, 113, 116]. Furthermore, characteristics that clearly discriminated the genus from other actinomycetes, such as *Frankia*'s morphology and unique whole-cell chemistry, were conserved among almost all *Frankia* strains. *Frankia* cell walls are of type III (meso-diaminopimelic acid, glutamic acid, alanine, glucosamine, and muramic acid), the most common type among actinomycetes [5], but whole-cell sugar patterns are unique to the genus: all *Frankia* contain 2-*O*-methyl-*D*-glucose which is not found in other actinomycetes [113, 140, 201]. Distinguishing morphological characteristics, particularly vesicles, may vary depending on culture conditions but only slightly between strains [201]. Consequently it was informally agreed among researchers in the field that species names would not be assigned "until better criteria for species definition" were found [113].

Approaches used to characterize *Frankia* have included DNA-DNA relatedness [2, 5, 17, 31, 65], carbon and nitrogen substrate utilization patterns [27, 119, 166, 180], isoenzyme variation [68], immunochemisty [11], pigment production and antibiotic resistance [60, 181], and DNA restriction patterns and fingerprinting [30, 31, 101, 102]. Phenotypes observed in symbiosis have also been considered, such as spore formation in nodules (sp+/sp-) [164, 187], host-specificity [13, 184], and mode of infection, either through intracellular penetration or intercellular infection (Figure 1.1). The molecular phylogeny of *Frankia* has been found to match the mode of host infection, with the infection mode of some promiscuous strains being determined by the host [137, 145].

1.3.1 *Frankia* host-infection groups

Actinorhizal plants belong to eight diverse families: Betulaceae, Casuarinaceae, Myricaceae, Eleaegnaceae, Rhamnaceae, Rosaceae, Coriariaceae and Datiscaceae (Table 1.1) [23, 26]. Cross-inoculation experiments on species from these families using cultured *Frankia* strains revealed the existence of three to four groups of isolates (Clusters I-IV) with different host affinities [13].

These groups are mirrored in nature, as when host species are examined *in the field*[7] strains from Cluster I are found in nodules from families within the order Fagales: Betulaceae, Casuarinaceae (with the exception of *Gymnostoma*) and Myricaceae. Strains infectious on Casuarinaceae belong to a subgroup 1c, while those from subgroup 1a are infectious on *Alnus* species [26, 148]. Strains from both of these subgroups are generally infectious on hosts from the Myricaceae [26, 62].

Cluster II *Frankia* are infective on families from three orders: Coriariaceae and Datiscaceae (Curcurbitales), Rosaceae (Rosales), and on members of the genus *Ceanothus* (Rhamnaceae). One of these was very recently isolated for the first time, following decades of failed attempts [82]. Cross-inoculation studies and direct molecular detection of strains in root nodules have revealed that *Frankia* from this cluster, while nodulating distantly related host species, are very closely related [25, 184, 190].

Strains from Cluster III infect members of the families Myricaceae (*Morella* only), Eleaegnaceae, Rhamnaceae (excluding *Ceanothus*), and *Gymnostoma* from the Casuarinaceae [26,47,142,143,157]. This group of strains has the widest distribution globally, and are also infrequently isolated from root nodules of other families. In such cases they usually do not reinfect the host from which they were isolated [26,71].

However, while strains are typically infective on plants from the host group from which they were isolated [23], some may also (infrequently) infect hosts from another group, as is the case with the flexible strain E15ab [36, 98]. A final group of strains, cluster IV, are unable to re-infect their original hosts or fix nitrogen [148].

[7]Some host species may be effectively nodulated by a greater range of strains in greenhouse trials than in their natural environments.

The relationships between these major *Frankia* host-infection groups (or clusters) and actinorhizal plant orders are summarized in Figure 1.1 and Table 1.1.

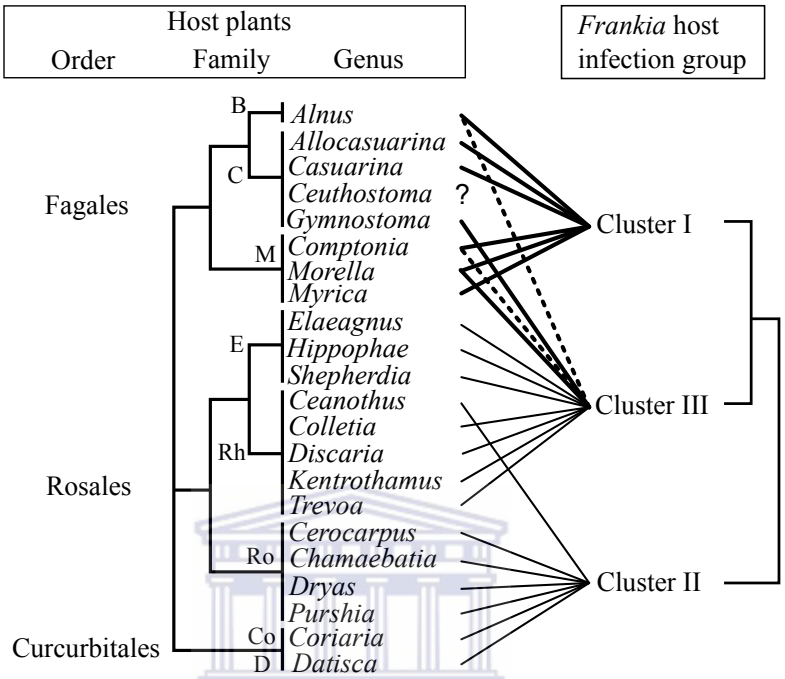


Figure 1.1: Relationships between actinorhizal plants (left) and *Frankia* (right). Families indicated as follows: B, Betulaceae; C, Casuarinaceae;, M, Myricaceae; E, Eleaegnaceae; Rh, Rhamnaceae; Ro, Rosaceae; Co, Coriariaceae; D, Datiscaceae. Oblique lines indicate relationships between hosts and host infection clusters. Modes of infection as follows: Heavy lines - root hair infection, lighter lines - intercellular penetration. Dotted lines indicate symbioses found in the laboratory but not in the field. *Morella* indicated in bold. Modified from Normand and Fernandez [146], and http://web.uconn.edu/mcbstaff/benson/Frankia/PhylogenyFrankia.htm.

1.3.2 *Frankia* phylogenetics

Phylogenetic methods, in particular, have proven useful in *Frankia* research, both in describing strains and in resolving the position of *Frankia* within the actinomycetes. The first study using 16S rDNA showed that *Frankia* was related to *Geodermatophilus obscurus* (with which it also has morphological similarities) which was transferred to the Frankiaceae, emending the family [86, 87]. This family was largely artificial, based only on 16S relatedness.

While no morphological traits had been found with which to discriminate members

of the major *Frankia* host-infection groups (described above) following the rejection of Becking's species, physiological differences between groups (in terms of carbon substrate utilization) were found [115]. These groups were subsequently shown to be phylogenetically distinct according to their 16S rDNA sequences [148]. This has been confirmed using several other genes and intergenic spacer regions: *gln*II, *nif*H, a 23S rDNA insertion sequence, the *nif*D/*nif*K IGS [90, 96, 141], and with combinations of markers: 16S rDNA and *gln*A sequences [48]; and *gyr*B, *nif*H and *gln*II [150].

Of particular interest, especially when studying the presence of *Frankia* in soil, are highly conserved markers absent or rarely found in other microbes (such as genes involved in nitrogen fixation or symbiosis) [85]. The *nif*H gene, for example, is conserved and has seen wide application in studies of *Frankia* diversity [67, 88, 149, 197], even though this gene is subject to horizontal gene transfer and phylogenies determined by it do not therefore, in some instances, reflect the actual evolutionary history and relatedness of strains [17, 197]. Similarly, domain III of the 23S rRNA gene contains a large actinomycete-specific insertion sequence (IS) not found in other prokaryotes [160]. This region is highly variable, with differences between strains being nearly as great as those found in ribosomal intergenic spacers [85, 96, 160]. Analysis of this region in *Frankia* generated a cluster structure which was consistent with the four host-infection group system proposed by Normand et al. [148], and also allowed strains from the *Alnus*-infective group to be further separated into, at first, four subgroups (II, IIIa, IIIb and IV) [96]. Later, three additional subgroups within IIIb were proposed [131, 132]. However, as group IIIb appeared to be rare, doubt was expressed as to whether these additional groups have much ecological significance [85]. The same author points out that the 23S rDNA IS might prove useful in discriminating strains from the *Elaeagnus*-infective cluster [85]. This IS has proven particularly useful in discriminating *Frankia* in field studies of both soil and nodules populations using in-situ hybridization, and a number of specific probes targeting different subgroups have been used [206, 207]. It has also been used to study the physiology of different *Frankia* lineages inoculated into soil [138]. Further studies are reviewed in Hahn et al. and Chaia et al. [44, 88].

Most recently, the order Frankiales was split into Frankiales, Geodermatophilales,

Acidothermales and Nakamurellales, each containing a single family [165]. As a result, *Frankia* is currently the sole genus within the Frankiaceae, the only family remaining in the order Frankiales. Sen's refinement of actinobacterial phylogeny was based on whole-genome approaches, which are considerably more robust than taxonomic methods relying on individual genes [165].

1.3.3 *Frankia* genospecies

Currently, bacterial species are delineated by whole-genome relatedness in combination with phenotypic tests [73, 173, 174, 189, 195]. A species (or genospecies) is defined as any group of strains that share 70% DNA-DNA homology with a type strain by whole-genome DNA-DNA hybridization (DDH), and in which the difference in melting temperature (ΔT_m) between homologous and heterologous hybrids is less than 5°C [189]. In actinorhizal research DDH is superior to 16S sequencing as the latter, while discriminating between HIGs, lacks the resolution to delineate strains at the species level [17, 65]. Other commonly used phylogenetic markers such as *nif*H, while having higher resolution than 16S, may be subject to lateral gene transfer and so may not reflect the actual relationships between strains [197, 205].

The application of DDH to *Frankia* research has been hampered by difficulties in obtaining sufficient gDNA from slow-growing isolates. All *Frankia* strains grow slowly (doubling times of 2 or more days for most strains) and produce comparatively little biomass in culture [113, 115]. Additionally, DDH requires that pure cultures of the strains in question be available, which is problematic in the case of *Frankia* as many strains have resisted all isolation attempts. *Frankia* isolation is notoriously difficult, and the majority of isolation attempts are unsuccessful [85]. Furthermore, differences in the size of the genomes being compared may affect DDH estimates [17], and *Frankia* have the largest differences in genome size for any prokaryotic genus described thus far, ranging from approximately five to ten Mbp [147]. Nevertheless, several studies have been conducted.

Chung Sun An et al. proposed two genogroups, one containing nine *Alnus*-HIG strains and the other containing two *Elaeagnus*-HIG strains. This study also

included eight strains which remained unassigned [5,6]. Fernandez assigned a large number of strains to three, five and one genospecies in the *Alnus*, *Elaeagnus* and *Casuarina* HIGs, respectively. Nine of the 43 strains used in the study remained unclassified [65]. Akimov et al. found nine genospecies amongst 28 strains, five within the *Alnus*- and four within the *Elaeagnus*-HIG [2].

Unfortunately, as these studies used almost completely different sets of strains, it is uncertain whether genospecies described in one study correspond to those described in the others. Fernandez and Akimov both used ArI3 and CpI1 and so both of their "genospecies 1", at least, are synonymous [2]. Bloom studied strains isolated from *Morella pensylvanica* and found at least three genospecies, but did not compare them to isolates obtained from other hosts [31]. Despite originating from the same host, the differences between some of Bloom's strains were greater than any described by An [6, 31]. Lumini later described a further three groups amongst *Elaeagnus*-infective strains [123].

With the development of new techniques in genomics the proposal of new bacterial species using methods other than DDH became acceptable, provided authors could demonstrate that results were congruent with those obtained from DDH [158, 173]. Bautista et al. assigned *Frankia* strains not previously analysed to nine previously-defined genospecies using amplified fragment length polymorphism (AFLP) and mutilocus sequence typing (MLST) [16, 17]. These techniques have the advantage of not needing genomic DNA from isolates, and can be applied directly to root nodules [16, 17].

It is thought that between twelve and twenty *Frankia* genospecies exist among currently available isolates, and that as many as 100 genospecies may exist in total [146]. While DDH suggests that these should be assigned *de facto* species status, there are no distinguishing phenotypic characteristics within these groups [17, 83, 201]. Consequently, only one species (*Frankia alni*) has been validly published to date, the type strain being CpI1 [83, 201].

1.4 Global distribution, taxonomy and symbiotic interactions within the Myricaceae

The family Myricaceae is contained within the Fagales, which is within the Eurosid I clade [7, 45, 170, 172, 177]. It has the widest geographical distribution of the actinorhizal families (Figure 1.2) and is represented on all continents except Australia and Antarctica, in a wide variety of climates ranging from subarctic zones to the tropics [23, 34, 46, 125]. While the family is comparatively small it has been the subject of several controversies concerning its taxonomy, the total number of species it contains, the first appearance of its pollen in the fossil record, and its geographical origin [92, 129].

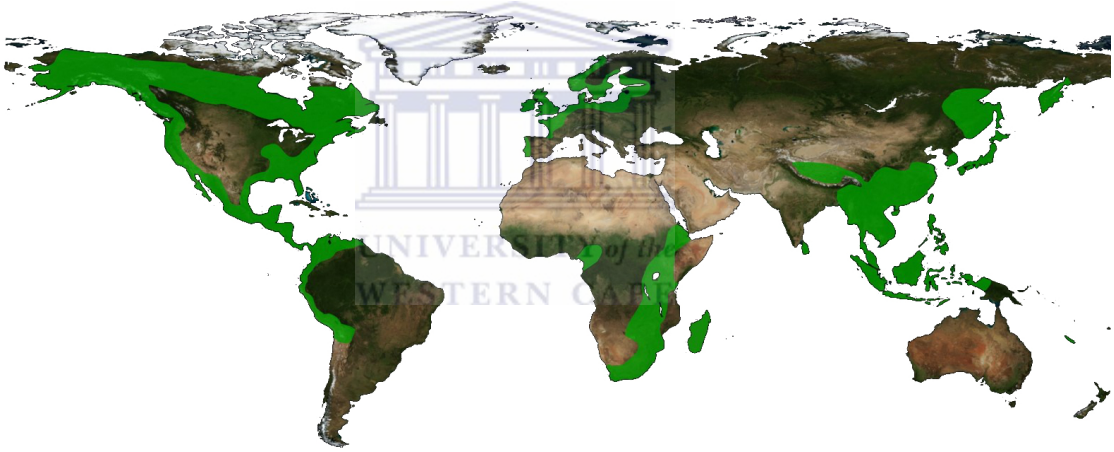


Figure 1.2: Global distribution of actinorhizal genera from the Myricaceae: *Myrica*, *Morella* and *Comptonia*. Distribution data retrieved from www.theplantlist.org and http://bgis.sanbi.org/. World map retrieved from www.nasa.gov.

Taxonomic difficulties within the Myricaceae are largely due to early distinctions between species being based entirely upon vegetative features from a limited number of specimens [46, 105]. Species described by Chevalier in 1901, often from single specimens, lack clearly discriminative morphological features when descriptions are applied to specimens in the field, as characteristics such as leaf morphology may vary with age and habitat [46, 105]. Consequently, the only genus commonly agreed upon by most authors is *Comptonia*[8], which is monotypic [46, 63, 98, 202].

[8] *Comptonia peregrina* was previously classified as *Myrica aspenifolia* [34]

Several studies have followed on from Chevalier's, and prior to the inclusion of *Canacomyrica*, either two [63] or three genera [46, 202] have been recognized within the Myricaceae [93, 98]. These depended on the rank assigned to *Myrica gale*, and whether *Morella* was considered a genus or awarded a subordinate rank [202]. Elias recognized *Comptonia* and *Myrica* (within which *Morella* was assigned subgenus status) containing one and seven species, respectively [63]. Wilbur recognized the genera *Comptonia*, *Myrica* and *Morella*, which correspond to Chevalier's *Comptonia*, *Gale* and *Myrica* [46, 202]. He also maintained Chavalier's subgroups and placed the Asian species in subgenus *Morella* and the African and American species, including *M. faya* and *M. californica*, in *Cerothamnus* (Tidestr.) [202]. An alternative structure to the family was proposed by Verdcourt and Polhill, who proposed the genera *Comptonia*, *Gale* and *Myrica* [193], but this proposal was rejected following the decision of the Committee for Spermatophyta in 1998. *Comptonia* was recognized, and *Myrica* and *Morella* were retained for temperate and tropical species, respectively [37, 106]. These assignments were on the basis of vegetative characteristics and are reviewed by Wilbur [202]. The enigmatic New Caledonian species *Canacomyrica*, considered a relic ancestor of the family, was later included [93]. The existence of three genera (four when *Canacomyrica* is included) as suggested by morphological data is supported by karyology and phylogenetics [92]. Myricaceae have a base count of 8 chromosomes, with numbers for currently recognized genera varying: *Morella* is diploid ($2n = 16$), *Comptonia* is tetraploid ($2n = 32$), *Myrica gale* is hexaploid or dodecaploid ($2n = 48, 92$), depending on the geographical origin of the specimens (North America vs Europe, respectively) according to Herbert et al. [93], Macdonald [126] and references contained therein [3, 121, 122, 151, 175]. The fourth genus, *Canacomyrica*, is diploid ($2n = 16$) [93].

Due to the overall uniformity and sometimes-overlapping morphologies displayed by specimens in the field, especially in leaf shape, the number and identity of species within individual genera that are recognized by different authors varies. For example, White (1993) and Killick (1998) differ greatly in the number of African species they recognize [106, 200]. White solved the problem of overlapping phenotypes by synonymizing several African *Morella* (such as *M. humilis*, *M. kraussiana* and *M. diversifolia*, which is possibly a hybrid of *M. kraussiana* and

M. quercifolia), arriving at a total of only six species [92, 200]. Killick rejected this proposal and maintained thirteen African species, ten of which occur within South Africa (Table 1.2) [105, 106]. In North America *M. heterophylla*'s existence as an independent species has been questioned [75, 202]. Taxonomic revision of the Myricaceae is ongoing, and for the purposes of the current study the conventions of Wilbur (1994) and Killick (1998) are followed [93, 98, 106, 202]. All African species are thus considered to be *Morella*.

While no comprehensive, family-wide molecular phylogeny has been generated the relationships between 13-species from *Myrica*, *Comptonia*, *Morella* have been established using the *rbc*L gene and 18S-23S ITS sequences [98]. This phylogeny included a South African species (*M. quercifolia*). Huguet et al. proposed *M. gale* be split into two new species based on an analysis of *rbc*L and ITS sequence data from North American and European populations which, as stated above, were already known to have different chromosome counts [98]. The same study also identified three possible subgroups within *Morella*, one of which would contain all African species and agrees with Chevalier's subsection *Africanae* [46, 98].

1.4.1 Symbiotic associations within the Myricaceae

Of the four genera within the Myricaceae, all members of the *Myrica*, *Comptonia*, *Morella* are nodulated [92,98]. Nodules have not been documented on the monotypic genus *Canacomyrica* in its natural habitat[9] [142]. Despite claims that this species is not actinorhizal, this has never been conclusively demonstrated [34, 72, 93].

Host-microsymbiont associations in actinorhizal symbioses may be determined in the field, with nodule populations identified using molecular techniques [47, 142, 196]. The effectiveness of collected nodules may also be determined directly in the field as had been done by Grobbelaar and Van Ryssen in 1967 [188], but studies combining both approaches have not been reported. Alternatively, cross-inoculation experiments may be performed in a greenhouse [24]. In such experiments actinorhizal host species are inoculated with either nodule suspensions prepared

[9]In two studies in which *Canacomyrica* nodulation trials were attempted seedlings grown under greenhouse conditions either did not germinate, or died before trials could be performed [34, 92]

from different actinorhizal hosts, or with pure cultures of *Frankia* isolates of varied provenance [13, 184]. The effectiveness (meaning the ability of the established nodule to fix atmospheric N_2) of any induced nodules is then monitored under controlled conditions. Such studies determine whether strains are (1) infective, (2) nitrogen-fixing in symbiosis, and also (3) how efficient the established symbiosis is in comparison with other strains when inoculated on a particular host [13, 24, 184].

Inoculation with nodule suspensions introduces three recognized difficulties: no matter how thorough nodule surface sterilization is, suspensions may contain non-*Frankia* contaminants that affect experimental findings, single nodules may contain more than one *Frankia* strain, and nodule homogenates may contain inhibitory plant metabolites toxic to the microsymbiont [31, 184, 191]. A fourth potential problem has recently come to light with the discovery that defensin protein Ag5, which is present in *Alnus* nodules, modifies *Frankia* membrane porosity [42]. For these reasons inoculation with pure cultures is preferred.

Cross-inoculation studies of Myricaceae using *Frankia* isolates

In experiments using pure strains as inoculum most host genera tested thus far form nodules with *Frankia* from a single host-specificity group only [13, 184]. Dwight Baker performed an extensive cross-inoculation trial using 50 *Frankia* isolates and six host species, finding that strains could be divided into four host-specificity groups: (1) strains nodulating *Alnus* (Betulaceae) and *Myrica*, (2) *Casuarina* and *Myrica*, (3) *Elaeagnus* and *Myria*, or (4) Elaeagnaceae only (*Elaeagnus*, *Hippophae* and *Shepherdia*) [13]. Normand et al. later found the *Alnus* and *Casuarina*-infective strains clustered together according to the 16S rDNA sequences, forming *Frankia* Cluster I, while the *Elaeagnus*-infective strains grouped within *Frankia* Cluster III [148].

Notably the only member of the Myricaceae included in Baker's study, *Myrica gale*, nodulated with strains from three of his four groups and was the only host species to nodulate with such a wide range of highly divergent microsymbionts [13]. Later studies with *Myrica gale* and *Morella cerifera* showed that both could nodulate effectively with every strain with which they were inoculated [62, 157]. When

M. pensylvanica was inoculated with a variety of strains[10] it was always nodulated and these nodules were effective [30]. As a result it was concluded that species from the Myricaceae, alone among actinorhizal families, were promiscuous hosts[11] and established symbioses with *Frankia* from all host infection groups [98].

The most recent cross-inoculation study involving the Myricaceae tested the compatibility of six *Frankia* strains against six *Morella* species, *Myrica gale* and *Comptonia peregrina*, and found that all of the *Morella* were promiscuous, nodulating with every strain tested [98]. *Myrica gale* was only nodulated by Cluster I strains, a finding at odds with previous cross-inoculation experiments but which agreed with what has been observed in natural populations of this species [184]. *Comptonia peregrina* also demonstrated promiscuity, but to a lesser degree than the *Morella* [98].

Myricaceae/*Frankia* associations in natural habitats

Field studies, which reveal the natural associations between host species and *Frankia*, have only been conducted on six *Morella* to date [47, 90, 97, 99, 100, 208]. These studies have shown that while *M. pensylvanica* is nodulated by different HIGs in its natural habitat, *Comptonia peregrina* and *M. gale* were nodulated by Cluster I strains only [47, 97, 100]. Later *M. rubra*, like *M. pensylvanica*, was also found to be nodulated by both Cluster I and Cluster III strains in natural settings [90]. A similar finding was reported for *M. californica* [99]. On the other hand *M. faya*[12], *M. rivas-martinezii* and *M. cerifera* are only nodulated by strains

[10] These were strains isolated from *M. pensylvanica* itself but were known to be highly divergent [30].

[11] It has been suggested that the Myricaceae is the most ancient of the actinorhizal families, and that this accounts for the phenomenon of promiscuity [129]. Eudicot pollen appears in the fossil record approximately 125 MYA or later [51, 128] and, while contentious, the oldest evidence for the existence of the Myricaceae seems to have appeared soon thereafter (94 MYA) [129]. It is thought that symbiotic *Frankia* also arose at about this time, and were infectious on an older lineage of actinorhizal plants; later establishing symbioses with the more recently-derived lineages which evolved towards higher specificity [22, 129].

[12] Interestingly, *Morella faya* is found readily nodulated in Hawaii, where it grows as an aggressive invader species in the absence of any known endemic actinorhizal plants [124]. Complicating the matter is the finding that uncultured nodule populations in Hawaii share no genotypes with those in the Canary islands (to which the *M. faya* is native) [99]. Several explanations have been put forward, including the existence of indigenous saprophytic *Frankia* populations in Hawaii, introduction with *M. faya* itself, or introduction by early settlers in Polynesian prehistory [124].

from Cluster III [99, 208].

In combination with cross-inoculation studies these findings demonstrate that members of the genus *Morella* are potentially promiscuous, but that the specificity of their symbiotic associations may be restricted in their natural habitats, at least in the populations studied [98]. There are indications that nodulating strains correlate with prevalent soil conditions [142], but whether this is due to absence of particular *Frankia* strains from the local environment, or to competition between (or host-dependent selection for) particular strains under defined soil conditions remains an open question.

1.4.2 South African *Morella*

The largest and most widespread of the actinorhizal genera within Myricaceae is *Morella*, with species indigenous to Africa, Asia and North and South America and possibly Europe [13], and with 33 currently accepted species [1,152]. The majority of these (19) are African endemics[14], and are concentrated (13 species) in the south of the continent [81]. Only those native to the Republic of South Africa are discussed here.

Despite the controversies surrounding the rest of the family, the *Morella* of South Africa are well described [77, 92]. In his 1969 revision of South African *Myrica*, Killick reduced the fifteen species of African Myricaceae (Chevalier's 13 species and two others (*M. mossii* and *M. rogersii*)), to nine [105]. An additional species (*M. microbracteata*) was later added [106]. There are thus ten currently recognized South African species, of which seven occur naturally within the Western Cape province (Table 1.2). Three of these species (*M. diversifolia*, *M. integra* and *M. kraussiana*) are endemic to the Cape flora, which is contained entirely within the province [46, 78, 105, 106].

A further explanation is that it was introduced in soil ballast, which was often deposited at ports during global oceanic trade during the 18th to 20th centuries [199].

[13]There is some question as whether *Morella faya* is indigenous to Portugal [124]

[14]Symbiosis involving *Alnus glutinosa*, *Coriaria myrtifolia* and the introduced *Elaeagnus angustifolia* have been investigated in North Africa. Most of these species are native to Africa, Asia and Europe and are thus indigenous, but not endemic. The *Morella* of southern and central Africa have the distinction of being the only known actinorhizal species endemic to the continent.

1.4. Global distribution, taxonomy and symbiotic interactions within the Myricaceae

South African *Morella* display significant variation in morphology, especially in terms of leaf shape, which is unusual given the morphological uniformity that characterizes the rest of the genus (Figure 1.3) [93,105]. This, along with ecological and occurrence data, helps with identification in the field as some species are very distinctive and unlikely to be misidentified [92].

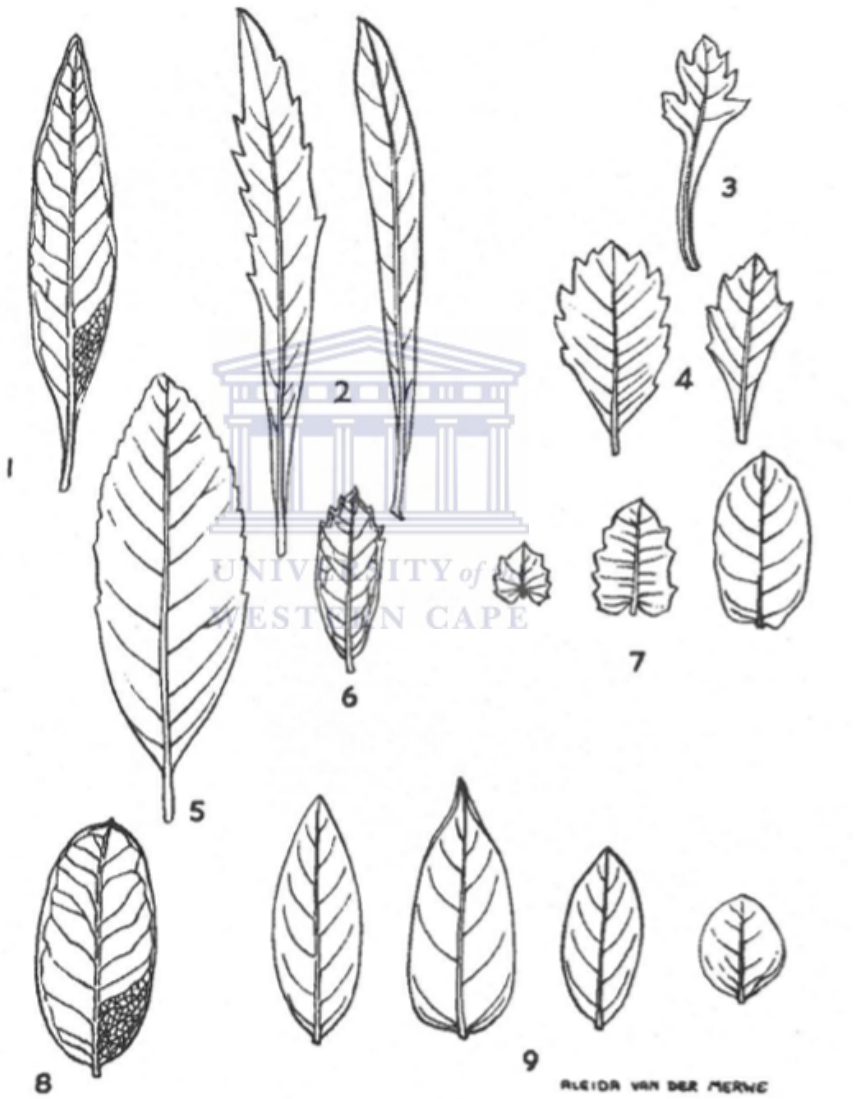
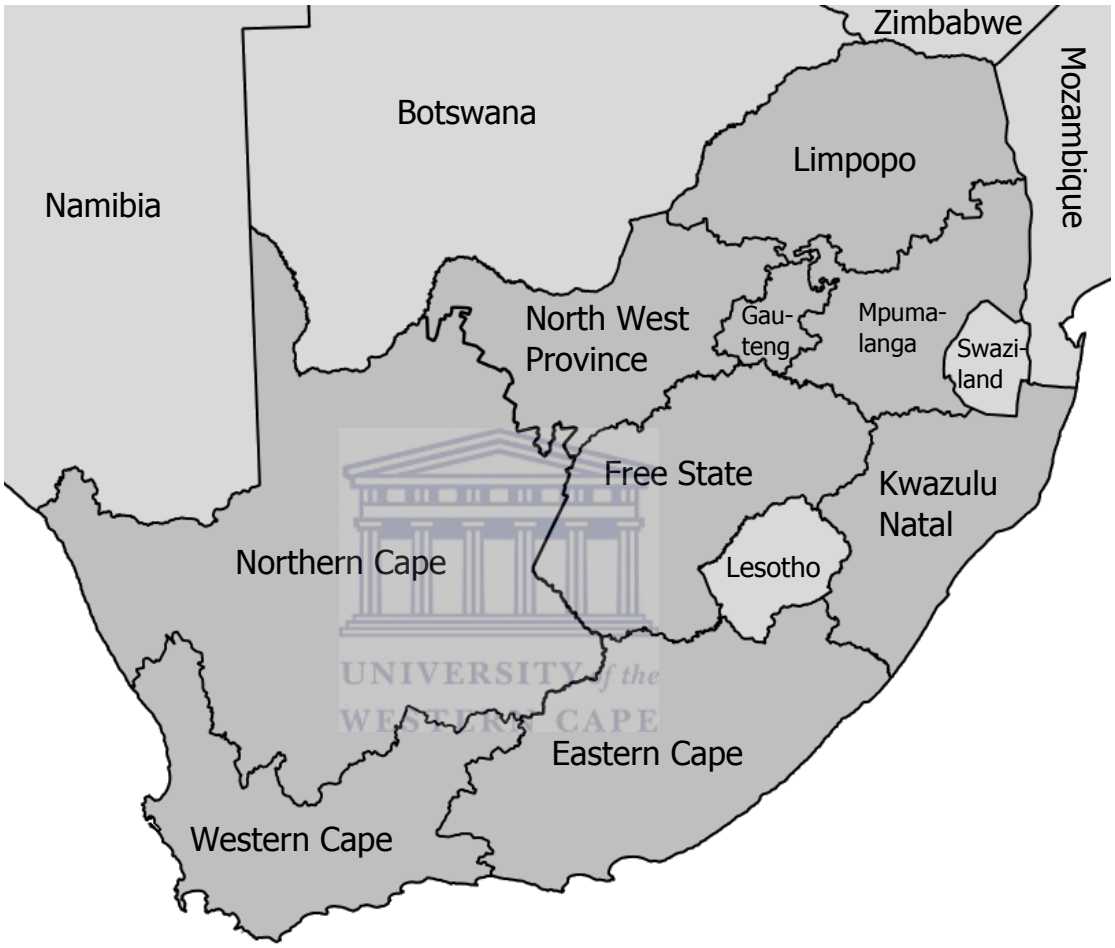


Figure 1.3: Illustrations of typical South African *Morella* leaf morphologies, adapted from Killick [105]. 1: *M. integra*, 2: *M. serrata*, 3: *M. quercifolia*, 4: *M. diversifolia*, 5: *M. pilulifera*, 6: *M. brevifolia*, 7: *M. cordifolia*, 8: *M. kraussiana*, 9: *M. humilis*. Absent is *M. microbracteata*, an endangered species indigenous to Kwazulu Natal.

Table 1.2: Distribution of ten indigenous *Morella* species within the provinces of South Africa. Provinces are shaded dark gray, species distributions are tabulated below.



Species	Distribution	Habitat	Habit
M. brevifolia	Kwazulu Natal	Montane	Shrub to 1 m
M. cordifolia[a]	Western to Eastern Cape	Coastal dunes	Prostrate shrub to 1 m
M. diversifolia[b]	Western Cape, Cape peninsula only	Montane	Shrub to 1 m
M. humilis[a]	Western and Eastern Cape	Montane	Shrub to 1 m
M. integra[b]	Western Cape	Riparian	Tree or shrub to 3 m
M. kraussiana[b]	Western Cape	Montane	Low shrub
M. microbracteata	Kwazulu Natal	Montane	Low shrub
M. pilulifera	Eastern Cape to Zimbabwe	Riparian	Tree to 12 m
M. quercifolia[a]	Western to Eastern Cape	Coastal lowlands	Shrub to 60 cm
M. serrata[a]	Western Cape. South Africa to Tanzania	Riparian	Tree or shrub to 6 m

[a] Indigenous to the Cape Flora of the Western Cape province of South Africa
[b] Endemic to the Cape Flora of the Western Cape province of South Africa

1.4.3 *Morella* and the Cape Flora of Southern Africa

The Western Cape province of South Africa is extremely rich in terms of plant biodiversity, especially for an area of temperate climate, and contains in its entirety the Cape Floristic Region (CFR) or Cape flora [76–78]. At 90,000 km^2, this region covers just 4% of the area of the southern African subcontinent, but houses more than 40% of the region's plant species (approximately 9000 of an estimated 20,500 species). Additionally, almost 70% of these species are endemic [78]. A combination of factors has driven plant diversification and endemism in the Cape flora [78, 110, 117, 162]. Historical factors, such as change of climate from tropical to temperate and sea-level fluctuations 5 million years ago, combined with the unusually complex topography, climate and geology of the region gave rise to a large number of habitats. The close proximity of starkly-contrasting habitats in turn created steep environmental gradients which, when plants lack effective seed dispersal mechanisms[15], limits their range [78, 110, 117].

The poor nutrient status of Cape soils in particular has given rise to an abundance of plants with specialized mechanisms for nutrient acquisition and sequestration. The region's substrata are predominantly sedimentary sandstones, shale and limestone, and most soils are acidic, aluminum-rich and oligotrophic, lacking phosphorus and nitrogen. Many are also poor in organic matter [77, 159, 203]. Adaptations to these conditions include proteoid root systems, insectivory[16], mycorrhizal associations, and various nitrogen-fixing associations [4, 78, 109, 111, 133].

In the Cape, nitrogen-fixing symbioses are abundant and are accounted for mostly by legumes: Fabaceae is the second largest family in the region after the Asteraceae, and is represented by approximately 760 species [78]. While this number seems high it is not unusually so, as the family is also well-represented in other regions with dry climates [76, 78]. Also present are four species of cycads (Zamiaceae) and seven species of *Morella* (Myricaceae), the only actinorhizal genus containing species endemic to Africa [78, 81]. Myricaceae were already present in the southwestern

[15] I observed interactions between fruit-bearing *M. diversifolia* and what appeared to be indigenous harvester ants, which were suggestive of myrmechory (a fairly common seed dispersal mechanism in fire-prone Cape ecosystems). Gulls are known to disperse *M. cordifolia* seeds, which are far larger than those of other Cape species.

[16] 20 of the 130 known sundew (*Drosera*) species are native to the Cape flora [192].

Cape region of South Africa by the Miocene (21-5 MYA) where they shared habitat with indigenous *Casuarina*, which became locally extinct following severe climate change during the Pleistocene [49, 50, 76, 117]. They appear to have subsequently undergone comparatively rapid diversification, in terms of both ecology and morphology, when compared to other members of the genus [93]. In addition to their association with *Frankia*, *Morella* have proteoid roots, an advantageous adaptation to phosphorus-limitation in soils of the Cape [59,120,194].

As with all Myricaceae, Cape species (with the possible exception of *M. cordifolia*) require an abundant supply of water, and their varied habitats reflect this [78]. Among the Cape varieties, *Morella quercifolia* may be found in coastal wetlands and mountain slopes, while *M. kraussiana* and *M. humilis* are found on mountains throughout the region where they draw moisture from near-constant precipitation during the region's dry summers. *M. diversifolia* is also a montane species, but is restricted entirely to the mountains of the Cape peninsula [91]. *Morella integra* and *M. serrata* are small trees and may be found in riparian habitats. In *M. integra* nodule-bearing roots are often completely submerged (author's personal observations), in which case the nodules possess negatively-geotrophic roots similar to those reported in *Myrica gale* [167, 182]. It is probable that the same may be found in *M. serrata*. Finally, *M. cordifolia* colonizes coastal sand dunes in much the same manner as the North American *M. cerifera* [92]. The ranges of some of these plants overlap, and several species may co-occur within a small area[17].

The effects of the region's geological history on diversification can be seen in the emergence and current distribution of *Morella diversifolia*, which is restricted to the mountains of the Cape peninsula [81, 91]. The existence of this species may in part be attributed to these mountains being cut off from the mainland $2-3$ MYA [117]. The effect of the various factors driving speciation in the Cape may also be seen in Cape *Morella* leaf morphology (Figure 1.3), variation in which is also extensive in many other Cape families [118, 144]. Of the Cape species, *M. cordifolia* and *M. quercifolia* are the most distinctive and easily identified in the field, although there may be significant phenotypic variation between isolated local populations

[17] An extreme case of this was encountered in the Cape Peninsula National Park, where *M. cordifolia*, *M. kraussiana*, *M. quercifolia* and *M. diversifolia* were found within 500 m of each other

(personal observations). Both species have leaves that differ significantly from the elliptical-oblanceolate shape typical of the Myriaceae. The remainder of the Cape species are more difficult to discriminate, as leaf shapes intergrade almost completely (personal observations). *M integra* for example infrequently has a small proportion of serrated leaves more reminiscent of *M. serrata*. Genetic variation between species from the Cape flora, however, is surprisingly low, and the Cape *Morella* are considered the most recently derived [92].

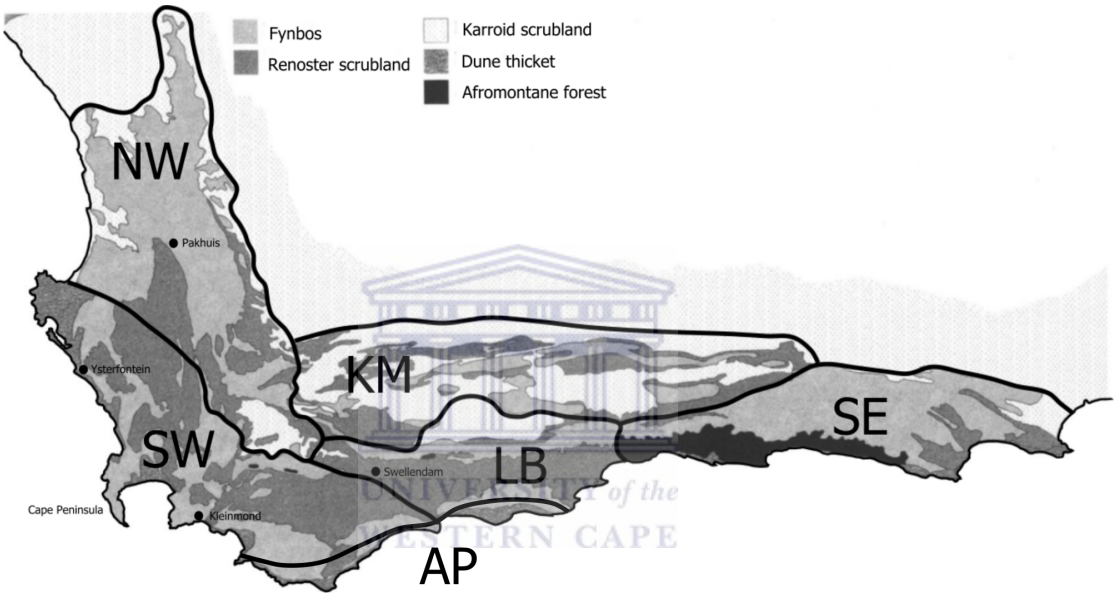


Figure 1.4: Regions containing flora with distinct characteristics within the Western Cape province of South Africa. Phytogeographic regions as follows: SW: South West, NW: North West, KM: Kleinmond Mountains, AP: Agulhas plain, LB: Langeberg, SE: South East. Modified from Killick [78, 105].

The following list of descriptions of Cape *Morella* is reproduced nearly verbatim from "Cape Plants. A conspectus of the Cape flora of South Africa" [78]. Regions referred to are presented in Figure 1.4.

M. cordifolia . . . . (L.) Killick (= *Myrica cordifolia* L.) Dioecious shrub to 3 meters. Leave imbricate, sessile, broadly ovate-cordate, toothed, margins revolute, gland-dotted. Flowers in axillary spikes. Fruits warty, c. $5-8$ mm diameter. May – Aug. Coastal sands and limestone, AP, SW, LB, SE (Yzerfontein to E Cape).

M. diversifolia . . . (Adamson) Killick (= *Myrica diversifolia* Adamson) Dioecious shrub to 1 meter. Leaves obovate, tapered below, more or less toothed, margins mostly revolute, gland-dotted. Flowers in axillary spikes. Fruits warty, c. 3 mm diameter. Aug. – Sept. Sandstone slopes, SW (Cape peninsula).

M. humilis . . . . . (Cham. ex Schltdl.) Killick (= *Myrica humilis* Cham. ex Schltdl.) Like ***M. kraussiana*** but spikes not robust, with smaller bracts. July – Nov. sandstone slopes, KM, LB, SE (Swellendam and Warmwaterberg to E Cape).

M. integra . . . . . (A. Chev.) Killick (= *Myrica integra* (A. Chev.) Killick) Dioecious shrub or tree to 3 meters. Leaves narrowly elliptic, attenuate below, sometimes toothed. Flowers in axillary spikes. Fruits warty, c. 3 mm diameter. Sept. – Apr. Rocky stream sides, NW, SW (Pakhuis to Kleinmond Mts).

M. kraussiana . . . (Buchinger ex Meisn.) Killick (= *Myrica krausianna* Buchinger ex Meisn.) Dioecious shrub to 1 meter. Leaves elliptic, rounded below, sometimes toothed above, margins revolute, gland-dotted. Flowers in robust axillary spikes with large, imbricate bracts. Fruits warty, $2-3$ mm diameter. Aug. – Oct. Sandstone slopes, AP, SW, LB (Cape peninsula to Swellendam).

M. quercifolia . . . (L.) Killick (= *Myrica quercifolia* L.) Dioecious spreading shrub to 60 cm. Leaves obovate, attenuate below, usually pinnatifid, gland-dotted. Flowers in axillary spikes. Fruits warty, $3-4$ mm diameter. July – Sept. Mostly coastal sandy and limestone flats and slopes., AP, NW, SW, LB (Namaqualand to E Cape).

M. serrata . . . . . (Lam.) Killick (= *Myrica serrata* Lam.) Like ***M. integra*** but leaves mostly toothed, conspicuously reticulate-veined above, usually gland-dotted beneath, margins more or less revolute. Aug. –Dec. Rocky stream sides, NW, SW, KM, LB, SE (Bainskloof to Mpumalanga and Caprivi).

1.4.4 Studies of South African *Morella* prior to 2011

Few studies have been undertaken specifically involving the *Morella* of southern Africa. Many species were included in the 1976 International Biological Program survey of nodule formation in non-leguminous angiosperms, which considerably expanded the number of known actinorhizal plants [34]. In that report the majority of newly-added species from the Myricaceae were African, mostly from South Africa [34]. Investigators contributing to this survey had demonstrated that all South African *Morella* fixed nitrogen in the field, and observed that the microsymbiont appeared typical of those infective on other hosts from the Myricaceae [33,188]. At the time little was known about the microsymbionts of non-legumes, and George Bond and others were interested in establishing the degree of diversity that existed amongst them. Several cross-inoculation studies from this time were carried out using *M. cordifolia* and *M. pilulifera* and inoculating with either crushed nodules or habitat soil (Table 1.3) [33, 58, 70, 127, 135, 191].

While varying degrees of compatibility among the various microsymbionts were found between African, European and North American species, no simple pattern

emerged. *M. cordifolia* became nodulated when inoculated from *M. gale*, *M. cerifera* and *M. pilulifera* but in all cases the nodules were small and ineffective, suggesting incompatibility with the symbiont [127]. In another study when *M. cordifolia* (and *M. cerifera*, a North American species) were inoculated with crushed *Myrica gale* nodules, neither benefited from the inoculation and very few nodules were produced [70]. Similarly, when *M. faya* was inoculated from *M. cordifolia* and *M. pilulifera* the resulting symbiosis was initially graded as poorly effective [127], although a later study of *M. faya* inoculated with *M. cordifolia* nodule homogenates contradicted this [135]. While the symbiosis was considered effective, the *M. cordifolia* endosymbiont did not perform as well as either *M. faya*'s natural symbiont or one from *M. cerifera*. Furthermore, *M. pilulifera*-induced nodulation, while effective, was slow to develop on *M. faya* and so it became apparent that degrees of compatibility existed [135]. Field-soil collected from under *M. cordifolia* was, however, able to produce highly effective nodules on *M. faya*, and in microscopic examination these were indistinguishable from those induced by *M. faya*'s normal microsymbiont [135]. *Myrica gale* did not form nodules when inoculated with either of the South African species, and while *M. cordfolia* nodulated rapidly when inoculated from *M. gale*, the nodules formed were ineffective [127]. Interestingly, *M. cordifolia* could only form truly effective nodules when inoculated with material originating from its own environment.

Bond concluded that geographically distinct *Myrica* have incompatible microsymbionts, particularly as the African genera failed to induce nodulation on *M. gale*. He also suggested that these distinctions could be explained by the taxonomic separateness (according to Chevalier [46]) between *M. gale* and the rest of his tested species, and later suggested that a degree of compatibility existed between Myricaceae species in terms of their microsymbionts, with the exclusion of *M. gale* [135]. The outcomes of these cross-inoculation trials are presented in Table 1.3, but it should be kept in mind that studies using nodule homogenates instead of pure cultures as inoculum should be interpreted cautiously [13, 24, 36, 108].

In the years immediately following these studies the subject of native microsymbiont diversity in southern African actinorhizal hosts was not pursued further, and until the current study the diversity of *Frankia* nodulating African *Morella* has remained

unexplored, and African strains nodulating species endemic to the continent have never obtained in pure culture.

Table 1.3: Cross-inoculation studies involving *Morella* species from Southern Africa between 1966 and 1976, adapted from Vandenbosch and Torrey [191].

	Inoculum[b]				
Host species	*M. gale* nodule	*M. cerifera* soil	*M. cordifolia*[c] soil	*M. pilulifera*[c] nodule	*M. faya* nodule
M. gale	E[a][127]	E [127]	N [127]	N [127]	-
M. cerifera	I [70]	-	-	E [127]	-
M. cordifolia	I [70]	I [127]	E [33, 127]	I [127]	-
M. faya	I [135]	E [135]	E [135]	-	E [135]
M. rubra	I [127]	I [127]	I [127]	I [127]	-
M. pilulifera	-	-	-	E [33]	-

[a] E, effective symbiosis; I, ineffective nodules; N, no nodules formed.
[b] Inoculation with either nodules homogenates or habitat soil from indicated species.
[c] South African species.

1.5 *Frankia* isolation and culture

To date isolates have been obtained from seven of the eight actinorhizal host families. Species include *Casuarina* [57, 84, 130, 161], Alnus [12, 21, 28], *Myrica gale* [9, 191], *Morella* [30, 31, 114], *Coriaria* [139] and *Gymnostoma* [157]. Until recently and in spite of repeated attempts, Cluster II strains nodulating Rosaceae, Coriariaceae, Datiscaceae and Ceanothus, while detected in soil, had proven impossible to cultivate *ex planta* [82]. An isolate from this group, the last of the host infection groups lacking pure-culture isolates, has now been recovered from *Coriaria myrtifolia* [82].

Despite these successes isolation of *Frankia* remains challenging; no universally-applicable isolation media has been reported and only a small number of attempts succeed [14, 116]. Most attempts are made using root nodules [116]. Only one study reports isolation of strains directly from soils [10].

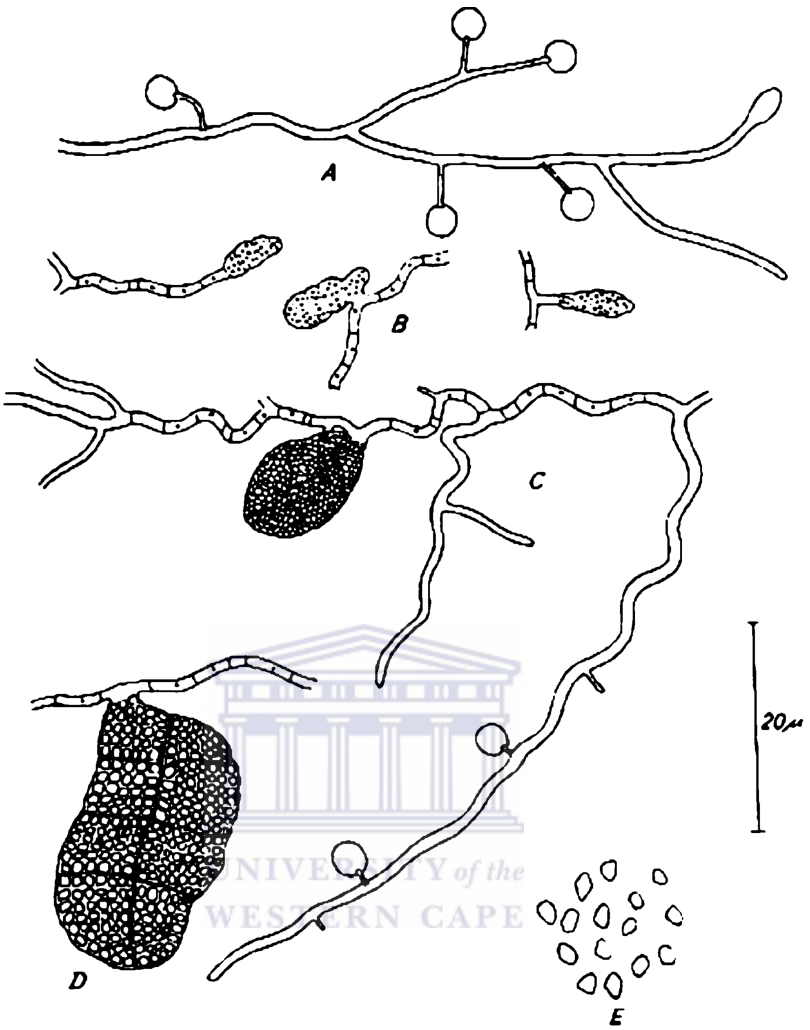


Figure 1.5: Illustration by Pommer showing morphological characteristics of his isolate [155]. *A:* hyphae with vesicles; *B:* thinner hyphae developing bubble-shaped structures (early sporangia development); *C:* hyphae divided by septa; *D:* sporangia increase in size; *E:* spores are released from sporangia. While Pommer never used the terms directly, it is clear that he observed sporangia, diazovesicles and spores typical of Cluster I *Frankia* [147].

1.5.1 Isolation procedures

Callaham isolated *Frankia* from the root nodules of *Comptonia peregrina* using mercuric chloride for surface-sterilization, enzymatic digestion of nodule tissue and a complex medium [40]. Following this success, methods for isolation and distribution of strains were quickly developed [9, 10, 28, 39]. Dozens of strains have subsequently become available, although the majority remain unreported [85]. Procedures for isolating and maintaining *Frankia* in culture have been reviewed by Lechevalier and Lechevalier [116], and are discussed here in brief.

Frankia is usually isolated from root nodules, which contain enriched populations of (usually) only one strain [24]. These may either be collected in the field, or from trap plants inoculated in a greenhouse with environmental soil samples [80]. While it is easier to surface-sterilize and process greenhouse-grown nodules, *Frankia* strains obtained in this manner lack the ecological context of those isolated directly from field-collected actinorhizal plants in their natural habitats [44, 116].

Field-collected nodules should be transported to the laboratory as rapidly as is practical, and stored in a cool environment until processed. Fresh nodules are cleaned of adherent soil and surface contaminants by rinsing or sonicating in sterile water. Surface sterilization is then performed using phenol and hydrogen peroxide or osmium tetroxide. Following sterilization nodules are thoroughly rinsed in sterile distilled water, after which a second sterilization step may be performed. At this point contaminating organisms may still be present, particularly in the nodule epiderm.

Methods are of two basic types: either *Frankia* cells are separated from homogenized plant material using techniques such as modified sucrose-fractionation method [10], differential centrifugation [21] or Sephadex fractionation [9,28] and used as inoculum, or nodules undergo microdissection whereafter *Frankia* outgrowth from segment surfaces is monitored [9, 10, 28]. Where nodule pieces are used they may be either embedded in solid media or incubated in submerged liquid media. In either case, outgrowth from the nodule surface becomes apparent in anything from five days to eight weeks [116]. Such outgrowths are then removed, homogenized, and inoculated into new media. Several rounds of homogenization and sub-culture normally follow.

Because *Frankia* vary widely in terms of nutritional requirements, growth rates and tolerance for oxygen (factors which cannot be known before a new strain has been isolated). The use of a variety of media and growth conditions is therefore advised. Some strains may require growth factors, but these requirements are not universal [39, 116, 166]. Most strains are insensitive to antifungals (particularly kanamycin), nalidixic acid and sodium azide, and so these compounds may be freely incorporated into isolation media, at least initially. Media deficient in fixed nitrogen are generally used, although nitrogen sources may be added once a strain has been obtained in pure culture [116]. Not all isolation attempts are successful [116].

Shin-ichi Suzuki used a solid medium base on gellan gum instead of agar for selective isolation of actinomyces, principally as he found it to stimulate the formation of aerial mycelium and spores in a variety of actinomycetes [176]. A medium using this solidifying agent was subsequently used to culture *Frankia* CcI3 to good effect [15]. For *Frankia* gellan gum based media offer several advantages: first, lower gel strength allows for easier penetration of the matrix by growing hyphae, and accelerated growth compared to agar-based media [15]. Secondly, as gellan gum is colorless, microscopic observation of growing cultures is possible [15]. Gellan gum has been suggested for use in isolation trials [44]. Chaia also predicted that the increased availability of *Frankia* genomes would allow media to be developed specifically for isolating *Frankia* from soil. Genomic information has since been useful in developing isolation media for Cluster II strains, and it will be interesting to see how this work will be expanded upon [82].

Strains isolated thus far show considerable variety and, while a large body of work still remains to be done, have allowed for description of the genus [201]. It is now apparent that isolation from some host species is more difficult than from others. Some strains are also harder than others to obtain in pure culture [116]. Limiting steps in isolation are *Frankia*'s slow growth rate (with doubling times of 14 hours to several days), frequent contamination by faster growing organisms, and the varying growth requirements of different strains. The most critical factors to success are therefore sterilization of the nodule surface to prevent contamination by faster growing organisms, and selection of the isolation media (reviewed in Benson and Sylvester [24]).

Non-*Frankia* actinomycetes have also been isolated from surface-sterilized nodules, some of which are nitrogen-fixing [74, 186]. *Micromonospora* is commonly found in legume nodules, and also common in actinorhiza [41, 43].

1.5.2 *Frankia* morphology in culture

Following isolation endophytes may be putatively identified according to *Frankia*'s morphological characteristics. All strains are slow growing, with doubling times of 2 to 5 days, and all produce little biomass in liquid culture [115]. Hyphae are septate, with diameters ranging from 0.5 to 2.0 μm and may be sparsely to extensively branched [201]. Among actinomycetes, *Frankia* alone will form sporangia in submerged liquid culture [116]. These may be positioned at hyphal tips or in an intercallary position and contain non-motile sporangiospores [24]. Furthermore, effective strains form "vesicles", a distinguishing characteristic unique to the genus, which contain the nitrogenase complex and supporting bio-machinery, and are the site of nitrogen fixation in both culture and symbiosis in most strains and under most conditions [113, 201], predominantly when grown under nitrogen limiting conditions. Vesicles (alternatively "diazovesicles") are roughly spherical with diameters of between 2 and 6 μm [201], and possess laminated lipid envelopes which protect the oxygen-labile nitrogenase from oxidative inactivation [153]. Some strains may also form vesicles routinely, but these are inactive under conditions of fixed nitrogen availability [85, 115]. Vesicles may be positioned either terminally or laterally, in the latter case they are attached to hyphae with a short stalk [24].

In addition to nitrogen-fixing strains, root nodule formation by non-nitrogen-fixing ("atypical") *Frankia* has been observed, and a number are available in pure culture (discussed by D. Hahn [85]). These do not produce vesicles in either culture or symbiosis, will not grow under nitrogen-limiting conditions, and are frequently unable to reinfect their original hosts during forced-nodulation experiments. There is some indication that they may be opportunistic, relying on infective/effective strains for penetration into a host's root [116]. There are also indications that under some conditions ineffective strains may be suppressed by hosts [89].

1.6 Problem statement and methods used

The diversity of *Frankia* associated with endemic African actinorhizal plants, specifically *Morella*, has never been explored. To address this knowledge gap the author designed the current study to investigate *Frankia* diversity in the root nodules of *Morella* species within their natural habitats (Chapter 2) and located in parks and reserves in the Western Cape province of South Africa. These protected areas were selected in order to minimize the risk of detecting foreign *Frankia*, potentially introduced in ecosystems impacted by human activity. As *Frankia* occupy two niches (host-plant nodules as well as soil, from which they are recruited), a study was conducted to determine the diversity of frankiae in rhizosphere soils at six sampling locations, for comparison with local nodule populations (Chapter 3). Both of these studies described *frankia* diversity using *nifH* gene fragments. This is because the nitrogenase reductase gene is required for functional symbiosis, and *Frankia* sequences are distinctive when compared to *nifH* from other organisms. Furthermore, a large *nifH* sequence database exists, which allows South African sequences to be compared to *Frankia* sequences sourced elsewhere. Finally, *Frankia* was isolated, and genetic and phenotypic characterizations performed (Chapters 4 and 5). Chapter 5 reports on the sequencing of three isolate genomes and the structure of their nitrogenase gene clusters in comparison to that of other strains from their respective host-infection groups. The arrangement of this gene cluster has been found to vary between, and be conserved within, the major *Frankia* clusters.

1.7 Conclusion

In this chapter a general introduction to the actinorhizal symbiosis was presented, which discussed relationships between both actinorhizal host species and their *Frankia* microsymbionts. A history of research into this subject involving indigenous South African actinorhizal species was also provided, as were strategies for isolating *Frankia* in pure culture. The information presented here provides background for the study's three main objectives: 1) to explore the diversity and natural associations of *Frankia* strains in *Morella* root nodules from the Cape flora of

South Africa, 2) and to explore *Frankia* diversity in rhizosphere soils from natural stands of *Morella*, comparing soil and nodule populations, and 3) to isolate and characterize endemic South African *Frankia*. As the following chapters are intended to be submitted as independent papers, some repetition of background information in each is unavoidable.

References

[1] The Plant List (2013). URL www.theplantlist.org

[2] Akimov, V.N., Dobritsa, S.V., Stupar, O.S.: Grouping of Frankia strains by DNA-DNA homology: how many genospecies are there in the genus *Frankia*? Developments in Plant and Soil Sciences **48**, 635–636 (1991)

[3] Al-Bermani, K.K.A., Al-Shammary, K.I.A., Gornall, R.J., Bailey, J.P.: Contributions to a cytological catalogue of the British and Irish flora. 3. Watsonia **19**(6), 269–271 (1993)

[4] Allsop, N., Stock, W.D.: Mycorrhizal status of plants growing in the Cape Floristic Region, South Africa. Bothalia **23**(1), 91–104 (1993)

[5] An, C.S., Riggsby, W.S., Mullin, B.C.: Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. International Journal of Systematic Bacteriology **35**, 140–146 (1985)

[6] An, C.S., Wills, J.W., Riggsby, W.S., Mullin, B.C.: Deoxynucleic acid base composition of 12 *Frankia* strains. Canadian Journal of Botany **61**(11), 2859–2862 (1983)

[7] Angiosperm Phylgeny Group: An ordinal classification for the families of flowering plants. Annals of the Missouri Botanical Garden **85**(4), 531–553 (1998)

[8] Arveby, A.S., Huss-Danell, K.: Presence and dispersal of infective *Frankia* in peat and meadow soils in Sweden. Biology and Fertility of Soils **6**(1), 39–44 (1988)

[9] Baker, D., Kidd, G.H., Torrey, J.G.: Separation of actinomycete nodule endophytes from crushed nodule suspensions by Sephadex fractionation. Botanical Gazette **140**, S49–S51 (1979)

[10] Baker, D., O'Keefe, D.: A modified sucrose fractionation procedure for the isolation of frankiae from actinorhizal root nodules and soil samples. Plant and Soil **78**(1-2), 23–28 (1984)

[11] Baker, D., Pengelly, W.L., Torrey, J.G.: Immunochemical analysis of relationships among isolated frankiae (Actinomycetales). International Journal of Systematic Bacteriology **31**(2), 148–151 (1981)

[12] Baker, D., Torrey, J.G.: The isolation and cultivation of actinomycetous root nodule endophytes. In: J.C. Gordon, C.T. Wheeler, J.G. Perry (eds.) Symbiotic Nitrogen Fixation in the Management of Temperate Forests, pp. 38–56. Forest Research Laboratory, Oregon State University (1979)

[13] Baker, D.D.: Relationships among pure cultured strains of *Frankia* based on host specificity. Physiologia Plantarum **70**, 245–248 (1987)

[14] Baker, D.D., Schwintzer, C.R.: Introduction. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 1–14. Academic Press, Inc., San Diego, California (1990)

[15] Bassi, C.A., Benson, D.R.: Growth characteristics of the slow-growing actinobacterium *Frankia* sp. strain CcI3 on solid media. Physiologia Plantarum **130**, 391–399 (2007)

[16] Bautista, G.H.H.: Phylogenomic study and specific depiction of Frankia genus. Special focus on non-cultivatable strains and ecological implications. Ph.D. thesis, Universite Claude Bernard, Lyon (2010)

[17] Bautista, G.H.H., Cruz, H.A., Nesme, X., Valdes, M., Mendoza, H.A., Fernandez, M.P.: Genomospecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of *Frankia* spp. Systematic and Applied Microbiology **34**, 200–206 (2011)

[18] Becking, J.H.: Frankiaceae Fam. Nov. (Actinomycetales) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. International Journal of Systematic Bacteriology **20**(2), 201–220 (1970)

[19] Becking, J.H., de Boer, W.E., Houwink, A.L.: Electron microscopy of the endophyte of *Alnus glutinosa*. Antonie van Leeuwenhoek **30**(1), 343–376 (1964)

[20] Benoit, L.F., Berry, A.M.: Methods for production and use of actinorhizal plants in forestry, low-maintenance landscapes, and revegetation. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, chap. 14, pp. 281–298. Academic Press, Inc., San Diego, California (1990)

[21] Benson, D.R.: Isolation of *Frankia* strains from Alder actinorhizal root nodules. Applied and Environmental Microbiology **44**(2), 461–465 (1982)

[22] Benson, D.R., Clawson, M.L.: Evolution of the actinorhizal plant symbiosis. In: E.W. Triplett (ed.) Prokaryotic nitrogen fixation: a model system for the analysis of a biological process, pp. 207–224. Horizon Scientific Press (2000)

[23] Benson, D.R., Dawson, J.O.: Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiologia Plantarum **130**, 318–330 (2007)

[24] Benson, D.R., Silvester, W.B.: Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiological Reviews **57**(2), 293–319 (1993)

[25] Benson, D.R., Stephens, D.W., Clawson, M.L., Silvester, W.B.: Amplification of 16S rRNA genes from *Frankia* strains in root nodules of *Ceanothus griseus*, *Coriaria arborea*, *Coriaria plumosa*, *Discaria toumatou*, and *Purshia tridentata*. Applied and Environmental Microbiology **62**(8), 2904–2909 (1996)

[26] Benson, D.R., Vanden Heuvel, B.D., Potter, D.: Actinorhizal symbioses: diversity and biogeography. In: M. Gillings (ed.) Plant Microbiology, pp. 102–133. BIOS Scientific Publishers Ltd., Oxford (2004)

[27] Benson, David R, Schultz, N.A.: Physiology and biochemistry of *Frankia* in culture. In: R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal plants, chap. 6, pp. 107–128. Academic Press, Inc., San Diego, California (1990)

[28] Berry, A., Torrey, J.G.: Isolation and characterization *in vivo* and *in vitro* of an actinomycetous endophyte from *Alnus rubra* Bong. In: J.C. Gordon, C.T.

Wheeler, J.G. Perry (eds.) Symbiotic Nitrogen Fixation in the Management of Temperate Forests, pp. 69–83. Forest Research Laboratory, Oregon State University (1979)

[29] Birgitta Bergman, B.O.: The *Gunnera*: *Nostoc* symbiosis. Biology and Environment: Proceedings of the Royal Irish Academy **102B**(1), 35–39 (2002)

[30] Bloom, R.A., Lechevalier, M.P., Tate 3rd, R.L.: Physiological, chemical, morphological, and plant infectivity characteristics of *Frankia* isolates from *Myrica pennsylvanica*: correlation to DNA restriction patterns. Applied and Environmental Microbiology **55**(9), 2161–2166 (1989)

[31] Bloom, R.A., Mullin, B.C., Tate, R.L.: DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pennsylvanica* (bayberry). Applied and Environmental Microbiology **55**(9), 2155–2160 (1989)

[32] Bond, G.: Evidence for fixation of nitrogen by root nodules of Alder (*Alnus*) under field conditions. New Phytologist **55**(2), 147–154 (1955)

[33] Bond, G.: Fixation of nitrogen in two African species of *Myrica*. New Phytologist **70**, 1–8 (1971)

[34] Bond, G.: The results of the IBP survey of non-legume root nodule formation in non-leguminous angiosperms. In: P.S. Nutman (ed.) Symbiotic Nitrogen Fixation in Plants, pp. 443–474. Cambridge University Press (1976)

[35] Bond, G., Fletcher, W.W., Ferguson, T.P.: The development and function of the root nodules of *Alnus*, *Myrica* and *Hippophae*. Plant and Soil **5**(4), 309–323 (1954)

[36] Bosco, M., Fernandez, M.P., Simonet, P., Materassi, R., Normand, P.: Evidence that some *Frankia* sp. strains are able to cross boundaries between *Alnus* and *Elaeagnus* host specificity groups. Applied and Environmental Microbiology **58**(5), 1569–1576 (1992)

[37] Brummitt, R.K.: Report of the committee for Spermatophyta: 48. Taxon **48**(2), 359–371 (1999)

[38] Brunchorst, D.: Uber einige wurzelanschwellungen, besonders diejenigen von Alnus und den Elaeagnaceen. Untersuch. Bot. Instit. Tubingen **2**, 150–177 (1888)

[39] Burgraaf, A.J.D., Quispel, A., Tak, T., Valstar, J.: Methods of isolation and cultivation of *Frankia* species from actinorhizas. Plant and Soil **61**, 157–168 (1981)

[40] Callaham, D., Del Tredici, P., Torrey, J.G.: Isolation and cultivation in vitro of the actinomycete causing root nodulation in *Comptonia*. Science **199**, 899–902 (1978)

[41] Carro, L., Cerda, E., Trujillo, M.E., Alonso-vega, P.: The genus *Micromonospora* is widespread in legume root nodules: the example of *Lupinus angustifolius*. The ISME Journal **4**(10), 1265–1281 (2010)

[42] Carro, L., Pujic, P., Alloisio, N., Fournier, P., Boubakri, H., Poly, F., Francois, P., Hocher, V., Merqaert, P., Balmand, S., Rey, M., Heddi, A., Normand, P.: *Alnus* peptides modify membrane porosity and induce the release of nitrogen-rich metabolites from nitrogen-fixing *Frankia*. ISME J **9**(8), 1723–1733 (2015)

[43] Carro, L., Pujic, P., Trujillo, M.E., Normand, P.: *Micromonospora* is a normal occupant of actinorhizal nodules. Journal of Biosciences **38**(4), 685–693 (2013)

[44] Chaia, E.E., Wall, L.G., Huss-Danell, K.: Life in soil by the actinorhizal root nodule endophyte *Frankia*. A review. Symbiosis **51**(3), 201–226 (2011)

[45] Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Donald, H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.l., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Sytsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedren, M., Gaut, B.S., Jansen, R.K., Kim, K.j., Wimpee, C.F.: Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Annals of the Missouri Botanical Garden **80**(3), 528–548 (1993)

[46] Chevalier, A.: Monographie des Myricacees; anatomie et histologie, organographie, classification et description des especes, distribution geographique. Memoires de la societe Nationale des Sciences Naturelles et Mathematiques Cherbourg **32**, 88–153 (1901)

[47] Clawson, M.L., Benson, D.R.: Natural diversity of *Frankia* strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Applied and Environmental Microbiology **65**(10), 4521–4527 (1999)

[48] Clawson, M.L., Bourret, A., Benson, D.R.: 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 (2004)

[49] Coetzee, J.A., Muller, J.: The phytogeographic significance of some extinct Gondwana pollen types from the Tertiary of the Southwestern Cape (South Africa). Annals of the Missouri Botanical Garden **71**(4), 1088–1099 (2014)

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

[51] Crane, P.R., Friis, E.M., Pedersen, K.R.: The origin and early diversification of the angiosperms. Nature **374**, 27–33 (1995)

[52] Cronquist, A.: An integrated system of classification for flowering plants. Columbia University Press, New York, N.Y. (1981)

[53] Cronquist, A.: In: A. Lawrence (ed.) The evolution and classification of flowering plants, 2nd edn., p. 396. The New York Botanical Garden, Kansas (1988)

[54] Dawson, J.O.: Ecology of actinorhizal plants. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing Actinorhizal Symbioses, chap. 8, pp. 199–234. Springer, Dordrecht, the Netherlands (2008)

[55] Diagne, N., Arumugam, K., Ngom, M., Nambiar-Veetil, M., Franche, C., Narayanan, K.K., Laplaze, L.: Use of Frankia and Actinorhizal Plants for

Degraded Lands Reclamation. BioMed Research International p. 948258 (2013)

[56] Diagne, N., Diouf, D., Svistoonoff, S., Kane, A., Noba, K., Franche, C., Bogusz, D., Duponnois, R.: *Casuarina* in Africa: Distribution, role and importance of arbuscular mycorrhizal, ectomycorrhizal fungi and *Frankia* on plant development. Journal of Environmental Management **128**, 204–209 (2013)

[57] Diem, H.G., Gauthier, D., Dommergues, Y.R.: Isolation of *Frankia* from nodules of *Casuarina equisetifolia*. Canadian Journal of Microbiology **28**, 526–531 (1982)

[58] Dillon, J.T., Baker, D.D.: Variations in nitrogenase activity among pure-cultured *Frankia* strains tested on actinorhizal plants as an indication of symbiotic compatibility. New Phytologist **92**(2), 215–219 (1982)

[59] Dinkelaker, B., Hengeler, C., Marschner, H.: Distribution and function of proteiod roots and other root clusters. Plant Biology **108**(3), 183–200 (1995)

[60] Dobritsa, S.V.: Grouping of *Frankia* strains on the basis of susceptibility to antibiotics, pigment and host specificity. International Journal of Systematic Bacteriology **48**, 1265–1275 (1998)

[61] Doyle, J.J.: Phylogenetic perspectives on the origins of nodulation. Molecular Plant-Microbe Interactions **24**(11), 1289–95 (2011)

[62] Du, D.Z., Baker, D.D.: Actinorhizal Host-Specificity of Chinese Frankia Strains. Plant and Soil **144**(1), 113–116 (1992)

[63] Elias, T.S.: The genera of Myricaceae in the Southeastern United States. Arnold Arboretum Journal **52**, 305–318 (1971)

[64] Erichsen-Brown, C.: Medicinal and other uses of North American plants: a historical survey with special reference to the Eastern Indian tribes. Dover Publications Inc., New York, N.Y. (1989)

[65] Fernandez, M.P., Meugnier, H., Grimont, P.A.D., Bardin, R.: Deoxyribonucleic acid relatedness among members of the genus *Frankia*. International Journal of Systematic Bacteriology **39**(4), 424–429 (1989)

[66] Fontaine, M.S., Young, P.H., Torrey, J.G.: Effects of long-term preservation of *Frankia* strains on infectivity, effectivity, and in vitro nitrogenase activity. Applied and Environmental Microbiology **51**(4), 694–698 (1986)

[67] Gaby, J.C., Buckley, D.H.: A comprehensive evaluation of PCR primers to amplify the *nif*H gene of nitrogenase. PLOS one **7** (2012)

[68] Gardes, M., Bousquet, J., Lalonde, M.: Isozyme Variation among 40 *Frankia* Strains. Applied and Environmental Microbiology **53**(7), 1596–1603 (1987)

[69] Gardner, I.C.: Mycorrhizae of actinorhizal plants. Journal of Applied Microbiology and Biotechnology **2**(1), 147–160 (1986)

[70] Gardner, I.C., Bond, G.: Host plant-endophyte adaptation in *Myrica*. Naturwissenschaften **53**(6), 161 (1966)

[71] Gauthier, D., Jaffre, T., Prin, Y.: Occurrence of both *Casuarina*-infective and *Elaeagnus*-infective *Frankia* strains within actinorhizae of *Casuarina collina*, endemic to New Caledonia. European Journal of Soil Biology **35**(1), 9–15 (1999)

[72] Gauthier, D., Jaffre, T., Prin, Y.: Abundance of *Frankia* from *Gymnostoma* spp. in the rhizosphere of *Alphitonia neocaledonica*, a non-nodulated Rhamnaceae endemic to New Caledonia. European Journal of Soil Biology **36**, 169–175 (2000)

[73] Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Peer, Y.V.D., Vandamme, P., Thompson, F.L., Swings, J.: Re-evaluating prokaryotic species. Nature Reviews Microbiology **3**(September), 733–739 (2005)

[74] Ghodhbane-Gtari, F., Essoussi, I., Chattaoui, M., Chouaia, B., Jaouani, A., Daffonchio, D., Boudabous, A., Gtari, M.: Isolation and characterization of

non-*Frankia* actinobacteria from root nodules of *Alnus glutinosa*, *Casuarina glauca* and *Elaeagnus angustifolia*. In: Symbiosis, vol. 50, pp. 51–57 (2010)

[75] Gleason, H.A., Cronquist, A.: Manual of vascular plants of Northeastern United States and adjacent Canada. Van Nostrand Reinhold, New York, N.Y. (1963)

[76] Goldblatt, P.: An analysis of the flora of Southern Africa: its characteristics, relationships, and orgins. Missouri Botanical Garden Press **65**(2), 369–436 (1978)

[77] Goldblatt, P., Manning, J.: Plant diversity of the Cape region of southern Africa. Annals of the Missouri Botanical Garden **89**(2), 281–302 (2002)

[78] Goldblatt, P., Manning, J.C.: Cape Plants. A conspectus of the Cape flora of South Africa. National Botanical Institute of South Africa (2000)

[79] Gordon, J.C., Dawson, J.O.: Potential uses of nitrogen-fixing trees and shrubs in commercial forestry. Botanical Gazette **140**, S88–S90 (1979)

[80] Gtari, M., Brusetti, L., Skander, G., Mora, D., Boudabous, A., Daffonchio, D.: Isolation of *Elaeagnus*-compatible *Frankia* from soils collected in Tunisia. FEMS Microbiology Letters **234**, 349–355 (2004)

[81] Gtari, M., Dawson, J.O.: An overview of actinorhizal plants in Africa. Functional Plant Biology **38**, 653–661 (2011)

[82] Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., Sbissi, I., Ayari, A., Yamanaka, T., Normand, P., Tisa, L.S., Boudabous, A.: Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Scientific Reports **5**, 13,112 (2015)

[83] Gtari, M., Tisa, L.S., Normand, P.: Diversity of *Frankia* strains, actinobacterial symbionts of actinorhizal plants. In: R. Aroca (ed.) Symbiotic endophytes, chap. 7, pp. 123–148. Springer, Berlin Heidelberg (2013)

[84] Guetsky, R., Gollop, N., Bernstein, N.: Genetic diversity of *Frankia* strains isolated from root nodules of *Casuarina* in Israel. Israel Journal of Plant Sciences **53**(2), 125–133 (2005)

[85] Hahn, D.: Polyphasic taxonomy of the genus *Frankia*. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing Actinorhizal Symbioses, chap. 2, pp. 25–47. Springer, Dordrecht, the Netherlands (2008)

[86] Hahn, D., Dorsch, M., Stackebrandt, Akkermans, A.D.L.: Synthetic oligonucleotide probes for identification of *Frankia* strains. Plant and Soil **118**, 211–219 (1989)

[87] Hahn, D., Lechevalier, M.P., Fischer, A., Stackebrandt, E.: Evidence for a close phylogenetic relationship between members of the genera *Frankia*, *Geodermatophilus*, and *Blastococcus* and emdendation of the family Frankiaceae. Systematic and Applied Microbiology **11**(3), 236–242 (1989)

[88] Hahn, D., Nickel, A., Dawson, J.: Assessing *Frankia* populations in plants and soil using molecular methods. FEMS Microbiology Ecology **29**, 215–227 (1999)

[89] Hahn, D., Starrenburg, M.J.C., Akkermans, A.D.L.: Variable compatibility of cloned *Alnus glutinosa* ecotypes against ineffective *Frankia* strains. Plant and Soil **107**, 233–243 (1988)

[90] He, X.H., Chen, L.G., Hu, X.Q., Asghar, S.: Natural diversity of nodular microsymbionts of *Myrica rubra*. Plant and Soil **262**, 229–239 (2004)

[91] Helme, N.A., Trinder-Smith, T.H.: The endemic flora of the Cape peninsula, South Africa. South African Journal of Botany **72**, 205–210 (2006)

[92] Herbert, J.: Systematics and Biogeography of Myricaceae. Ph.D. thesis, University of St. Andrews (2005)

[93] Herbert, J., Chase, M.W., Moller, M., Abbott, R.J.: Nuclear and plastid DNA sequences confirm the placement of the enigmatic *Canacomyrica monticola* in Myricaceae. Taxon **55**(2), 349–357 (2006)

[94] Hiltner, L.: Über die Bedeutung der Wurzelknöllchen von *Alnus glutinosa* für die Stickstoffernährung dieser Pflanze. Landw. Versuchsstat **46**, 153–161 (1895)

[95] Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., Gherbi, H., Queiroux, C., Da Silva, C., Wincker, P., Normand, P., Bogusz, D.: Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. Plant Physiology **156**(2), 700–711 (2011)

[96] Honerlage, W., Hahn, D., Zepp, K., Zeyer, J., Normand, P.: A hypervariable 23S rRNA region provides a discriminating target for specific characterization of uncultured and cultured *Frankia*. Systematic and Applied Microbiology **17**(3), 433–443 (1994)

[97] Huguet, V., Batzli, J.M.C.C., Zimpfer, J.F., Normand, P., Dawson, J.O., Fernandez, M.P., Lyon, C.B., Cedex, V.: Diversity and specificity of *Frankia* strains in nodules of sympatric *Myrica gale*, *Alnus incana*, and *Shepherdia canadensis* determined by rrs gene polymorphism. Applied Environmental Microbiology **67**(5), 2116–2122 (2001)

[98] Huguet, V., Gouy, M., Normand, P., Zimpfer, J.F., Fernandez, M.P.: Molecular phylogeny of Myricaceae: a reexamination of host-symbiont specificity. Molecular phylogenetics and evolution **34**(3), 557–568 (2005)

[99] Huguet, V., Land, E.O., Casanova, J.G., Zimpfer, J.F., Fernandez, M.P.: Genetic diversity of *Frankia* microsymbionts from the relict species *Myrica faya* (Ait.) and *Myrica rivas-martinezii* (S.) in Canary Islands and Hawaii. Microbial Ecology **49**, 617–625 (2005)

[100] Huguet, V., Mergeay, M., Cervantes, E., Fernandez, M.P.: Diversity of *Frankia* strains associated to *Myrica gale* in Western Europe: impact of host plant (*Myrica* vs. *Alnus*) and of edaphic factors. Environmental Microbiology **6**(10), 1032–1041 (2004)

[101] Igual, M., Valverde, A., Mateos, P.F., Rodr, C., Mart, E., Cervantes, E., Vel, E.: Genomic fingerprinting of *Frankia* strains by PCR-based techniques. Assessment of a primer based on the sequence of 16S rRNA gene of *Escherichia coli*. Plant and Soil **254**, 115–123 (2003)

[102] Jeong, S.C., Myrold, D.D.: Genomic fingerprinting of *Frankia* microsymbionts from *Ceanothus* copopulations using repetitive sequences and polymerase chain reactions. Canadian Journal of Botany **77**, 1220–1230 (1999)

[103] Jeong, S.C., Myrold, D.D.: Recent progress in the evolution and ecology of actinorhizal symbioses. Plant Pathology Journal **19**(1), 1–8 (2003)

[104] Jeong, S.C., Ritchie, N.J., Myrold, D.D.: Molecular phylogenies of plants and *Frankia* support multiple origins of actinorhizal symbioses. Molecular phylogenetics and evolution **13**(3), 493–503 (1999)

[105] Killick, D.B.J.: The South African species of *Myrica*. Bothalia **10**(1), 5–17 (1969)

[106] Killick, D.J.B., Polhill, R.M., Verdcourt, B.: New combinations in African Myricaceae. Kew Bulletin **53**(4), 993–995 (1998)

[107] Koch, R.: Die Aetiologie der Tuberculose. Berliner Klinische Wochenschrift **19**(15), 221–230 (1882)

[108] Lalonde, M.: Immunological and ultrastructural demonstration of nodulation of the European *Alnus glutinosa* (L.) Gaertn. host plant by an actinomycetal isolate from the North American *Comptonia peregrina* (L.) Coult. root nodule (1979)

[109] Lambers, H., Bishop, J.G., Hopper, S.D., Laliberté, E., Zúñiga-Feest, A.: Phosphorus-mobilization ecosystem engineering: the roles of cluster roots and carboxylate exudation in young P-limited ecosystems. Annals of Botany **110**(2), 329–48 (2012)

[110] Lambers, H., Brundrett, M.C., Raven, J.A., Hopper, S.D.: Plant mineral nutrition in ancient landscapes: high plant species diversity on infertile soils is linked to functional diversity for nutritional strategies. Plant and Soil **334**, 11–31 (2010)

[111] Lambers, H., Raven, J.A., Shaver, G.R., Smith, S.E.: Plant nutrient-acquisition strategies change with soil age. Trends in Ecology and Evolution **23**(2), 95–103 (2008)

[112] Lechevalier, M.P.: The taxonomy of the genus *Frankia*. Plant and Soil **78**(1), 1–6 (1984)

[113] Lechevalier, M.P.: Taxonomy of the Genus *Frankia* (Actinomycetales). International Journal of Systematic Bacteriology **44**(1), 1–8 (1994)

[114] Lechevalier, M.P., Lechevalier, H.: The taxonomic position of the actinomycetic endophytes. In: J.C. Gordon, C.T. Wheeler, J.G. Perry (eds.) Symbiotic Nitrogen Fixation in the Management of Temperate Forests, pp. 111–122. Forest Research Laboratory, Oregon State University (1979)

[115] Lechevalier, M.P., Lechevalier, H.A.: Taxonomy of *Frankia*. In: L. Ortiz-Ortiz, L.F. Bojalil, V. Yakoleff (eds.) Biological, biochemical and biomedical aspects of actinomycetes, pp. 575–582. Academic Press, Inc., Orlanda, Fla. (1984)

[116] Lechevalier, M.P., Lechevalier, H.A.: Systematics, isolation and culture of *Frankia*. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 35–60. Academic Press, Inc., New York, N.Y. (1990)

[117] Linder, H.P.: The radiation of the Cape flora, southern Africa. Biological Reviews **78**, 597–638 (2003)

[118] Linder, H.P.: Evolution of diversity: the Cape flora. Trends in Plant Science **10**(11), 536–541 (2015)

[119] Lopez, M.F., Young, P., Torrey, J.G.: A comparison of carbon source utilization for growth and nitrogenase activity in two *Frankia* isolates. Canadian Journal of Microbiology **32**(4), 353–358 (1986)

[120] Louis, I., Racette, S., Torrey, J.G.: Occurence of cluster roots on *Myrica cerifera* L. (Myricaceae) in water culture in relation to phosphorus nutrition. New Phytologist **115**(2), 311–317 (1990)

[121] Love, A.: Chromosome number reports LXIX. Taxon **29**, 703–730 (1980)

[122] Love, A.: Chromosome number reports LXXIV. Taxon **30**, 119–128 (1982)

[123] Lumini, E., Bosco, M., Fernandez, M.P.: PCR-RFLP and total DNA homology revealed three related genomic species among broad-host *Frankia strains*. FEMS Microbiology Ecology **21**, 303–311 (1996)

[124] Lutzow-Felling, C.J., Gardner, D.E., Markin, G.P., Smith, C.W.: *Myrica faya*: review of the biology, ecology, distribution, and control, including an annotated bibliography. Cooperative National Park Resources Studies Unit, University of Hawaii at Manoa, Department of Botany (1995)

[125] MacDonald, A.D.: Myricaceae: floral hypothesis for *Gale* and *Comptonia*. Canadian Journal of Botany **55**, 2636–2651 (1977)

[126] MacDonald, A.D.: The morphology and relationships of the Myricaceae. In: P.R. Crane, S. Blackmore (eds.) Evolution, Systematics and fossil history of the Hamamelidae, Vol 2, pp. 147–165. Clarendon Press, Oxford (1989)

[127] Mackintosh, A.H., Bond, G.: Diversity in the nodular endophytes of *Alnus* and *Myrica*. Phyton Buenos Aires **27**, 79–90 (1970)

[128] Magallon, S., Crane, P.R., Herendeen, P.S.: Phylogenetic pattern, diversity, and diversification of Eudicots. Annals of the Missouri Botanical Garden **86**(2), 297–372 (1999)

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

[130] Mansour, S.R., Dewedar, A., Torrey, J.G.: Isolation, culture, and behavior of *Frankia* strain HFPCgI4 from root nodules of *Casuarina glauca*. Botanical Gazette **151**(4), 490–496 (1990)

[131] Maunuksela, L.: Molecular and physiological characterization of rhizosphere bacteria and *Frankia* in forest soils devoid of actinorhizal plants. Ph.D. thesis, University of Helsinki (2001)

[132] Maunuksela, L., Zepp, K., Koivula, T., Zeyer, J., Haahtela, K., Hahn, D.: Analysis of *Frankia* populations in three soils devoid of actinorhizal plants. FEMS Microbiology Ecology **28**(1), 11–21 (1999)

[133] Meeks, J.C.: Symbiosis between nitrogen-fixing cyanobacteria and plants. Bioscience **48**(4), 266–276 (1998)

[134] Meyen, J.: Über das Hervorwachsen parasitischer Gebilde aus den Wurzeln anderer Pflanzen. Flora (Jena) **12**, 48–64 (1829)

[135] Mian, S., Bond, G., Rodriguez-Barrueco, C.: Effective and ineffective root nodules in *Myrica faya*. Proceedings of the Royal Society of London B: Biological Sciences **194**(1116), 285–293 (1976)

[136] Miettinen, P.: The response of a free-living and an endophytic *Frankia* to extreme environmental conditions. Symbiosis **15**(1-2), 121–134 (1993)

[137] Miller, I.M., Baker, D.D.: Nodulation of actinorhizal plants by *Frankia* strains capable of both root hair infection and intercellular penetration. Protoplasma **131**, 82–91 (1986)

[138] Mirza, B.S., Welsh, A.K., Hahn, D.: Growth of *Frankia* strains in leaf litter-amended soil and the rhizosphere of a nonactinorhizal plant. FEMS Microbiology Ecology **70**(1), 132–141 (2009)

[139] Mirza, M.S., Hahn, D., Akkermans, A.D.L.: Isolation and characterization of *Frankia* Strains from *Coriaria nepalensis*. Systematic and Applied Microbiology **15**(2), 289–295 (1992)

[140] Mort, A., Normand, P., Lalonde, M.: $2-O-$methyl$-$D$-$glucose, a key sugar in the taxonomy of *Frankia*. Canadian Journal of Microbiology **29**, 993–1002 (1983)

[141] Nalin, R., Normand, P., Simonet, P., Domenach, A.M.: Polymerase chain reaction and hybridization on DNA extracted from soil as a tool for *Frankia* spp. population distribution studies in soil. Canadian Journal of Botany **77**, 1239–1247 (1999)

[142] Navarro, E., Jaffre, T., Gauthier, D., Gourbiere, F., Rinaudo, G., Simonet, P., Normand, P.: 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 (1999)

[143] Navarro, E., Nalin, R., Gauthier, D., Normand, P.: The nodular microsymbionts of *Gymnostoma* spp. are *Elaeagnus*-infective *Frankia* strains. Applied Environmental Microbiology **63**(4), 1610–1616 (1997)

[144] Nicotra, A.B., Leigh, A., Boyce, K., Jones, C.S., Niklas, K.J., Royer, D.L., Tsukaya, H.: The evolution and functional significance of leaf shape in the angiosperms. Functional Plant Biology **38**, 535–552 (2011)

[145] Normand, P., Cournoyer, B., Gouy, M., Biometric, T.D.: Molecular phylogeny of the symbiotic actinomycetes of the genus *Frankia* matches host-plant infection processes. Molecular and Biological Evolution **10**(6), 1303–1316 (1993)

[146] Normand, P., Fernandez, M.P.: Evolution and Diversity of *Frankia*. In: K. Pawlowski (ed.) Prokaryotic Symbionts in Plants, vol. 8, pp. 103–125. Springer-Verlag, Berlin Heidelberg (2009)

[147] Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choisne, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltsman, E., Huang, Y., Kopp, O.R., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastronunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D.: Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research **17**, 7–15 (2007)

[148] Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J.O., Evtushenko, L., Misra, A.K.: Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. International Journal of Systematic Bacteriology **46**(1), 1–9 (1996)

[149] Normand, P., Simonet, P., Bardin, R.: Conservation of *nif* sequences in *Frankia*. Molecular and General Genetics **213**(2-3), 238–246 (1988)

[150] Nouioui, I., Ghodhbane-Gtari, F., Beauchemin, N.J., Tisa, L.S., Gtari, M.: Phylogeny of members of the *Frankia* genus based on *gyrB*, *nifH* and *glnII* sequences. Antonie van Leeuwenhoek **100**(4), 579–587 (2011)

[151] Oginuma, K., Tanaka, R.: Karyomorphological studies on three species of Myrica. The Journal of Japanese Botany **62**(6), 183–188 (1987)

[152] Parra-O, C.: New combinations in South American Myricaceae. Brittonia **54**(4), 322–326 (2002)

[153] Parsons, R., Silvester, W.B., Harris, S., Gruijters, W.T.M., Bullivant, S.: *Frankia* vesicles provide inducible and absolute oxygen protection for nitrogenase. Plant Physiology **83**, 728–731 (1987)

[154] Pawlowski, K., Bisseling, T.: Rhizobial and actinorhizal symbioses: what are the shared features? The Plant cell **8**(October), 1899–1913 (1996)

[155] Pommer, E.: Über die isolierung des endophyten aus den wurzelknöllchen *Alnus glutinosa* Gaertn. und uber erfolgreiche reinfektionsversuche. Ber. Dtsh. Bot. Ges. **72**, 138–150 (1959)

[156] Quispel, A.: Discoveries, discussions and trends in research on actinorhizal root nodule symbioses before 1978. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of *Frankia* and Actinorhizal Plants, pp. 15–34. Academic Press, Inc., San Diego, California (1990)

[157] Racette, S., Torrey, J.G.: The isolation, culture and infectivity of a *Frankia* strain from *Gymnostoma papuanum* (Casuarinaceae). Plant and Soil **118**, 165–170 (1989)

[158] Rademaker, J.L.W., Hoste, B., Louws, F.J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P., De Bruijn, F.J.: Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. International Journal of Systematic and Evolutionary Microbiology **50**(2), 665–677 (2000)

[159] Richards, A.M.B., Stock, W.D., Cowling, R.M.: Soil nutrient dynamics and community boundaries in the fynbos vegetation of South Africa. Plant Ecology **130**(2), 143–153 (2011)

[160] Roller, C., Ludwig, W., Schleifer, K.H.: Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes. Journal of General Microbiology **138**(6), 1167–1175 (1992)

[161] Rosbrook, A., Reddell, P.: Isolation of *Frankia* from root nodules of three species of *Casuarina*. Soil Biology and Biochemistry **27**(4), 427–429 (1995)

[162] Schnitzler, J., Barraclough, T.G., Boatwright, J.S., Goldblatt, P., Manning, J.C., Powell, M.P., Rebelo, T., Savolainen, V.: Causes of plant diversification in the Cape biodiversity hotspot of South Africa. Systematic biology **60**(3), 343–57 (2011)

[163] Schwencke, J., Caru, M.: Advances in actinorhizal symbiosis: host plant-Frankia interactions, biology, and applications in arid land reclamation. A review. Arid Land Research and Management **15**(4), 285–327 (2001)

[164] Schwintzer, C.R.: Spore-positive and spore-negative nodules. In: C. Schwintzer, J. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, chap. 9, pp. 177–194. Academic Press, Inc., San Diego, California (1990)

[165] Sen, A., Duabin, V., Abrouk, D., Gifford, I., Berry, A.M., Normand, P.: 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 Nakamu. International Journal of Systematic and Evolutionary Microbiology (2014)

[166] Shipton, W.A., Burggraaf, A.J.P.: A comparison of the requirements for various carbon and nitrogen sources and vitamins in some *Frankia* isolates. Plant and Soil **69**, 149–161 (1982)

[167] Silvester, W.B., Berg, R.H., Schwintzer, C.R., Tjepkema, J.D.: Oxygen responses, hemoglobin, and the structure and function of vesicles. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing Actinorhizal Symbioses, pp. 105–141. Springer, Dordrecht, the Netherlands (2008)

[168] Skead, C.J.: In: J.C. Manning, N.C. Anthony (eds.) Historical plant incidence in southern Africa, p. 378. South African National Biodiversity Institute, Pretoria (2009)

[169] Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., DePamphilis, C.W., Wall, P.K., Soltis, P.S.: Polyploidy and angiosperm diversification. American Journal of Botany **96**(1), 336–348 (2009)

[170] Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hhan, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J., Nixon, K.C., Farris, J.S.: Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. Botanical Journal of the Linnean Society **133**(4), 381–461 (2000)

[171] Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., Martin, P.G.: Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proceedings of the National Academy of Sciences **92**(7), 2647–2651 (1995)

[172] Soltis, D.E., Soltis, P.S., Nickrent, D.L., Johnson, L.A., Hahn, W.J., Hoot, S.B., Sweere, J.A., Kuzoff, R.K., Kron, K.A., Chase, M.W., Swensen, S.M., Zimmer, E.A., Chaw, S.M., Gillespie, L.J., Kress, W.J., Sytsma, K.J.: Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Annals of the Missouri Botanical Garden **84**(1), 1–49 (1997)

[173] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, A.D., Kampfer, P., Maiden, M.C.J., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B.: Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. International Journal of Systematic and Evolutionary Microbiology **52**, 1043–1047 (2002)

[174] Stackebrandt, E., Goebel, B.M.: Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology **44**(4), 846–849 (1994)

[175] Stokes, J.: Cytological studies of the Myricaceae. Botanical Gazette **99**(2), 387–399 (1937)

[176] Suzuki, S.I.: Establishment and use of gellan gum media for selective isolation and distribution survey of specific rare actinomycetes. Actinomycetologica **15**(2), 55–60 (2001)

[177] Swensen, S.M.: The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. American Journal of Botany **83**(11), 1503–1512 (1996)

[178] Swensen, S.M., Benson, D.R.: Evolution of actinorhizal host plants and *Frankia* endosymbionts. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing Actinorhizal Symbioses, chap. 4, pp. 73–104. Springer, Dordrecht, the Netherlands (2008)

[179] Tarrant, R.F., Trappe, J.M.: The role of *Alnus* in improving the forest environment. Plant and Soil **Special Volume**, 335–346 (1971)

[180] Tisa, L., McBride, M., Ensign, J.C.: Studies of growth and morphology of *Frankia* strains EAN1pec, Eul1c, Cpl1, and ACN1ag. Canadian Journal of Botany **61**(11), 2768–2773 (1983)

[181] Tisa, L.S., Chval, M.S., Krumholz, G.D., Richards, J.: Antibiotic resistance patterns of *Frankia* strains. Canadian Journal of Botany **77**(9), 1257–1260 (1999)

[182] Tjepkema, J.D.: The role of oxygen diffusion from the shoots and nodule roots in nitrogen fixation by root nodules of *Myrica gale*. Canadian Journal of Botany **61**(11), 1365–1371 (1978)

[183] Torrey, J.G.: Nitrogen fixation by actinomycete-nodulated angiosperms. BioScience **28**(9), 586–592 (1978)

[184] Torrey, J.G.: Cross-Inoculation Groups within *Frankia* and Host-Endosymbiont Associations. In: C. Schwintzer, J. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 83–106. Academic Press, Inc., San Diego, California (1990)

[185] Torrey, J.G., Tjepkema, J.D.: Symbiotic nitrogen fixation in actinomcete-nodulated trees. Botanical Gazette **140**(Supplement), i–ii (1979)

[186] Valdes, M., Perez, N., Estrada-de los Santos, P., Caballero-Mellado, J., Pena-Cabriales, J.J., Normand, P., Hirsch, A.M.: Non-*Frankia* actinomycetes isolated from surface-sterilized roots of *Casuarina equisetifolia* fix nitrogen. Applied Environmental Microbiology **71**(1), 460–466 (2005)

[187] Van Dijk, C.: Spore formation and endophyte diversity in root nodules of *Alnus glutinosa*. New Phytologist **81**, 601–615 (1978)

[188] Van Ryssen, F.W.J., Grobbelaar, N.: The nodulating and nitrogen fixing ability of South African *Myrica* species. South African Journal of Science **66**(1), 22–25 (1970)

[189] Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., Swings, J.: Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiological reviews **60**(2), 407–438 (1996)

[190] Vanden Heuvel, B.D., Benson, D.R., Bortiri, E., Potter, D.: 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). Canadian Journal of Microbiology **50**, 989–1000 (2004)

[191] Vandenbosch, K.A., Torrey, J.G.: Host-endophyte interactions in effective and ineffective nodules induced by the endophyte of *Myrica gale*. Canadian Journal of Botany **61**, 2898–2909 (1983)

[192] Various: Notes on African plants. Bothalia **38**(1), 141–159 (2008)

[193] Verdcourt, B., Polhill, R.: Proposals to conserve the names *Myrica* and *Gale* (Myricaceae) with conserved types. Taxon **46**(2), 347–348 (1997)

[194] Watt, M., Evans, J.R.: Proteoid Roots. Physiology and development. Plant physiology **121**(2), 317–323 (1999)

[195] Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, G.E., Stackebrandt,

E., Starr, M.P., Truper, H.G.: Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic Bacteriology **37**(4), 463–464 (1987)

[196] Welsh, A.K., Dawson, J.O., Gottfried, G.J., Hahn, D.: Diversity of *Frankia* populations in root nodules of geographically isolated Arizona Alder trees in central Arizona (United States). Applied and Environmental Microbiology **75**(21), 6913–6918 (2009)

[197] Welsh, A.K., Mirza, B.S., Rieder, J.P., Paschke, M.W., Hahn, D.: Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Systematic and Applied Microbiology **32**(3), 201–210 (2009)

[198] Wheeler, C.T., Akkermans, A.D.L., Berry, A.M.: *Frankia* and actinorhizal plants: a historical perspective. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing Actinorhizal Symbioses, pp. 1–24. Springer, Dordrecht, the Netherlands (2008)

[199] Wheeler, C.T., Miller, I.M.: Current and Potential uses of Actinorhizal Plants in Europe. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, chap. 18, pp. 365–389. Academic Press, Inc., San Diego, California (1990)

[200] White, F.: African Myricaceae and the history of the Afromontane flora. Opera Botanica **121**, 173–188 (1993)

[201] Whitman, W.B., Parte, A., Goodfellow, M., Kämpfer, P., Busse, H.J., Trujillo, M.E., Ludwig, W., Suzuki, K.: Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria. v. 5, pts. 1-2. Springer (2012)

[202] Wilbur, R.L.: The Myricaceae of the United States and Canada: genera, subgenera, and series. Sida **16**(1), 93–107 (1994)

[203] Witkowski, E.T.F., Mitchell, D.T.: Variations in soil phosphorus in the fynbos biome, South Africa. Journal of Ecology **75**, 1159–1171 (1987)

[204] Woronin, M.: Über die bei der Schwarzerle (*Alnus glutinosa*) und der gewöhnlichen Gartenlupine (Lupinus mutabilis) auftretenden Wurzel-anschwellungen

anschwellungen. Memoires de l'Academie inperiale des sciences de St. Petersbourg **10**(6), 1–13 (1866)

[205] Zehr, J.P., Jenkins, B.D., Short, S.M., Steward, G.F.: Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental Microbiology **5**(7), 539–554 (2003)

[206] Zepp, K., Hahn, D., Zeyer, J.: Evaluation of a 23S rRNA insertion as target for the analysis of uncultured *Frankia* populations in root nodules of alders by whole cell hybridization. Systematic and Applied Microbiology **20**(1), 124–132 (1997)

[207] Zepp, K., Hahn, D., Zeyer, J.: In-situ analysis of introduced and indigenous *Frankia* populations in soil and root nodules obtained on *Alnus glutinosa*. Soil Biology and Biochemistry **29**(9), 1595–1600 (1997)

[208] Zimpfer, J.F., Kennedy, G.J., Smyth, C.A., Hamelin, J., Navarro, E., Dawson, J.O.: Localization of *Casuarina*-infective *Frankia* near *Casuarina cunninghamiana* trees in Jamaica. Canadian Journal of Botany **77**(9), 1248–1256 (1999)

2 Natural diversity of *Frankia* in field collected nodules from Cape *Morella* species

Abstract

Despite the considerable number of endemic actinorhizal *Morella* species on the African continent, the diversity of their natural *Frankia* microsymbionts has hitherto not been investigated. In order to address this I investigated *Morella* of the Cape Flora of Southern Africa. Root nodules were collected from natural stands of six endemic *Morella* species at 23 sites within the Western Cape province of South Africa, and the diversity of *Frankia* populations investigated by means of comparative *nif*H gene sequence analysis. Amplification of partial *nif*H gene fragments from uncultured strains in 202 root nodules yielded 26 unique 606 bp sequences. Phylogenetic analysis of 512 bp *nif*H gene fragments from this data set aligned with database sequences from *Frankia* isolates and uncultivated nodular strains assigned Cape sequences to *Frankia* Cluster I and Cluster III, with sequences from both clusters present in three of the examined host species. Nineteen sequences were assigned to five previously described subclusters within Cluster III and seven sequences to two subclusters within Cluster I, one of which has been described previously and the other subcluster novel to the current study. Identical sequences were recovered from nodules collected at geographically distant locations, suggesting a cosmopolitan distribution within the region for strains from both clusters, with the exception of two subclusters associated with *M. diversifolia* only and restricted to the mountains of the Cape Peninsula. Soil pH correlated with strain presence, with Cluster I sequences recovered from nodules originating from acidic soils

exclusively. As with other members of the genus, some *Morella* of the Cape flora of southern Africa are promiscuous in their natural habitats, with host infection group influenced by habitat edaphic conditions. A possible exception is *M. cordifolia*, which appears to associate with *Frankia* from Cluster III exclusively. This chapter reports on the first investigation of *Frankia* nodulating endemic African *Morella* in their natural habitats.

2.1 Introduction

Frankia are geographically widespread gram-positive soil actinomycetes capable of entering into nitrogen fixing symbioses with specific host plants, in what are known as "actinorhizal" symbioses. Host plants benefit from this association by gaining the ability to grow in marginal, nitrogen-limited, soils and are thus often pioneer species in newly formed or exposed soils [5]. Africa hosts an abundance of actinorhizal species, and is home to perhaps half of all *Morella* species globally [1,23]. Despite this, within Africa endemic actinorhizal plants (all of which belong to the genus *Morella*) and their microsymbionts have been largely neglected. *Morella*, the only actinorhizal genus endemic to Africa (*Alnus glutinosa* is indigenous to North Africa, Asia and Europe), currently contains 30 accepted species [1]. Seven of these are found within the Cape flora, with three endemic to it, where they occupy diverse habitats including coastal dunes and riparian zones [22, 24]. Isolation on montane islands, where fog provides moisture during the Cape's dry summers, has likely contributed to the extensive diversification within the genus [24]; a possible example being *M. diversifolia*, which is restricted to the mountains of the Cape peninsula [29][1].

Early investigations found South African *Morella* to be nodule-bearing, and the nodules to be the site of nitrogen fixation. The endophyte, an actinomycete, appeared typical of those infective on hosts from the *Myricaceae* [7, 59]. These studies contributed to the 1976 International Biological Programme (IBP) survey of non-legume nitrogen fixing plants, which considerably expanded the list of known actinorhizal species [8]. In the following years the subject of native endophyte

[1]Note that *M. diversifolia*'s range may to be more extensive according to nine observations made between 1963 and 1984 (www.sibis.sanbi.org).

diversity in southern African actinorhizal hosts was not pursued, partly because at that time *Frankia* strains had not yet been reproducibly isolated [9]. Additionally, while a considerable export market had previously existed for the wax of *M. cordifolia*, which was used in candle making in the 19th century [52], *Morella* were no longer of economic importance in the region.

Prior to the successful isolation of CpI1 in 1978, Becking proposed species names for ten of the as-yet uncultured endophytes [4], based on host species and various characteristic phenotypes in symbiosis. His system was rejected once isolates became available, and cross-inoculation experiments demonstrated that some strains could infect hosts from multiple families [2, 9, 18]. The family *Frankiaceae*, from the order *Frankiales* ord. nov., currently contains a single genus in which only one species is formally recognized (*Frankia alni*) [50, 64]. *Frankia* are informally classified into four clusters, the first three of which correlate with the host genera upon which they are infective (and effective). These clusters, or host infection groups, are supported by Normand's phylogenetic analysis of the 16S rRNA gene [44], and also by other phylogenetic markers such as the *nif*H, *gln*A and *gln*II genes [15, 37, 60], *nif*D-*nif*K and 16S-23S intergenic spacer regions [36, 45], and an actinomycete-specific 23S rDNA insertion sequence [32, 47]. Strains from *Frankia* Cluster I infect *Alnus*, *Comptonia*, *Morella* and *Myrica* with a subgroup of closely related strains infectious on *Allocasuarina* and *Casuarina*. Cluster II infects *Coriaria*, *Datisca* and *Dryas*. Representatives of this cluster had until very recently defeated all attempts at isolation [26]. Cluster III strains form effective nodules on species of *Colletia*, *Elaeagnus*, *Gymnostoma*, *Hippophae*, *Myrica*, *Morella* and *Shepherdia*. A fourth and final cluster consists of "atypical" strains which are unable to re-infect their original hosts and/or lack the ability to fix nitrogen [64].

Because of their ability to effectively nodulate with strains from both Clusters I and III, alternatively the *Alnus*-, and *Elaeagnus*- Host Infection Groups (HIGs), when inoculated under laboratory conditions, the genus *Morella* is considered promiscuous [34, 40, 57, 61]. Although the host-endophyte associations of comparatively few *Morella* species have been studied under natural field conditions, these associations were typically found to be more specific than those found in greenhouse trials. It is thought that this may be a consequence of prevalent edaphic conditions affecting

the availability or infectivity of *Frankia* strains [11, 13, 28, 33, 34, 65].

Despite evidence of radiation within *Morella* of the Cape region being comparatively recent and there being a low degree of molecular variation between Cape species [30], they display a surprisingly large degree of variation in both ecology and morphology. The variety of distinct leaf shapes is particularly remarkable considering the lack of variation among other *Morella* species, in which leaf morphology is typically elliptical-oblanceolate [30] (see Figure 2.1). Cape species occupying distinct niches [22], dictated by their need for abundant water. *Morella cordifolia* and *M. quercifolia* are lowland species, the former stabilizing coastal dunes in the same role as the North American *M. cerifera*, whilst the latter inhabits coastal flats, marshes and wetlands. *M. integra* or *M. serrata*, both of which are small trees, colonize perennial stream and river banks. *M. kraussiana* is found on the mountains of the Western Cape Province, which have semi-constant cloud cover during the region's drought-prone summers. *M. diversifolia*, which is possibly a hybrid of *M. kraussiana* and *M. quercifolia*, occupies the same niche as *M. kraussiana* but is restricted to the mountains of the Cape peninsula [29]. Despite disagreements concerning classifications within the Myricaceae [30], all of the Cape *Morella* species are currently accepted [1].

The diversity of *Frankia* in symbiosis with African *Morella* has never been explored. This fact provided the motivation for the current chapter, the aim of which was to explore the diversity of *Frankia* associated with South African *Morella* in their natural environments using the nitrogenase reductase gene as phylogenetic marker. Six species were examined in the field, and the populations associated with each were compared to those of the other local species, as well as actinorhizal hosts from elsewhere in the world. Soil characteristics were correlated to *Frankia* lineages present in nodules at individual sites.

2.2 Materials and Methods

2.2.1 Identification of Cape *Morella* species in the field

In addition to their varied habitats, *Morella* of the Cape flora have an unusual variety of distinctive leaf shapes compared to other *Morella* [30]. Species were identified by comparison with type material and in consultation with botanical descriptions [22]. Where doubt existed as to the identity of *M. integra* and *M. serrata* I examined the undersides of leaves microscopically for balloon glands, considering trees in which they were present to be *M. integra*. The leaf morphology of *M. diversifolia*, *M. kraussiana* and *M. humilis* intergrade, making identification in the field problematic. Where doubt existed concerning the identity of *M. diversifolia* and *M. kraussiana*, I considered plants in which the majority of leaves were toothed to be *M. diversifolia*. *Morella humilis* was not encountered. See Figure 2.2 for an example of an indeterminate species. All populations from which nodules were collected were photographed and their GPS coordinates recorded.

2.2.2 Nodule collection and processing

Root nodules were collected from six southern African *Morella* 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 Tables 2.1[2] and A.1, and Figure 2.3[3], between April of 2012 and December of 2013. Of these species, *M. diversifolia*, *M. integra* and *M. kraussiana* are endemic to the Cape flora [22]. Some populations of *M. cordifolia*, *M. diversifolia* and *M. kraussiana* were not nodulated, insofar as could be determined without extensive excavation and damage to the plant. Nodules were stored on ice or in habitat soil during transport to the laboratory, where they were processed immediately upon arrival.

Individual nodules were cleaned of adherent soil particles. A single lobe was then

[2]Nodules and soil from site A4 were collected by Gerda Theron.

[3]Species occurrence data may be conveniently viewed using the South African Biodiversity Information Facility mapping tool at http://sibis.sanbi.org/

Figure 2.1: Typical leaf morphologies for Cape *Morella* species. 1: *M. serrata*, 2: *M. integra*, 3a: *M. cordifolia* leaf arrangement on stalks, 3b: *M. cordifolia*, 4: *M. quercifolia*, 5: *M. diversifolia*, 6: *M. kraussiana*

Figure 2.2: Unidentified *Morella* species growing in the Cape Point Nature Reserve, possibly *M. kraussiana* or *M. diversifolia*. No root nodules were recovered from this specimen.

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 for 30-45 min with periodic manual agitation, after which they were rinsed with copious amounts of sterile distilled water. Nucleic acids were extracted from nodule lobes using a MoBio PowerPlant Pro kit (MoBio, Carlsbad, CA) according to the manufacturer's instructions. Due to the diminutive size of most nodule lobes the epidermis could not be removed prior to homogenization. In order to ensure that *nif*H amplicons arose from nodule endophytes, and not surface contaminants, a control was performed in which root material adjacent to several nodules was subjected to the same cleaning and DNA extraction procedures as nodule lobes.

2.2.3 Soil pH determination

Soil pH was determined as previously described [49]. Measurements were taken in triplicate on individually prepared 1:5 suspensions of soil in 0.01 N $CaCl_2$ and distilled water.

2.2.4 PCR conditions

Amplification of *nif*H gene fragments from nodule extracts were performed using primers nifHf1 (5′-GGC AAG TCC ACC ACC CAG C) and nifHr (5′-CTC GAT GAC CGT CAT CCG GC) [62] in a reaction volume of 25 μL. Each 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 (5U μL^{-1}; TaKaRa) and 1 μL of template. An initial 5-min incubation at 95°C was followed by 35 rounds of temperature cycling (94°C for 30 s, 60°C for 30 s, 72°C for 45 sec) and a final 5-min extension at 72°C. Reactions (5 μL) were checked for amplification products by gel electrophoresis (1% (w/v) agarose gel in TAE buffer, stained with ethidium bromide (0.5 μg mL^{-1})).

2.2.5 Sequencing and sequence analysis

Amplified *nif*H 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, in both directions using primers nifHf1 and nifHr, at the Stellenbosch University Central Analytical Facility (http://academic.sun.ac.za/saf). Chromatograms were visually assessed, sequences corrected by hand where necessary, and assembled in CLC Main Workbench (version 6.2.1). Representatives of each unique *nif*H sequence were deposited in GenBank under accession numbers KP342075-KP342100 (*nif*H).

2.2.6 Phylogenetic analysis

Non-identical 606 bp partial *nif*H sequences (hereafter referred to as Cape *Frankia* Nitrogenase, 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). This data set comprised 26 sequences from Cape *Morella* nodules and 250 database sequences (75 from pure cultures and 175 obtained from nodules) representing the clusters defined by Higgins et al. (KL1-KL5) [31], Kennedy et al. (KL6-KL8) [38] and Welsh et al. (EI-EIV, AI-AV) [62]. 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[4].

[4]The effect of trimming CFN sequences from 606 bp to 512 bp on uniqueness and assignment to Cape Clusters is illustrated in Figure A.1

The resultant 94 sequence data set was analyzed using neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods as previously described [62]. Neighbour joining was completed in PAUP (version 4.0b) [56]. 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) [16, 27]. 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) [41, 54]. 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" [56]. 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 [55] and custom scripts, and displayed in Figtree (version 1.4) (http://tree.bio.ed.ac.uk/software/figtree/).

2.3 Results

Using the *Frankia* nitrogenase-specific primer pair nifHf1/nifHr, PCR products of the expected size (641 bp) were obtained from 202 of 208 (97%) *Morella* nodule lobe extracts, collected from natural stands of all six species of *Morella* examined (Table 2.2). Nodules from which PCR products were not obtained were collected from *M. kraussiana* on the northern plateau of Table Mountain exclusively (not indicated in Figure 2.3). These nodules were subsequently tested for the presence of Cluster II *Frankia* using purpose-designed primers, which were also found to be absent (results not shown).

Sequence analysis revealed within this data set the presence of twenty-six non-identical 606 bp *nif*H sequences, namely CFN1-26. At 97% sequence similarity, these CFNs clustered into a total of seven *Frankia* genotypes, which are hereafter referred to as Cape Clusters (CC) 1-7. In-silico translations of the non-identical 606 bp CFNs revealed identical peptide sequences within each CG, with the sole exception of CFN5 whose translation product differed from the rest of CC-3 at one position (Table 2.2). NCBI nucleotide BLAST identified all 26 of the 606 bp sequences as *Frankia* nitrogenase reductases (*nif*H). Phylogenetic analysis clearly assigned all *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 2.4 & 2.5). Two of seven CC genotypes were assigned to *Frankia* Cluster I, and the remaining five to Cluster III, Table 3.5, Figures 2.4 & 2.5). The dominant genotypes (CC-1, CC-3 & 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, usually restricted to one host species each.

Within the *Alnus*-HIG, genotype CC-2 belonged to subgroup AV, as defined by Welsh et al. [62]. Genotype CC-1 is novel to my study and is most closely related to KL8, described by Kennedy et al. [38]. Within the *Elaeagnus* HIG the majority of sequences (119/127) fell within EIII, described by Welsh et al. [62]. The remaining 8 sequences constituted genotypes clustering within EI (CC-6, CC-7) and EII (CC-5).

Table 2.1: Details of species occurrence, with samples collection within 100 m of indicated coordinates for each site (Figure 2.3), soil pH (mean of three readings, with standard deviation in parentheses) by the $CaCl_2$ method, number of nodules for which *nif*H amplification was successful collected per locality†, and *Frankia* host infection groups‡ found at each site.

Species	Site	Latitude	Longitude	pH ($CaCl_2$)	nod†	HIG‡
M. cordifolia	A1.1	33° 11' 33.40" S	18° 4' 19.00" E	8.31 (0.07)	2	III
	A1.2	33°11'51.70" S	18°4'41.80" E	8.43 (0.13)	4	III
	A1.3	33°8'28.40" S	18°6'12.90" E	N/D	1	III
	A1.4	33°14'41.00" S	18°11'32.40" E	N/D	2	III
	A1.5	33°13'40.00" S	18°9'19.30" E	7.70 (0.24)	4	III
	A2	33°45'14.49" S	18°26'31.70" E	N/D	1	III
	A4	34°23'20.00" S	21°25'29.14" E	8.16 (0.02)	8	III
	B4	34°4'54.50" S	18°27'59.00" E	8.16 (0.02)	36	III
	B7	34°21'27.70" S	18°55'28.00" E	7.06 (0.17)	10	III
M. quercifolia	A1.5	33° 13' 40.00 "S	18° 9' 19.30" E	7.70 (0.24)	15	III
	B1.1	34°13'59.90" S	18°22'51.10" E	N/D	2	III
	B3.1	34°5'31.30" S	18°25'30.80" E	N/D	1	I
	B5	33°56'11.00" S	18°37'7.90" E	7.45 (0.07)	21	III
M. diversifolia	B2	33°58'16.90" S	18°25'15.20" E	3.45 (0.01)	18	I/III
	B1.2	34°13'13.60" S	18°22'49.70" E	4.48 (0.03)	13	III
M. kraussiana	B1.3	34°13'10.70" S	18°22'46.00" E	4.46 (0.02)	2	I
	B3.2	34°5'33.80" S	18°25'35.60" E	N/D	1	I
	B3.3	34°5'34.60" S	18°25'59.60" E	4.48 (0.03)	3	I
M. serrata	B6	33°58'23.20" S	18°56'13.40" E	5.21 (0.08)	6	I
M. integra	A3.1	32°6'55.00" S	19°3'54.30" E	4.17 (0.06)	18	I/III
	A3.2	32°6'6.80" S	19°4'3.10" E	3.83 (0.07)	3	III
	A3.3	32°6'10.60" S	19°3'52.80" E	4.04 (0.02)	8	I
	A3.4	32°21'41.25" S	19°4'19.52" E	N/D	23	I
Total					202	

Table 2.2: Occurrence of 606 bp *nif*H sequences in nodules of Cape *Morella* species. Sequences are listed by Cape Genotype 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 (see chapter 5). 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	umique *nif*H	*M. cordifolia*	*M. diversifolia*	*M. quercifolia*	*M. integra*	*M. kraussiana*	*M. serrata*	Total	Translation	Isolates
CC-1	CFN1	-	-	-	42	-	6	48	1	FMi1, FMi2
	CFN3	-	7	1	2	6	-	16	1	FMk1, FMq1
	CFN20	-	-	-	1	-	-	1	1	
	CFN22	-	-	-	1	-	-	1	1	
	CFN23	-	-	-	1	-	-	1	1	
CC-2	CFN10	-	6	-	-	-	-	6	2	
	CFN16	-	2	-	-	-	-	2	2	
CC-3	CFN5	12	-	1	-	-	-	13	3a	FMc1, FMc2, FMc3
	CFN6	11	2	6	-	-	-	19	3b	
	CFN7	6	-	2	4	-	-	12	3b	
	CFN8	-	9	-	-	-	-	9	3b	
	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	
CC-4	CFN2	20	-	-	-	-	-	20	4	FMc6, FMc7
	CFN4	-	-	20	-	-	-	20	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		

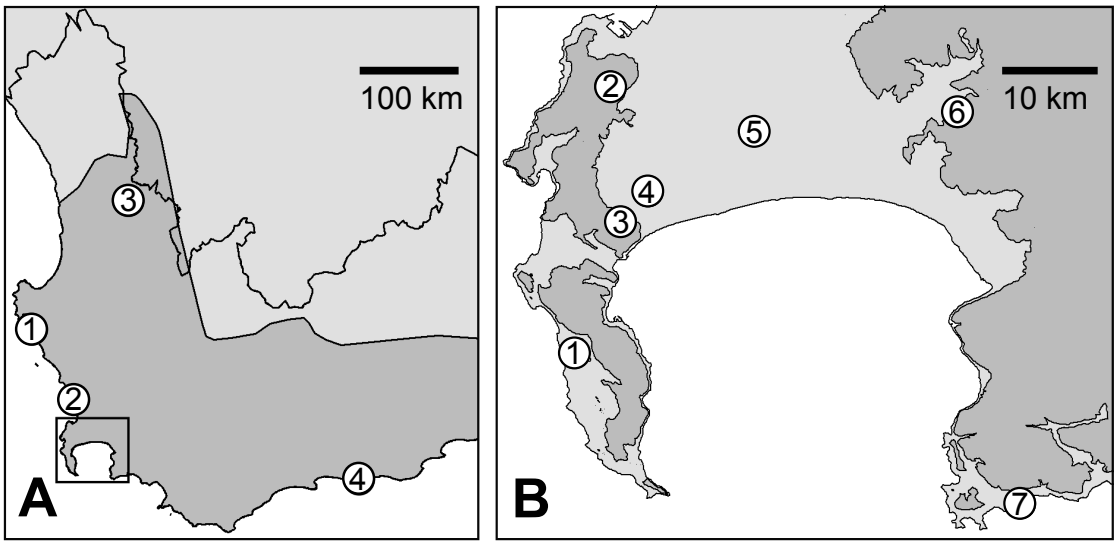


Figure 2.3: Map of *Morella* sampling locations within the Western Cape Province of South Africa, with the Cape Floristic Region indicated in dark gray (panel A). Site A1 is West Coast National Park, site A2 is the Blouberg Provincial Nature Reserve, site A3 is the Cedeberg wilderness area, site A4 is the Stilbaai Nature Reserve. The Cape peninsula and greater False Bay area are indicated by the boxed inset (panel B). Areas at altitudes above 100 m are indicated in dark gray and correspond to the Table Mountain range on the Cape Peninsula in the West, and the Boland and Hex River mountain ranges 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 2.1. Nodules recovered from each sites are detailed in the Appendix (Table A.1).

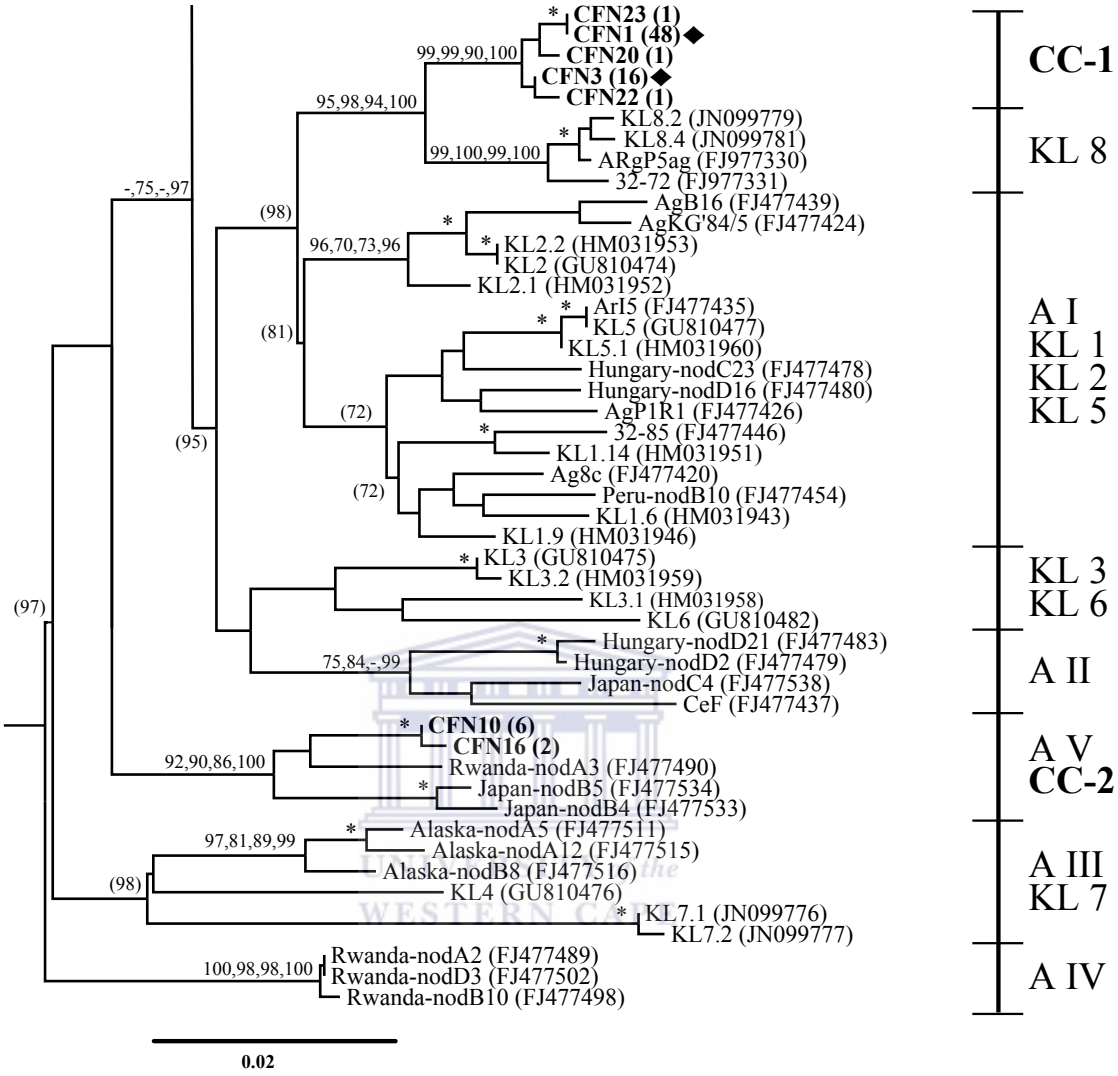


Figure 2.4: Phylogenetic tree generated using 512 bp *nif*H gene sequences from *Frankia* isolates and nodule-derived sequences, cut to display sequences from Cluster I [44] (the *Alnus* host infection group) only. Sequences are labeled with country of origin, genotype name or strain designation and GenBank accession number. Sequences encountered in this study are designated in bold, with the number of nodules in which each was found indicated in brackets. Diamonds indicate Cape *Frankia* isolates corresponding to nodule-derived *nif*H sequences. Numbers at nodes reflect bootstrap (BS) measures from neighbor 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 assigned by Welsh et al. (AI-AV) [62], Kennedy et al. (KL1-5, KL6-8) [31, 38] and Cape Clusters (CC) as defined in this study are indicated on the right. An uncultured *Frankia* strain from *Datisca* nodules ($X76398$) was included as the outgroup (not shown).

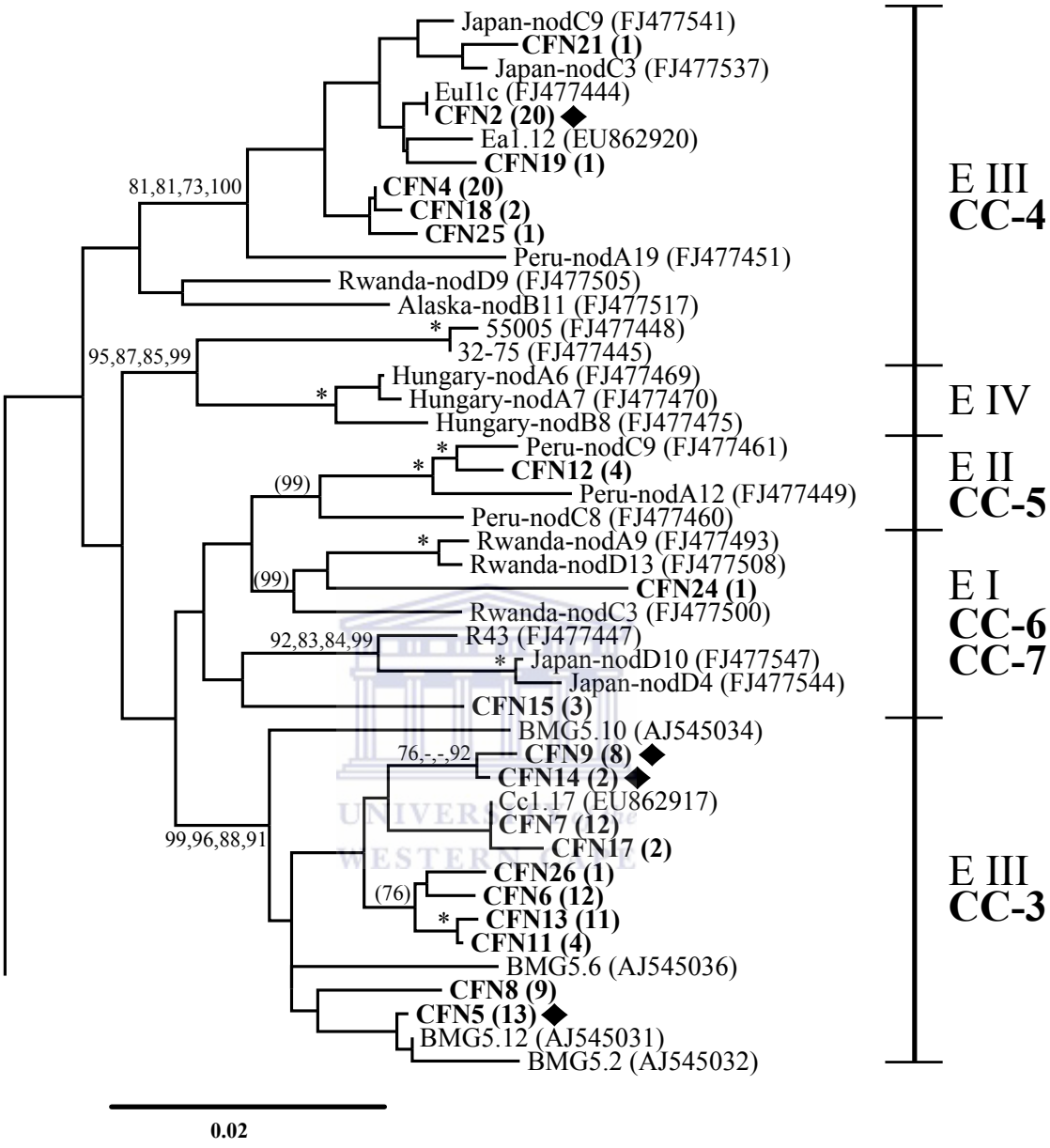


Figure 2.5: Phylogenetic tree generated using 512 bp *nif*H gene sequences from *Frankia* isolates and nodule-derived sequences, displaying sequences from Cluster III [44] (the *Elaeagnus* host infection group) only. Labels are country of origin, strain designation and GenBank accession number. Cape sequences are indicated in bold, with the number of nodules in brackets. Diamonds indicate Cape *Frankia* isolates corresponding to nodule-derived *nif*H sequences. Numbers at nodes reflect bootstrap (BS) measures from neighbor joining, maximum likelihood and maximum parsimony, and posterior probabilities (PP) from Bayesian analysis, respectively. Where support was from Bayesian analysis only, PP values are indicated in parentheses. Asterisks indicate nodes with support from at least 3 of the 4 phylogenetic methods. *Elaeagnus* subgroups as assigned by Welsh et al. [62] (EI-EIV), and Cape Clusters (CC) as defined in this study, are indicated on the right. EuI1c's genome is now available (CP002299) and contains no nif*H* gene. The Genbank sequence FJ77444 was mislabeled, and in fact refers to strain EAN1pec.

2.4 Discussion

From the 202 successful reactions, a total of 26 unique 606 bp *nif*H sequences were obtained. These were identified as *Frankia* nitrogenase reductases by BLAST analysis, and given the identifiers CFN1-26 (Table 3.5). Average similarity between non-identical Cape *Frankia nif*H sequences was 95.25% (SD 0.021) and cluster analysis in MOTHUR assigned all CFNs to one cluster at ⩾92% identity. When clustered at ⩾97% similarity Cape *nif*H sequences were assigned to a total of 7 genotypes: CC-1 to CC-7. A maximum-dissimilarity threshold of 3% is commonly used to investigate *Frankia* diversity, and allows for differentiation between Cluster I and Cluster III strains with reliable assignment to subclusters within these groups [31, 39, 42, 62].

A preliminary phylogenetic analysis was performed in order to determine appropriate conditions for further analysis. Two CCs containing seven CFNs were assigned to Cluster I, and five CCs containing nineteen CFNs to Cluster III (Table 3.5). While both major *Frankia* host infection groups were detected, overall strain richness was comparatively low, which is in keeping with what has been found in other environments [33, 46, 60, 62]. For phylogenetic analysis CFNs were aligned with database *nif*H sequences from *Frankia* isolates and nodule-derived sequences and trimmed to 512 bp to allow the incorporation of the greatest number of database sequences possible. Following this, the number of unique Cape *Morella* derived sequences was reduced from 26 to 25, with CFN1 and CFN23 identical throughout 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 (data not shown).

Phylogenetic analysis of diverse *Frankia*-derived *nif*H sequences reliably assigns strains to the same host infection clusters as found in 16S rRNA gene analyses (see chapter 5 and [62]). Additionally, similar *nif*H-based tree topologies have been found irrespective of the methodology used [62]. Subclusters within these groups, however, have been previously assigned based on tree topology and node support measures rather than strict sequence similarity. Consequently, clusters outlined by Welsh et al. [62] are defined at a range of similarity values from 93%

to 99% instead of a rigid similarity threshold. In line with this the common practice of assigning microbial genetic marker sequences (usually 16S rDNA) to operational taxonomic units has been criticized, principally because of the lack of support from an underlying theory of microbial speciation [21]. This is illustrated by *Frankia* phylogenies, as it is known that defined similarity values (or OTU assignments) for both *nif*H and 16S rDNA sequences, as well as other commonly used markers, do not necessarily agree with *Frankia* genomospecies as defined in DNA/DNA hybridization and AFLP studies [3, 19, 62]. Nevertheless, cutoffs of 97% similarity for *nif*H sequences are commonly used in investigations of *Frankia* diversity. In-silico translation of CFNs demonstrated that all sequences within each respective genotype defined at this threshold, with the exception of CFN5 in CC-3, coded for identical peptide sequences (Table 2.2). While this approach holds for Cape sequences, when a larger data set incorporating *nif*H sequences from globally sourced *Frankia* strains was analyzed clusters defined at the same cutoff contained a variety of translation products (results not shown). Genotype assignments at this threshold may therefore have deeper ecological significance when strains from the same region are examined.

Identical *nif*H sequences were found in nodules from geographically distant sites, with 8 of the 13 general localities sharing identical CFNs with between one and three other locations (Supplementary Table A.1). 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. 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, Hex river and Table Mountain ranges in the south to the Cederberg range in the north. 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*).

Soil pH and other edaphic factors are known to influence both strain presence in nodules and the degree of nodulation [17, 43, 53, 66]. 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 [66]. Similarly, In my study *M. cordifolia* was found in neutral to alkaline coastal soils (Table 2.1, Figure 2.3), was the most widely and intensively sampled host species with 68 nodules recovered from 8 widely-dispersed coastal sites over more than half of the species's natural range. Sequences from this species accounted for 11 of 26 unique CFNs (Table 3.5), but it 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 2.3) yielded a Cluster I sequence (CFN3), indicating that this species too is promiscuous in the field. This sample was recovered from the same soil horizon as that of *M. kraussiana* at the same site, for which soil pH was found to be acidic (Table 2.1), supporting suggestions that soil pH is a determining factor in *Frankia* strain selection [33].

Within *Frankia* Cluster I, strains 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, this group was found in nodules from soils found to be acidic exclusively (Table 2.1), and was dominant in *M. integra* collected at riparian sites in the Cederberg mountains (Figure 2.3, site A2), although Cluster III strains were also found in nodules from this species. Sequence CFN1 was also detected in six nodules from *M. serrata* (Figure 2.3, 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 [31]. 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 2.3, site A3), exclusively. Similarly, sequence CFR2 was found in *M. kraussiana*, *M. diversifolia* and *M. quercifolia* nodules from acidic soils on Table Mountain, as well as *M. integra*

nodules collected in the Cederberg (Figure 2.3, site A3). *Myrica gale* was found to be nodulated by Cluster I strains exclusively, and typically inhabits acidic, water-logged habitats [11] similar to those inhabited by *M. integra* and *M. serrata*.

Genotypes CC-2, CC-5 and CC-6, from *Frankia* Clusters I and III, were detected in *M. diversifolia* nodules only. This species is restricted to the mountains of the Cape peninsula [22, 29] and it is possible that these genotypes are associated uniquely with the species. Furthermore, Cluster III sequence CFN8 was found in this species, and while belonging to CC-3, was not detected in the nodules of any other host species. More intensive sampling would be necessary to establish whether specific relationships between these genotypes and *M. diversifolia* exist. 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. Despite intense effort, I was unable to isolate *Frankia* strains representing genotype CC-2, and no isolates from genotype AV (within which CC-2 falls, Figure 2.4) are known to exist [62].

Overall Cluster I sequences displayed low diversity, a finding often encountered in studies of this host infection group [33, 38, 63]. Cluster I strains are apparently selective for specific hosts, as demonstrated in sympatric *Alnus* stands, and where *Alnus* and *Myrica gale* grow together, although the underlying basis for this selectivity is unknown [33, 46]. Outside of the Myricaceae, overall *Frankia* diversity in root nodules from naturally occurring populations of actinorhizal host species is typically low [46, 62]. In one case only three *Frankia* strains were detected in twelve *Alnus* species in an established Alder stand [46]. Each species was nodulated by a single strain only, suggesting highly selective relationship between even closely related species of this genus and their microsymbionts. This is similar to what I found in *M. kraussiana*, nodules from which contained a single sequence only, CFN10, despite being collected from several well-separated sites. It also suggests that the effect of selection by host species is large, as the *Alnus* species were all growing in the same soil (assuming identical soil conditions across the population) [46]. Furthermore, Cluster I 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 [33, 35]. While Cluster I strains are cosmopolitan, local dominance by sub-groups within the cluster found in other studies, and clear selection for specific groups within the Cape region that are rarely encountered elsewhere is evidence for greater selectivity between strains from this cluster and their hosts than is found with Cluster III strains. Notably, I did not encounter strains clustering with *Casuarina* genotypes (*nif*H subgroup AII according to Welsh et al. [62]) in my survey, which thus provides additional support for the claim that *Morella* do not serve as a reservoir for *Casuarina* strains in the absence of their normal hosts [51].

Kennedy and Higgins [31] reported a new Cluster I *Frankia* lineage (KL8) in their survey of *Alnus* nodules across North America. In my analysis *Frankia* isolates ARgP5ag and 32-72 clustered within this genotype (but were excluded from Kennedy's analysis) and separately from CC-1, at 3% sequence identity and with good support from all four treeing methods (Figure 2.4). ARgP5ag is currently the sole member of *Frankia* genomospecies G3 [3, 19]. 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 is a new group of *Frankia* within the otherwise well described Cluster I.

Conversely, Cluster III strains were more diverse, with 18 unique sequences from five Cape Genotypes distributed amongst three of the four clusters identified by Welsh et al. [62]. The majority of sequences were assigned to subcluster EIII, with only four sequences assigned to each of EI and EII (Figure 2.5, Table 3.5). As with the Cluster I sequences, unique genotypes were associated with *M. diversifolia* at site B1 (Figure 2.3) exclusively, and were found in several nodules, as well as soil collected at this site (see chapter 4). In my analysis subcluster EIII was resolved into two separate groups (CC-3 and CC-4), with CFN sequences present in both. One of these groups, CC-3, was composed principally of sequences originating from north Africa, which in the previous study had been assigned to EIII but with poor support from treeing methods [62]. Furthermore, *Frankia* strain CC1.17, assigned to EII by Welsh et al. [62], clustered with the African sequences according to both *nif*H 3%-dissimilarity cluster assignments and treeing methods (Figure 2.5).

Welsh et al. [62] found local dominance of *Frankia* clusters in soils collected from 5

sites spread across the world, with clusters present in one soil rarely detected in others. My findings are similar, as sequences from cluster EIII were dominant, with sequences assigned to clusters EI and EII only rarely detected. Studies specifically focused on *Frankia* from the *Elaeagnus* HIG are more limited, but mirror what has been found with the better-studied *Alnus* HIG in this regard [12, 20, 42, 43].

Morella is the only extant native actinorhizal genus in southern Africa, with native *Casuarina* becoming locally extinct following the Tertiary period [14], Cape Genotype I *Frankia* have thus, over the course of the last four to eight million years, only entered into nodular symbiosis with *Morella*. It will be interesting to determine whether they are infective on "traditional" Cluster I hosts, especially *Casuarina*. As representatives of this cluster were successfully isolated in the current study (see Chapter 4), future genomic comparisons and cross-inoculation studies will a provide deeper understanding of host specificity and the molecular mechanisms of infection for these strains.

Neither the influence of geographic location on *Frankia* occurrence nor strain diversities per host species were investigated in depth as they could not be detached from the numerous potentially confounding factors, including the limited number of populations sampled for most host species, differences in populations age, and varying soil conditions between sites.

Six of the 208 nodules failed to yield amplification products. All of these were collected from the same site on the slopes of Devil's Peak on the Table mountain range, where *M. kraussiana* grew among *Colletia* species (Rosaceae). I considered the possibility that these plants were nodulated by strains from *Frankia* Cluster II, and so designed primers specifically targeting this group (in-silico analyses indicated that primers nifHf1 and nifHr do not amplify Cluster II *nif*H). Amplification with the new primers was unsuccessful (results not shown). Root nodules are known be formed on actinorhizal hosts by atypical Cluster IV *Frankia*, which would not be detected with my methodology, and other actinomycetes are commonly found in actinorhizal nodules and may be involved in their induction [10, 25, 58]. *M. kraussiana* nodules may be mycorrhizal, as *Morella* have previously been found to have mycorrhizal associations, despite also possessing cluster roots [6, 48]. While

I observed cluster roots on *M. cordifolia* none were found on the remaining five species, including *M. kraussiana* at Devil's Peak site.

2.5 Conclusion

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 subcluster closely related to KL8. Cluster III strains were more diverse, with a total of 18 unique sequence distributed across five genotypes, clustering within EI, EII and EIII, and the majority of these falling within subcluster EIII. I found three *Morella* species (*M. integra*, *M. diversifolia* and *M. quercifolia*) to be promiscuous in the field, two (*M. serrata*, *M. kraussiana*) to be 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. This work constitutes the first investigation into the diversity of *Frankia* infective on endemic actinorhizal plants of southern Africa.

References

[1] The Plant List (2013). URL www.theplantlist.org

[2] Baker, D.D.: Relationships among pure cultured strains of *Frankia* based on host specificity. Physiologia Plantarum **70**, 245–248 (1987)

[3] Bautista, G.H.H., Cruz, H.A., Nesme, X., Valdes, M., Mendoza, H.A., Fernandez, M.P.: Genomospecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of *Frankia* spp. Systematic and Applied Microbiology **34**, 200–206 (2011)

[4] Becking, J.H.: Frankiaceae Fam. Nov. (Actinomycetales) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. International Journal of Systematic Bacteriology **20**(2), 201–220 (1970)

[5] Benson, D.R., Vanden Heuvel, B.D., Potter, D.: Actinorhizal symbioses: diversity and biogeography. In: M. Gillings (ed.) Plant Microbiology, pp. 102–133. BIOS Scientific Publishers Ltd., Oxford (2004)

[6] Berliner, R., Torrey, J.G.: On tripartite *Frankia*-mycorrhizal associations in the Myricaceae. Canadian Journal of Botany **67**, 1708–1712 (1989)

[7] Bond, G.: Fixation of nitrogen in two African species of *Myrica*. New Phytologist **70**, 1–8 (1970)

[8] Bond, G.: The results of the IBP survey of non-legume root nodule formation in non-leguminous angiosperms. In: P.S. Nutman (ed.) Symbiotic Nitrogen Fixation in Plants, pp. 443–474. Cambridge University Press (1976)

[9] Callaham, D., Del Tredici, P., Torrey, J.G.: Isolation and cultivation in vitro of the actinomycete causing root nodulation in *Comptonia*. Science **199**, 899–902 (1978)

[10] Carro, L., Pujic, P., Trujillo, M.E., Normand, P.: *Micromonospora* is a normal occupant of actinorhizal nodules. Journal of Biosciences **38**(4), 685–693 (2013)

[11] Clawson, M.L., Benson, D.R.: Natural diversity of *Frankia* strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Applied and Environmental Microbiology **65**(10), 4521–4527 (1999)

[12] Clawson, M.L., Caru, M., Benson, D.R.: Diversity of *Frankia* strains in root nodules of plants from the families Elaeagnaceae and Rhamnaceae. Applied and Environmental Microbiology **64**(9), 3539–3543 (1998)

[13] Clawson, M.L., Gawronski, J., Benson, D.R.: Dominance of *Frankia* strains in stands of *Alnus incana* subsp. *rugosa* and *Myrica pensylvanica*. Canadian Journal of Botany **77**, 1203–1207 (1999)

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

[15] Cournoyer, B., Lavire, C.: Analysis of *Frankia* evolutionary radiation using *glnII* sequences. FEMS Microbiology Letters **177**, 29–34 (1999)

[16] Darriba, D., Taboada, G.L., Doallo, R., Posada, D.: jModelTest 2: more models, new heuristics and parallel computing. Nature Methods **9**, 772 (2012)

[17] Dawson, J.O., Klemp, M.T.: Variation in the capacity of black alder to nodulate in central Illinois soils. In: Proceedings of the Central Hardwoods Forest Conference IV, pp. 255–260. Department of Forestry, University of Tennessee. Knoxville, TN (1987)

[18] Dillon, J.T., Baker, D.D.: Variations in nitrogenase activity among pure-cultured *Frankia* strains tested on actinorhizal plants as an indication of symbiotic compatibility. New Phytologist **92**(2), 215–219 (1982)

[19] Fernandez, M.P., Meugnier, H., Grimont, P.A.D., Bardin, R.: Deoxyribonucleic acid relatedness among members of the genus *Frankia*. International Journal of Systematic Bacteriology **39**(4), 424–429 (1989)

[20] Gauthier, D., Jaffre, T., Prin, Y.: Occurrence of both *Casuarina*-infective and *Elaeagnus*-infective *Frankia* strains within actinorhizae of *Casuarina collina*, endemic to New Caledonia. European Journal of Soil Biology **35**(1), 9–15 (1999)

[21] Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Peer, Y.V.D., Vandamme, P., Thompson, F.L., Swings, J.: Re-evaluating prokaryotic species. Nature Reviews Microbiology **3**(September), 733–739 (2005)

[22] Goldblatt, P., Manning, J.C.: Cape Plants. A conspectus of the Cape flora of South Africa. National Botanical Institute of South Africa (2000)

[23] Gtari, M., Brusetti, L., Skander, G., Mora, D., Boudabous, A., Daffonchio, D.: Isolation of *Elaeagnus*-compatible *Frankia* from soils collected in Tunisia. FEMS Microbiology Letters **234**, 349–355 (2004)

[24] Gtari, M., Dawson, J.O.: An overview of actinorhizal plants in Africa. Functional Plant Biology **38**, 653–661 (2011)

[25] Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Beauchemin, N., Tisa, L.S.: Phylogenetic perspectives of nitrogen-fixing actinobacteria. Archives of Microbiology **194**, 3–11 (2012)

[26] Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., Sbissi, I., Ayari, A., Yamanaka, T., Normand, P., Tisa, L.S., Boudabous, A.: Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Scientific Reports **5**, 13,112 (2015)

[27] Guidon, S., Gascuel, O.: A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology **52**, 696–704 (2003)

[28] He, X.H., Chen, L.G., Hu, X.Q., Asghar, S.: Natural diversity of nodular microsymbionts of *Myrica rubra*. Plant and Soil **262**, 229–239 (2004)

[29] Helme, N.A., Trinder-Smith, T.H.: The endemic flora of the Cape peninsula, South Africa. South African Journal of Botany **72**, 205–210 (2006)

[30] Herbert, J.: Systematics and Biogeography of Myricaceae. Ph.D. thesis, University of St. Andrews (2005)

[31] Higgins, L.M., Kennedy, P.G.: Symbiotic *Frankia* bacteria in *Alnus* forests in Mexico and the United States of America: is geographic location a good predictor of assemblage structure? Botany **90**, 423–431 (2012)

[32] Honerlage, W., Hahn, D., Zepp, K., Zeyer, J., Normand, P.: A hypervariable 23S rRNA region provides a discriminating target for specific characterization of uncultured and cultured *Frankia*. Systematic and Applied Microbiology **17**(3), 433–443 (1994)

[33] Huguet, V., Batzli, J.M.C.C., Zimpfer, J.F., Normand, P., Dawson, J.O., Fernandez, M.P., Lyon, C.B., Cedex, V.: Diversity and specificity of *Frankia* strains in nodules of sympatric *Myrica gale*, *Alnus incana*, and *Shepherdia canadensis* determined by rrs gene polymorphism. Applied Environmental Microbiology **67**(5), 2116–2122 (2001)

[34] Huguet, V., Gouy, M., Normand, P., Zimpfer, J.F., Fernandez, M.P.: Molecular phylogeny of Myricaceae: a reexamination of host-symbiont specificity. Molecular phylogenetics and evolution **34**(3), 557–568 (2005)

[35] Huguet, V., Mergeay, M., Cervantes, E., Fernandez, M.P.: Diversity of *Frankia* strains associated to *Myrica gale* in Western Europe: impact of host plant (*Myrica* vs. *Alnus*) and of edaphic factors. Environmental Microbiology **6**(10), 1032–1041 (2004)

[36] Jamann, S., Fernandez, M.P., Normand, P.: Typing method for N_2-fixing bacteria based on PCR-RFLP - application to the characterization of *Frankia* strains. Molecular Ecology **2**, 17–26 (1993)

[37] Jeong, S.C., Ritchie, N.J., Myrold, D.D.: Molecular phylogenies of plants and *Frankia* support multiple origins of actinorhizal symbioses. Molecular phylogenetics and evolution **13**(3), 493–503 (1999)

[38] Kennedy, P.G., Schouboe, J.L., Rogers, R.H., Weber, M.G., Nadkarni, N.M.: *Frankia* and *Alnus rubra* canopy roots: an assessment of genetic diversity, propagule availability, and effects on soil nitrogen. Microbial Ecology **59**, 214–220 (2010)

[39] Lipus, A., Kennedy, P.G.: *Frankia* assemblages associated with *Alnus rubra* and *Alnus viridis* are strongly influenced by host species identity. International Journal of Plant Sciences **172**(3), 403–410 (2011)

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

[41] Miller, M.A., Pfeiffer, W., Schwartz, T.: Creating the CIPRES Science Gateway for inference of large phylogenetic trees (2010)

[42] Mirza, B.S., Welsh, A., Rieder, J.P., Paschke, M.W., Hahn, D.: Diversity of frankiae in soils from five continents. Systematic and Applied Microbiology **32**(8), 558–570 (2009)

[43] Navarro, E., Jaffre, T., Gauthier, D., Gourbiere, F., Rinaudo, G., Simonet, P., Normand, P.: 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 (1999)

[44] Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J.O., Evtushenko, L., Misra, A.K.: Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. International Journal of Systematic Bacteriology **46**(1), 1–9 (1996)

[45] Perez, N.O., Olivera, H., Vasquez, L., Valdes, M.: Genetic characterization of Mexican *Frankia* strains nodulating *Casuarina equisetifolia*. Canadian Journal of Botany **77**, 1214–1219 (1999)

[46] Pokharel, A., Mirza, B.S., Dawson, J.O., Hahn, D.: *Frankia* populations in soil and root nodules of sympatrically grown *Alnus* taxa. Microbial Ecology **61**, 92–100 (2011)

[47] Roller, C., Ludwig, W., Schleifer, K.H.: Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes. Journal of General Microbiology **138**(6), 1167–1175 (1992)

[48] Rose, S.L.: Mycorrhizal associations of some actinomycete nodulated nitrogen-fixing plants. Canadian Journal of Botany **58**, 1449–1454 (1980)

[49] Schofield, R.K., Wormald Taylor, A.: The measurement of soil pH. Soil Science Society of America Journal **19**(2), 164–167 (1955)

[50] Sen, A., Duabin, V., Abrouk, D., Gifford, I., Berry, A.M., Normand, P.: 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 Nakamu. International Journal of Systematic and Evolutionary Microbiology (2014)

[51] Simonet, P., Navarro, E., Rouvier, C., Reddell, P., Zimpfer, J., Dommergues, Y., Bardin, R., Combarro, P., Hamelin, J., Domernach, A., Gourbiere, F., Prin, Y., Dawson, J.O., Normand, P.: Co-evolution between *Frankia* populations and host plants in the family *Casuarinaceae* and consequent patterns of global dispersal. Environmental Microbiology **1**(6), 525–533 (1999)

[52] Skead, C.J.: In: J.C. Manning, N.C. Anthony (eds.) Historical plant incidence in southern Africa, p. 378. South African National Biodiversity Institute, Pretoria (2009)

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

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

[55] Sukumaran, J., Holder, M.T.: DendroPy: a Python library for phylogenetic computing. Bioinformatics **26**, 1569–1571 (2010)

[56] Swofford, D.L.: PAUP. Phylogenetic analysis using parsimony (and other methods). Version 4 (2003)

[57] Torrey, J.G.: Cross-Inoculation Groups within *Frankia* and Host-Endosymbiont Associations. In: C. Schwintzer, J. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 83–106. Academic Press, Inc., San Diego, California (1990)

[58] Valdes, M., Perez, N., Estrada-de los Santos, P., Caballero-Mellado, J., Pena-Cabriales, J.J., Normand, P., Hirsch, A.M.: Non-*Frankia* actinomycetes isolated from surface-sterilized roots of *Casuarina equisetifolia* fix nitrogen. Applied Environmental Microbiology **71**(1), 460–466 (2005)

[59] Van Ryssen, F.W.J., Grobbelaar, N.: The nodulating and nitrogen fixing ability of South African *Myrica* species. South African Journal of Science **66**(1), 22–25 (1970)

[60] Vanden Heuvel, B.D., Benson, D.R., Bortiri, E., Potter, D.: 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). Canadian Journal of Microbiology **50**, 989–1000 (2004)

[61] Vandenbosch, K.A., Torrey, J.G.: Host-endophyte interactions in effective and ineffective nodules induced by the endophyte of *Myrica gale*. Canadian Journal of Botany **61**, 2898–2909 (1983)

[62] Welsh, A.K., Dawson, J.O., Gottfried, G.J., Hahn, D.: Diversity of *Frankia* populations in root nodules of geographically isolated Arizona Alder trees in central Arizona (United States). Applied and Environmental Microbiology **75**(21), 6913–6918 (2009)

[63] Welsh, A.K., Mirza, B.S., Rieder, J.P., Paschke, M.W., Hahn, D.: Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Systematic and Applied Microbiology **32**(3), 201–210 (2009)

[64] Whitman, W.B., Parte, A., Goodfellow, M., Kämpfer, P., Busse, H.J., Trujillo, M.E., Ludwig, W., Suzuki, K.: Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria. v. 5, pts. 1-2. Springer (2012)

[65] Zimpfer, J.F., Kennedy, G.J., Smyth, C.A., Hamelin, J., Navarro, E., Dawson, J.O.: Localization of *Casuarina*-infective *Frankia* near *Casuarina cunninghamiana* trees in Jamaica. Canadian Journal of Botany **77**(9), 1248–1256 (1999)

[66] Zitzer, S.F., Dawson, J.O.: Soil properties and actinorhizal vegetation influence nodulation of *Alnus glutinosa* and *Elaeagnus angustifolia* by *Frankia*. Plant And Soil **140**(2), 197–204 (1992)

3 Diversity of frankiae in root-associated soils from Cape *Morella*

Abstract

Diversity of *Frankia* in six soils associated with four Cape *Morella* species was investigated using clone libraries of *nif*H fragments ($n = 433$ clones). Analyses of these libraries revealed the presence of frankiae from the *Alnus* and *Elaeagnus* host infection groups, with the majority of clones ($n = 409$) assigned to previously defined Cape Clusters (CC). Of these, 254 clones were identical to sequences recovered from root nodules within the Cape region, but the presence of clusters in soil did not always precisely correspond to those found in nodules collected at the same site. Sequences representing CC-1 were absent from all soil libraries, despite this being the dominant *Alnus*-HIG subgroup in the region, and present in nodules collected from three of the six sampling locations. An additional *Alnus*-HIG subcluster was detected in three libraries, including both constructed from alkaline soils, but was never detected in Cape *Morella* nodules. These findings contrast sharply with those of previous studies in which nodular sequences have rarely corresponded to those from soil libraries. Simultaneously, they demonstrate that some frankiae lineages known to be abundantly present in nodules may not be detected in soils from which those nodules are collected.

3.1 Introduction

Frankia occur naturally in two ecological niches: either as microsymbionts within the root nodules of actinorhizal plant hosts or as free-living soil microorganisms [3,14,50]. The first indications that the organisms were not restricted to nodules came after inoculation of actinorhizal hosts with soil suspensions unexpectedly developed nodules, clearly indicating that *Frankia* was present in the inoculum [36]. *Frankia* populations in root nodules are usually enriched cultures of single strains; assessing the diversity of strains associated with a host is comparatively straightforward using PCR-based and in situ hybridization techniques [26,27,31,51,53,54]. In contrast to this, investigating *Frankia* ecology in soil is challenging for a variety of reason [2]. The total number of bacteria in soil communities is large, an early estimate and 1.5×10^{10} cells per gram of dry soil [49]. *Frankia* make up a comparatively small part of this population, with estimates varying from 1.7×10^3 [29] to 0.2×10^5 [33] or 1.6×10^6 cells per gram of soil [39]. This low abundance makes direct observation difficult, and until recently the principle method used to investigate *Frankia* soil populations has been the classical plant-trap assay. This method allows strains present in soil to nodulate an actinorhizal host, after which they may be recovered from the enriched nodular populations. Plant-trap bioassays have provided a wealth of ecological information; we know that *Frankia* is found not only in soils under actinorhizal hosts [42], but is also found where hosts have been absent for extended periods [4,52,56], under non-actinorhizal plants [12,19,23], in circumpolar soils [17], mine spoils [55], and in lake beds and stream sediments [18]. When used in combination with molecular techniques, trap-assays have shown that infective *Frankia* populations may change with soil depth [30], and that the number of genomic units (GU), a measure synonymous with cell numbers for filamentous organisms, in soil greatly exceed nodulating units (NU) estimated by trap-assay [29].

In contrast to the many reports of plant-trap assay studies, direct detection and enumeration of *Frankia* in soil using molecular methods is more seldom reported [14, 16, 29, 34]. Such studies include estimation of population sizes using qPCR [38–40] and in situ hybridization [23, 25, 27, 29], enumeration using booster-PCR [33], and estimation of population diversity using clone libraries [26, 32, 34]. Where these methods have been used in combination with trap-assays it has been found that

bioassays do not describe the true diversity of *Frankia* soil populations [26]. It is also apparent that the identity of the host species, even within promiscuous plant families, determines the identity of *Frankia* recruited [25]. Trap-assays may therefore reflect the symbiotic preferences of the host plant rather than the true diversity of *Frankia* soil populations [25]. *Frankia* populations in soil from natural stands of actinorhizal plants have rarely been directly investigated using molecular methods [14, 15, 34].

The purpose of this study was to explore the diversity of *Frankia* present in soils from natural stands of selected *Morella* species (*M. cordifolia*, *M. integra*, *M. diversifolia* and *M. kraussiana*) in natural settings and at six sites. Nodular *Frankia* diversity for hosts at these sites was determined in the previous study (Chapter 2), and found to be correlated with soil pH. In order to directly compare *Frankia* diversity in soil with that found in nodules, *nif*H gene fragments were used to create clone libraries. An attempt to directly classify *Frankia* soil populations demonstrated that when primers used in specific PCR amplification of Frankia in nodules was applied to soil the majority of recovered sequences were unrelated to the Frankiaceae [32]. As this was also found to be the case with the primer set used in my nodular diversity study, a semi-nested PCR approach was used to retrieve *Frankia*-specific sequences from soil. These amplicons were then used in the creation of clone libraries, which were analyzed using several phylogenetic methods and compared to sequences retrieved from databases and Cape *Morella* nodules. Soils were selected in order to specifically address the finding that Cluster I strains were only recovered from nodules of hosts growing in acidic sols.

3.2 Materials and Methods

3.2.1 Soil selection and sample preparation

Soils collected from six sites (detailed in Table 3.1) were selected based on a combination of the diversity of *Frankia* detected in their nodules, and soil pH. Soils selected were as follows: two coastal alkaline soils associated with *M. cordifolia* nodulated by Cluster III *Frankia* (used to create the West coast (WC) and Zandvlei

(ZV) libraries). Four acidic soils were selected: two associated with *M. diversifolia* (Bonteberg (BB) library) and *M. integra* nodulated by strains from clusters I and III (Cederberg (CW) library), one in which *M. diversifolia* was nodulated by Cluster III *Frankia* only (Table Mountain (TM) library) and finally a soil in which *M. kraussiana* was nodulated by a single Cluster I strain only (Steenberg (SB) library).

Methods which separate cells from the soil matrix prior to lysis do not work well for filamentous organisms such as *Frankia* [15], and so direct-lysis was employed. Five 0.25 g soil specimens were taken from 1 kg of well-mixed rhizosphere and bulk soil from each of six *Morella* root systems at the sites detailed in Table 3.1 using a Powersoil® extraction kit (MoBio, Carlsbad, CA, USA). Cells were disrupted in a Bullet Blender Storm®. Nucleic acids were eluted in 30 μl of sterile distilled water and stored at -20°C until use.

3.2.2 PCR conditions

Equimolar nucleic acid solutions from each individual soil were pooled and used as template for PCR amplification of 606 bp *nif*H gene fragments using primers nifHf1 (5′-GGC AAG TCC ACC ACC CAG C) and nifHr (5′-CTC GAT GAC CGT CAT CCG GC) [51] in a reaction volume of 50 μL. Each 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, 2.5 μL BSA (10 μg mL^{-1}), 0.2 μL ExTaq (5U μL^{-1}; TaKaRa) and 2 μL of template. An initial 5-min incubation at 95°C was followed by 35 rounds of temperature cycling (94°C for 30 s, 60°C for 30 s, 72°C for 45 sec) and a final 5-min extension at 72°C. Reaction products were analyzed by gel electrophoresis (1% w/v agarose gel in TAE buffer) and amplicons of the expected size (645 bp) were isolated from the gel and purified using an UltraClean® GelSpin® DNA Extraction Kit (MoBio, Carlsbad, CA, USA), eluted in 10 μl and then further diluted into 100 μL with sterile distilled water. One microliter of diluted PCR product was used as template for a semi-nested PCR reaction, using the same protocol described above, but with nifH269r (5′-CCG GCC TCC TCC AGG TA) replacing nifHr as reverse primer, and elongation time reduced to 30 s. Three amplifications were performed for each soil and the products pooled. PCR amplicons were analyzed by

gel electrophoresis (1.5% w/v agarose gel in TAE buffer, stained with ethidium bromide (0.5 μg mL^{-1})) and amplicons of approximately 280 bp were isolated from the gel and purified using and UltraClean® GelSpin® DNA Extraction Kit (MoBio, Carlsbad, CA, USA) with elution in 10 μL of sterile distilled water. Purified fragments were ligated into pDrive® (Qiagen) according to the manufacturer's instructions and transformed into *Escherichia coli* DH5α. Following blue/white selection, clones were analyzed by PCR for the presence of inserts of the expected size using the conditions described above, except that primers M13f and M13r were used. Nif*H* clones generated from *Frankia* isolates representing the major Cape clusters were used as individual positive controls, using the same semi-nested methodology used to create the soil clone libraries.

Reaction products containing amplicons of the correct size were purified with with ExoSAP-IT® 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, with incubation at 37°C for 45 min, followed by deactivation at 80°C for 15 min. Sequencing was performed at the Stellenbosch University Central Analytical Facility (http://academic.sun.ac.za/saf) in both directions using the M13f and M13r primers listed above. Chromatograms were visually assessed, corrected by hand where necessary, and assembled using CLC Main Workbench version 6.2.1 (Qiagen).

3.2.3 Phylogenetic analysis

Assembled sequences were identified by nucleotide BLAST analysis (Supplementary Table A.2). Unique *nif*H gene fragments from Cape *Morella* root nodules (n = 26, pure *Frankia* cultures from databases (n = 38) and unique sequences from each soil library (BB = 24, CW = 17, SB = 35, TM = 29, WC = 28, ZV = 38) were aligned and trimmed to 255 bp corresponding to positions 64 to 319 of the *Frankia* ACN14a *nifH gene* (NC008278), using CLC Main Workbench version 6.2.1(Qiagen)[1]. Following this identical sequences were once again removed, database sequences identical to either nodule or soil derived *nif*H fragments were

[1]The effect of trimming CFN sequences from 606 bp to 263 and 255 bp on uniqueness and assignment to Cape Clusters is illustrated in Figure A.1. The shorter sequence was selected to allow for the inclusion of more database sequences in the final analysis.

retained. The resulting six data sets (comprising 74, 69, 86, 78, 76 and 87 non-identical sequences, respectively) were analyzed using Neighbor Joining, Maximum Likelihood, Maximum Parsimony and Bayesian analysis as described previously [51]. Briefly, Neighbour joining was completed in PAUP (version 4.0b) [46]. Models of sequence evolution and set values for the proportion of invariant sites and gamma shape parameters were estimated for each individual data set using jModeltest2 (version 2:1:4) [9, 13]. Settings for jModeltest2 included 11 substitution schemes, 88 candidate models, rate variation I+G, nCat=4, and a ML-optimized base tree for likelihood calculation, with tree topology best of NNI and SPR. ML analysis was completed using the RAxML-HPC2 program on the CIPRES computer cluster (www.phylo.org) [24, 44]. 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" [46]. Bootstrapping included 1000 replicates and a full heuristic search. Bayesian analysis was carried out using MR-BAYES (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 [45], support measures from each tree concatenated on branches of a single NJ tree using purpose-written scripts, and resulting trees displayed in Figtree version 1.4 [37].

3.2.4 Clone coverage

Clone library coverage (C) was estimated as described by Mullins et al. [28]. Coverage estimates the proportion of a hypothetical library containing all target sequences/OTUs present in the starting material that would be contained within smaller real-world libraries, such as those created in this study. It is calculated as follows:

$$C = 1 - (n_1/N) \tag{3.1}$$

Where n_1 is the number of taxons occurring only once ($n = 1$), N is the total size of the library. Either unique clones or OTUs at a defined level of sequence similarity (such as *nif*H OTUs defined at 97% identity) may be considered individual taxons [43]. Coverage was calculated at both levels and is presented in Table 3.2.

3.2.5 Intra-library diversity assessment

The combined aligned library of unique 255 bp *nif*H gene fragments from all soil libraries (B=24, C=17, S=35, T=29, W=28, Z=38), as well as those from each individual library (seven data sets in total), were grouped into operational taxonomic groups (OTUs) using the average-neighbor algorithm in MOTHUR (version 1.32.0) and at every threshold the data could describe. Clone library sequences grouping at 97% identity were assigned to previously identified Cape Clusters (CC), *Alnus*-infective group AI [51], or to "Soil Groups" (SG) not corresponding to any previously described cluster represented by sequences included in the soil library data sets. Phylogenetic analyses were carried out for individual libraries as detailed in Section 3.2.3.

3.2.6 Inter-library diversity comparisons

Inter-library diversity was assessed by pairwise comparison of community membership (richness) and community structure; the analyses were performed in EstimateS (Version 9, R. K. Colwell, http://purl.oclc.org/estimates). Shared species for each pair of libraries were estimated using both Jaccard's and Sorensen's qualitative sample similarity indices. These indices take into account only the presence/absence of a taxon, and not its abundance [8]. Library structures were compared using both the Bray-Curtis (Sorensens' quantitative sample similarity) and the Morisita-Horn [22] similarity indices. The latter is highly sensitive to the abundance of the most abundant individual sequence or OTU, but is considered a superior measure of community structure as it is not unduly influenced by sample size or species richness [47]. Analyses were performed for both unique sequences and OTUs defined at 97% sequence identity using MOTHUR's average-neighbor algorithm.

3.3 Results

Amplicons of the expected size (approx. 650 bp) were obtained by PCR of nucleic acid extracts from all six soils. Using these PCR products as template, a semi-nested *nif*H gene-targeted PCR produced products of the expected size (approx 270 bp), which were used to generate gene clone libraries. Randomly selected clones (n=68-77 per library) containing inserts of the correct size contained partial frankiae *nif*H gene sequences, as demonstrated by BLAST searches in Genbank (Supplementary Table A.2) and by sequence classification in MOTHUR using a custom *nif*H taxonomic database. The specificity of the nifHf1/nifHr269 primer pair for frankiae has previously been demonstrated [25, 34]; in-silico analyses demonstrated that this primer set also primes all frankiae *nif*H gene sequences detected in *Morella* nodules in the current study (results not shown).

Of the 433 clones from six soil libraries, 254 (a majority of 58%) were identical to sequences recovered from Cape *Morella* nodules (Table 3.5. The number of unique nodular sequences, when trimmed to 255 bp, was reduced from 26 to 17, with several distinct sequences being identical over the length of the shorter fragment (Table 3.5). As a result of this eight of the 26 sequences recovered from nodules, and possibly as many as 14, were detected in soil libraries, although not always in the same soil from which the nodular sequences were recovered (Figures 3.5, 3.6, 3.7, 3.8, 3.9 and 3.10).

When all six libraries were considered a total of seven genotypes (OTUs defined at 97% similarity) were detected. Of these, two were assigned to the *Elaeagnus* and five to the *Alnus* host infection groups, respectively (Tables 3.3 and 3.4). Three clusters corresponded to Cape Genotypes recovered from *Morella* nodules; one (CC-2) assigned to the *Alnus*-infective cluster, and CC-3 and CC-4 to the *Elaeagnus*-infective cluster. Sequences assigned to these three clusters comprised the majority of the combined libraries (409 of 433 clones). The remaining 24 sequences were assigned to four clusters, with the majority (12) clustering within AI as defined by Welsh et al. [51] (Supplementary Table A.2). Sequence W14 was identical to a *nif*H gene of *Frankia* CPI1, isolated from *Comptonia peregrina* [5]. Cluster AI *Frankia* were not detected in nodules from Cape *Morella* species, but

were present in three soils (Table 3.4). Conversely, CC-1 was absent from all libraries, despite its presence in nodules from three of the sites (SB, CW and BB. Figures 3.7, 3.6 and 3.5). As positive controls created from FMi1 gDNA performed as expected, the absence of this cluster from the soil libraries cannot be ascribed to methodological error, at least where PCR is concerned.

Pairwise comparisons of unique *nif*H sequences shard between clone libraries are presented in Figures 3.1 and 3.2). Pairwise comparisons of soil genotypes are presented in Figures 3.3 and 3.4.

Table 3.1: Details of sites from which soils were collected for creation of clone libraries. Hosts present at each site, library designation, soil pH (mean of three readings) by the $CaCl_2$ method, and *Frankia* host infection groups found in nodules from each site are indicated. Sites (†) are indicated in Figure 1.3.

Species	Site†	Library	Latitude	Longitude	pH ($CaCl_2$)	HIG
M. cordifolia	A1.5	ZV	33°13'40.00" S	18°9'19.30" E	7.70 (0.24)	III
	B4	WC	34°4'54.50" S	18°27'59.00" E	8.16 (0.02)	III
M. diversifolia	B2	BB	33°58'16.90" S	18°25'15.20" E	3.45 (0.01)	I/III
	B1.2	TM	34°13'13.60" S	18°22'49.70" E	4.48 (0.03)	III
M. kraussiana	B3.2	SB	34°5'34.60" S	18°25'59.60" E	4.48 (0.03)	I
M. integra	A3.1	CW	32°6'55.00" S	19°3'54.30" E	4.17 (0.06)	I/III

Table 3.2: Soil library metrics detailing numbers of clones per library as well as sequences or genotypes (OTUs defined at 97% similarity) represented only once in each library. Coverage as calculated in Equation (3.1) is reported for both unique sequences and genotypes. Genotype totals across all combined libraries are indicated by (†).

soil library	total clones	unique sequences	unique singletons	coverage (C)	genotypes 97%	genotype singletons	genotype coverage (C)
BB	76	24	21	0.72	4	1	0.99
CW	77	17	15	0.81	1	0	1.0
SB	68	35	30	0.56	3	0	1.0
TM	69	29	23	0.67	2	0	1.0
WC	74	28	24	0.66	4	1	0.99
ZV	69	38	29	0.55	3	0	1.0
Totals	433	155	80		7†	2	

Table 3.3: Genotypes (*nif*H gene fragments clustered at 97% sequence similarity) shared between soils and nodules collected at the same sites. Phylogenetic analysis assigned CC-2 to the *Alnus*-HIG, and CC-3 and CC-4 to the *Elaeagnus*-HIG, as indicated in Chapter 2.

	CC-2		CC-3		CC-4	
Site	nod	soil	nod	soil	nod	soil
BB	8	60	2	13	-	-
CW	-	-	2	-	-	77
SB	-	-	-	42	-	22
TM	-	-	9	57	-	12
WC	-	-	4	62	-	7
ZV	-	-	13	30	22	29

Table 3.4: Genotypes (*nif*H gene fragments clustered at 97% sequence similarity) unique to either soils or *Morella* nodules collected at the same sites (Tables 1.3 and 2.3). All sequences unique to soil libraries were assigned to the *Alnus*-HIG, either to cluster AI or to soil genotypes not clustering within previously identified groups.

	Unique nodule genotypes				Unique soil genotypes			
Site	CC-1	CC-5	CC-6	CC-7	AI	SG-1	SG-2	SG-3
BB	7	-	1	-	-	2	1	-
CW	16	-	-	-	-	-	-	-
SB	3	-	-	-	4	-	-	-
TM	-	4	-	-	-	-	-	-
WC	-	-	-	-	4	-	-	1
ZV	-	-	-	1	10	-	-	-

Table 3.5: Soil derived 255 bp *nif*H sequences identical to those recovered from *Morella* nodules. Partial *nif*H gene fragment from soil libraries identical to nodule-derived sequences. Representative clones from each library are indicated, with counts per library given in brackets (see also Supplementary Table A.2). Nodular *nif*H sequences sharing 100% sequence identity following reduction from 606 bp to 255 by are indicated†.

	Soil clone library sequences						
Nodular sequences†	BB	CW	SB	TM	WC	ZV	Total
CFN5	B20 (6)	-	S2 (17)	T1 (35)	W6 (41)	Z3 (10)	109
CFN2,CFN4,CFN18	-	C58 (58)	S9 (13)	T19 (3)	W12 (4)	Z29 (4)	82
CFN10,CFN16	B3 (46)	-	-	-	-	-	46
CFN9,CFN26	-	-	-	-	-	Z8 (6)	6
CFN14	-	-	-	T23 (1)	W4 (3)	-	4
CFN8	B26 (3)	-	-	-	-	-	3
CFN11,CFN13	-	-	-	-	W21 (2)	-	2
CFN7,CFN17	-	-	-	T4 (2)	-	-	2
Totals	55	58	30	41	50	20	254

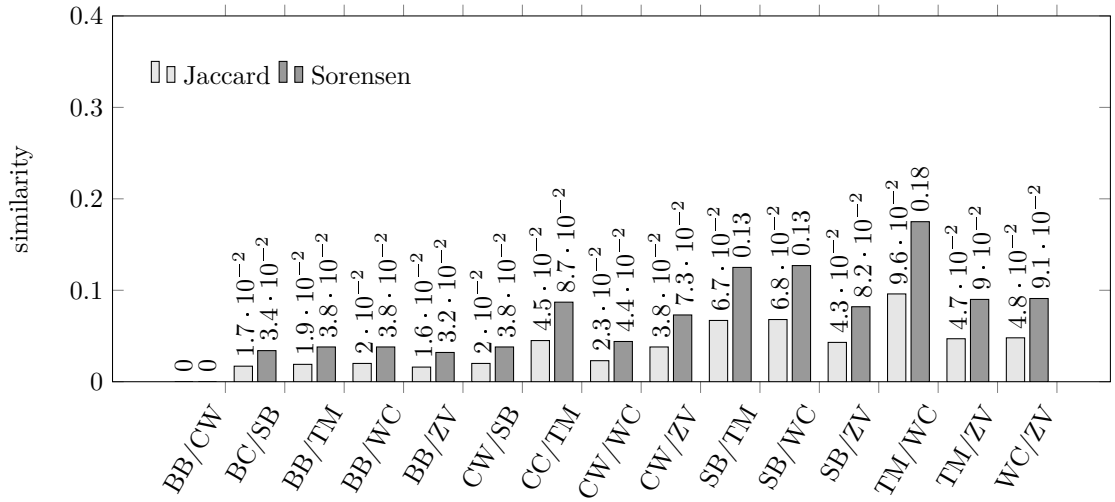


Figure 3.1: Pairwise comparisons of *nif*H gene fragment community membership (sequence presence/absence) in soil libraries, computed at distance "unique". Values range from 0 to 1, higher values indicate greater community similarity according to the Jaccard and Sorensen indices.

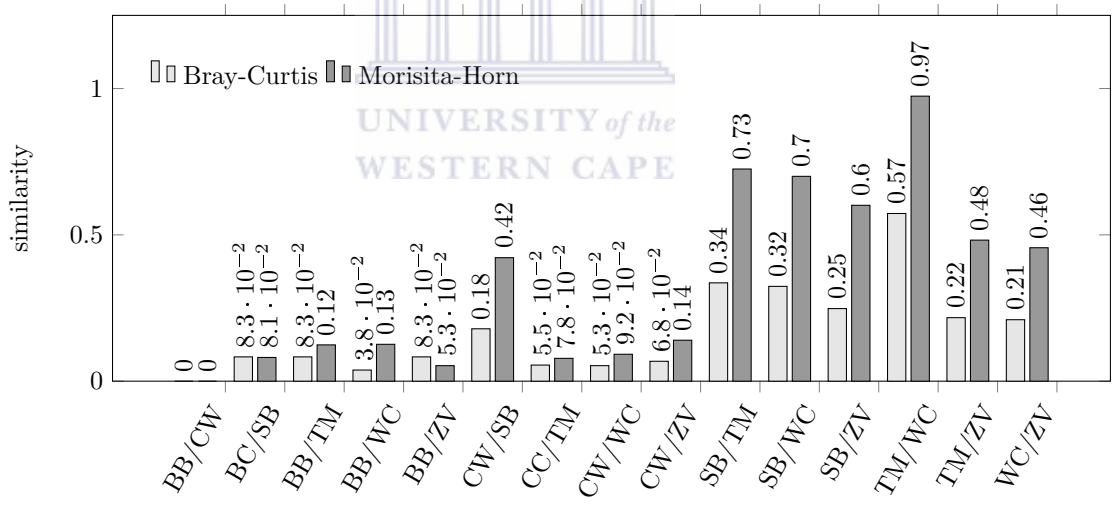


Figure 3.2: Pairwise comparisons of *nif*H gene fragment soil library structure (sequence presence/absence and relative abundance), computed at distance "unique". Values range from 0 to 1, higher values indicate greater community similarity according to the Bray-Curtis and Morisita-Horn indices.

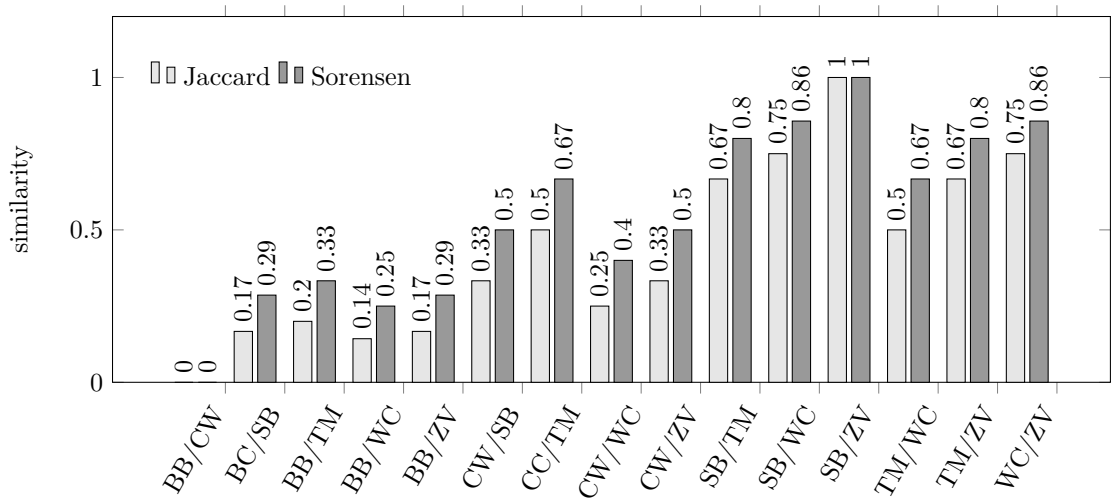


Figure 3.3: Pairwise comparisons of *nif*H gene fragment soil library membership (cluster presence/absence), with clusters defined at 97% identity. Values range from 0 to 1, higher values indicate greater community similarity according to the Jaccard and Sorensen indices.

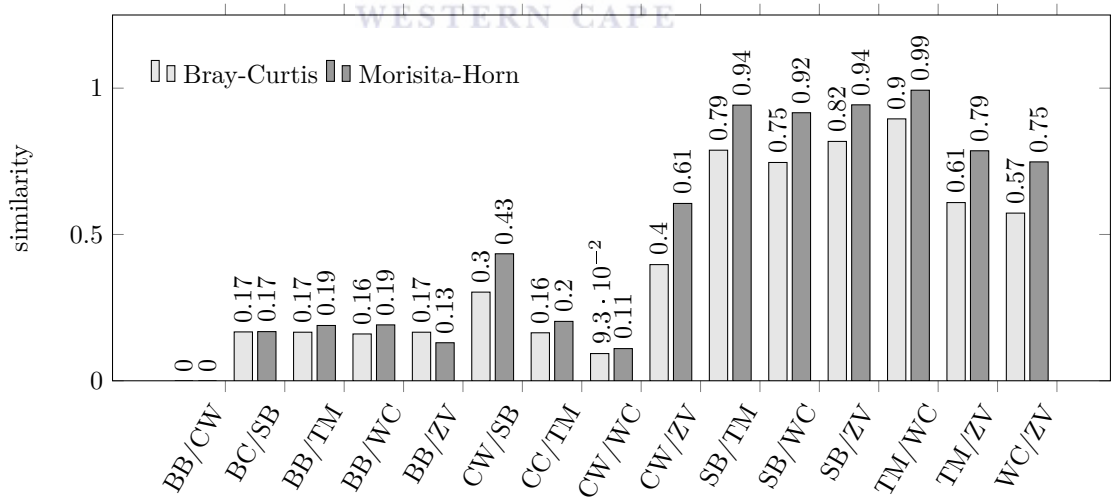


Figure 3.4: Pairwise comparisons of *nif*H gene fragment soil library structure (cluster presence/absence and relative abundance), computed at distance "97%". Values range from 0 to 1, higher values indicate greater community similarity according to the Bray-Curtis and Morisita-Horn indices.

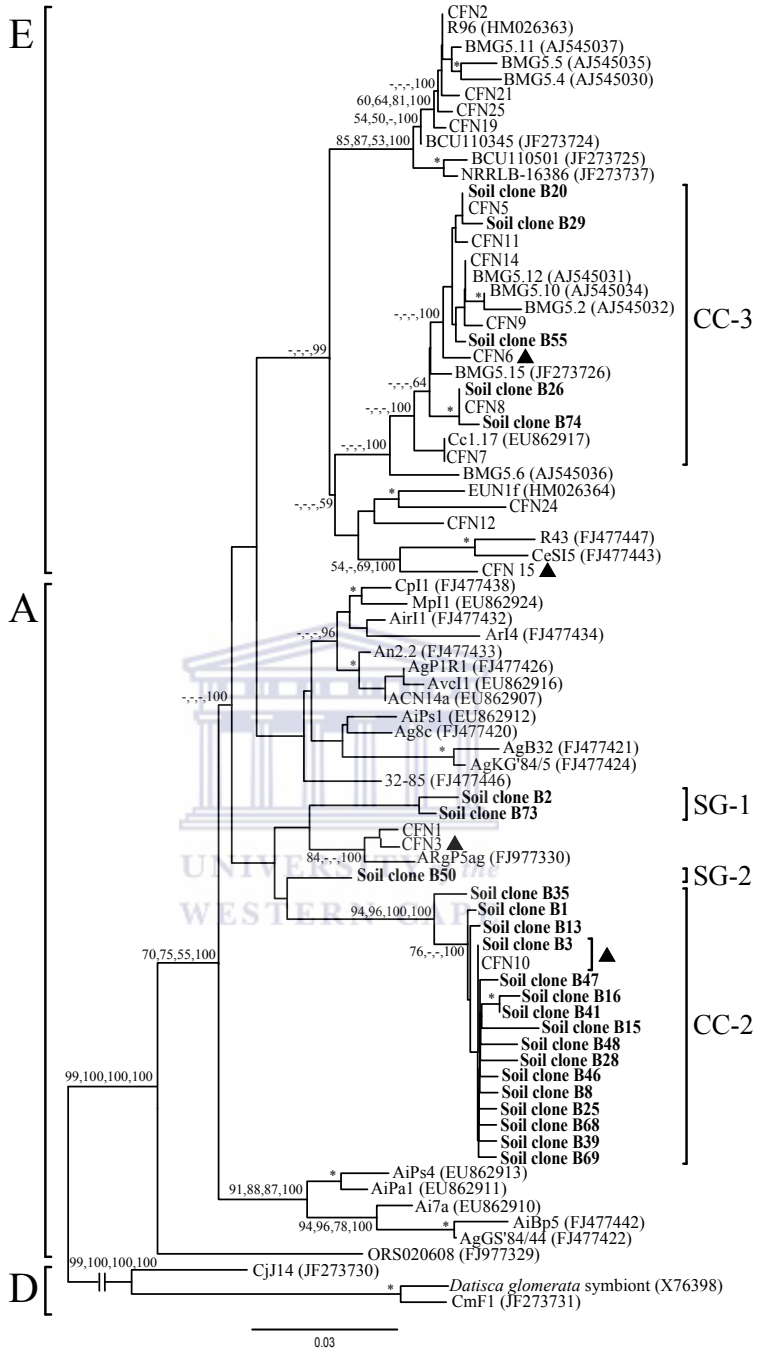


Figure 3.5: NJ tree of 255 bp *nif*H fragments from clone library created from *M. diversifolia* rhizosphere soil collected at site B1 (Bonteberg, BB). CFR nitrogenase fragments recovered from rhizosphere soil are indicated in bold. Sequences found in *M. diversifolia* nodules from this site are indicated with triangles (▲). Database sequences are labeled with strain designation and GenBank accession number. Soil library genotypes defined at 97% similarity are indicated in the right margin. "CC" indicates Cape Clusters. "SG" (soil genotypes) denotes genotypes recovered from soil but never from CFR nodules. The three major *Frankia* infectious clusters are indicated in the left hand margin. A: *Alnus*-infective/Cluster I, E: *Elaeagnus*-infective/Cluster III, D: *Datisca*-infective/Cluster II. Support from Neighbor Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (B) analyses are indicated on nodes.

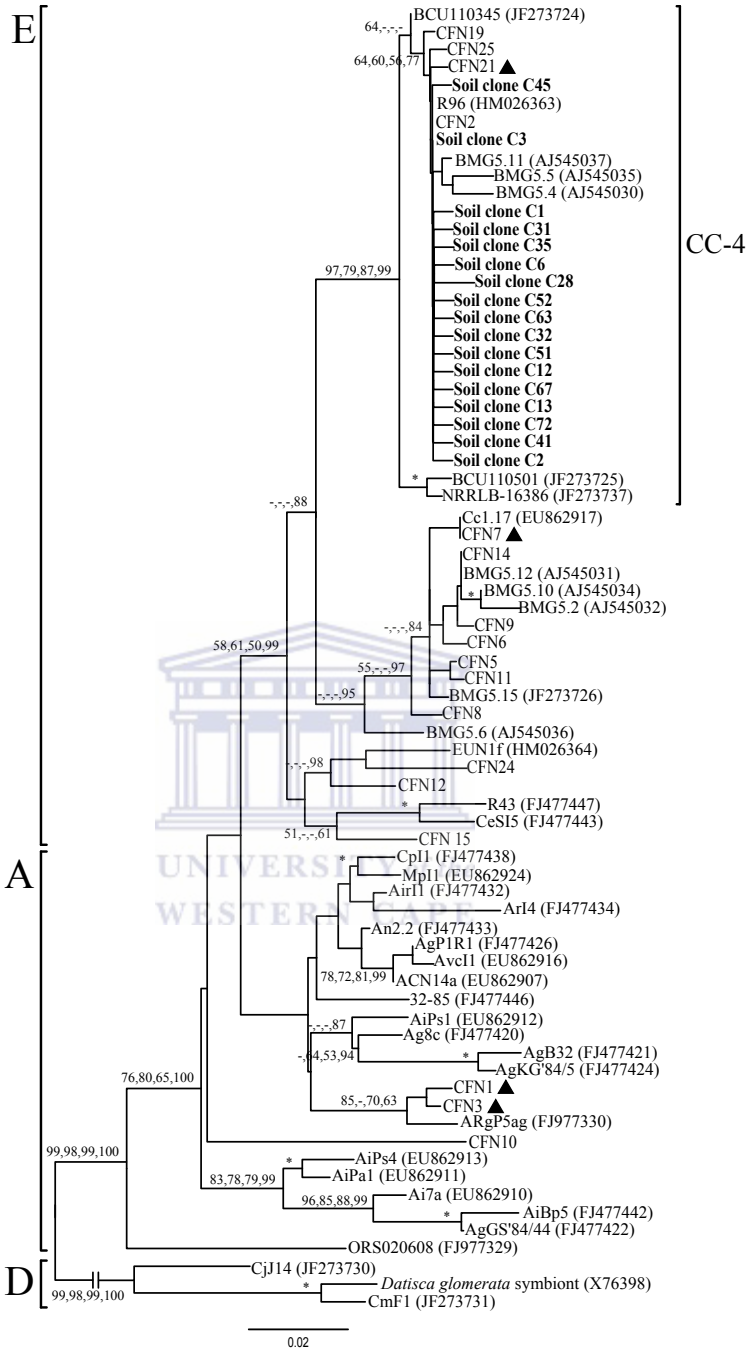


Figure 3.6: NJ tree of 255 bp *nif*H fragments from clone library created from *M. integra* rhizosphere soil collected at site A3 (Cederberg, CW). CFR nitrogenase fragments recovered from rhizosphere soil are indicated in bold. Sequences found in *M. integra* nodules from this site are indicated with triangles (▲). Database sequences are labeled with strain designation and GenBank accession number. Soil library genotypes defined at 97% similarity are indicated in the right margin. "CC" indicates Cape Cluster. The three major *Frankia* infectious clusters are indicated in the left hand margin. A: *Alnus*-infective/Cluster I, E: *Elaeagnus*-infective/Cluster III, D: *Datisca*-infective/Cluster II. Neighbor Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (B) analyses are indicated on nodes.

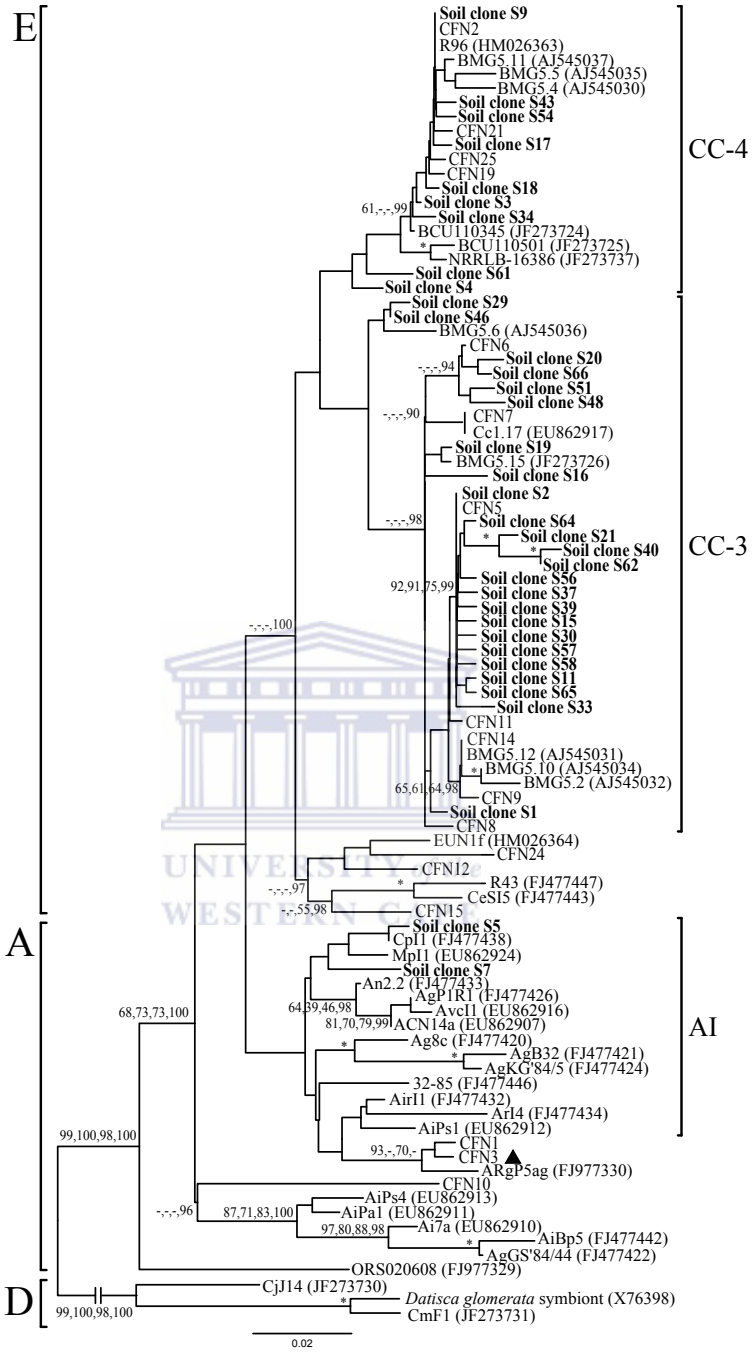


Figure 3.7: NJ tree of 255 bp *nif*H fragments from clone library created from *M. kraussiana* rhizosphere soil collected at site B3 (Steenberg, SB). CFR nitrogenase fragments recovered from rhizosphere soil are indicated in bold. The sequence found in *M. kraussiana* nodules from this site is indicated with a triangle (▲). Database sequences are labeled with strain designation and GenBank accession number. Soil library genotypes defined at 97% similarity are indicated in the right margin. "CC" indicates Cape Cluster. The three major *Frankia* infectious clusters are indicated in the left hand margin. A: *Alnus*-infective/Cluster I, E: *Elaeagnus*-infective/Cluster III, D: *Datisca*-infective/Cluster II. Support from Neighbor Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (B) analyses are indicated on nodes.

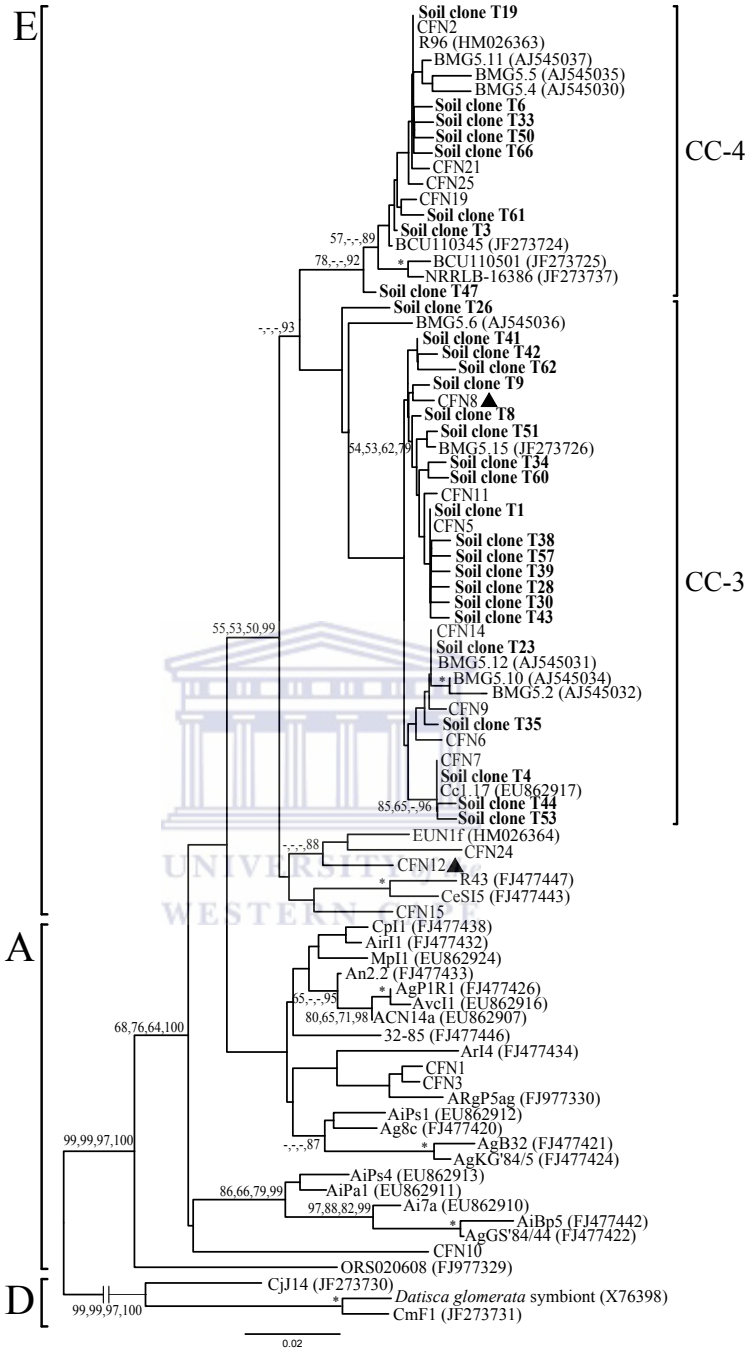


Figure 3.8: NJ tree of 255 bp *nif*H fragments from clone library created from *M. diversifolia* rhizosphere soil collected at site B2 (Table Mountain, TM). CFR nitrogenase fragments recovered from rhizosphere soil are indicated in bold. Sequences found in *M. diversifolia* nodules from this site are indicated with triangles (▲). Database sequences are labeled with strain designation and GenBank accession number. Soil library genotypes defined at 97% similarity are indicated in the right margin. "CC" indicates Cape Cluster. The three major *Frankia* infectious clusters are indicated in the left hand margin. A: *Alnus*-infective/Cluster I, E: *Elaeagnus*-infective/Cluster III, D: *Datisca*-infective/Cluster II. Support from Neighbor Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (B) analyses are indicated on nodes.

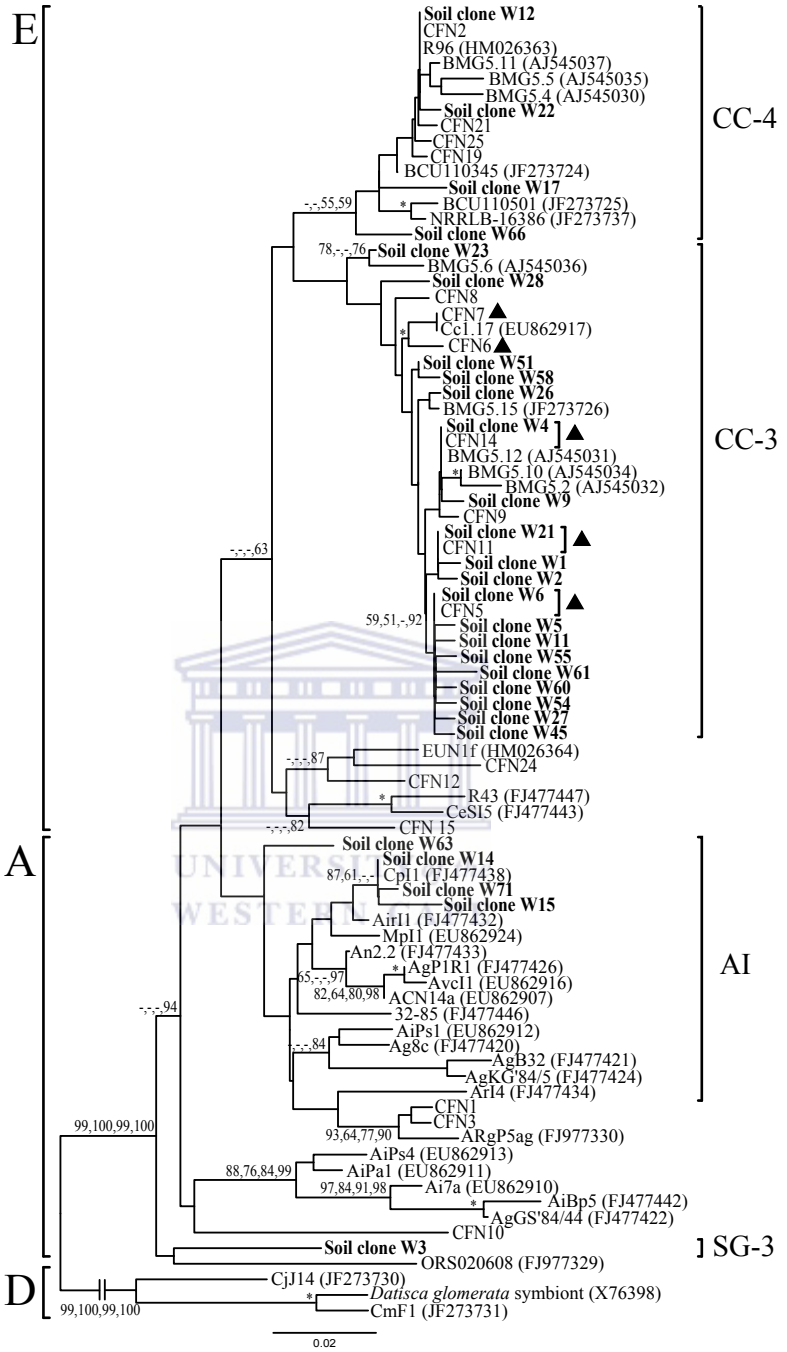


Figure 3.9: NJ tree of 255 bp *nif*H fragments from clone library created from *M. cordifolia* rhizosphere soil collected at site A1 (West Coast National Park, WC). CFR nitrogenase fragments recovered from rhizosphere soil are indicated in bold. Sequences found in *M. diversifolia* nodules from this site are indicated with triangles (▲). Database sequences are labeled with strain designation and GenBank accession number. Soil library genotypes defined at 97% similarity are indicated in the right margin. "CC" indicates Cape Cluster. The three major *Frankia* infectious clusters are indicated in the left hand margin. A: *Alnus*-infective/Cluster I, E: *Elaeagnus*-infective/Cluster III, D: *Datisca*-infective/Cluster II. Support from Neighbor Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (B) analyses are indicated on nodes.

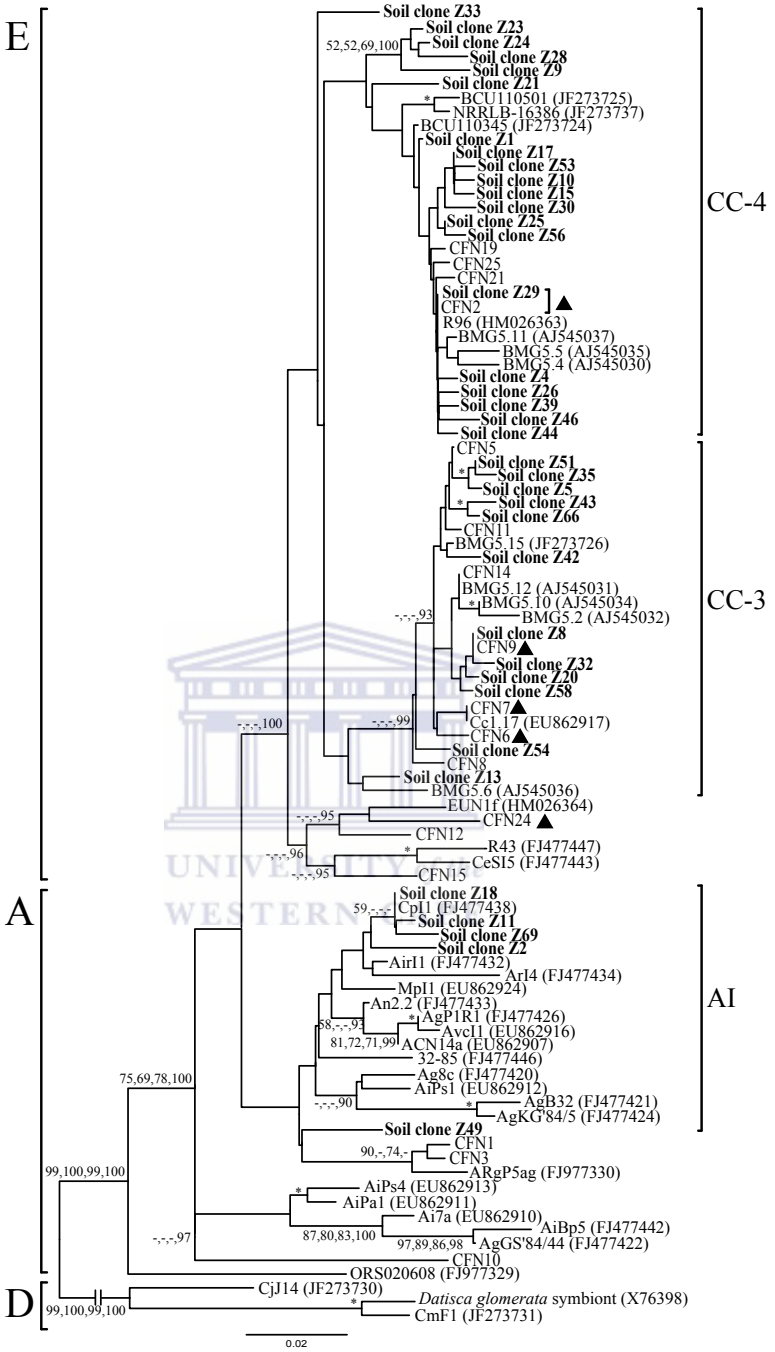


Figure 3.10: NJ tree of 255 bp *nif*H fragments from clone library created from *M. cordifolia* rhizosphere soil collected at site B4 (Zandvlei, ZV). CFR nitrogenase fragments recovered from rhizosphere soil are indicated in bold. Sequences found in *M. cordifolia* nodules from this site are indicated with triangles (▲). Database sequences are labeled with strain designation and GenBank accession number. Soil library genotypes defined at 97% similarity are indicated in the right margin. "CC" indicates Cape Cluster. The three major *Frankia* infectious clusters are indicated in the left hand margin. A: *Alnus*-infective/Cluster I, E: *Elaeagnus*-infective/Cluster III, D: *Datisca*-infective/Cluster II. Support from Neighbor Joining (NJ), Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian (B) analyses are indicated on nodes.

3.4 Discussion

Clone libraries containing *Frankia nif*H gene fragments were created from nucleic acid extracts of soils collected from the roots systems of specific *Morella* populations[2] in which all seven of the *Frankia* clusters detected in Cape nodules were represented (Tables 3.1, 3.3 and 3.4). As *Alnus*-HIG *Frankia* strains were found in nodules from acidic soil only, while those from the *Elaeagnus*-HIG were also found in neutral and alkaline soils (Chapter 1), the soils were selected in such a way as to reflect the various combinations of soil pH and presence in nodules of either one or both *Frankia* HIGs at well-separated sites (Table 3.1). Frankiae diversity within individual soil libraries could then be compared both between libraries and in light of strains known to be present in each respective soil, as evidenced by their presence in root nodules from the collection site.

Library construction started with amplification of a 606 bp fragment followed by semi-nested PCR producing a 263 bp fragment. A reaction producing a longer PCR product (400 bp) was rejected as the reverse primer failed to prime most of the Cape *nif*H sequences generated during the nodule survey (results not shown). This emphasizes the necessity of conducting preliminary analysis in newly-investigated environments if sequence information is available, as many of the *Alnus*-HIG sequences detected in the nodular diversity study described in Chapter 2 would not have been detected using this primer set. Amplicons of the expected size, created with the nifHf1/nifHr269 primer set, were cloned into pDrive. Clones selected at random ($n = 68 - 77$ clones per library) were identified as *nif*H sequences from *Frankia* by BLAST analysis in Genbank (Supplementary Table A.2). The specificity of the nifHf1/nifHr269 primer pair for *Frankia* sequences in soil has previously been demonstrated [26, 34]. In order to confirm that Cape *Frankia* from both clusters I and III could be amplified by the semi-nested methodology, the nif*H* gene was cloned from each of the isolates (Chapter 4) and used individually as positive controls. All controls performed as expected.

Cluster analysis assigned soil sequences from the combined library data set to

[2]It was impossible to identify individual plants in certain populations of *M. diversifolia*, *M. quercifolia* and *M. cordifolia* because of the density of surrounding brush.

a total of seven clusters, defined at 97% similarity. The majority of sequences (411 of 433) were assigned to three previously-described clusters (CC-2, CC-3 and CC-4) dominant in nodules from Cape *Morella* (Tables 3.3 and3.4). Four additional clusters (defined at 97% sequence similarity) not found in Cape nodules and represented by 24 sequences were detected in four of the six soils (Table 1.2). Of sequences assigned to Cape Genotypes, a majority (254 of 409) were identical to nodular sequences (Table 3.5). Multiple identical sequences being present in each respective library (see Supplementary Table A.2) It is possible that at least some of the closely related sequences present as singletons or in low abundance within each library are the result of PCR artifacts. This would have the effect of artificially inflating the apparent number of strains when unique sequences are considered. The use of clusters/OTUs constrains this apparent higher/artefactual diversity [1], as more than 7 substitutions would be necessary in order for a 255 bp sequence to be assigned to another cluster at the defined threshold.

Phylogenetic analysis, following trimming of the 263 bp fragments to 255 bp, confirmed the positions of these clusters within the host infection groups, and the positions of strains and nodular sequences and overall tree topology remaining largely consistent regardless of the phylogenetic method employed (see Figures 3.5, 3.6, 3.7, 3.8, 3.9 and 3.10). Reduction in the size of nifH fragments (522 bp for nodular and 255 bp for soil library sequences) has previously been shown to impact phylogenetic analyses minimally [27]. Discrepancies between trees in the position of some shared strains or groups, such as ARgP5 and the closely related Cape sequences CFN1 and CFN3 (as in Figures 3.6 and 3.7 for example), could therefore be attributed to long branch attraction. Individual Clusters AI, SG-1, SG-2 and SG-3 present only in soil were all assigned to the *Alnus*-HIG. Of these rarely-detected sequences 18 of 22 fell into *Frankia* group AI [51]. Group AI (containing strains ACN14a, Arl3, CpI1 and Mpl3) has previously been detected in soils from five continents and actinorhizal nodules from Europe, Asia and North America [27, 51]. The remaining three clusters contained 4, 1 and 1 sequences respectively (Tables 3.3 and 3.4). Due to low support from all four phylogenetic methods the exact positions of these SG (Soil Genotype) clusters within the *Alnus*-HIG could not be ascertained with any certainty (Figures 3.5 and 3.9). Clone W3, for instance, clusters with a *Casuarina*-infective strain but with low support and sequence similarity (Figure 3.9.

Database entries with which the SG sequences (B2, B50, B73 and W3) had highest identity are presented in Supplementary Table A.2.

When individual sites are considered, and soil libraries and nodular microsymbionts compared, the following picture emerges: generally libraries were dominated by sequences CC-3 and CC-4 from the *Elaeagnus*-HIG, although clusters present in nodules were not always represented in soils (Table 3.3. The CC-1 genotype was absent from all libraries, most notably from library CW (Figure 3.6). *Morella integra* from this site was nodulated mostly by strains from CC-1 (16 of 18 nodules). While the library consisted solely of sequences assigned to CC-4, the remaining two nodules contained CC-3 sequences only. None of the soil sequences represent clusters nodulating *M. integra* at this site (Table 3.3). Similarly library SB, created from soil associated with *M. kraussiana*, was dominated by CC-3 and contained three other genotypes, including sequences representing *Frankia* cluster AI. Only six nodules from *M. kraussiana*, three from the SB site, yielded *nif*H sequences. These were always identical (CFN3, Tables 1.2 and 1.3), despite the nodules being collected at three widely separated sites (Table 1.1, Figure 1.2). From this observation it appeared that this species has strong microsymbiont preferences, but this was not reflected in the SB soil library.

Where the *Alnus*-HIG is concerned, clone library BB stands in contrast to libraries CW and SB. This library, created from soil recovered from a *Morella diversifolia* root environment, contained four genotypes and was the only library to contain an *Alnus*-HIG genotype (CC-2) related to those found in nodules from the *Morella* nodule data set (Table 1.3). This genotype was only associated with plants from this location and was restricted to *M. diversifolia*, despite *M. kraussiana* also being present at the site. The library was dominated by clone B4 ($n = 46$), which was identical to CFN10 found in 8 of 18 nodules from this population (Figure 3.5, Supplementary Table A.2). CC-1, CC-3 and CC-4 were also present in nodules, but only the CC-3 was represented in the library, and in low abundance. Two rare genotypes, SG-1 and SG-2 were also found in this library but their position within the *Alnus*-HIG is unclear.

Library TM, created from an acidic soil in which *M. diversifolia* nodulated by

Elaeagnus-HIG *Frankia* only was growing, contained sequences from two major Cape Clusters (Figure 3.8, Table 3.3). Sequences from CC-3 were most abundant, with T35 being represented 35 times (Supplementary Table A.2). This agrees with what was found in root nodules at this site as 9 of 13 nodular sequences were assigned to this cluster (Table 1.3), although none of the soil clones were identical to those found in nodules. CC-4 was also present in the soil library but absent from nodules, whereas *nif*H gene fragments from CC-5 were present in nodules but not in soils (Table 3.4).

The remaining soil libraries, WC and ZV, were both created from alkaline dune soils collected from the root systems of *M. cordifolia* (Table 3.1. This host was the most well-sampled in terms of both number of nodules collected and the number of sites examined, which extended across half of its range (Table 1.1, Figure 1.2). It is notable in that it was, alone among the Cape *Morella*, nodulated by strains from the *Elaeagnus*-HIG exclusively. It was also only found in neutral to alkaline soils; *Alnus*-HIG strains were never detected in nodules from these soils, regardless of host species (Chapter 1). *Elaeagnus*-HIG cluster CC-3 and CC-4 were well represented in both libraries (Table 3.3), and both were present in nodules from the Zandvlei (ZV) site. These libraries shared sequences from cluster AI sequence (clones W14 and Z18, Supplementary Table A.2, Figures 3.9 and 3.10) identical to the *nif*H of *Frankia* strain CpI1. Another clone, W3, clustered with *Casuarina*-infective *Frankia*, but with low support; probable evidence of long-branch attraction considering the comparatively large sequence dissimilarity.

If we accept the assertion that, ignoring potentially complicating factors such as biases in DNA extraction [11] and PCR [1, 35], clone libraries accurately represent the diversity and relative abundance of the most abundant microorganisms in an environment under investigation (there is evidence both for and against this [7, 21, 48]), the structure and diversity of the respective soil libraries may be compared. Firstly, clone coverage demonstrated that while the number of unique sequences in each respective soil library was comparatively low (between 55 and 81 percent of total expected in a library of infinite size), each library was complete with respect to the number of genotypes it contained (Table 3.2). As CC-1 was absent from all soil libraries but definitely present in the soils, as evidenced by its presence

in nodules from three of the sites (Table 3.3), it is obvious that estimations of genotype coverage present a misleading picture and must be interpreted cautiously.

Comparisons were made taking both presence/absence of individual sequences and clusters (defined at 97% sequence similarity) between libraries (Figures 3.1 and 3.3), as well as community similarity at these two thresholds (Figures 3.2 and 3.4). In terms of genotype presence libraries TM, SB, WC and ZV were most similar according to both Jaccard and Sorensen metrics, which reflects the presence of identical OTU's in all libraries. This is also apparent from Table 3.5, as numerically dominant sequences were identical to nodule sequences and shared between these libraries. Bray-Curtis and Morisita-Horn measures of cluster presence and relative abundance also demonstrated these soil communities to be similar. Most dissimilar were the CW and BB libraries, which shared no genotypes. This is surprising as *M. integra* and *M. diversifolia* from these sites were each nodulated by CC-1 and CC-3, and shared a sequence from each (CFN3 and CFN6.

It is clear that nodulation is not correlated with the population size of strains available in soil. Discrepancies between nodular and soil populations as evidenced by clone libraries have been highlighted in the past [27, 34], and interpretation of results is complicated by the fact that nodules may be induced by individual hyphae, whole colonies or individual spores [27]. Infectious hyphae from symbiotic strains may be present at lower titers than strains better adapted to life as saprophytes in the environment under investigation. Alternatively, infectious spores may not be amenable to lysis. MoBio's PowerSoil kit (used in the current study) effectively extracts *Frankia* DNA from soil [39], and this was most recently demonstrated in meta-genomes created from Pleistocene permafrost [20]. The kit has also been shown to effectively extract DNA from *Bacillus* spores introduced into soil [10]. It hasn't been specifically tested on *Frankia* spores, which are possibly the only form strains from cluster CC-1 take outside of host nodules.

The absence from clone libraries of sequences known to be present in nodules originating from the same soils raises questions as to their usefulness in assessing the real diversity of *Frankia* in soil. As analysis of clone coverage did not indicate that additional clusters were to be expected, the absence of a dominant Cape

genotype (CC-1) would not have been suspected were evidence for its existence not available from another source (root nodules). This highlights difficulties often reported in the analysis of *Frankia* populations in soil, particularly in relating actively growing (and therefore numerically dominant) soil populations to infectious populations responsible for nodulating hosts in that soil. The presence in soil of *nif*H sequences related to frankiae not detected in nodules within the wider environment (in this case the Cape flora) suggests the presence of saprophytic populations worthy of further investigation.

3.5 Conclusion

The diversity of *Frankia* populations in soils associated with natural stands of actinorhizal plants has seldom been directly investigated. *Frankia* occupy two niches, namely root nodules and soil. The effects of both predominant environmental conditions and the genetics of both host and microsymbiont are factors known to influence nodulation and the population dynamics of *Frankia* in soil. The situation is further complicated by the varied ability of different *Frankia* groups to grow saprophytically, and to use carbon substrates available in the form of decaying plant material [25]. These range from the inability of *Casuarina*-infective strains to persist in soils absent their hosts, to members of the *Elaeagnus*-HIG which are cosmopolitan [6, 25, 26, 41]

While progress has recently been made towards understanding *Frankia* ecology in soil habitats, the questions of what proportion of the total *Frankia* soil population is infectious, and whether infectious *Frankia* found in the nodules of actinorhizal plants in their natural environments are actively growing components of soil microbial communities, remain open. Where soil clone libraries have been used to address these questions in the past populations in root nodules have been found to differ from those detected in soil [27, 34].

Soil and nodule populations of *Elaeagnus*-HIG strains in Cape habitats were to a large extent in agreement with the dominant clusters in nodules, (CC-3 and CC-4) being identical to those found in the six soils investigated. The situation among *Alnus*-HIG strains is less clear. The dominant *Alnus*-HIG genotype in nodules

of Cape *Morella* is CC-1, but this genotype was conspicuously absent from soil clone libraries, even at sites where the genotype was present in nodules. Whether this was because these strains were completely absent from the soils in question, whether they were out competed by *Elaeagnus* strains (as has been found to occur in moist soils [40]), or were present as spores which survived lysis is unknown. Unfortunately, as this genotype was not detected in any of the soils its absence from nodules collected from neutral and alkaline soils could not be commented upon, despite the experimental design being set up to specifically address this issue. What is clear is that this genotype would have remained undetected if its presence in the environment had not already been established by a survey of nodular diversity. In sharp contrast to CC-1, another cluster from the *Alnus*-HIG (CC-2) was detected in nodules from only one *Morella* species in the Cape, and was dominant in the clone library made from soil collected from the site at which it grew. This suggests different saprophytic potentials for clusters of *Frankia*, even within the same host infection group. Furthermore, the detection of *nif*H sequences identical to those from *Alnus*-HIG subgroup AI in soils, and the absence of strains from this subgroup from Cape *Morella* nodules raises questions as to the infectious capability of strains from this cluster (which has been detected in soils collected globally).

The ecology and population dynamics of *Frankia* in soil are not easily investigated. Future studies should focus on determination of absolute population sizes and growth states of different specific clusters in soil through the use of methods like qPCR and in situ hybridization. Furthermore, *Frankia* isolates from the Cape environment should be used in microcosm experiments to determine population dynamics in host rhizospheres under conditions simulating those found in Cape habitats.

References

[1] Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., Polz, M.F.: PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology **71**(12), 8966–8969 (2005)

[2] Baker, D.D.: Opportunities for autoecological studies of *Frankia*, a symbiotic actinomycete. In: Y. Okami, T. Beppu, H. Ogawara (eds.) Biology of Actinomycetes, pp. 271–276. Japanese Scientific Societies Press, Tokyo (1988)

[3] Benson, D.R., Vanden Heuvel, B.D., Potter, D.: Actinorhizal symbioses: diversity and biogeography. In: M. Gillings (ed.) Plant Microbiology, pp. 102–133. BIOS Scientific Publishers Ltd., Oxford (2004)

[4] Burleigh, S.H., Dawson, J.O.: Occurrence of *Myrica*-nodulating *Frankia* in Hawaiian volcanic soils. Plant and Soil **164**, 283–289 (1994)

[5] Callaham, D., Del Tredici, P., Torrey, J.G.: Isolation and cultivation in vitro of the actinomycete causing root nodulation in *Comptonia*. Science **199**, 899–902 (1978)

[6] Chaia, E.E., Wall, L.G., Huss-Danell, K.: Life in soil by the actinorhizal root nodule endophyte *Frankia*. A review. Symbiosis **51**(3), 201–226 (2011)

[7] Chandler, D.P., Brockman, F.J., Fredrickson, J.K.: Use of 16S rDNA clone libraries to study changes in a microbial community resulting from ex situ perturbation of a subsurface sediment. FEMS Microbiology Reviews **20**(34), 217–230 (1997)

[8] Chao, A., Chazdon, R.L., Shen, T.J.: A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters **8**(2), 148–159 (2005)

[9] Darriba, D., Taboada, G.L., Doallo, R., Posada, D.: jModelTest 2: more models, new heuristics and parallel computing. Nature Methods **9**, 772 (2012)

[10] Dineen, S.M., Aranda, R., Anders, D.L., Robertson, J.M.: An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. Journal of Applied Microbiology **109**(6), 1886–1896 (2010)

[11] Frostegard, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., Nesme, X., Simonet, P.: Quantification of bias related to the extraction of DNA directly from soils. Applied and Environmental Microbiology **65**(12), 5409–5420 (1999)

[12] Gauthier, D., Jaffre, T., Prin, Y.: Abundance of *Frankia* from *Gymnostoma* spp. in the rhizosphere of *Alphitonia neocaledonica*, a non-nodulated Rhamnaceae endemic to New Caledonia. European Journal of Soil Biology **36**, 169–175 (2000)

[13] Guidon, S., Gascuel, O.: A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology **52**, 696–704 (2003)

[14] Hahn, D., Nickel, A., Dawson, J.: Assessing *Frankia* populations in plants and soil using molecular methods. FEMS Microbiology Ecology **29**, 215–227 (1999)

[15] Hilger, A.B., Myrold, D.D.: Method for extraction of *Frankia* DNA from soil. Agriculture, Ecosystems and Environment **34**(1-4), 107–113 (1991)

[16] Huss-Danell, K., Myrold, D.D.: Intrageneric variation in nodulation of *Alnus*: Consequences for quantifying *Frankia* nodulation units in soil. Soil Biology and Biochemistry **26**(5), 525–531 (1994)

[17] Huss-Danell, K., Sverrisson, H., Hahlin, A.S., Danell, K.: Occurrence of *Alnus*-infective *Frankia* and *Trifolium*-infective *Rhizobium* in circumpolar soils. Arctic, Antarctic, and Alpine Research **31**(4), 400–406 (1999)

[18] Huss-Danell, K., Uliassi, D., Renberg, I.: River and lake sediments as sources of infective *Frankia* (*Alnus*). Plant and Soil **197**(1), 35–39 (1997)

[19] Jeong, S.C., Myrold, D.D.: Population size and diversity of *Frankia* in soils of *Ceanothus velutinus* and Douglas-fir stands. Soil Biology and Biochemistry **33**(7-8), 931–941 (2001)

[20] Krivushin, K., Kondrashov, F., Shmakova, L., Tutukina, M., Petrovskaya, L., Rivkina, E.: Two metagenomes from late Pleistocene Northeast Siberian permafrost. Genome Announcements **3**(1), 4–5 (2015)

[21] Leigh, M., Taylor, L., Neufeld, J.: Clone libraries of ribosomal RNA gene sequences for characterization of bactrial and fungal communities. In: K. Timmis (ed.) Handbook of Hydrocarbon and Lipid Microbiology, pp. 3971–3990. Springer-Verlag, Berlin Heidelberg (2010)

[22] Magurran, A.E.: Ecological diversity and its measurement. Princeton University Press, Princeton, New Jersey (1988)

[23] Maunuksela, L., Zepp, K., Koivula, T., Zeyer, J., Haahtela, K., Hahn, D.: Analysis of *Frankia* populations in three soils devoid of actinorhizal plants. FEMS Microbiology Ecology **28**(1), 11–21 (1999)

[24] Miller, M.A., Pfeiffer, W., Schwartz, T.: Creating the CIPRES Science Gateway for inference of large phylogenetic trees (2010)

[25] Mirza, B.S., Welsh, A., Rasul, G., Rieder, J.P., Paschke, M.W., Hahn, D.: Variation in *Frankia* populations of the *Elaeagnus* host infection group in nodules of six host plant species after inoculation with soil. Microbial Ecology **58**(2), 384–393 (2009)

[26] Mirza, B.S., Welsh, A., Rieder, J.P., Paschke, M.W., Hahn, D.: Diversity of frankiae in soils from five continents. Systematic and Applied Microbiology **32**(8), 558–570 (2009)

[27] Mirza, B.S., Welsh, A.K., Hahn, D.: Growth of *Frankia* strains in leaf litter-amended soil and the rhizosphere of a nonactinorhizal plant. FEMS Microbiology Ecology **70**(1), 132–141 (2009)

[28] Mullins, T.D., Britschgi, T.B., Krest, R.L., Giovannoni, S.J.: Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnology and Oceanography **40**(1), 148–158 (1995)

[29] Myrold, D.D., Huss-Danell, K.: Population dynamics of *Alnus*-infective *Frankia* in a forest soil with and without host trees. Soil Biology and Biochemistry **26**(5), 533–540 (1994)

[30] Nalin, R., Normand, P., Domenach, A.M.: Distribution and N_2-fixing activity of *Frankia* strains in relation to soil depth. Physiologia Plantarum **99**(4), 732–738 (1997)

[31] Nickel, A., Hahn, D., Zepp, K., Zeyer, J.: In situ analysis of introduced *Frankia* populations in root nodules of *Alnus glutinosa* grown under different water availability. Canadian Journal of Botany **77**(9), 1231–1238 (1999)

[32] Normand, P., Chapelon, C.: Direct characterization of *Frankia* and of close phyletic neighbors from an *Alnus viridis* rhizosphere. Physiologia Plantarum **99**, 722–731 (1997)

[33] Picard, C., Ponsonnet, C., Paget, E., Nesme, X., Pascal, S.: Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Applied and Environmental Microbiology **58**(9), 2717–2722 (1992)

[34] Pokharel, A., Mirza, B.S., Dawson, J.O., Hahn, D.: *Frankia* populations in soil and root nodules of sympatrically grown *Alnus* taxa. Microbial Ecology **61**, 92–100 (2011)

[35] Polz, M.F., Cavanaugh, C.M.: Bias in template-to product ratios in multi-template PCR. Applied and Environmental Microbiology **64**(10), 3724–3730 (1998)

[36] Quispel, A.: Discoveries, discussions and trends in research on actinorhizal root nodule symbioses before 1978. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of *Frankia* and Actinorhizal Plants, pp. 15–34. Academic Press, Inc., San Diego, California (1990)

[37] Rambaut, A.: FigTree v1.2.3. Institute of Evolutionary Biology, University of Edinburgh (2009). Available from http://tree.bio.ed.ac.uk/software/figtree

[38] Samant, S., Amann, R., Hahn, D.: Evaluation of the 23S rRNA gene as target for qPCR based quantification of *Frankia* in soils. Systematic and Applied Microbiology **37**(3), 229–234 (2014)

[39] Samant, S., Sha, Q., Iyer, A., Dhabekar, P., Hahn, D.: Quantification of *Frankia* in soils using SYBR Green based qPCR. Systematic and Applied Microbiology **35**(3), 191–197 (2012)

[40] Samant, S.S., Dawson, J.O., Hahn, D.: Growth responses of indigenous *Frankia* populations to edaphic factors in actinorhizal rhizospheres. Systematic and Applied Microbiology pp. 3–7 (2015)

[41] Simonet, P., Navarro, E., Rouvier, C., Reddell, P., Zimpfer, J., Dommergues, Y., Bardin, R., Combarro, P., Hamelin, J., Domernach, A., Gourbiere, F., Prin, Y., Dawson, J.O., Normand, P.: Co-evolution between *Frankia* populations and host plants in the family *Casuarinaceae* and consequent patterns of global dispersal. Environmental Microbiology **1**(6), 525–533 (1999)

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

[43] Stackebrandt, E., Goebel, B.M.: Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology **44**(4), 846–849 (1994)

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

[45] Sukumaran, J., Holder, M.T.: DendroPy: a Python library for phylogenetic computing. Bioinformatics **26**, 1569–1571 (2010)

[46] Swofford, D.L.: PAUP. Phylogenetic analysis using parsimony (and other methods). Version 4 (2003)

[47] Taylor, L.: Synoptic dynamics, migration and the Rothamsted insect survey: Presidential address to the British Ecological Society, December 1984. Journal of Animal Ecology **55**, 1–38 (1986)

[48] Timke, M., Wang-lieu, N.Q., Altendorf, K., Lipski, A.: Community structure and diversity of biofilms from a beer bottling plant as revealed using 16S rRNA gene clone libraries. Applied and Environmental Microbiology **71**(10), 6446–6452 (2005)

[49] Torsvik, V., Goksoyr, J., Daae, F.L., Torsvik, V., Goksyr, J., Daae, F.L.: High diversity in DNA of soil bacteria. Applied and Environmental Microbiology **56**(3), 782–787 (1990)

[50] Valdes, M.: Frankia ecology. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing actinorhizal symbioses, chap. 3, pp. 49–71. Springer, Dordrecht, the Netherlands (2008)

[51] Welsh, A.K., Mirza, B.S., Rieder, J.P., Paschke, M.W., Hahn, D.: Diversity of frankiae in root nodules of *Morella pensylvanica* grown in soils from five continents. Systematic and Applied Microbiology **32**(3), 201–210 (2009)

[52] Wollum II, A.G., Youngberg, C.T., Chichester, F.W.: Relation of previous timber stand age to nodulation of *Ceanothus velutinus*. Forest Science **14**(1), 114–118 (1968)

[53] Zepp, K.: Analysis of *Frankia* populations by in situ hybridization with 23S rRNA targeted oligonucleotide probes. Ph.D. thesis, Swiss Federal Institute of Technology (1996)

[54] Zepp, K., Hahn, D., Zeyer, J.: In-situ analysis of introduced and indigenous *Frankia* populations in soil and root nodules obtained on *Alnus glutinosa*. Soil Biology and Biochemistry **29**(9), 1595–1600 (1997)

[55] Zimpfer, J.F., Smyth, C.A., Dawson, J.O.: The capacity of Jamaican mine spoils, agricultural and forest soils to nodulate *Myrica cerifera*, *Leucaena leucocephala* and *Casuarina cunninghamiana*. Physiologia Plantarum **99**, 664–672 (1997)

[56] Zitzer, S.F., Dawson, J.O.: Soil properties and actinorhizal vegetation influence nodulation of *Alnus glutinosa* and *Elaeagnus angustifolia* by *Frankia*. Plant And Soil **140**(2), 197–204 (1992)

4 Isolation, characterization and culture of endemic CFR *Frankia* strains

Abstract

Ten *Frankia* strains were isolated from the root nodules four *Morella* species growing in their natural habitats in the Western Cape region of Southern Africa. Isolates displayed phenotypic characteristics typical of the genus, including branched hyphae, vesicles and multilocular sporangia. Furthermore, phylogenetic analyses based on 16S rRNA and partial *nif*H sequences confirmed the isolates as members of *Frankia* host-infection groups I and III, and more specifically as members of Cape Genotypes CC-1, CC-2 and CC-3, the dominant *Frankia* genotypes found in root nodules of Cape *Morella*. Utilization of sole carbon sources was typical for strains from each HIG; Cluster III strains were pigmented and demonstrated the ability grow on simple sugars, propionate or pyruvate as sole carbon sources. Isolates from Cluster I were unpigmented and grew on pyruvate and propionate only. A multilocus sequence analysis demonstrated that Cape strains belong to at least three genomospecies, two of which appear to be novel; one to Africa and the other to the Cape flora. This chapter reports isolation and characterization of *Frankia* isolated from actinorhizal species endemic to the African continent[1].

[1] *Frankia* symbiosis involving *Alnus glutinosa*, *Coriaria myrtifolia* and the introduced *Elaeagnus angustifolia* have been investigated in North Africa. Most of these species are native to Africa, Asia and Europe and are thus indigenous, but not endemic. The *Morella* of southern and central Africa have the distinction of being the only species endemic to the continent.

4.1 Introduction

Frankia is a free-living nitrogen-fixing soil actinomycete which enters into symbioses with members of 22 genera of actinorhizal plants from eight families of dicotyledons [14,57,71]. Despite a century of research into "non-legume" symbioses, the determination that root nodules were the site of nitrogen fixation and the discovery that the endophyte within these nodules was an actinomycete, for the longest time the exact identity of nodulating organism remained uncertain [10,19]. Principally, this was because *Frankia* had proven difficult to isolate and was thus considered an obligate symbiont[2]. It was not until 1978 that reproducible isolation was reported, and the organism found to be slow growing with fastidious growth requirements [13,23,62]. The isolate, CpI1 from *Comptonia peregrina*, was able to both re-infect actinorhizal hosts and induce nitrogen fixation in root nodules, thus fulfilling Koch's postulate [23,42,44,74]. Following this initial success isolation methods were quickly developed and strains have subsequently been recovered from root nodules of seven of the eight actinorhizal host families [3,11,21,34,47]. Despite these successes isolation remains challenging; no universally-applicable isolation media have been reported, and only a small number of attempts succeed [6,47]. Illustrating this, a strain from Cluster II (members of which were for many years thought to be obligate symbionts) has only recently been isolated for the first time [34]. Most attempts are made using root nodules, only one study reports isolation of strains directly from soils, *Frankia*'s second natural niche [3].

Morphologically *Frankia* have three basic cell types: (1) filamentous and sparsely-branched hyphae usually 0.5-2.0 μm in diameter, (2) terminal or intercalary multilocular sporangia which may vary in size and shape, and finally (3) vesicles (or "diazovesicles") which are the site of nitrogen fixation in most strains[3]. These last two cell types are unique to the genus, and display morphological variation both between strains and under different growth conditions. Despite this, and the

[2]E. Pommer had reported isolation from *Alnus glutinosa* in 1959 [60], but this report went largely ignored and his strains were subsequently lost. Later examination of his laboratory records revealed, from the morphological characteristics of his isolates, that he had doubtlessly succeeded in isolating *Frankia* [13,23,47,61]

[3]While *Casuarina* strains form vesicles in culture, they appear to be absent in root nodules even under nitrogen-fixing conditions, and nitrogenase has been found in hyphae in *Elaeagnus* [16,64,82]

fact that *Frankia* cells walls contain 2-*O*-methyl-D-mannose which is not found in other actinomycetes [53], it has proven more difficult to reliably discriminate between strains within the genus, using morphological and physiological traits alone, than it has to discriminate between *Frankia* and other actinomycetes [35, 46]. Under ideal conditions *Frankia*'s doubling time is slow: $14 - 24$ h for fast-growing strains [13]. Growth periods of months are sometimes required to obtain sufficient biomass for physiological experiments, and comprehensive physiological tests are often impractical or give anomalous results [15, 47]. Additionally, characteristics such as pigment production may vary depending on the composition of growth media and are not taxonomically useful [47].

Early physiological and serological characterization did allow two groups of *Frankia* to be identified, which differed in their ability to grow on various substrates and which loosely correspond to the *Alnus* and *Elaeagnus* HIGs [4, 45]. Following this, cross-inoculation experiments suggested that isolates fell into three to four host-infection groups [5]. These host-infection groups (or clusters) were later shown to be phylogenetically distinct according to 16*S* rRNA [59] and were defined as follows: cluster 1, *Frankia* infective on *Alnus*, *Casuarina*, *Myrica* and *Morella*; cluster 2, *Frankia* infective on the Rosaceae, and on *Coriaria* and *Datisca*; cluster 3, *Elaeagnus*-infective strains (also infective on *Morella* and *Gymnostoma*); and finally cluster 4, strains which are unable to re-infect their original hosts or are incapable of fixing nitrogen [25, 55, 59]. This four-cluster classification scheme has subsequently been confirmed using other genes and intergeneic spacer regions: *gln*II, *nif*H, a 23S rRNA insertion sequence, and the *nif*D/*nif*K IGS [38, 39, 54], and with combinations of markers: 16S rRNA and *gln*A sequences [26]; and *gyr*B, *nif*H and *gln*II citeNouioui2011. Finally, strains have been discriminated using multilocus sequence analysis (MLSA) of *atp*D, *dna*A, *fts*Z, *pgk* and *rpo*B gene fragments [8]. Furthermore, at least 12 genomospecies, defined using DNA-DNA hybridization (DDH) and amplified fragment length polymorphism (AFLP), have been identified [1, 9, 29]. While DDH suggests that these should be assigned species status, there are no distinguishing phenotypic characteristics within these groups [9, 35, 76]. Consequently, only one species (*Frankia alni*) has been validly published to date and no type strain has, as yet, been designated [35, 76].

The aim of the present work is to report the isolation and characterization of endemic *Frankia* strains from field-collected root nodules of four *Morella* species native to the Cape region of Southern Africa: *M. cordifolia*, *M. quercifolia*, *M. kraussiana* and *M. integra*. The isolated strains were characterized according to the following: morphology, utilization of single carbon sources, single gene phylogenies (16S rRNA, *nif*H), phylogeny according to an actinomycete-specific 23S insertion sequence, and multilocus analysis for genomospecies assignment.

4.2 Materials and Methods

4.2.1 Isolation and cultivation of Cape *Frankia*

Isolation of *Frankia* from root nodules

Fresh root nodules were collected from six *Morella* species within the Cape region of Southern Africa, as described in Chapter 2. In the laboratory nodules were washed in a stream of tap water to remove adherent soil, after which individual lobes were excised with a straight-edged scalpel blade. These lobes were cleaned in distilled water and examined under a dissecting microscope for adherent soil particles. Next, nodule lobes were surface sterilized in the same manner as those used for DNA extraction in Chapter 2, except that surface sterilization in 30% hydrogen peroxide was extended to $30 - 40$ minutes. Approximately 1 mm of tissue was removed from the base of each lobe (opposite the root hair) before washing each lobe twice more in sterile dH_2O. All steps following this were carried out aseptically. Surface-sterilized lobes underwent microdissection in which apical sections of the lobe cortex were excised. Following this two approaches were used: firstly, lobe sections were inoculated into 6 mL of (defined propionate medium (DPM) [3], *Frankia* defined medium (FDM) [7] without a fixed nitrogen source and supplemented with various combinations of cyclohexamide (50 μg mL^{-1}), nalidixic acid (10 μg mL^{-1}) and sodium azide (5 μg mL^{-1}) [3,47,50]. Where propionate was used as sole carbon source, sodium azide was excluded from the media. Secondly, lobe sections were streaked across, or incubated directly on top of, gellan gum solid media. This solidifying agent has been shown to improve *Frankia* growth

on solid media [7]. Solid media were prepared both with and without ammonium chloride as fixed nitrogen source. Plates and tubes were incubated in the dark at 28°C until colony outgrowth from nodule surfaces or on the surface of solid media was observed. Filamentous outgrowths were excised under a dissecting microscope, homogenized in 1 mL of sterile dH_2O 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.

Growth media, routine culture procedures and *Frankia* reference strain

After isolation and dereplication Cape *Frankia* monocultures were maintained in the absence of antibiotic agents, either on plates or in 100 mL medium in 250 mL Erlenmeyer flasks. Isolates were cultured in DPM [3], modified FDM [7] containing (in 1000 mL dH_2O) 2.5 g sodium pyruvate; 2.5 g peptone; 0.5 g NH_4Cl; 100 mL of a 10× stock salts solution containing (in 1000 mL dH_2O) 1.0 g $CaCl_2{\cdot}2H_2O$; 2.0 g $MgSO_4{\cdot}7H_2O$, 1.0 mL of micronutrient solution containing (in 100 mL dH_2O) 0.75 g Na_2-EDTA$\cdot 2H_2O$, 0.56 g $FeSO_4{\cdot}7H_2O$, plus 0.02 g $Na_2MoO_4{\cdot}2H_2O$; and finally 100 mL of 10× buffer stock containing (per 1000 mL dH_2O) 6.8 g KH_2PO_4 plus 11.56 g MOPS, or on gellan-gum-based solid media, identically formulated but containing (per 1000 mL) 6.0 g $CaCl_2{\cdot}2H_2O$; 2.0 g $MgSO_4{\cdot}7H_2O$ and 2.5 g gellan gum (Sigma-Aldrich Corp., St. Louis, MO) [7]. Cold buffer stocks and filter-sterilized carbon sources were added to the medium immediately after autoclaving, and the final mixture was cooled in a water bath to 70°C before pouring. This formulation is identical to CB medium, except for the carbon substrates [7]. Cultures were incubated in the dark at 28°C and liquid cultures were manually agitated once per week; cultures on plates were inspected every second day and resealed with Parafilm® whenever necessary. A *Frankia* strain isolated from *Alnus rubra*, DSM-44251, was obtained from the Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (http://www.dsmz.de/) and used as reference for characterization of cell morphology. While it is reported that this strain only grows in liquid culture [36], it also grew well on the gellan-

gum-based media used to maintain cultures of the Cape isolates.

Single carbon substrate growth determinations

In order to establish Cape *Frankia* isolates' ability to grow on single carbon sources, a plating experiment based on that of Bassi et al. was performed [7]. Flocs from healthy, 30-day old *Frankia* cultures grown in liquid DPM were collected by settling and pipetting. These were rinsed twice prior to homogenization, and twice after, in sterilized dH_2O in order to remove residual propionate. Thereafter they were incubated for 24 hours to allow the cultures to metabolize residual propionate. Mycelia were homogenized with a 5 mL Dounce homogenizer, producing hyphal fragments in the range of 5-20 μm. These were serially-diluted and spread-plated onto gellan-gum medium prepared as described above, except that the following filter-sterilized substrates were added as single carbon source at 5 g/L: D-glucose, D-fructose, D-maltose, D-mannose, sucrose, D-trehalose, D-galactose, D-mannitol, D-sorbitol, Na-pyruvate and Na-propionate. For each individual strain tested all media were inoculated from the same source of homogenized mycelia. A medium containing no carbon source was included as a control. Plates were incubated in the dark at 28°C. Plates were inspected microscopically immediately after inoculation, and strains producing colonies larger than 100 μm from hyphal fragments after 21 days were considered capable of growth on the carbon source in question.

4.2.2 Light and electron microscopy

Photomicrographs of water preparations and colonies growing on the surface of gellan-gum plates were made using bright-field and phase-contrast microscopy using an Axiovert® inverted microscope (Carl Zeiss Microscopy, LLC, USA). Scanning electron microscopy of colonies growing on the surface of FDM gellan-gum plates was carried out at the Electron Microscopy Unit, University of Cape Town.

4.2.3 DNA extraction

Single *Frankia* colonies were selected from the surface of gellan-gum plates and total DNA extracted using a MoBio PowerPlant Pro kit (MoBio, Carlsbad, CA) according to the manufacturer's instructions.

4.2.4 PCR conditions

Amplification of 16S rRNA, an actinomycete-specific IS in domain III of the 23S rRNA gene, and *nif*H gene fragments from *Frankia* isolate DNA was performed using primers described in Table 4.1. For each strain the product of a single extraction was used as template for all PCRs. Reactions targeting 16S rRNA were carried out in a volume of 25 μL and contained 2 μL dNTPs (2.5 mM each), 2.5 μL 10$\times$ 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 (5U μL^{-1}; TaKaRa) and 1 μL of template. For 16S an initial 5-min incubation at 95°C was followed by 35 rounds of temperature cycling (94°C for 30 s, 60°C for 90 s, 72°C for 45 s) and a final 5-min extension at 72°C. Reactions targeting 23S rRNA IS were carried out in a volume of 25 μL and contained 2 μL dNTPs (2.5 mM each), 2.5 μL 10$\times$ 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 (5U μL^{-1}; TaKaRa) and 1 μL of template. An initial 5-min incubation at 95°C was followed by 35 rounds of temperature cycling (94°C for 30 s, 54°C for 30 s, 72°C for 45 s) and a final 5-min extension at 72°C. Amplification of *nif*H gene fragments was performed in a reaction volume of 25 μL. Each reaction contained 2 μL dNTPs (2.5 mM each), 2.5 μL 10$\times$ 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 (5U μL^{-1}; TaKaRa) and 1 μL of template. An initial 5-min incubation at 95°C was followed by 35 rounds of temperature cycling (94°C for 30 s, 60°C for 30 s, 72°C for 45 sec) and a final 5-min extension at 72°C. For multilocus sequence analysis, amplification of *atp*D, *dna*A, *fts*Z, *pgk* and *rpo*B gene fragments from *Frankia* isolates was performed using primer sets D1, D2, D3, D3 and D5 (Table 4.1). These primers are modifications of those designed by Bautista [8] from which the T3 and T7 RNA polymerase promoter regions have been removed. Reactions were carried out in a volume of 25 μL. Each

reaction contained 2 μL dNTPs (2.5 mM each), 2.5 μL 10$\times$ 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 (5U μL^{-1}; TaKaRa) and 1 μL of template. An initial 5-min incubation at 95°C was followed by 30 rounds of temperature cycling (94°C for 30 s, 60°C for 30 s, 72°C for 45 sec) and a final 5-min extension at 72°C. All reactions were checked for amplification products of the expected size by gel electrophoresis: 5 μL reaction product, 1% $^{w}/_{v}$ agarose gel in TAE buffer, stained with ethidium bromide (0.5 μg mL^{-1}).

4.2.5 Sequence analysis

PCR amplicons of the expected size for each target locus 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 in both directions at the Stellenbosch University Central Analytical Facility (http://academic.sun.ac.za/saf). Chromatograms were visually assessed and sequences corrected by hand where necessary. Assembly was performed in CLC Main Workbench (version 6.2.1). For MLSA alignments were concatenated using CLC Main Workbench in the following order: *atp*D, *dna*A, *fts*Z, *pgk*, *rpo*B. 16S rRNA gene sequences were deposited in GenBank under accession numbers KP342101-KP342110 and partial *nif*H sequences under accession numbers KP342111-KP342120. 23S rRNA IS rRNA gene sequences were deposited in GenBank under accession numbers KU174954-KU174963 and partial *nif*H sequences under accession numbers KP342111-KP342120. Individual gene fragments from the MLSA analysis were deposited in GenBank under accession numbers KU174964-KU175013.

4.2.6 Phylogenetic analysis

16S, *nif*H and 23S rRNA IS

1381 bp 16S rRNA amplification products from Cape *Frankia* isolates were aligned with those of 34 isolates retrieved from public databases and trimmed to 1202 bp using CLC Main Workbench (version 6.2.1). Sequences were checked for the presence of chimeras using DECIFER [77], and BLAST analysis performed to confirm the identity of the isolates. A NJ tree was constructed using PAUP (version 4.0b) [72] to determine their positions within previously defined *Frankia* clusters [59]. 144 bp partial gene sequences, spanning the actinomycete-specific insertion sequence of domain III of the 23S rRNA gene, from 36 isolates representing *Frankia* groups I to VI as defined by Zepp et al. [81] were retrieved from public databases and aligned with sequences from Cape *Frankia* isolates in CLC Main Workbench (version 6.2.1). A NJ tree was constructed using PAUP (version 4.0b) [72]. Bootstrap values for the 16S rRNA and 23S rRNA IS trees are expressed as percentages and were determined from 1000 replicates [28]. Partial *nif*H sequences from Cape *Frankia* isolates were identified by nucleotide BLAST analysis and analyzed in comparison to previously reported *Frankia* strains, as well as in comparison to a larger data set including sequences recovered from root nodules, as described in Chapter 2. Non-identical *nif*H amplification fragments from Cape *Frankia* isolates were aligned with *Frankia* isolates *nif*H sequences downloaded from public databases, using CLC Main Workbench (version 6.2.1). The data set composed of *Frankia* isolates consisted of 6 non-identical sequences from Cape isolates and 47 sequences from database sequences (*Frankia* pure cultures and actinomycete BMG5.6). The alignment was trimmed to 512 bp, corresponding to positions 227 to 738 of the *Frankia* ACN14a *nif*H gene (NC008278). The resultant 54 sequence data set was analyzed using neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods as previously described [75]. Neighbour joining was completed in PAUP (version 4.0b) [72]. A GTR+I+G model of sequence evolution and set values for the proportion of invariant sites and gamma shape parameter were estimated in jModeltest (version 2.1.4) [27, 37]. 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) [51, 69]. 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" [72]. 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 [70] and displayed in Figtree version 1.4 [63].

Multilocus sequence analysis (MLSA)

Individual *atp*D, *dna*A, *fts*Z, *pgk* and *rpo*B sequences from Cape *Frankia* isolates were identified using nucleotide BLAST analysis. Pairwise distances for individual gene fragments and concatenated sequences were computed in MOTHUR (version 1.32.0). For multilocus analysis (MLSA) and putative assignment of Cape *Frankia* to genomospecies a maximum likelihood analysis of concatenated sequences was carried out with 50 *Frankia* strains from databases, 10 Cape *Frankia* isolates and *Acidothermus cellulolyticus* as outgroup, using the RAxML-HPC2 program on the CIPRES computer cluster (www.phylo.org) [51, 69]. 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. The resultant phylogenetic tree was prepared in Figtree (version 1.4) (http://tree.bio.ed.ac.uk/software/figtree/) [63]. Splits decomposition and Pairwise Homoplasic Index (PHI test) analyses were carried out in Splits Tree 4 [20, 40].

Table 4.1: PCR primers used in characterization of *Frankia* isolates

Target	Primer	Sequence (5'⟶3')	Tm	Ref.
*nif*H	nifHf1	GGC AAG AAG TTC ACC ACC CAG C	63.5	[75]
	nifHr	CTC GAT GAC CGT CAT CCG GC	63.0	
16S rRNA	FGPS4-281$_{bis}$	ATG GAR AGY TTG ATC CTG GCT CA	-	[56]
	FGPS-1509'-153	AAG GAG GGG ATC CAG CCG CA	65.7	
23S rRNA IS	23Fra	ATC GCA TGC CTA CTA CC	53.1	[81]
	23InsVFra	CAG GCG TAG TCG ATG G	53.6	
*atp*D	D1f	ACC GGS ATC AAG GTC ATC GAC	-	[8]
	D1r	CCG AGG ATG GCG ATG ATG TC	60.7	
*dna*A	D2f	GAG GAR TTC ACC AAC GAC TTC AT	-	
	D2r	CRG AAG TGC TGG CCG ATC TT	-	
*fts*Z	D3f	CCG TCA ACC GGA TGA TCG AA	60.2	
	D3r	GCS GCC TTG ATC TCG AAC AG	61.4	
pgk	D4f	TGA GGA CGA TCG ACC ACC TGC	64.2	
	D4r	CGC SAG GAA GGT GAA GCA CAT	-	
*rpo*B	D5r	TAC GGC GTC TCG ATG AAS CC	61.4	
	D5f	CGA CCA CTT CGG CAA CCG	-	

4.3 Results

4.3.1 *Frankia* isolation and morphological characteristics

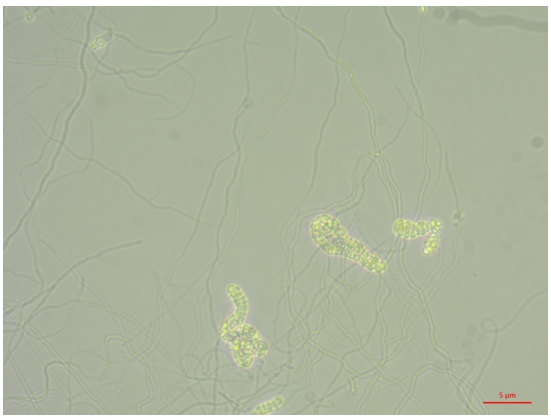
Following dereplication a total of ten isolates were obtained from nodules from four of the six host species examined: seven from *M. cordifolia* and one each from *M. kraussiana*, *M. integra* and *M. quercifolia*. Strains from *M. cordifolia* were isolated in DPM and FDM without a source of fixed nitrogen. Strains from the remaining three species were isolated on modified FDM gellan gum supplemented with $NH_4C\ell$. Isolates had typical *Frankia* morphological features, including branched hyphae, multilocular sporangia and vesicles (Table 4.2, Figures 4.3a, 4.1b and 4.1). Scanning electron microscopy showed spores to be irregular in shape and hyphae to be 0.5 to 1.2 μm in diameter (Figure 4.2). Only isolates from

M. cordifolia produced pigments.

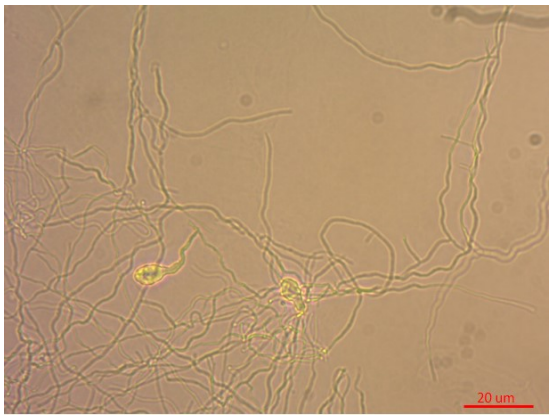
Table 4.2: Morphological characteristics of Cape *Frankia* isolates.

*nif*H		Soluble	Sporangia[b]		Vesicles[c]	
Genotype	Isolate[a]	pigment	LC	Gum	+$NH_4C\ell$	-$NH_4C\ell$
CC-3	FMc1	+ (pink)	+	+	+	++
	FMc2	+ (blue)	+	+	+	++
	FMc3	+ (black)	+	+	+	++
	FMc4	+ (black)	+	+	+	++
	FMc5	+ (black)	+	+	+	++
CC-4	FMc6	+ (yellow)	+	+	+	++
	FMc7	+ (orange)	+	+	+	++
CC-1	FMi1	-	+	+	-	+
	FMk1	-	+	+	-	+
	FMq1	-	+	+	-	+

[a] Named according to original host: M. cordifolia isolate = FMc etc.
[b] Presence of sporangia in submerged DPM liquid culture (LC) or on gellan gum plates containing pyruvate as carbon source.
[c] Vesicle production on gellan gum plates containing pyruvate and with or without a fixed nitrogen source ($NH_4C\ell$).

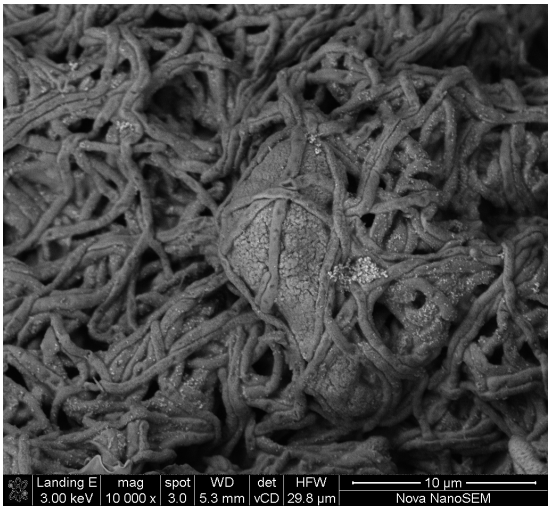


(a) *Frankia* DSMZ 44251 sporangia

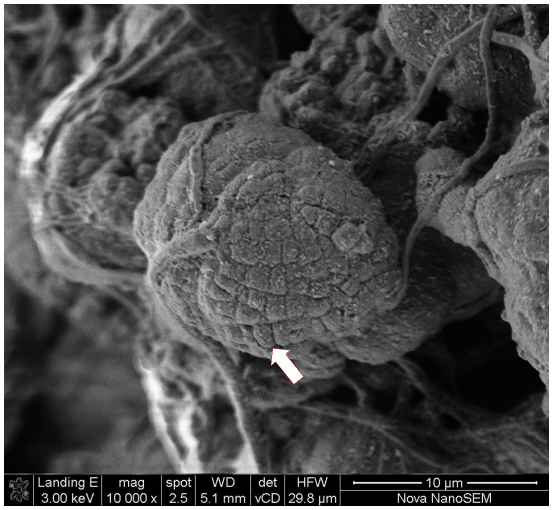


(b) *Frankia* FMi1 sporangia

Figure 4.1: Characteristic *Frankia* sporangia from DSMZ 44251 in comparison to those of a Cape *Frankia* isolate growing in DPM submerged liquid culture.

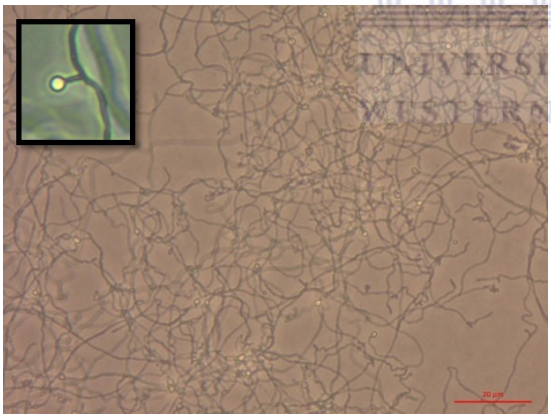


(a) *Frankia* FMi1 hyphae and immature sporangium

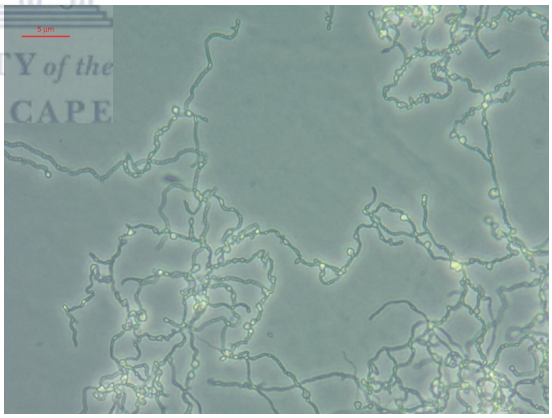


(b) *Frankia* FMc1 hyphae and mature sporangium containing irregular spores (arrow)

Figure 4.2: Scanning electron microscopy (SEM) showing typical *Frankia* cell types in two Cape isolates grown on the surface of FDM gellan-gum supplemented with $NH_4C\ell$.



(a) *Frankia* FMc6 hyphae with vesicles



(b) *Frankia* FMc5 abnormal hyphal swellings

Figure 4.3: Characteristic morphological forms of *Frankia* in two Cape isolates growing in DPM submerged liquid culture. A magnified view of a stalked vesicle is provided as inset to panel (a). Unusual hyphal swelling in an 84 day old culture is shown in panel (b).

4.3.2 Sole carbon substrate utilization

The ability of *Frankia* isolates to grow using individual sugars (glucose, fructose, maltose, mannose, cellobiose, trehalose and galactose), sugar alcohols (mannitol,

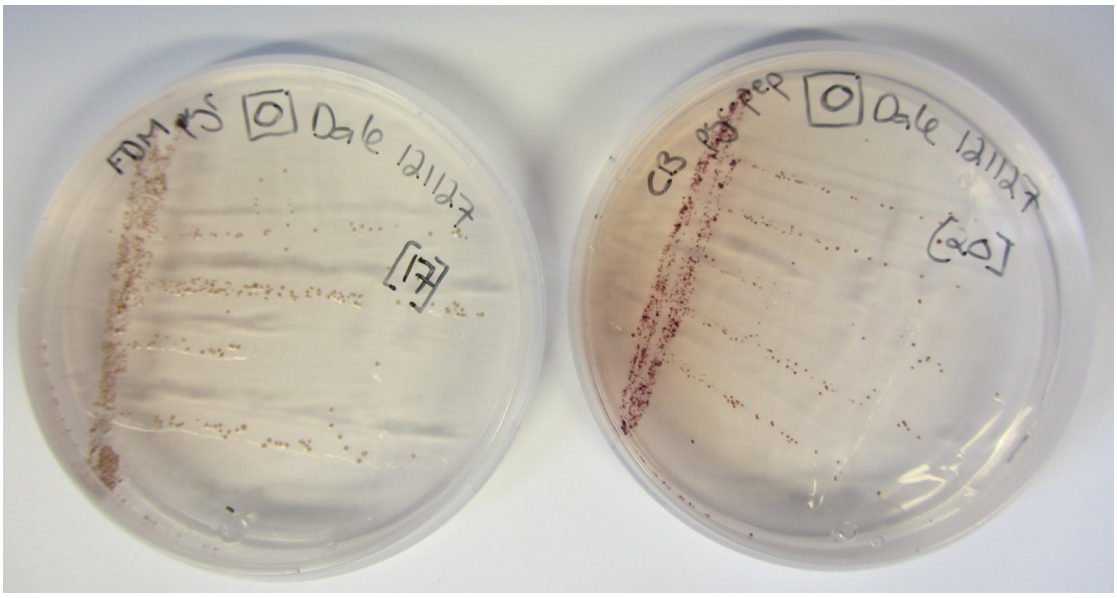
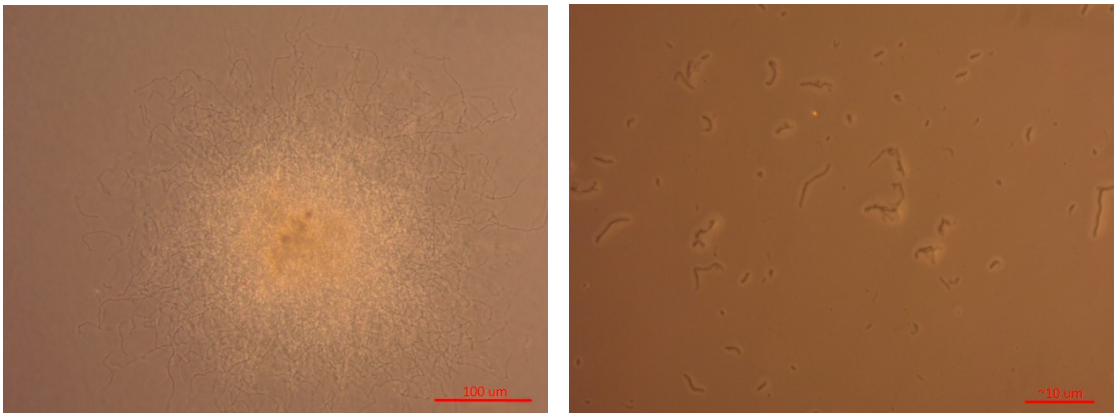


Figure 4.4: Colonies of *Frankia* isolate FMc2 on gellan gum based CB media containing either pyruvate or combined pyruvate and peptone showing differences in pigment production under different nutrient regimes. Cultures are 90 days old.

sorbitol) or organic acids (pyruvate, propionate) as sole carbon source was tested on gellan gum solid media. Not all Cape *Frankia* isolates were tested as the experiment required inoculum raised in liquid culture and FMc1, FMc4, FMc7 and FMk1 had stopped growing in liquid culture after several rounds of sub-culturing. Serially diluted homogenates from liquid cultures were inspected microscopically and spread plated in such a way as to give good separation between colonies arising from hyphal fragments no larger than 20 μm. Strains producing colonies in excess of 100 μm after 21 days were considered capable of using the substrate in question as sole carbon source (Figure 4.5a). When *Frankia* hyphae from each of the tested strains were inoculated onto solid media lacking a carbon substrate no hyphal elongation or colony formation was observed after 21 days (Figure 4.5b). Results of the carbon source utilization test are reported in Table 4.3.



(a) *Frankia* colony on 0.5 g/l Cellobiose (b) Hyphal fragments - no carbon source

Figure 4.5: Growth of *Frankia* isolate FMc6 on gellan gum with or without a carbon source after 21 days, illustrating efficacy of the adapted culture technique for determining the ability of an isolate to grow on single substrates.

Table 4.3: Carbon source utilization by Cape *Frankia* isolates. Strains producing ten or more colonies larger than 100 μm per plate after 21 days were considered capable of growth on the substrate indicated.

		Sugars[a]								SA[b]		OA[c]	
Genotype	Isolate	Glu	Fru	Mal	Man	Suc	Cel	Tre	Gal	Mat	Sor	Pyr	Pro
CC-3	FMc2	-	-	-	-	-	+	-	-	-	-	+	+
	FMc3	-	-	-	-	-	+	-	-	-	-	+	+
	FMc5	-	-	-	-	-	+	-	-	-	-	+	+
CC-4	FMc6	-	+	-	-	-	+	-	-	-	-	+	+
CC-1	FMi1	-	-	-	-	-	-	-	-	-	-	+	+
	FMq1	-	-	-	-	-	-	-	-	-	-	+	+

[a] Glu, Glucose; Fru, Fructose; Mal, Maltose; Man, Mannose; Suc, Sucrose; Cel, Cellobiose; Tre, Trehalose; Gal, Galactose; Mat, Mannitol; Sor, Sorbitol; Pyr, Pyruvic acid; Pro, Propionic acid

[b] SA, Sugar alcohols; OA, Organic acids

4.3.3 PCR amplification, phylogenetic and multilocus analysis

Partial *nif*H sequences were recovered from ten clonally pure Cape *Frankia* isolates. The identity of the products were confirmed by BLAST analysis and their relationship to other strains and sequences from actinorhizal nodules by phylogenetic analysis (Figures 4.7, 4.6). Isolates were found to represent each of the numerically dominant CC genotypes found in the the field study reported in chapter 2, namely CC-1, CC-3 and CC-4. In each case *nif*H sequences from the isolates were identical to those found in nodules of the host species from which they were isolated (Figure 4.6). Almost full length 16S rRNA amplicons of the expected size were recovered from nine of the ten Cape isolates. Nucleotide BLAST analysis confirmed their identity as *Frankia* and phylogenetic analysis assigned them to the same host infection groups as *nif*H analysis (Figures 4.8 and 4.7). Amplification of the actinomycete-specific 23S rRNA insertion sequence was successful for all Cape isolates, and phylogenetic analysis revealed that Cape strains from Cluster I did not fall into any of the previously described subgroups based on this marker (Figure 4.9). Cape strains were most similar to groups VI (the only *Elaeagnus*-HIG group previously defined using the 23S rRNA domain III IS) and subgroup IIIb within the *Alnus*-HIG, according to signature sequences within the IS which have previously been used in probe design [80, 81, 81]. These sequences are presented in Figures A and A. The 16S rRNA, 23S rRNA and *nif*H gene fragments analyzed differed slightly in their ability to discriminate between Cape isolates, with ten strains represented by six, seven and six unique sequences for each of these markers, respectively (Figures 4.7, 4.9 & 4.9). Where identical sequences were recovered from different isolates 16S and nifH always agreed, but differed slightly from the 23S insertion sequence with regard to FMc2, FMc6 and FMc7. Slight differences in topology were detected between individual trees depending on target gene, but the overall positions of Cape isolates were congruent and always agreed in the assignment of isolates to major host infection groups.

For MLSA analysis PCR products of the expected size were recovered from ten *Frankia* isolates for each reactions targeting *atp*D, *dna*A, *fts*Z, *pgk* and *rpo*B gene fragments retrieved from nine Cape *Frankia* isolates. Gene fragment identities were

confirmed by BLAST analysis. Five gene fragments from 49 *Frankia* strains, ten isolates from Cape *Morella*, and *Acidothermus cellulolyticus* were concatenated with the product averaging 3149 bp. As previously demonstrated, phylogenetic analysis of this five-gene concatenation yielded a robust maximum likelihood phylogeny, with strong support for nodes at branches grouping strains from the four major *Frankia* clusters (Figure 4.10) [8]. Cluster I (*Alnus*-infective) strains grouped with support for the node of 98 and an overall mean distance of 0.063. Cluster III (*Elaeagnus*-infective) strains grouped with node support of 100, and at an overall mean distance of 0.056. Within each major cluster strains from previously identified genomospecies grouped with a maximum average dissimilarity of 0.03. Overall mean distances between all strains, from all four major *Frankia* clusters, was 0.094. The largest distance between two individual *Frankia* strains was 84.4 (Dg and Ea8.4, Table 4.5). The strain most similar to Cape isolates within CC-3 (G, Figure 4.10) was BMG5.3, with identities of $98.2 - 99.8\%$. Within CC-4 (G6) FMc6 was closest to Ea36.7, and FMc7 was closest to Hr75.2 with similarities of 99.4% and 99.8% respectively. Cluster I strains FMi1, FMk1 and FMq1 were most similar to ARgP5$_{AG}$, with similarities of $94.3 - 94.4\%$. The PHI test showed no evidence for recombination when individual gene fragments were considered ($p0.05$ was considered statistically significant). Significant evidence of recombination was found when all five loci were considered together ($p = 1.746 \times 10^{-7}$). Split decomposition analysis revealed a bushy network, suggestive of homologous recombination (Figure 4.11). Three groups were distinguished amongst Cape isolates, two among the *Elaeagnus*-infective and one within the *Alnus*-infective HIGs (clusters III and I, respectively). These putative genomospecies corresponded to groups CC-2, CC-3 and CC-4, as defined by analysis of *nif*H clusters (Figure 4.6).

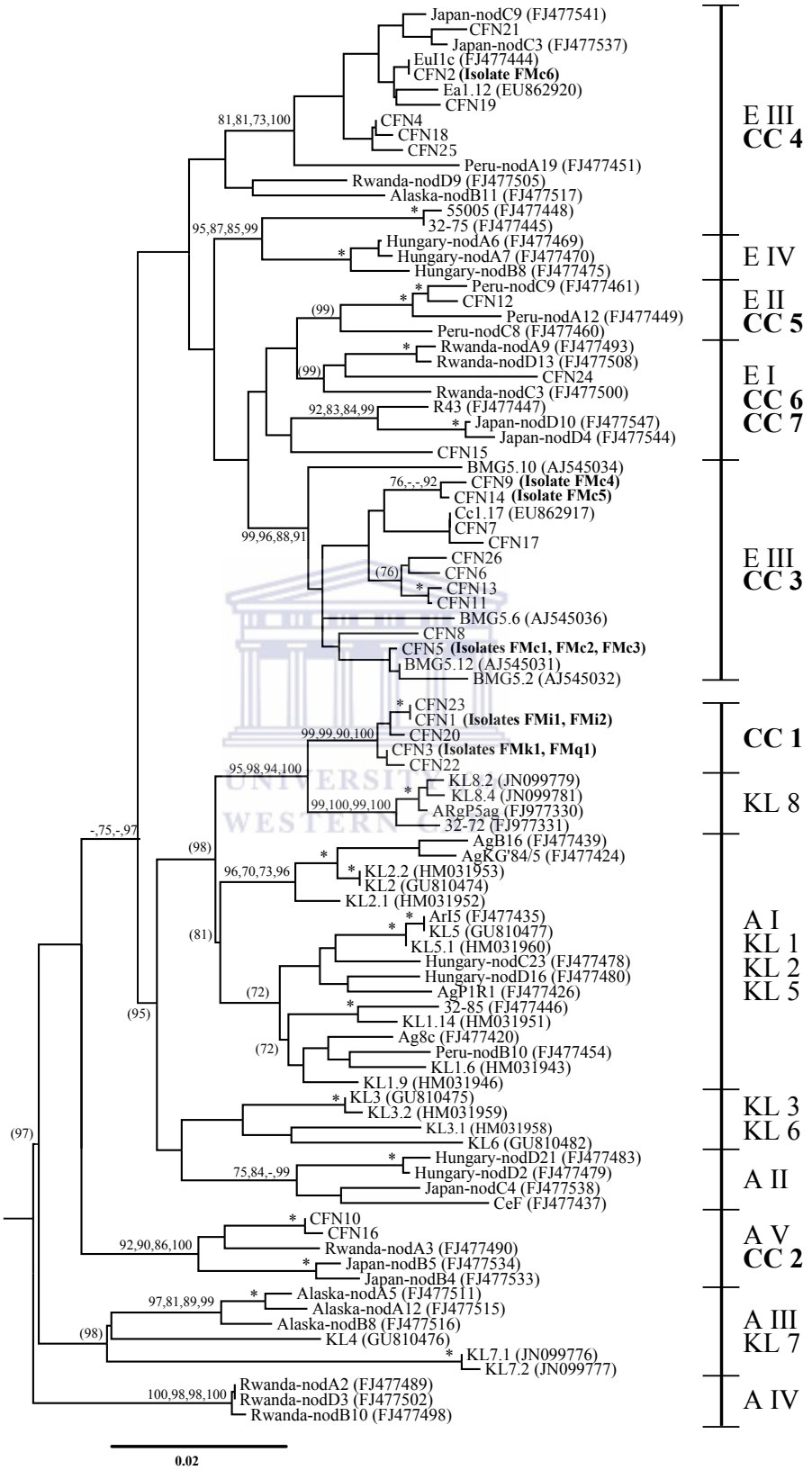


Figure 4.6: Phylogenetic tree indicating the positions of *Frankia* isolates in relation to partial *nif*H sequences from nodules. Cape isolates are bracketed in bold alongside their corresponding CFN (Cape *Frankia* Nitrogenase). Accession numbers from database sequences are indicated in parentheses. EuI1c's genome is now available (CP002299) and contains no nif*H* gene. The Genbank sequence FJ77444 was mislabeled in the original publication, and in fact refers to strain EAN1pec.

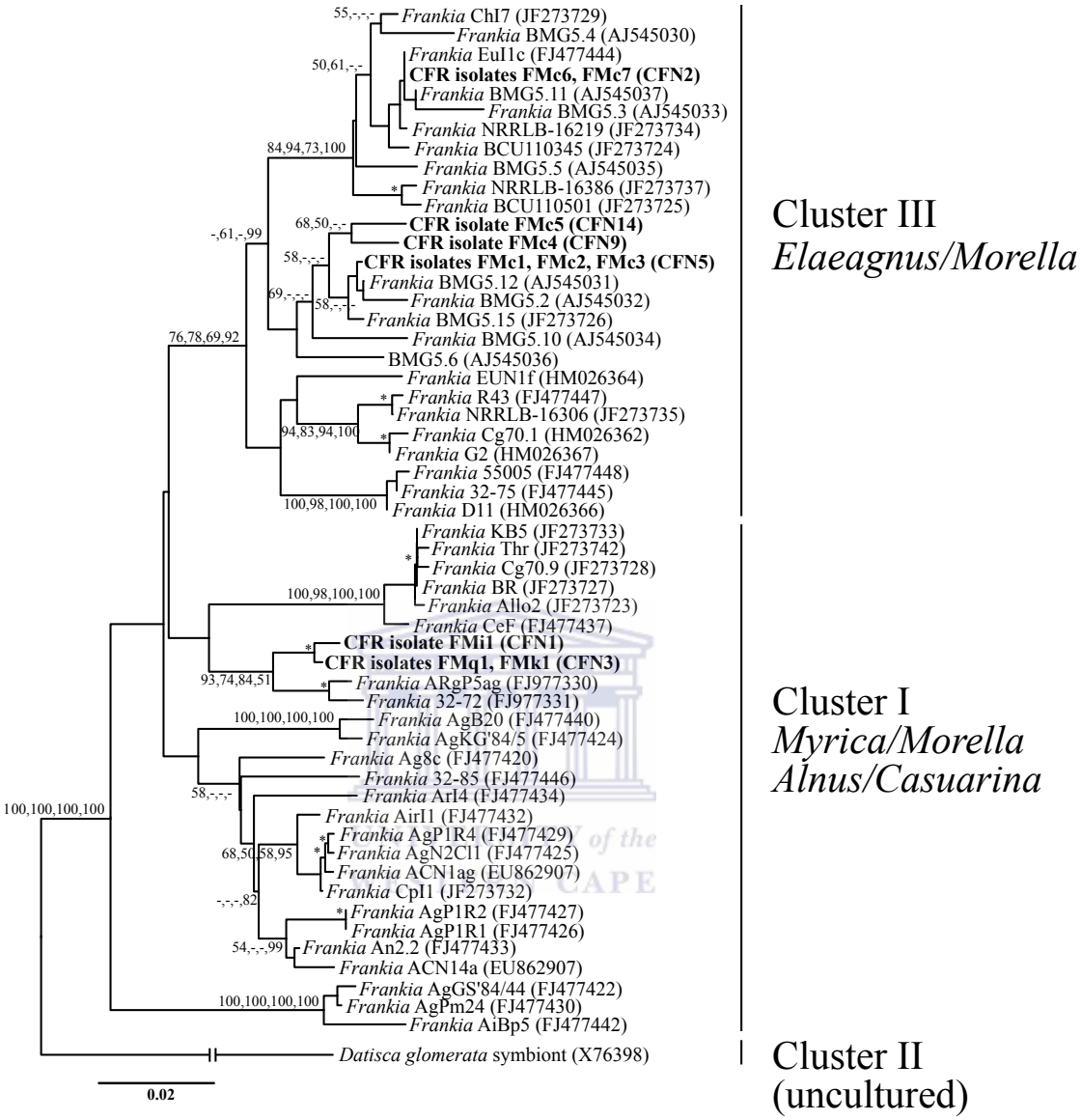


Figure 4.7: Phylogenetic tree indicating the positions of Cape *Frankia* isolates in relation to isolated strains according to *nif*H. Cape isolates are indicated in bold alongside their corresponding CFN (Cape *Frankia* Nitrogenase). Accession numbers from database sequences are indicated in parentheses. EuI1c's genome is now available (CP002299) and contains no nif*H* gene. The Genbank sequence FJ77444 in fact refers to strain EAN1pec, which was mislabeled in the original publication.

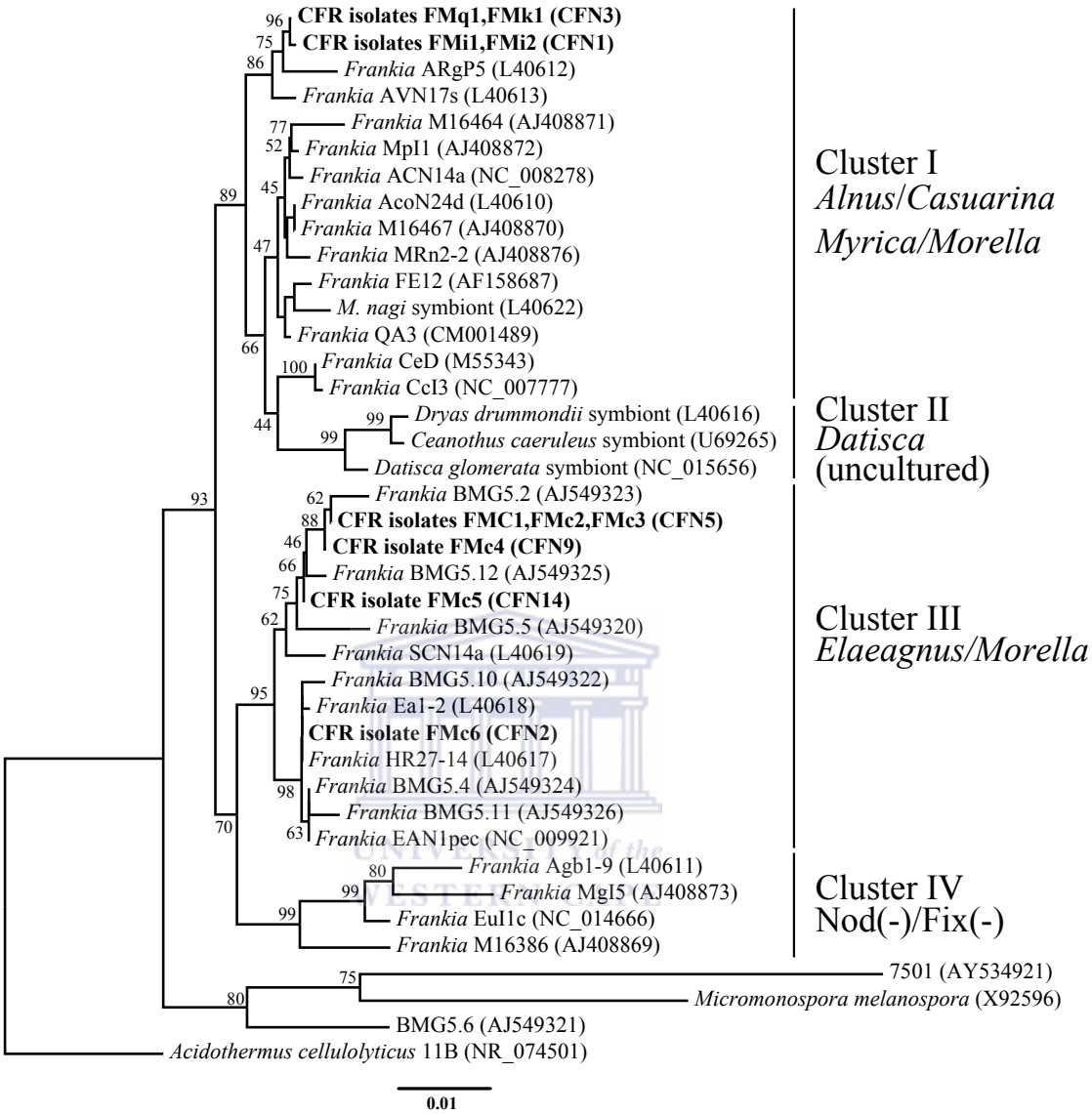


Figure 4.8: Neighbor-joining phylogenetic tree of 1202 bp partial 16S ribosomal RNA gene sequences showing positions of *Frankia* strains within clusters according to Normand et al. [59]. Isolates are labeled with trivial designations and GenBank accession number. Cape isolates are indicated in bold, 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.

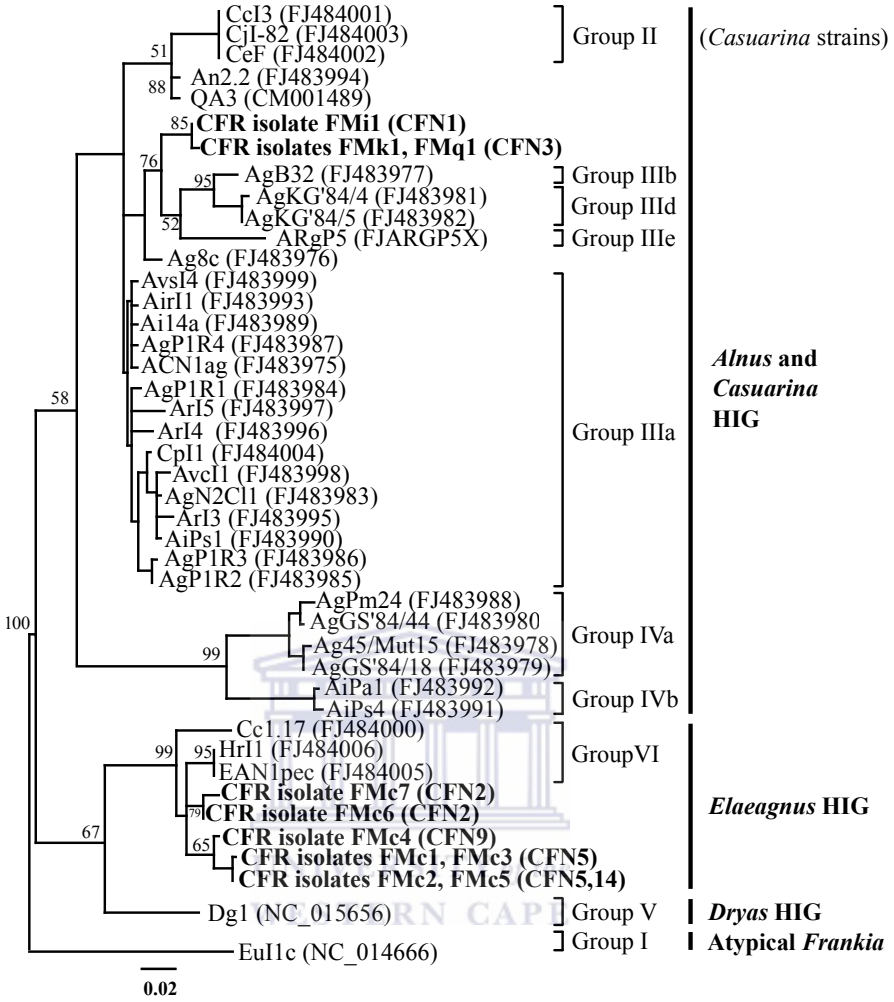


Figure 4.9: Neighbor-joining phylogenetic tree of 144 bp actinomycete-specific 23S ribosomal RNA gene insertion sequences showing positions of *Frankia* strains within clusters according to Normand et al. [59]. Isolates are labeled with trivial designations and GenBank accession number. Cape isolates are indicated in bold with trivial strain designation, and their corresponding *nif*H sequence. Dg1 is an uncultivated strains from *Datisca glomerata* nodules. Accession numbers are shown in parentheses, *Frankia* host-infection groups and 23S-based subgroups proposed by Hahn et al [81] are indicated in the margin. *Frankia* EuI1c was used to root the tree.

Table 4.4: Pairwise distances between *Frankia* isolates according to the 3149bp five-gene alignment used for MLSA. Percentage similarity presented in upper triangle, number of nucleotide difference presented in lower triangle.

GS[a]	Strain	FMc1	FMc2	FMc3	FMc4	FMc5	FMc6	FMc7	FMi1	FMk1	FMq1
G*	FMc1		99.921	99.865	99.341	98.222	92.132	92.109	88.607	88.433	88.417
	FMc2	3		99.897	99.397	98.222	92.164	92.140	88.615	88.441	88.449
	FMc3	5	4		99.341	98.166	92.109	92.085	88.560	88.449	88.457
	FMc4	22	19	21		98.150	92.188	92.164	88.607	88.369	88.377
	FMc5	57	57	59	60		92.529	92.506	88.790	88.512	88.465
G6	FMc6	248	247	249	247	236		99.976	89.306	89.131	89.171
	FMc7	249	248	250	248	237	1		89.282	89.108	89.147
G**	FMi1	358	359	361	360	354	337	338		98.087	98.103
	FMk1	364	365	365	368	362	343	344	60		99.674
	FMq1	365	364	364	367	364	341	342	60	11	

[a] Genomospecies defined at 97% sequence identity

Table 4.5: *Frankia* genomospecies as determined by DDH, AFLP and MLSA

Trivial designation	Original host	Country of origin	Ref.	Genomospecies		
				Original DDH[a]	AFLP[b]	MLSA[c]
Alnus and Myrica strains						
AcoN24d	*Alnus cordata*	France	[29]	GS1	G1	G1
$Ag24_2 51$	*Alnus glutinosa*	France	[29]	GS1	G1	G1
ArI3	*Alnus rubra*	USA	[29]	GS1	G1	G1
ARgN22d	*Alnus rugosa*	Canada	[29]	GS1	G1	G1
ACN1AG	*Alnus crispa*	Canada	[2, 29]	GS1	G1	G1
CpI1	*Comptonia peregrina*	USA	[2, 18, 29]	GS1	G1	G1
MpI1	*Morella pensylvanica*	USA	[2]	GS1	G1	G1
AirI1	*Alnus incana* ssp. *rugosa*	USA	[2]	GS1	G1	G1
AvcI1	*Alnus viridis* ssp. *crispa*	Canada	[2, 18]	GS1	G1	G1
M16467	*Morella pensylvanica*	USA	[18]	GS1	G1	G1
Ar24H5	*Alnus rubra*	France	[9]	-	G1	G1
I38	*Alnus incana*	France	[9]	-	G1	G1
M16477	*Morella pensylvanica*	USA	[9]	-	G1	G1
ACN14a	*Alnus crispa*	Canada	[9]	-	G1	G1
AcVcI	*Alnus cordata*	France	[29]	NC[d]	G1	G1
$Ac2_{18}$	*Alnus cordata*	France	[8]	-		G1
$Ac23_{23}$	*Alnus cordata*	France	[8, 29]	-		G1
$Ac24I_5$	*Alnus cordata*	France	[8]	-		G1
$Ai96_6$	*Alnus incana*	USA	[8]	-		G1
$Av200_{nod}$	*Alnus viridis*	France	[8]	-		G1
$Av201_{nod}$	*Alnus viridis*	France	[8]	-		G1
$Av59_7$	*Alnus viridis*	-	[8]	-		G1
$Mg60_{2AG}$	*Alnus glutinosa*	France	[8, 29]	-		G1
AVN17o	*Alnus viridis*	France	[29]	GS2	G2	G2
$Ac23_4 0$	*Alnus crispa*	France	[29]	GS2	G2	G2
AVL3	*Alnus viridis*	France	[9]	-	G2	G2
$ARgP5_{AG}$	*Alnus rugosa*	Canada	[29]	GS3	G3	G3
FMi1	*Morella integra*	South Africa	This study	-		G**
FMk1	*Morella kraussiana*	South Africa	This study	-		G**
FMq1	*Morella quercifolia*	South Africa	This study	-		G**
AJ01	*Alnus japonica*	Japan	[8]	-		G?
Ag21D1	*Alnus glutinosa*	France	[8]	-		G?

Table 4.5 *continued on following page...*

Table 4.5 – *Continued*

Trivial designation	Original host	Country of origin	Ref.	Genomospecies		
				Original DDH[a]	AFLP[b]	MLSA[c]
Casuarinaceae strains						
BR	*Casuarina equisetifolia*	Brazil	[29]	GS9	G4	G4
CcI3	*Casuarina cunninghamiana*	USA	[29]	GS9	G4	G4
CjI-82	*Casuarina junghuniana*	Thailand	[29]	GS9	G4	G4
TA	*Allocasuarina torulosa*	Australia	[29]	GS9	G4	G4
$Cg70_4$	*Casuarina glauca*	India	[9]	-	G4	G4
$Cg70_5$	*Casuarina glauca*	India	[9]	-	G4	G4
$Cg70_3$	*Casuarina glauca*	India	[9]	-	**G5**	**G4**
CcI2	*Casuarina cunninghamiana*	USA	[9, 29]	NC	**G5**	**G4**
Elaeagnaceae/Rhamnaceae strains						
FMc1	*Morella cordifolia*	South Africa	This study	-	-	G*
FMc2	*Morella cordifolia*	South Africa	This study	-	-	G*
FMc3	*Morella cordifolia*	South Africa	This study	-	-	G*
FMc4	*Morella cordifolia*	South Africa	This study	-	-	G*
FMc5	*Morella cordifolia*	South Africa	This study	-	-	G*
BMG5.3	*Elaeagnus angustifolia*	Tunisia	[8, 33]	-	-	G*
FMc6	*Morella cordifolia*	South Africa	This study	-	-	G6
FMc7	*Morella cordifolia*	South Africa	This study	-	-	G6
ChI_7	*Colletia hystrix*	Chile	[8, 24]	-	-	G6
$Ea1_{12}$	*Elaeagnus angustifolia*	France	[29]	**GS4**	**G6**	**G6**
$Ea35_2$	*Elaeagnus angustifolia*	Italy	[8, 41]	-	-	G6
$Ea36_7$	*Elaeagnus angustifolia*	France	[8, 29]	-	-	G6
$Ea48_1$	*Elaeagnus angustifolia*	France	[8]	-	-	G6
$Ea48_4$	*Elaeagnus angustifolia*	France	[8, 41]	**GS4**	**G6**	**G6**
$Ea7_1$	*Elaeagnus angustifolia*	France	[8, 41]	**GS4**	**G6**	**G6**
EaCm51	*Elaeagnus angustifolia*	France	[8, 29]	**GS4**	**G6**	**G6**
EAN1pec	*Elaeagnus angustifolia*	USA	[8]	**GS5**		**G6**
ORS060501	*Colletia spinosissima*	Argentina	[29]	NC	G6	G6
$Hr75_2$	*Hippophae rhamnoides*	France	[9]	-	G6	G6
GFN14a	*Hippophae rhamnoides*	China	[29]	NC	G9	ND
CH37	*Hippophae rhamnoides*	France	[9]	-	G9	G9
EUN1f	*Elaeagnus umbellata*	USA	[29]	GS6	G9	G9
$Cg70_1$	*Casuarina glauca*	India	[8]	-	-	G9
Ea8.4	*Elaeagnus angustifolia*	France	[8, 41]	-	-	G?
Atypical strains						
MgI5	*Myrica gale*	USA	-	-	G7	G?
Pti1	*Purshia tridentata*	USA	[66]	-	G7	G?
Cn3	*Coriaria nepalensis*	-	[8, 52]	-	-	G?
EuI1a	*Elaeagnus umbellata*	USA	[2, 8]	-	-	G?
G2	*Casuarina equisetifolia*	Guadeloupe	[2]	-	G8	G8

Incompatibilities between genomospecies are indicated in bold
[a] Fernandez 1989, [b] Bautista et al. 2009, [c] Bautista 2010
[d] Strain included in DDH study, but Not Classified

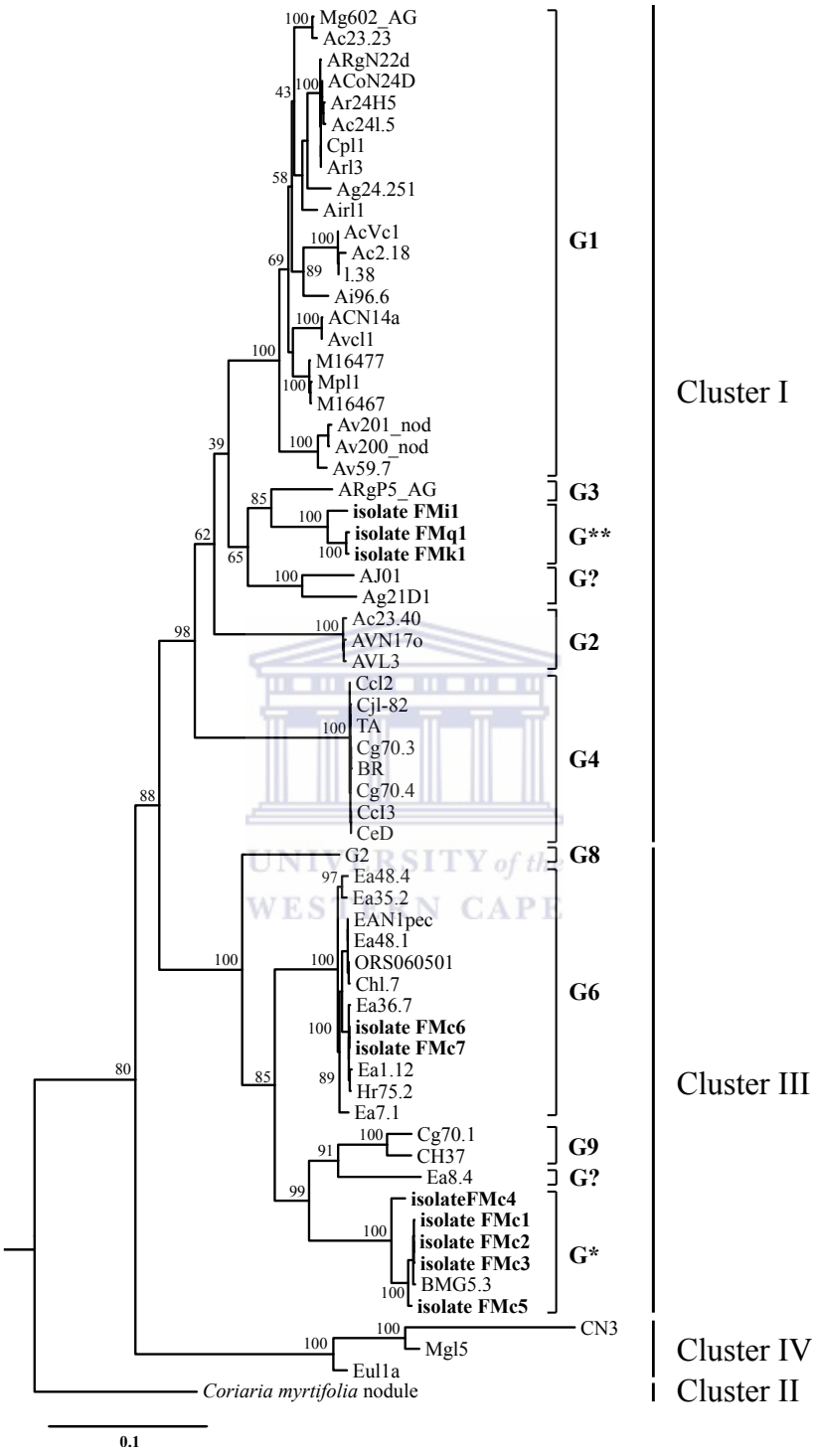


Figure 4.10: Maximum-likelihood based tree of concatenated *atp*D, *dna*A, *pgk*, *rpo*B and *fts*Z gene fragments from *Frankia* strains. Cape *Frankia* isolates are indicated in bold. Host infection clusters according to Normand et. al. [58] are indicated on the right. Strains assignments to genomospecies within host infection groups (G) (as defined by Bautista et. al [9]) are indicated. G? indicates strains not yet assigned to genomospecies, G** and G* indicate potentially novel genomospecies containing Cape isolates. Bootstrap support measures (BS) are indicated on the nodes (100 replicates). *A. cellulolyticus* was used as outgroup.

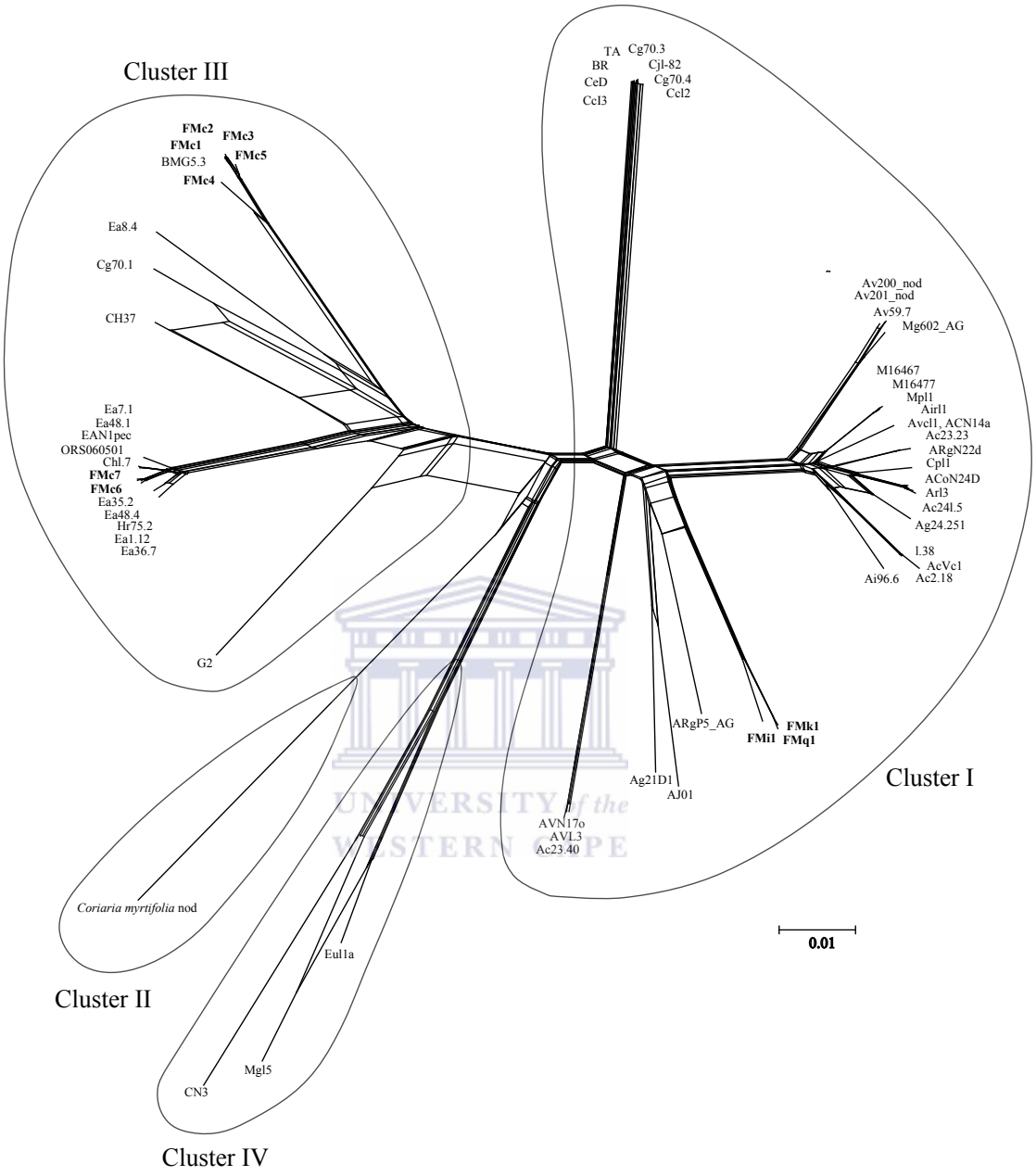


Figure 4.11: NeighourNet graph of concatenated *atp*D, *dna*A, *pgk*, *rpo*B and *fts*Z gene fragments from *Frankia* strains constructed in SplitsTree 4.0. Cape *Frankia* isolates are indicated in bold. Host infection clusters according to Normand et al. [58] are indicated. Split decomposition analysis resulted in a bushy network, which with PHI test analysis ($p = 1.745 \times 10^{-9}$) indicates significant evidence of recombination. This is particularly apparent among Cluster III strains.

4.4 Discussion

A total of ten *Frankia* strains were recovered from four species of southern African *Morella*: FMc1-FMc7 from *M. cordifolia*, FMi1 from *M. integra*, FMk1 from *M. kraussiana* and FMq1 from *M. quercifolia*. While *Frankia* strains have been isolated in Africa, they have never previously been obtained from species endemic to the continent [32, 33]. All seven strains from *M. cordifolia* were found to belong to the *Elaeagnus*-HIG (Cluster III) by 16S rRNA sequencing (Figure 4.8), and to fall into CC-3 and CC-4 by *nif*H sequencing (Figures 4.6, 4.7). Similarly, the remaining three strains were assigned to CC-1 within *Frankia* Cluster I (the *Alnus*-HIG), the numerically dominant *Alnus*-HIG subcluster In Cape *Morella* nodules (see Chapter 2).

While several isolation protocols were attempted, none was universally successful for all host species. Isolates were recovered from *M. cordifolia* using either propionate-based liquid medium (DPM) without a source of fixed nitrogen and containing cyclohexamide and nalidixic acid, or a similarly-formulated medium containing sodium azide and with pyruvate replacing propionate. Isolation attempts using these same liquid media failed to recover strains from *M. serrata*, *M. integra*, *M. quercifolia*, *M. kraussiana* and *M. diversifolia*. The isolation strategy was therefore modified empirically by varying carbon sources, eliminating nalidixic acid and sodium azide, adding $NH_4C\ell$ and attempting isolation using solid media. Ultimately I succeeded in isolating a single Cluster I (*Alnus*-HIG) strain from each of *M. integra*, *M. quercifolia* and *M. kraussiana* using gellan gum based solid media containing pyruvate, peptone and $NH_4C\ell$, and without antimicrobial compounds. Addition of $NH_4C\ell$ was the critical factor, and many isolates (all the same strain) were quickly recovered from *M. integra* nodules once this was included in the media. In the case of *M. diversifolia* from which isolates were not obtained: once the epidermis and periderm were removed from roots surrounding nodules the cortex was found to be blood-red, possibly indicating the presence of anthocyanins. This was not seen in roots of the other hosts. Differences in root chemistry may explain the failure to recover strains from this species as plant compounds (such as phenolics) are known to interfere with *Frankia* isolation [11, 47]. After isolation and dereplication *Frankia* monocultures were initially maintained in liquid DPM

and on FDM gellan gum-based plates containing pyruvate, peptone and $NH_4C\ell$.

All ten isolates displayed typical *Frankia* phenotypes in liquid culture, including branched hyphae, vesicles, and terminal and interlocular sporangia with non-motile spores. Sporangia and vesicles were also produced by colonies growing on solid media (Table 4.2, Figures 4.3a, 4.1b and 4.2). There were, however, variations in these phenotypes between isolates from the different host infection groups. Sporangia from Cluster III were typically oval to round, whereas Cluster I sporangia were invariably irregularly shaped, and more similar to those seen in *Frankia* ArI3 (DMSZ 44251, https://www.dsmz.de/), an isolate from *Alnus rubra* (Figures 4.1b) and 4.2). In Cluster III strains vesicle production was constitutive with vesicles present even in young cultures in the presence of fixed nitrogen, although in lower numbers per field than under nitrogen-limiting conditions (Table 4.2). Cluster I strains produced vesicles only when no source of fixed nitrogen was available, and always to a lesser degree than Cluster III strains. Similarly, all Cluster III strains were pigmented while those from Cluster I were not (Table 4.2). Pigment production in individual Cluster III strains varied with growth substrate, as can be seen with FMc2 grown on either pyruvate or pyruvate supplemented with peptone (Figure 4.4. Pigment colour-variation under different growth conditions is a general characteristic in actinomycetes, as is the fact that their ability to produce pigments may be lost [76]. It is thus not generally considered a taxonomically useful trait in *Frankia* [47, 76]. However, two groups of strains (FMc1, FMc2, FMc3 from CC-3; and FMc6, FMc7 from CC-4) were each identical in terms of their *nif*H and 16S rRNA gene sequences, but consistently produced different soluble pigments in DPM liquid culture (pink, blue and black for isolates from CC-3, yellow and orange for those from CC-4, respectively) and were thus considered different strains for the purposes of this study. With extended growth on FDM, strain FMc5 displayed an extraordinary phenotype with swollen hyphae (Figure 4.3b). These are similar to the "terminal and intercalary swellings" reported for *Frankia* CgI4 in some media [49]. It is possible that this phenotype is indicative of some form of nutrient sequestration under starvation conditions, *Frankia* is known to accumulate glycogen and trehalose as a storage compound [12, 22]. When colonies displaying this phenotype were sub-cultured their hyphae reverted to more typical forms.

Frankia strains vary in their ability to grow on carbon sources which, together with their slow doubling times, has implications for their maintenance in culture. Cluster I strains seem to be limited to growth on organic acids and Tween, whereas Cluster III strains may also use simple sugars and sugar alcohols [12,46,48,67,73,76]. All strains tested thus far, with the exception of EAN1pec[4], are able to grow on propionate [12, 48, 73, 76]. Because of this, and while propionate is not always the best substrate for rapid growth, it is frequently used in isolation media [47, 73]. Transferring newly-isolated cultures to media promoting improved growth requires that the ability of new isolates to utilize various carbon substrates first be determined [12]. As Cape isolates representing both the *Alnus*- and *Elaeagnus*-infective *Frankia* clusters were available it was of interest to determine whether carbon substrate utilization patterns mirrored those of previously-tested strains. As determining growth rates in filamentous organisms presents unique challenges [12] and time did not allow for deeper investigation, only the ability to grow on a given substrate was examined. Using a modification of the method employed by Bassi et al. [7], no hyphal elongation was detected on plates lacking a carbon source (Figure 4.5). Cape strains from Cluster I were limited to growth on propionate and pyruvate (Table 4.3). Those from Cluster III were divided into two groups, corresponding to those determined by phylogenetic analysis of their *nif*H genes. In addition to organic acids, strains from CC-3 grew on cellobiose, while FMc6 from CC-4 also utilized fructose. Interestingly, no strains were capable of growth on trehalose, a common storage compound in *Frankia* which is readily assimilated by ArI3 [12, 48]. Not all Cape isolates were tested as, after several rounds of sub-culturing, some strains (FMc1, FMc4, FMc7 and FMk1) would no longer grow in liquid culture and could thus not be used in carbon source utilization experiments. Lechevalier and Lechevalier report that strains isolated on any given medium may spontaneously stop growing after being sub-cultured on that same medium [47].

The identities of nine of the ten Cape *Frankia* isolates were confirmed by BLAST analysis of near-complete 16S rRNA genes, and all isolates were confirmed as *Frankia* through analysis of their 23S rRNA IS sequences (Figures 4.8 & 4.9). Phylogenetic

[4]And possibly a newly-isolated Cluster II strain, provided the organism is grown at an appropriate pH [34]

analysis of these sequences and of the *nif*H gene confirmed the isolates' placement within the *Alnus* (Fmi1, FMk1, FMq1) and *Elaeagnus* (FMc1-FMc7) host infection groups, respectively.*Frankia* strains FMc1, FMc2 and FMc3 were identical in terms of both their 16S and *nif*H sequences but displayed significant phenotypic variation, including distinct differences in pigment production, substrate utilization and sporulation intensity. Previous studies have reported that individual genetic markers may not have sufficent discriminatory power to distinguish phenotypically distinct strains, and this appears to be the case with at least some of the Cape isolates [17]. The 23S rRNA IS was sequenced principally in order to design lineage-specific probes for in-situ hybridization aimed at detecting co-population within individual nodules from the Cape flora, which unfortunately could not be accomplished during this project. It was found that existing probe sets targeting *Frankia* subgroup IIIb and VI signature sequences within the 23S rRNA IS would be applicable for detecting strains from CC-1 and CC-4 (Cape *Frankia* clusters defined by *nif*H), but that new probes would be needed to detect strains from CC-3 (figures A & A).

Multilocus sequence analysis has advantages over analyses making use of only one marker, both in that it has higher discriminatory power and that the effects of horizontal gene transfer in individual taxa, which can lead to aberrant phylogenies in single-marker analyses (as seen with the *nif*H gene), are buffered in MLSA [30, 31, 75]. The method is also recognized as a promising alternative to DDH, which remains a current standard for delineation of prokaryote species [30, 68]. An MLSA scheme has recently been developed specifically targeting *Frankia*, and which is reported to assign strains to recognized genomospecies in agreement with both DDH and AFLP [8, 9]. Gene regions used in this were selected as they were highly discriminative, widely spaced in each genome, ubiquitous in *Frankia*, and present as single copies in all of the strains upon which primer design was based [31, 78]. Additionally, each gene was found to be under neutral selection [8]. Unlike DDH and AFLP, MLSA has the advantage of being sequence based; comparisons between strains can therefore be carried out without the need for a library of organisms within a laboratory (no such *Frankia* strain library exists in Southern Africa). In addition, sequence-based data is readily transferable between laboratories whereas results from fingerprinting methods are not [8, 31].

While *Frankia* are easily discriminated from other actinobacteria, and there is clearly a high degree of diversity within the genus, only one species (*Frankia alni*) has been validly published [76]. This is because, despite strains having been assigned to genomospecies by DDH (the gold standard for bacterial species delineation) no incontrovertible phenotypic traits discriminating between these genospecies have been reported to date [1, 2, 8, 9, 18, 29, 29]. For the present genospecies stand in place of bono fide species assignments in *Frankia* taxonomy. Despite its apparent utility a number of difficulties with this MLSA scheme exist, but were not considered detrimental to current study. For example, *Casuarina*-infective strains from AFLP genomospecies G4 and G5 are not discriminated by MLSA (Figure 4.10, Table 4.5) [8, 9, 29]. Furthermore, strains from GS4 and GS5 within *Frankia* Cluster III were assigned to G6 by MLSA analysis [8], despite strains from GS5 (comprising EAN1pec, HRX401a and TX31e^{HR}) displaying low DNA re-association (38-49%) to strain Ea1$_{12}$ from GS4, the basis for their assignment to separate genomospecies [29]. In agreement with MLSA, evolutionary genomic distance values from AFLP data suggest that these strains should be included in the same genomospecies [8, 9]. Large differences in *Frankia* DNA/DNA re-association values in what otherwise appear to be genetically similar strains is understandable as *Frankia* have the largest variation in genome size for any bacterial genus described to date [43, 58, 65]. *Casuarina*-infective strains were not isolated in the current study, nor were they detected in nodules from Cape *Morella*. In order to both improve upon the resolution of the phylogenetic analysis, and to determine the relationships of Cape *Frankia* isolates to strains from previously identified genospecies, Bautista's MLSA method was employed in the current study.

In agreement with single-gene markers, MLSA assigned isolates from Cape *Morella* to *Frankia* Cluster I and Cluster III, although small discrepancies were found between phylogenies as determined by 16S rRNA, 23S rRNA, *nif*H and MLSA, specifically in the relationships between Cape isolates and BMG strains from North Africa (Figures 4.10, 4.7 and 4.8). Individual genospecies within *Frankia* are described at average sequence distances of up to 0.03 (97% identity) [8] and are presented in Table 4.4. Using this threshold Cape isolates were assigned to three genospecies, and these genospecies agreed with the groupings identified by the *nif*H phylogenetic analysis (Figures 4.10 and 4.11. Cluster III strains were assigned

to either G6 (corresponding to CC-3), a large group of globally- strains including EAN1pec, or to an as-yet undescribed genomospecies (G*) containing BMG5.3 (a northern African strain isolated from roots nodules formed on *Elaeagnus angustifolia* after inoculation with Tunisian soils [33], corresponding to CC-4). Cluster I strains from CC-1 (FMi1, FMq1 and FMk1) were grouped into a unique genomospecies (G**) most similar to ARgP5, with an average similarity of 94.4 (Figures 4.10 and 4.11, Table 4.5). Cape isolates FMc6 and FMc7 fall into group G6, which contains multiple DDH genomospecies which cannot be discriminated by MLSA [8]. A comparison of these strains with other members of Bautista's genomospecies G6 using DDH, or whole-genome MLSA, will be interesting. As isolate placement within genomospecies by MLSA was consistent with the *nif*H groupings defined in the nodular diversity study (Figure 4.6 and Chapter 2), the names of these groups (CC-1, CC-3 and CC-4) were retained. Isolates FMc1, FMc2 and FMc3, which were identical in terms of their *nif*H, 16S rRNA and 23S rRNA sequences, were also very closely related according to MLSA, with only three to five differences over the length of the entire 3149 bp alignment (Table 4.4). Similarly, FMc6 and FMc7 differed at only one position. FMk1 and FMq1, which are assumed to be the same organism but which were isolated from different hosts at the same site, differed at 11 positions. These differences could possibly be attributed to sequencing errors. If this is the case strains from these groups may in fact be identical in terms of their core genomes, despite apparent phenotypic differences, reducing the total number of isolates from ten to six (two from Cluster I and four from Cluster III).

While phylogenetic methods have allowed for direct detection of microorganisms in the environment, pure cultures of isolates still play a role in understanding the ecology of microbial ecosystems, and their interactions with higher organisms [79]. The ability of the strains isolated in this study to re-infect their original hosts, and to establish effective symbioses remains to be demonstrated. Based on the identification of three genospecies amongst Cape *Frankia* isolates, three strains (FMi1 from CC-1, FMc2 from CC-3 and FMc6 from CC-4) representing numerically dominant strains from *Frankia* Cluster I (*Alnus*-infective) and Cluster III (*Elaeagnus*-infective) in nodules from Cape *Morella* were selected for whole genome sequencing.

4.5 Conclusion

Ten clonally pure *Frankia* strains, seven from Cluster III and three from Cluster I, representing each of the numerically dominant CC genotypes found in the nodules of Cape *Morella* species (CC-1, CC-3 and CC-4), were isolated from nodules of *M. cordifolia* (FMc1-FMc7), *M. integra* (FMi1), *M. kraussiana* (FMk1) and *M. quercifolia* (FMq1). 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. All ten isolates displayed phenotypes typical of *Frankia*, but differed in production of vesicles, pigments and their ability to grow on individual carbon sources. 16S rRNA gene and 23*S* rRNA gene IS BLAST analysis confirmed the identity of the isolates as *Frankia*, and phylogenetic analysis assigned them to the same host-infection groups indicated by *nif*H analysis. A multilocus sequence analysis assigned strains to three genospecies, two (from *Frankia* Cluster III) previously described and one (from Cluster I) unique to the Cape flora. These genospecies agreed with groupings determined through phylogenetic analysis of the *nif*H gene, and a representative of each was selected for whole genome sequencing. The strains described in this chapter are the first to be isolated from actinorhizal hosts endemic to the African continent.

References

[1] Akimov, V.N., Dobritsa, S.V., Stupar, O.S.: Grouping of Frankia strains by DNA-DNA homology: how many genospecies are there in the genus *Frankia*? Developments in Plant and Soil Sciences **48**, 635–636 (1991)

[2] An, C.S., Riggsby, W.S., Mullin, B.C.: Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. International Journal of Systematic Bacteriology **35**, 140–146 (1985)

[3] Baker, D., O'Keefe, D.: A modified sucrose fractionation procedure for the isolation of frankiae from actinorhizal root nodules and soil samples. Plant and Soil **78**(1-2), 23–28 (1984)

[4] Baker, D., Pengelly, W.L., Torrey, J.G.: Immunochemical analysis of relationships among isolated frankiae (Actinomycetales). International Journal of Systematic Bacteriology **31**(2), 148–151 (1981)

[5] Baker, D.D.: Relationships among pure cultured strains of *Frankia* based on host specificity. Physiologia Plantarum **70**, 245–248 (1987)

[6] Baker, D.D., Schwintzer, C.R.: Introduction. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 1–14. Academic Press, Inc., San Diego, California (1990)

[7] Bassi, C.A., Benson, D.R.: Growth characteristics of the slow-growing actinobacterium *Frankia* sp. strain CcI3 on solid media. Physiologia Plantarum **130**, 391–399 (2007)

[8] Bautista, G.H.H.: Phylogenomic study and specific depiction of Frankia genus. Special focus on non-cultivatable strains and ecological implications. Ph.D. thesis, Universite Claude Bernard, Lyon (2010)

[9] Bautista, G.H.H., Cruz, H.A., Nesme, X., Valdes, M., Mendoza, H.A., Fernandez, M.P.: Genomospecies identification and phylogenomic relevance of AFLP analysis of isolated and non-isolated strains of *Frankia* spp. Systematic and Applied Microbiology **34**, 200–206 (2011)

[10] Becking, J.H., de Boer, W.E., Houwink, A.L.: Electron microscopy of the endophyte of *Alnus glutinosa*. Antonie van Leeuwenhoek **30**(1), 343–376 (1964)

[11] Benson, D.R.: Isolation of *Frankia* strains from Alder actinorhizal root nodules. Applied and Environmental Microbiology **44**(2), 461–465 (1982)

[12] Benson, D.R., Schultz, N.A.: Physiology and biochemistry of *Frankia* in culture. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 107–128. Academic Press, Inc., San Diego, California (1990)

[13] Benson, D.R., Silvester, W.B.: Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiological Reviews **57**(2), 293–319 (1993)

[14] Benson, D.R., Vanden Heuvel, B.D., Potter, D.: Actinorhizal symbioses: diversity and biogeography. In: M. Gillings (ed.) Plant Microbiology, pp. 102–133. BIOS Scientific Publishers Ltd., Oxford (2004)

[15] Benson, David R, Schultz, N.A.: Physiology and biochemistry of *Frankia* in culture. In: R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal plants, chap. 6, pp. 107–128. Academic Press, Inc., San Diego, California (1990)

[16] Berg, H., McDowell, L.: Cytochemistry of the wall of infected cells in *Casuarina* actinorhizae. Canadian Journal of Botany **66**, 2038–2047 (1988)

[17] Berneche-D'Amours, A., Ghinet, M.G., Beaudin, J., Brzezinski, R., Roy, S.: Sequence analysis of rpoB and rpoD gene fragments reveals the phylogenetic diversity of actinobacteria of genus *Frankia*. Canadian Journal of Microbiology **57**, 244–249 (2011)

[18] Bloom, R.A., Mullin, B.C., Tate, R.L.: DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pennsylvanica* (bayberry). Applied and Environmental Microbiology **55**(9), 2155–2160 (1989)

[19] Bond, G., Fletcher, W.W., Ferguson, T.P.: The development and function of the root nodules of *Alnus*, *Myrica* and *Hippophae*. Plant and Soil **5**(4), 309–323 (1954)

[20] Bruen, T.C., Phillipe, H., Bryant, D.: A simple and robust statistical test for detecting the presence of recombination. Genetics **172**, 2665–2681 (2005)

[21] Burgraaf, A.J.D., Quispel, A., Tak, T., Valstar, J.: Methods of isolation and cultivation of *Frankia* species from actinorhizas. Plant And Soil **61**, 157–168 (1981)

[22] Burleigh, S.H., Dawson, J.O.: Desiccation tolerance and trehalose production in *Frankia* hyphae. Soil Biology and Biochemistry **26**(5), 593–598 (1994)

[23] Callaham, D., Del Tredici, P., Torrey, J.G.: Isolation and cultivation in vitro of the actinomycete causing root nodulation in *Comptonia*. Science **199**, 899–902 (1978)

[24] Caru, M.: Characterization of native *Frankia* strains isolated from chilean shrubs. Plant and Soil **157**, 137–145 (1993)

[25] Clawson, M.L., Benson, D.R.: Natural diversity of *Frankia* strains in actinorhizal root nodules from promiscuous hosts in the family Myricaceae. Applied and Environmental Microbiology **65**(10), 4521–4527 (1999)

[26] Clawson, M.L., Bourret, A., Benson, D.R.: 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 (2004)

[27] Darriba, D., Taboada, G.L., Doallo, R., Posada, D.: jModelTest 2: more models, new heuristics and parallel computing. Nature Methods **9**, 772 (2012)

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

[29] Fernandez, M.P., Meugnier, H., Grimont, P.A.D., Bardin, R.: Deoxyribonucleic acid relatedness among members of the genus *Frankia*. International Journal of Systematic Bacteriology **39**(4), 424–429 (1989)

[30] Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Peer, Y.V.D., Vandamme, P., Thompson, F.L., Swings, J.: Re-evaluating prokaryotic species. Nature Reviews Microbiology **3**(September), 733–739 (2005)

[31] Glaeser, S.P., Kampfer, P.: Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Systematic and Applied Microbiology **38**(4), 237–245 (2015)

[32] Gomaa, A.M., Abo-Aba, S.E.M., Awad, N.S.: Isolation, characterization and genetic differentiation of *Frankia* sp . isolated from ecologically different Egyptian locations. Research Journal of Cell and Molecular Biology **2**(1), 6–17 (2008)

[33] Gtari, M., Brusetti, L., Skander, G., Mora, D., Boudabous, A., Daffonchio, D.: Isolation of *Elaeagnus*-compatible *Frankia* from soils collected in Tunisia. FEMS Microbiology Letters **234**, 349–355 (2004)

[34] Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., Sbissi, I., Ayari, A., Yamanaka, T., Normand, P., Tisa, L.S., Boudabous, A.: Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Scientific Reports **5**, 13,112 (2015)

[35] Gtari, M., Tisa, L.S., Normand, P.: Diversity of *Frankia* strains, actinobacterial symbionts of actinorhizal plants. In: R. Aroca (ed.) Symbiotic endophytes, chap. 7, pp. 123–148. Springer, Berlin Heidelberg (2013)

[36] Guetsky, R., Gollop, N., Bernstein, N.: Genetic diversity of *Frankia* strains isolated from root nodules of *Casuarina* in Israel. Israel Journal of Plant Sciences **53**(2), 125–133 (2005)

[37] Guidon, S., Gascuel, O.: A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology **52**, 696–704 (2003)

[38] He, X.H., Chen, L.G., Hu, X.Q., Asghar, S.: Natural diversity of nodular microsymbionts of *Myrica rubra*. Plant and Soil **262**, 229–239 (2004)

[39] Honerlage, W., Hahn, D., Zepp, K., Zeyer, J., Normand, P.: A hypervariable 23S rRNA region provides a discriminating target for specific characterization

of uncultured and cultured *Frankia*. Systematic and Applied Microbiology **17**(3), 433–443 (1994)

[40] Huson, D.H., Bryant, D.: Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution **23**(2), 254–267 (2006)

[41] Jamann, S., Fernandez, M.P., Normand, P.: Typing method for N_2-fixing bacteria based on PCR-RFLP - application to the characterization of *Frankia* strains. Molecular Ecology **2**, 17–26 (1993)

[42] Koch, R.: Die Aetiologie der Tuberculose. Berliner Klinische Wochenschrift **19**(15), 221–230 (1882)

[43] Kucho, K.I., Yamanaka, T., Sasakawa, H., Mansour, S.R., Uchiumi, T.: Different dynamics of genome content shuffling among host-specificity groups of the symbiotic actinobacterium *Frankia*. BMC genomics **15**(1), 609–621 (2014)

[44] Lalonde, M.: Immunological and ultrastructural demonstration of nodulation of the European *Alnus glutinosa* (L.) Gaertn. host plant by an actinomycetal isolate from the North American *Comptonia peregrina* (L.) Coult. root nodule (1979)

[45] Lechevalier, M.P.: The taxonomy of the genus *Frankia*. Plant and Soil **78**(1), 1–6 (1984)

[46] Lechevalier, M.P.: Taxonomy of the Genus *Frankia* (Actinomycetales). International Journal of Systematic Bacteriology **44**(1), 1–8 (1994)

[47] Lechevalier, M.P., Lechevalier, H.A.: Systematics, isolation and culture of *Frankia*. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of Frankia and Actinorhizal Plants, pp. 35–60. Academic Press, Inc., New York, N.Y. (1990)

[48] Lopez, M.F., Young, P., Torrey, J.G.: A comparison of carbon source utilization for growth and nitrogenase activity in two *Frankia* isolates. Canadian Journal of Microbiology **32**(4), 353–358 (1986)

[49] Mansour, S.R., Dewedar, A., Torrey, J.G.: Isolation, culture, and behavior of *Frankia* strain HFPCgI4 from root nodules of *Casuarina glauca*. Botanical Gazette **151**(4), 490–496 (1990)

[50] Meesters, T.M., van Genesen, S.T., Akkermans, A.D.L.: Growth, acetylene reduction activity and localization of nitrogenase in relation to vesicle formation in *Frankia* strains Cc1.17 and Cp1.2. Archives of Microbioly **143**, 137–142 (1985)

[51] Miller, M.A., Pfeiffer, W., Schwartz, T.: Creating the CIPRES Science Gateway for inference of large phylogenetic trees (2010)

[52] Mirza, M.S., Hahn, D., Akkermans, A.D.L.: Isolation and characterization of *Frankia* Strains from *Coriaria nepalensis*. Systematic and Applied Microbiology **15**(2), 289–295 (1992)

[53] Mort, A., Normand, P., Lalonde, M.: $2-O-$methyl$-$D$-$glucose, a key sugar in the taxonomy of *Frankia*. Canadian Journal of Microbiology **29**, 993–1002 (1983)

[54] Nalin, R., Normand, P., Simonet, P., Domenach, A.M.: Polymerase chain reaction and hybridization on DNA extracted from soil as a tool for *Frankia* spp. population distribution studies in soil. Canadian Journal of Botany **77**, 1239–1247 (1999)

[55] Navarro, E., Nalin, R., Gauthier, D., Normand, P.: The nodular microsymbionts of *Gymnostoma* spp. are *Elaeagnus*-infective *Frankia* strains. Applied Environmental Microbiology **63**(4), 1610–1616 (1997)

[56] Normand, P., Chapelon, C.: Direct characterization of *Frankia* and of close phyletic neighbors from an *Alnus viridis* rhizosphere. Physiologia Plantarum **99**, 722–731 (1997)

[57] Normand, P., Fernandez, M.P.: Evolution and Diversity of *Frankia*. In: K. Pawlowski (ed.) Prokaryotic Symbionts in Plants, vol. 8, pp. 103–125. Springer-Verlag, Berlin Heidelberg (2009)

[58] Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choisne, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltsman, E., Huang, Y., Kopp, O.R., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastronunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D.: Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research **17**, 7–15 (2007)

[59] Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J.O., Evtushenko, L., Misra, A.K.: Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. International Journal of Systematic Bacteriology **46**(1), 1–9 (1996)

[60] Pommer, E.: Über die isolierung des endophyten aus den wurzelknöllchen *Alnus glutinosa* Gaertn. und uber erfolgreiche reinfektionsversuche. Ber. Dtsh. Bot. Ges. **72**, 138–150 (1959)

[61] Quispel, A.: Discoveries, discussions, and trends in research on actinorhizal root nodule symbioses before 1978. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of *Frankia* and Actinorhizal Plants, chap. 2, pp. 15–34. Academic Press, Inc., Dordrecht, the Netherlands (1990)

[62] Quispel, A.: Discoveries, discussions and trends in research on actinorhizal root nodule symbioses before 1978. In: C.R. Schwintzer, J.D. Tjepkema (eds.) The Biology of *Frankia* and Actinorhizal Plants, pp. 15–34. Academic Press, Inc., San Diego, California (1990)

[63] Rambaut, A.: FigTree v1.2.3. Institute of Evolutionary Biology, University of Edinburgh (2009). Available from http://tree.bio.ed.ac.uk/software/figtree

[64] Saskawa, H., Hiyoshi, T., Sugiyama, T.: Immuno-gold localization of nitrogenase in root nodules of Elaeagnus pungens Thunb. Plant and Cell Physiology **29**(7), 1147–1152 (1988)

[65] Sellstedt, A., Richau, K.H.: Aspects of nitrogen-fixing actinobacteria, in particular free-living and symbiotic *Frankia*. FEMS Microbiology Letters **342**(2), 179–186 (2013)

[66] Shi, Y., Ruan, J.: DNA base composition and homology values in the classification of some *Frankia* strains. Acta Microbiologica Sinica **32**, 133–136 (1992)

[67] Shipton, W.A., Burggraaf, A.J.P.: Aspects of the cultural behaviour of *Frankia* and possible ecological implications. Canadian Journal of Botany **61**, 2783–2792 (1982)

[68] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, A.D., Kampfer, P., Maiden, M.C.J., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B.: Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. International Journal of Systematic and Evolutionary Microbiology **52**, 1043–1047 (2002)

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

[70] Sukumaran, J., Holder, M.T.: DendroPy: a Python library for phylogenetic computing. Bioinformatics **26**, 1569–1571 (2010)

[71] Swensen, S.M., Benson, D.R.: Evolution of actinorhizal host plants and *Frankia* endosymbionts. In: K. Pawlowski, W.E. Newton (eds.) Nitrogen-fixing Actinorhizal Symbioses, chap. 4, pp. 73–104. Springer, Dordrecht, the Netherlands (2008)

[72] Swofford, D.L.: PAUP. Phylogenetic analysis using parsimony (and other methods). Version 4 (2003)

[73] Tisa, L., McBride, M., Ensign, J.C.: Studies of growth and morphology of *Frankia* strains EAN1pec, Eul1c, Cpl1, and ACN1ag. Canadian Journal of Botany **61**(11), 2768–2773 (1983)

[74] Torrey, J.G.: Nitrogen fixation by actinomycete-nodulated angiosperms. BioScience **28**(9), 586–592 (1978)

[75] Welsh, A.K., Dawson, J.O., Gottfried, G.J., Hahn, D.: Diversity of *Frankia* populations in root nodules of geographically isolated Arizona Alder trees in central Arizona (United States). Applied and Environmental Microbiology **75**(21), 6913–6918 (2009)

[76] Whitman, W.B., Parte, A., Goodfellow, M., Kämpfer, P., Busse, H.J., Trujillo, M.E., Ludwig, W., Suzuki, K.: Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria. v. 5, pts. 1-2. Springer (2012)

[77] Wright, E.S., Yilmaz, L.S., Noguera, D.R.: DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Applied Environmental Microbiology **78**(3), 717–725 (2012)

[78] Zeigler, D.R.: Gene sequences useful for predicting relatedness of whole genomes in bacteria. International Journal of Systematic and Evolutionary Microbiology **53**, 1893–1900 (2003)

[79] Zengler, K.: Central role of the cell in microbial ecology. Microbiology and Molecular Biology Reviews **73**(4), 712–29 (2009)

[80] Zepp, K.: Analysis of *Frankia* populations by in situ hybridization with 23S rRNA targeted oligonucleotide probes. Ph.D. thesis, Swiss Federal Institute of Technology (1996)

[81] Zepp, K., Hahn, D., Zeyer, J.: Evaluation of a 23S rRNA insertion as target for the analysis of uncultured *Frankia* populations in root nodules of alders by whole cell hybridization. Systematic and Applied Microbiology **20**(1), 124–132 (1997)

[82] Zhongze, Z., Torrey, J.G.: Biological and cultural characteristics of the effective *Frankia* strain HFPCcl3 (Actinomycetales) from *Casuarina cunninghamiana* (Casuarinaceae). Annals of Botany **56**, 367–378 (1985)

5 Genome sequencing of *Frankia* strains FMc2, FMc6 and FMi1

Abstract

Frankia strains enter into "actinorhizal" symbioses with woody trees and shrubs, upon which they form N_2-fixing root nodules. Actinorhizal plants, through this association, gain the ability to grow well on marginal soils. Four host-infection *Frankia* from both Cluster I and Cluster III nodulate *Morella* of the Cape flora of South Africa, and strains representing three genospecies are have been isolated. To date no *Frankia* isolated from endemic African hosts had been sequenced, and so draft genomes were generated for three strains, FMc2, FMc6 and FMi1, representing each of Cape genospecies. Basic assembly and annotation were carried out, and the structure of each strain's nitrogenase gene cluster ($\approx$ 20 kb each containing 17 genes associated with N_2-fixation) was determined. Slight variations in the arrangement of ORFs occur within these gene clusters, and are conserved in strains from each respective host-infection group. The arrangement of nitrogenase gene clusters from Cape Cluster III and Cluster I *Frankia* were found to precisely mirror those arrangements found in other strains from their corresponding host-infection groups, despite significant sequence divergence in individual genes.

5.1 Introduction

Actinomycetes from the genus *Frankia* pursue lifestyles as free-living saprophytes, or as microsymbionts of actinorhizal plant species [4]. The distribution of host species, their associations with *Frankia* from different host-infection groups, and the ability of *Frankia* to grow in soil independently of their hosts vary greatly [3, 6, 16, 17, 25]. With the advent of affordable whole-genome sequencing technologies, the genomic underpinnings of *Frankia* lifestyles are currently being investigated. Draft genomes from representatives of each of the four *Frankia* host-infection groups are now available [8–11, 15, 26, 30, 34]. Strains QA3 [26], Thr [11], CcI3 [18], CcI6 [15], BMG5.23 [9] and ACN14a [18] represent Cluster I. Two genomes, Dg1 and BMG5.1, represent Cluster II [10, 21]. EAN1pec [18], BMG5.12 [19] and BCU110501 [34] represent Cluster III. Finally, atypical strains from Cluster IV are represented by DC12 [30] and CN3 [8]. Various other strains, including EuI1c, CpI1, R43, Allo2, AvcI1 an ACN1ag, are either incompletely sequenced or available as permanent drafts (available at https://gold.jgi.doe.gov/).

Availability of these databases has allowed for comparative analysis of *Frankia* genomes, and for determination of genome characteristics within each host-infection group (reviewed by Tisa et al. [31]). With the sequencing of the first three *Frankia* genomes, one from each of Clusters I, II and III, it was found that chromosome sizes varied greatly. Sizes were found to correlate with the host-range specificity of each strain; narrow host-range *Causarina*-infective strain CcI3, which does not survive in soil in the absence of its host, had the smallest genomes at 5.4 Mbp, while EAN1pec, a cosmopolitan, saprophytic strain had the largest at 8.9 Mbp [18, 29]. As more strains were sequenced this genome size/host-range hypothesis was found to hold, with genomes from Cluster III being the largest, Cluster II the smallest (*Datisca glomerata* obligate symbiont Dg1 = 5.1 Mbp) and with Cluster I typically somewhere in the middle [21, 31]. Genome analysis also revealed a high degree of plasticity in *Frankia* (the extent of which varied depending host-infection group) as well as significant secondary metabolisms with the potential for synthesis of diverse natural products and stress-tolerance factors [18, 31, 33]. Genomic data has also allowed for prediction of gene expression levels, which shed light on *Frankia* lifestyles and the differences between strains retaining the capacity to

live saprophytically, and those moving towards obligatory symbioses [28]. The components of bacterial nitrogenase complexes, required for N_2-fixation and thus effective symbioses, are encoded by and array of *nif* genes. These not only code for the subunits of dinitrogenase reductase (*nif*H) and the dinitrogenase complex itself (*nif*D and *nif*K), but also for genes required in its maturation, such as *ora* and *orb* which encode ferredoxin oxidoreductase (OR) subunits, the *nif*E, *N* and *nif*X genes which encode for FeMo-cofactor biosynthesis, the ferredoxin gene (fdx) and that for homocitrate synthase (*nif*V) [7, 20, 22, 23]. In *Frankia* (as in many nitrogen-fixing prokaryotes) these genes are organized in highly-conserved islands, the arrangement of which corresponds to the host infection group to which each respective *Frankia* isolate belongs [20, 23].

The *Frankia* strains selected for sequencing represent isolates from three Cape *Morella* species, and represent two genospecies from the *Elaeagnus*-HIG, and one from the *Alnus*-HIG. DNA sequence information generated by this study will provide insight into *Frankia* biology. The aim of this work was to primarily to produce high quality draft genomes for three *Frankia*, representing the three numerically dominant subgroups found in nodules collected from Cape *Morella*. A second aim was basic comparison of Cape *Frankia* genomes to those of previously sequenced strains.

5.2 Materials and Methods

Frankia strains representing three genomospecies were isolated from *Morella* of the Cape flora of southern Africa. In order to identify genes involved in symbiosis and of potential biotechnological interest, draft genome assemblies were constructed. Genomic DNA was extracted from 90 day old liquid cultures of FMc2, FMc6 and FMi1 using the Joint Genomics Institute's 2012 CTAB method for bacterial DNA extraction (http://jgi.doe.gov/). Four genome libraries were created from each of three *Frankia* isolates using a Nextera XT kit and version 2.0 and 3.0 Illumina chemistry and sequenced at the UWC high-throughput sequencing facility using an Illumina MiSeq platform [2][1]. Composition and quality of high-throughput

[1]Libraries were created and sequencing performed by Bronwyn Kirby at UWC's next-generation sequencing facility.

sequencing libraries with read lengths of 250 and 300-bp were assessed using fastqc (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). Draft genomes were assembled using the a5 microbial genome assembly pipeline [32], with Geneious version 9.0.2 (http://www.geneious.com, [12]), and with Tadpole [5] following removal of contaminating phage and plasmid sequences using BBmap version 35 [5], and trimming of low-quality bases using prinseq [24]. Libraries contaminated by non-*Frankia* bacterial gDNA were excluded from analysis. Individual libraries created from the same aliquot of gDNA were found to contain different contaminants. Genome sizes and coverage were estimated from k-mer frequency using BBmap, and from the final assemblies. RNAmmer 1.2 and tRNA-scan-SE 1.4 were used to predict rRNAs and tRNAs, respectively [13, 14]. Automated genome annotation was performed using the RAST annotation server [1].

ORF prediction, sequence analysis and annotation of contigs containing nitrogenase gene clusters was conducted in Geneious version, and checked by hand 9.0.2 (http://www.geneious.com, [12]). Secondary metabolite clusters were identified using antiSMASH version 3.0.5 [35].

5.3 Results

High molecular weight, high quality gDNA was extracted from three *Frankia* isolates and used to create pair-end genomic libraries. Isolate FMc2 was assembled from 3,780,418 reads totaling 655 Mbp (with 66.20% passing error-correction). FMc6 sequencing resulted in 3,684,716 reads totaling 597 Mbp (57.48% passing error-correction). Isolate FMi1 was assembled from 1,656,853 reads totaling 277 Mbp (62.99% remaining after error-correction). Genome GC contents were 71.3%, 71.5% and 71.2%, respectively. Genome assembly statistics are presented in Table 5.1. Summary characteristics for each genome are presented in Tables 5.2 and 5.3.

Anti-SMASH 3.0.5 predicted complete synthetic clusters with high homology to known biosynthetic pathways. FMi1 pathways include those producing naringenin (100%), frakiamicin (92%) and coelibaction (90%). FMc2 contains predicted clusters antibiotics caerulomycin (68%) and frankiamicin (92%), as well as geosmin and the C_{40} carotenoid sioxanthin. FMc6 frankiamicin (92%), coelibactin (90%)

and sioxanthin (75%). Furthermore, 125, 154 and 143 incomplete and putative secondary metabolite gene clusters were predicted per genome, respectively. These included type I, II and III polyketide synthase (PKS) pathways, non-ribosomal polyketide synthetases (NRPS), lantipeptides, flavanones, carotenoids, siderophores and plant hormones.

A single contig containing the full nitrogenase gene cluster was recovered for each strain with the exception of FMi1, in which the contig terminated short of the ferredoxin gene (Figures A.4, A.5 and A.6). FMc2 nitrogenase gene cluster was located on a 85, 280 bp contig, contained 17 open reading frames and measured 17, 275 bp from the start codon of ORF 1 to the end of ORF 17. FMc6 nitrogenase gene cluster was located on a 91, 214 bp contig, contained 17 open reading frames and measured 17, 193 bp from the start codon of ORF 1 to the end of ORF 17. The FMi1 nitrogenase gene cluster was located on a 84, 130 bp contig, contained 17 nif-related open reading frames (having nifV at the 5′, but lacking ferredoxin) and measured 18, 882 bp from the start codon of ORF 1 to the end of the ORF 17. Nucleotide similarities and amino acid sequence similarities of each gene within the nitrogenase cluster to those of previously sequenced *Frankia* strains from Clusters I, II and III are presented in Supplementary Tables A.3 & A.4 (isolate FMc2), Tables A.5 & A.6 (isolate FMc2), and Tables A.7 & A.8 (isolate FMc2).

Table 5.1: Genome assembly statistics for isolates FMc2, FMc6 and FMi1

	coverage[a]		contig lengths		counts	
Isolate	k-mer	assembly	max	median	contigs	scaffolds
FMc2	34×	57×	206719	44665	703	521
FMc6	45×	65×	88537	28944	716	667
FMi1	26×	32×	158039	29971	736	693

Average genome coverage as estimated by k-mer frequency of from a5 pipeline assemblies.

Table 5.2: Genome characteristics of Cape *Frankia* isolates FMc2, FMc6 and FMi1

		genome[b] (Mbp)				rRNA	tRNA
Isolate	HIG[a]	k-mer	assembly	GC%	ORFs[c]	genes	genes
FMc2	III	8.4	9.1	71.3	8057	9	48
FMc6	III	8.8	9.1	71.5	7908	8	45
FMi1	I	7.4	7.7	71.2	6860	5	59

Host infection groups.
Genome sizes as estimated by k-mer frequency of from a5 pipeline assemblies.
Open reading frames.

Table 5.3: Subsystem component counts as estimated by RAST

	FMi1	FMc2	FMc6	EAN1pec	CcI3	ACN14a
Cofactors, Vitamins, Prosthetic Groups, Pigments	385	386	394	422	333	252
Cell Wall and Capsule	71	106	75	107	83	81
Virulence, Disease and Defense	67	75	60	57	37	41
Potassium metabolism	17	15	17	0	9	23
Miscellaneous	40	41	34	50	41	7
Phages, Prophages, Transposable elements, Plasmids	26	4	13	0	0	0
Membrane Transport	135	184	159	42	25	34
Iron acquisition and metabolism	40	51	64	27	9	17
RNA Metabolism	90	91	99	139	140	94
Nucleosides and Nucleotides	104	122	121	94	79	98
Protein Metabolism	324	315	332	267	255	234
Cell Division and Cell Cycle	38	43	37	30	28	27
Motility and Chemotaxis	4	0	0	18	0	0
Regulation and Cell signaling	69	64	58	74	78	48
Secondary Metabolism	16	45	33	38	45	11
DNA Metabolism	168	127	137	117	100	109
Fatty Acids, Lipids, and Isoprenoids	426	398	444	243	132	143
Nitrogen Metabolism	24	19	18	17	18	21
Dormancy and Sporulation	10	7	10	7	9	8
Respiration	134	178	136	166	116	118
Stress Response	126	147	149	88	69	90
Metabolism of Aromatic Compounds	18	54	43	48	26	5
Amino Acids and Derivatives	435	466	516	444	313	378
Sulfur Metabolism	66	109	83	67	27	48
Phosphorus Metabolism	41	39	41	15	37	31
Carbohydrates	527	526	627	481	230	211
Total	3401	3612	3700	3058	2239	2129

5.4 Discussion

Three isolates were submitted to the UWC next-generation sequencing facility in December 2013 as single vials of high molecular weight gDNA, and four libraries (A-D) were created from each vial. Some libraries from each isolates were subsequently found to be contaminated with phage, commercial plasmid DNA and/or non-*Frankia* bacterial DNA. These contaminants were present in various combinations in some, but not all, libraries generated from each respective organism. For example, whereas FMc2 library *C* was contaminated by *Geobacillus*, library *A* was not. For each individual *Frankia* strain sequenced all libraries were produced from the same stock of genomic DNA. It is therefore clear that the contaminants (bacterial, commercial cloning vector, and lambda phage DNA) were introduced at the sequencing facility, by facility staff, during library preparation. Plasmid and phage sequences were successfully removed from contaminated data sets. Libraries found to contain contaminating bacterial gDNA (confirmed by the recovery of a full-length 16S and 23S rDNA sequences and BLAST analysis of selected ORFs) were excluded from further analysis.

While interrogation of the data was complicated by contamination from multiple sources, inconsistent processing of raw reads by the sequencing facility, and the high proportion of low-quality reads, basic summary metrics such as genome size, GC-content, number of tRNAs, rRNAs and ORFs could be estimated. These estimates should not be considered reliable. Likewise, while all contigs were annotated, each would need to be examined independently to determine whether they are from the genome of the sequenced target organism or from some contaminant. Large gaps existed in each of the assemblies with between 182 and 196 genes were predicted to be missing, but these are conservative estimates. Basic genome characteristics from the three Cape *Frankia* are in agreement with what has been found previously. Strains FMc2 and FMc6, from the *Elaeagnus* host infection group, have the largest genomes, while FMi1 has a genome of intermediate size at 7 Mbp. All three strains have GC contents in excess of 70%. Each strain has a significant secondary metabolism, and includes antibiotics, cellulases, plant hormones and siderophores.

Where the organization of *nif* gene clusters is concerned, Oh et al. found that these

have remained comparatively stable, despite the plasticity of *Frankia* genomes and their associated changes in content and size [18, 20, 27]. When variations in gene arrangement do occur in this expression island they are conserved among strains within individual host-infection groups (see Supplementary Figures A.4, A.5 and A.6). My results agree with this finding, as in strains FMc2 and FMc6 the arrangement of nitrogenase genes was identical to that of other previously sequenced strains from the *Elaeagnus* host infection group (Figure A.6). Likewise the arrangement of FMi1's nitrogenase gene cluster was identical those of other *Alnus*-HIG strains (Figure A.4), although the 3' ferredoxin gene was absent due to termination of the contig. Individual genes within this cluster vary between strains with respect to their pair-wise sequence-similarity, with *nif*H, *nif*D, *nif*K and *nif*E showing the highest degree of conservation (Supplementary Tables A.3, A.4, A.5, A.6, A.7 and A.8). It is possible that other genes from this cluster will prove useful in the design of highly discriminative probe and primer sets for population studies of specific *Frankia* subgroups in natural habitats.

While some useful data was recovered in this study it would be preferable to repeat the sequencing. High quality drafts of these genomes will be of interest to those interested in comparative genomics, genome features/components responsible for *Frankia* adaptation to symbiotic and saprophytic lifestyles, and novel bioactive compounds.

5.5 Conclusion

The aim of this work was to produce high quality draft genomes for three *Frankia*, representing the three numerically dominant subgroups (one strain from Cluster I and two representing Cluster III) found in nodules collected from Cape *Morella*. These genomes were then to be compared to those from previously sequenced strains. While the genomes had characteristics typical of *Frankia* from their respective host infection groups, and some analysis of *nif* gene cluster arrangements was performed, deeper analysis was complicated by the poor quality of the generated data. Genomes from Cape *Frankia* are still of considerable interest and sequencing should therefore be repeated.

References

[1] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O.: The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**(1), 75 (2008)

[2] Bennet, S.: Miscellaneous. Pharmacogenetics **5**(4), 433–438 (2006)

[3] Benson, D.R., Dawson, J.O.: Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. Physiologia Plantarum **130**, 318–330 (2007)

[4] Benson, D.R., Silvester, W.B.: Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiological Reviews **57**(2), 293–319 (1993)

[5] Bushnell, B.: BBMap. URL sourceforge.net/projects/bbmap. Unpublished software

[6] Chaia, E.E., Wall, L.G., Huss-Danell, K.: Life in soil by the actinorhizal root nodule endophyte *Frankia*. A review. Symbiosis **51**(3), 201–226 (2011)

[7] Dos Santos, P.C., Dean, D.R., Ribbe, M.W.: Formation and insertion of the nitrogenase iron-molybdenum cofactor. Chemical Reviews **104**(2), 1159–1173 (2004)

[8] Ghodhbane-Gtari, F., Beauchemin, N., Bruce, D., Chain, P., Chen, A., Davenport, K.W., Deshpande, S., Detter, C., Furnholm, T., Goodwin, L., Gtari, M., Han, C., Han, J., Huntemann, M., Pagani, I., Pati, A., Pitluck, S., Santos, C.L., Sen, A., Sur, S., Szeto, E., Tavares, F.: Draft genome sequence of *Frankia* sp. strain CN3, an stypical, noninfective (Nod) ineffective (Fix) isolate from *Coriaria nepalensis*. Genome Announcements **1**(2), 1–2 (2013)

[9] Ghodhbane-Gtari, F., Hurst, S.G., Oshone, R., Morris, K., Abebe-Akele, F., Thomas, W.K., Ktari, A., Salem, K., Gtari, M., Tisa, L.S.: Draft genome

sequence of *Frankia* sp. strain BMG5.23, a salt-tolerant nitrogen-fixing actinobacterium isolated from the root nodules of *Casuarina glauca* grown in Tunisia. Genome Announcements **2**(3), e00,520–14 (2014)

[10] Gtari, M., Ghodhbane-Gtari, F., Nouioui, I., Ktari, A., Hezbri, K., Mimouni, W., Sbissi, I., Ayari, A., Yamanaka, T., Normand, P., Tisa, L.S., Boudabous, A.: Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Scientific Reports **5**, 13,112 (2015)

[11] Hurst, S.G., Oshone, R., Ghodhbane-Gtari, F., Morris, K., Abebe-Akele, F., Thomas, W.K., Ktari, A., Salem, K., Mansour, S., Gtari, M., Tisa, L.S.: Draft genome sequence of *Frankia* sp. strain Thr, a nitrogen-fixing actinobacterium isolated from the root nodules of *Casuarina cunninghamiana* grown in Egypt. Genome Announcements **2**(3), e00,493–14 (2014)

[12] Kearse, M., Moir, R., Wilson, A., Stones-havas, S., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A.: Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**(12), 1647–1649 (2012)

[13] Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W.: RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research **35**(9), 3100–3108 (2007)

[14] Lowe, T.M., Eddy, S.R.: tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Research **25**(5), 955–964 (1997)

[15] Mansour, S.R., Oshone, R., Hurst IV, S., Morris, K., Thomas, K.W., Tisa, L.S.: Draft genome sequence of *Frankia* sp. strain CcI6, a salt-tolerant nitrogen-fixing actinobacterium isolated from the root nodule of *Casuarina cunninghamiana*. Genome Announcements **2**(1), 1–2 (2014)

[16] Mirza, B.S., Welsh, A., Rieder, J.P., Paschke, M.W., Hahn, D.: Diversity of frankiae in soils from five continents. Systematic and Applied Microbiology **32**(8), 558–570 (2009)

[17] Normand, P., Fernandez, M.P.: Evolution and Diversity of *Frankia*. In: K. Pawlowski (ed.) Prokaryotic Symbionts in Plants, vol. 8, pp. 103–125. Springer-Verlag, Berlin Heidelberg (2009)

[18] Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choisne, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltsman, E., Huang, Y., Kopp, O.R., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastronunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D.: Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research **17**, 7–15 (2007)

[19] Nouioui, I., Beauchemin, N., Cantor, M.N., Chen, A., Detter, J.C., Furnholm, T., Ghodhbane-Gtari, F., Goodwin, L., Gtari, M., Han, C., Han, J., Huntemann, M., Hua, S.X., Ivanova, N., Kyrpides, N., Markowitz, V., Mavrommatis, K., Mikhailova, N., Nordberg, H.P., Ovchinnikova, G., Pagani, I., Pati, A., Sen, A., Sur, S., Szeto, E., Thakur, S., Wall, L., Wei, C.L., Woyke, T., Tisa, L.S.: Draft genome sequence of *Frankia* sp. strain BMG5.12, a nitrogen-fixing actinobacterium isolated from Tunisian soils. Genome Announcements **1**(4), e00,468–13 (2013)

[20] Oh, C.J., Kim, H.B., Kim, J., Kim, W.J., Lee, H., An, C.S.: Organization of *nif* gene cluster in *Frankia* sp. EuIK1 strain, a symbiont of *Elaeagnus umbellata*. Archives of Microbiology **194**(1), 29–34 (2012)

[21] Persson, T., Benson, D.R., Normand, P., Heuvel, B.V., Pujic, P., Chertkov, O., Teshima, H., Bruce, D.C., Detter, C., Tapia, R., Han, S., Han, J., Woyke, T., Pitluck, S., Pennacchio, L., Nolan, M., Ivanova, N., Pati, A., Land, M.L., Pawlowski, K., Berry, A.M.: Genome sequence of "*Candidatus Frankia datiscae*" Dg1, the uncultured microsymbiont from nitrogen-fixing root nodules of the dicot *Datisca glomerata*. Journal of Bacteriology **193**(24), 7017–7018 (2011)

[22] Rubio, L.M., Ludden, P.W.: Maturation of nitrogenase: A biochemical puzzle. Journal of Bacteriology **187**(2), 405–414 (2005)

[23] Rubio, L.M., Ludden, P.W.: Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Annual Review of Microbiology **62**(1), 93–111 (2008)

[24] Schmieder, R., Edwards, R.: Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**(6), 863–864 (2011)

[25] Schwencke, J., Caru, M.: Advances in actinorhizal symbiosis: host plant-Frankia interactions, biology, and applications in arid land reclamation. A review. Arid Land Research and Management **15**(4), 285–327 (2001)

[26] Sen, A., Beauchemin, N., Bruce, D., Chain, P., Chen, A., Davenport, W., Deshpande, S., Detter, C., Furnholm, T., Ghodbhane-gtari, F., Goodwin, L., Gtari, M., Han, C., Han, J., Nouioui, I., Pagani, I., Pati, A., Pitluck, S., Santos, C.L., Sur, S., Szeto, E., Tavares, F.: Draft genome sequence of *Frankia* sp. strain QA3, a nitrogen-fixing actinobacterium isolated from the root nodule of *Alnus nitida*. Genome Announcements **1**(2), 13–14 (2013)

[27] Sen, A., Duabin, V., Abrouk, D., Gifford, I., Berry, A.M., Normand, P.: 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 Nakamu. International Journal of Systematic and Evolutionary Microbiology (2014)

[28] Sen, A., Sur, S., Bothra, A.K., Benson, D.R., Normand, P., Tisa, L.S.: The implication of life style on codon usage patterns and predicted highly expressed genes for three *Frankia* genomes. Antonie van Leeuwenhoek **93**(4), 335–46 (2008)

[29] Simonet, P., Navarro, E., Rouvier, C., Reddell, P., Zimpfer, J., Dommergues, Y., Bardin, R., Combarro, P., Hamelin, J., Domernach, A., Gourbiere, F., Prin, Y., Dawson, J.O., Normand, P.: Co-evolution between *Frankia* populations and host plants in the family *Casuarinaceae* and consequent patterns of global dispersal. Environmental Microbiology **1**(6), 525–533 (1999)

[30] Tisa, L.S., Beauchemin, N., Cantor, M.N., Furnholm, T., Ghodhbane-gtari, F., Goodwin, L., Copeland, A., Gtari, M., Huntemann, M., Ivanova, N., Kyrpides,

N., Markowitz, V., Mavrommatis, K., Mikhailova, N., Nouioui, I., Oshone, R., Ovchinnikova, G., Pagani, I., Palaniappan, K., Pati, A., Sen, A., Shapiro, N., Szeto, E., Wall, L., Wishart, J., Woyke, T.: Draft genome sequence of *Frankia* sp. strain DC12, an atypical, noninfective, ineffective isolate from *Datisca cannabina*. Genome Announcements **3**(4), e00,889–15 (2015)

[31] Tisa, L.S., Beauchemin, N., Gtari, M., Sen, A., Wall, L.G.: What stories can the *Frankia* genomes start to tell us? Journal of Biosciences **38**(4), 719–726 (2013)

[32] Tritt, A., Eisen, J.A., Facciotti, M.T., Darling, A.E.: An integrated pipeline for de novo assembly of microbial genomes. PLoS ONE **7**(9), e42,304 (2012)

[33] Udwary, D.W., Gontang, E.A., Jones, A.C., Jones, C.S., Schultz, A.W., Winter, J.M., Yang, J.Y., Beauchemin, N., Capson, T.L., Clark, B.R., Esquenazi, E., Eusta, A.S., Freel, K., Gerwick, L., Gerwick, W.H., Gonzalez, D., Liu, W.t., Malloy, K.L., Maloney, K.N., Nett, M., Nunnery, J.K., Penn, K., Prieto-davo, A., Simmons, T.L., Weitz, S., Wilson, M.C., Tisa, L.S., Dorrestein, P.C., Moore, B.S.: Significant natural product biosynthetic potential of actinorhizal symbionts of the genus *Frankia*, as revealed by comparative genomic and proteomic analyses. Applied Environmental Microbiology **77**(11), 3617–3625 (2011)

[34] Wall, L.G., Beauchemin, N., Cantor, M.N., Chaia, E., Chen, A., Detter, J.C., Furnholm, T., Ghodhbane-Gtari, F., Goodwin, L., Gtari, M., Han, C., Han, J., Huntemann, M., Hua, S.X., Ivanova, N., Kyrpides, N., Markowitz, V., Mavrommatis, K., Mikhailova, N., Nordberg, H.P., Nouioui, I., Ovchinnikova, G., Pagani, I., Pati, A., Sen, A., Sur, S., Szeto, E., Thakur, S., Wei, C.L., Woyke, T., Tisa, L.S.: Draft genome sequence of *Frankia* sp. strain BCU110501, a nitrogen fixing actinobacterium isolated from nodules of *Discaria trinevis*. Genome Announcements **1**(4), e00,503–13 (2013)

[35] Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Muller, R., Wohlleben, W., Breitling, R., Tanako, E., Medema, M.H.: antiSMASH 3.0 - a comprehensive resource for the genome

mining of biosynthetic gene clusters. Nucleic Acids Res **W1**, W237–W243 (2015)

6 Summary and future prospects

The study described in this thesis had three objectives: 1) to explore the diversity and natural associations of *Frankia* strains in *Morella* root nodules from the Cape flora of South Africa, 2) and to explore *Frankia* diversity in rhizosphere soils from natural stands, comparing soil and nodule populations, and 3) to isolate and characterize endemic South African *Frankia*.

The investigation addressing the first objective (Chapter 2) found that Cape *Morella* growing in their natural habitats may be nodulated by *Frankia* from both Cluster I and Cluster III, as demonstrated by partial *nif*H gene sequencing. Three of six species examined (*M. integra*, *M. diversifolia* and *M. quercifolia*) are promiscuous in the field, with sequences from both host-infection groups present in nodules collected from each species. Two species (*M. serrata*, *M. kraussiana*) were nodulated by Cluster I strains only, but as few nodules were recovered from these species definite conclusions as to the full range of their natural associations could not be drawn. Sequences recovered from these species were identical to those found in *M. integra* nodules. *Morella cordifolia* was the most well-sampled species, both in terms of the number of populations examined and their geographical distribution (more than half of the species' natural range). It was nodulated by Cluster III strains exclusively, and it is possible that this association is specific.

Cluster I strains were represented by seven unique *nif*H sequences in two genotypes (CC-1 & CC-2), the numerically dominant one (CC-1) not previously described and thus possibly unique to the Cape flora. Cluster III strains were more diverse, with a

total of eighteen unique sequences distributed across five genotypes. These clustered within three of four previously identified and globally distributed Cluster III subgroups.

A notable finding was that *Morella* species were only nodulated by Cluster I strains when growing in acidic soils, with these strains completely absent in nodules recovered from neutral or alkaline soils. This work constituted the first investigation into the diversity of *Frankia* nodulating indigenous actinorhizal plants of southern Africa.

For the second objective (Chapter 3) *Frankia nif*H clone libraries were created from six soils associated with nodulated *Morella* populations growing in both acidic and alkaline soils, principally in order to address the apparent absence of Cluster I strains from nodules collected from alkaline soils. Populations of *Cluster* III *nif*H genotypes in these libraries agreed, to a large extent, with the dominant genotypes found in the nodules collected throughout the region. Genotypes found in nodules were not, however, always represented in libraries created from root-associated soils collected at the same site. Three rarely-detected Cluster III genotypes (CC-5, CC-6 and CC-7) were found in a total of six root nodules but were not detected in any of the soils.

Among Cluster I strains the dominant *nif*H genotype in Cape *Morella* nodules (CC-1) was absent from all six clone libraries, despite being present in nodules from three of the sites from which the soils were collected. Whether this was because strains composing this genotype were absent from the soils in question, were out-competed in soil by *Cluster* III strains to the extent that they were present in undetectable quantities, or were present only as spores which survived lysis, was not established. Unfortunately, as this genotype was not detected in any of the soils its absence from nodules collected from neutral and alkaline soils cannot be commented upon. What is clear is that CC-1 would have remained undetected had my study relied on soil clone libraries alone, and had its presence not already been established in my survey of nodular diversity. This illustrates the danger of relying on the ability of clone libraries to detect even major components of microbial communities in a given environment, and when library coverage appears

to be high or near-complete.

In contrast to CC-1, the second Cape Cluster I *Frankia* subgroup (CC-2), which was detected in nodules from *Morella diversifolia*, was present detected using soil clone libraries, but only in the library associated with this species. This suggests varying abilities of Cape Cluster I *Frankia* strains to grow saprophytically, and also very local geographical distributions for at least some genotypes. A third Cluster I genotype represented by partial *nif*H sequences identical to those of *Frankia* from Cluster I subgroup AI, which has been detected in soils collected globally, was detected in three of the Cape soils. Two of these soils were coastal alkaline sands hosting *M. cordifolia*. This subgroup was not detected in Cape *Morella* root nodules, which raises several questions: do these sequences represent bona fide *Frankia* strains, and if so are why were they not detected in any of the collected nodules? Are there conditions under which they nodulate Cape *Morella*, or are they exclusively saprophytic? If representatives of this group are recovered directly from soil this question may be addressed experimentally.

With respect to this study's third objective (Chapter 4), clonally pure *Frankia* strains were isolated from four host species: *M. cordifolia* (FMc1-FMc7), *M. integra* (FMi1), *M. kraussiana* (FMk1) and *M. quercifolia* (FMq1). The effectiveness of a gellan gum-based media for selective *Frankia* isolation from nodules was demonstrated. The strains recovered are the first to be isolated from actinorhizal hosts endemic to the African continent, and represent the numerically dominant Cape nodular genotypes from each of the *Alnus* (CC-1) and *Elaeagnus* (CC-3 and CC-4) host-infection groups. 16S rDNA and 23S rDNA IS BLAST analysis confirmed the identity of the isolates as *Frankia*, and phylogenetic analysis assigned them to the same host-infection groups suggested by *nif*H analysis. All ten isolates also displayed cell morphologies and phenotypes typical of *Frankia*, but varied in their respective abilities to produce vesicles, pigments and to grow on individual carbon sources. A multilocus sequence analysis assigned Cape *Frankia* isolates to three genomospecies, two of which have been described previously, and one of which is unique to the Cape flora. Genospecies groupings agreed with those determined by a larger phylogenetic analysis of the *nif*H gene. Representatives of each genospecies were selected for genomic sequencing (Chapter 5), which revealed

characteristics such as GC content, ORF count, and that the genome size of each was in line with what has been found for strains from their respective host-infection groups. FMc2 and FMc6 (Cluster III) have genomes of approximately 9 Mbp while FMi1 (Cluster I) is smaller at 7.1 Mbp. The organization of each isolate's nitrogenase gene cluster was found to be typical of those found in other strains belonging to their respective host-infection groups.

While I endeavored to carry out as comprehensive an investigation into the diversity of *Frankia* associated with *Morella* of the Cape flora as possible, the study had several limitations. For some host species sampling depth was insufficient; very few nodules were recovered from *M. kraussiana* and *M. serrata*, and only two nodulated populations of *M. diversifolia* were found. *Morella humilis*, the seventh and final Cape species, was never encountered. While *Frankia* strains representing both major host-infection groups were recovered, these represent only six of twenty-eight *nif*H sequences found in nodules. My isolates represent the three numerically dominant genotypes found in the survey of nodular diversity, but strains from genotype CC-2 proved impossible to isolate. During the field survey effects such as stand-age, which is known to influence assemblage composition, could not be taken into account. The ability of Cape *Frankia* to re-infect their original hosts was not determined, nor were they demonstrated to fix nitrogen in culture (despite possessing complete nitrogenase gene clusters). Comprehensive characterization of *Frankia* isolates, including growth characteristics, antibiotic resistance patterns, cell wall composition and total cellular sugars could not be carried out. Likewise, while I attempted to determine absolute *Frankia* populations sizes in soil using a qPCR protocol targeting *nif*H, this method proved unreliable with real-world samples and the experiments could not be completed[1]. This work was intended to supplement the diversity information retrieved by the soil clone libraries. Insufficient time remained for any significant processing of the genomic data, aside from basic assembly and annotation.

This dissertation constitutes the seminal investigation into natural symbiotic associations between *Frankia* and indigenous African *Morella* species. It reports the first isolation of *Frankia* from *Morella* on the continent, increases the number of

[1]Details of this work are not included in this dissertation.

Morella species whose *Frankia* associations have been investigated in natural settings globally from six to twelve, and expands our understanding of *Morella*/*Frankia* interactions. It lays the groundwork for future research involving indigenous actinorhizal species in southern Africa.

Future prospects

As isolates representing most of the major *Frankia* genotypes found in Cape *Morella* were recovered, it is now possible to perform comprehensive cross-inoculation experiments with indigenous strains. Whether my strains are able to re-infect the species from which they were isolated remains to be established, as do the relative efficiencies of any symbioses established with individual hosts by different strains. Beyond such basic studies, several avenues of investigation present themselves: it will be valuable to evolutionary biologists to learn, for example, whether *M. cordifolia* has evolved towards exclusivity with *Frankia* from Cluster III. If this species proves to be promiscuous in greenhouse trials, the mechanism behind its selectivity in the field should be investigated. Furthermore, my three Cluster I *Frankia* isolates represent a previously-undescribed genospecies; as these strains have evolved in symbiosis with Cape *Morella* exclusively it will be of considerable interest to determine whether they are infective on hosts from "traditional" Cluster I host genera. The ability of these strains to infect *Casuarina* in particular should be explored. Cross-inoculation studies and genomic comparisons with other Cluster I isolates can potentially shed light on the mechanisms of host/symbiont selection.

Frankia isolates from the Cape environment could be used in microcosm experiments to explore the ability of different Cape *Frankia* lineages to grow saprophytically, and to explore population dynamics in host rhizospheres under conditions simulating those found in Cape habitats. The ecology and population dynamics of *Frankia* in soil, both in the presence and absence of hosts, are particularly complex research problems. Future studies should focus on determination of population sizes and growth states of different *Frankia* clusters in soil through the use of methods such as qPCR and in situ hybridization. Such attempts should avoid SYBR green and use probe-based methodologies, possibly targeting 23S rDNA IS sequences.

While time did not permit for proper interrogation of the genomic data generated during this study, it is clear from preliminary analysis that each of the three strains chosen for sequencing possess significant ability to produce secondary metabolites, including plant hormones, pigments, degradative enzymes and antibiotics[2]. Whole genome comparisons between FMc6 and other members of genospecies G6, such as EAN1pec, would provide insight into genome plasticity in geographically widespread strains from this group. Similarly, FMc2 should be compared with the various Cluster III BMG isolates from Tunisia as they appear to belong to an African genospecies containing related strains. The genetics of FMi1 should be of particular interest due to the unique nature of CC-1 strains from southern Africa.

Factors such as stand age, season and a wide range of soil properties are known to influence *Frankia* assemblages. An expanded field study could take these factors into account, and also gather more information on the associations of the species examined in this study. The symbiotic partnerships of four South African species have yet to be investigated (*M. pilulifera*, *M. humilis*, *M. microbracteata* and*M. brevifolia*). The spore-positive/spore-negative status of nodules should also be investigated, particularly in species known to be nodulated by strains from genotype CC-1.

Finally, as suggested in Chapter 1, the insecticidal and medicinal applications of *M. quercifolia* should be explored, as should the potential application of *M. pilulifera* in South African silviculture. It would also be worthwhile to establish which *Frankia* strain(s) make the best microsymbiont partners for this species, and would serve as suitable inocula in plantation settings.

[2]Details not included in this dissertation.

Conference contributions

Biodiversity SA Conference

Cape Town, South Africa. December 2013 (oral presentation)

"Actinorhizal symbiosis in the Cape floristic region: diversity of *Frankia* nodulating six species of *Morella*."

D. A. Wilcox, B. Kirby K, D. A. Cowan

17th International Symposium on the Biology of Actinomycetes

Kusadasi, Turkey. October 2014 (poster presentation)

"Diversity of *Frankia* populations in the root nodules of actinorhizal *Morella* of the Cape flora of South Africa"

D. A. Wilcox[3], B. Kirby, D. A. Cowan

Recipient of the "Stanley Williams" young scientist award in the category "selective isolation" (personal communication: Professor Gilles van Wezel and Professor Atac Uzel, ISBA17 Chair and Chair of the Turkish National Organizing Committee, respectively).

18th International Meeting on *Frankia* and Actinorhzal Plants

Montpellier, France. August 2015 (Oral presentation)

"Diversity of *Frankia* associated with *Morella* species of the Cape flora of southern Africa."

D. A. Wilcox, D. A. Cowan

[3]The authorship as indicated on the now-defunct conference website (www.isba17.com, archived at https://archive.org/web/) and in the book of conference abstracts is incorrect. B. Kirby was removed from my study, at my insistence, after she changed the authorship on, and accepted undue credit for, my original research without my knowledge. She contributed nothing to the isolation of my *Frankia* strains, and was not attached to my project at the time I did the work. The effect of her actions was that conference attendees wrongfully considered her the major contributor to the study (personal communication: ISBA17 conference organizers).

A Supplementary material

Table A.1: CFN sequences detected in nodules at individual sites, some having multiple *Morella* species

	Site	A1						A2	A3				A4	B1			B2	B3			B4	B5	B6	B7	
	Subsite	1.1	1.2	1.3	1.4	1.5			3.1	3.2	3.3	3.4		1.1	1.2	1.3		3.1	3.2	3.3					
	Host	Mc	Mc	Mc	Mc	Mq	Mc	Mc	Mi	Mi	Mi	Mi	Mc	Mq	Md	Mk	Md	Mq	Mk	Mk	Mc	Mq	Ms	Mc	Totals
CC-1	CFN1	-	-	-	-	-	-	-	12	-	7	23	-	-	-	-	-	-	-	-	-	-	6	-	48
	CFN3	-	-	-	-	-	-	-	1	-	1	-	-	-	7	2	-	1	1	3	-	-	-	-	16
	CFN20	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	CFN22	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	CFN23	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
CC-2	CFN10	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	6
	CFN16	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	2
CC-3	CFN5	2	1	1	2	1	2	-	-	-	-	-	1	-	-	-	-	-	-	-	2	-	-	1	13
	CFN6	-	-	-	-	6	-	-	-	-	-	-	7	-	2	-	-	-	-	-	1	-	-	3	19
	CFN7	-	3	-	-	2	2	-	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	1	12
	CFN13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	11
	CFN8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	9
	CFN9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8	-	-	-	8
	CFN11	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4
	CFN14	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
	CFN17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	2
	CFN26	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
CC-4	CFN2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-	-	20
	CFN4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-	20
	CFN18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	2
	CFN19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
	CFN21	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	CFN25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
CC-5	CFN12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	4
CC-6	CFN15	-	-	-	-	-	-	-	-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	3
CC-7	CFN24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Totals		2	4	1	2	15	4	1	18	3	8	23	8	2	18	2	13	1	1	3	36	21	6	10	202

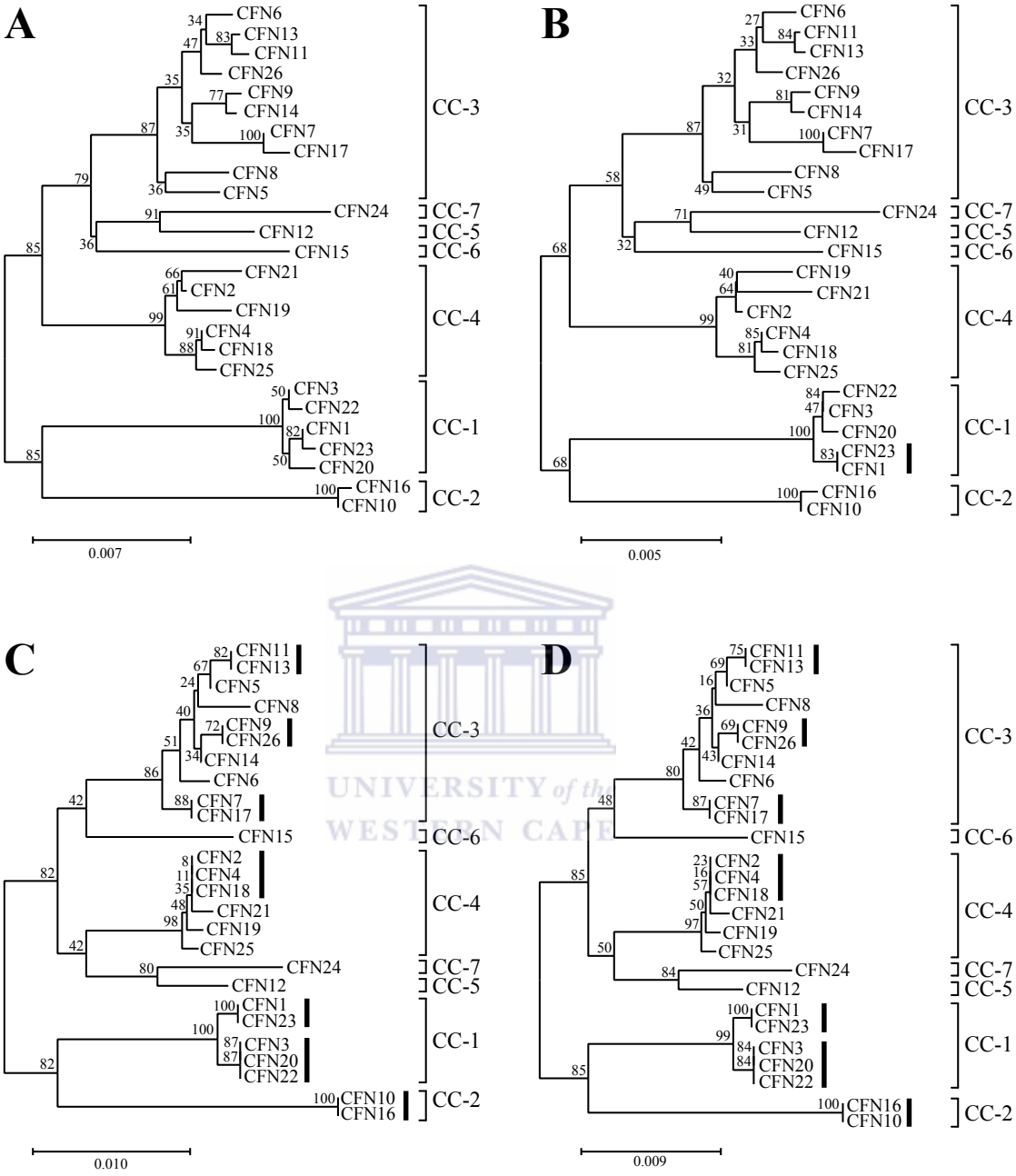


Figure A.1: Effect of trimming on nodular *Frankia nif*H uniqueness and cluster assignment. A, 606bp; B, 512bp; C, 263 bp; D, 255 bp. A & B are lengths considered during the survey of nodular diversity (Chapter 2). B & C were lengths considered in the study of soil diversity(Chapter 3). Black bars indicate sequences identical at the specified trimming threshold. Cape Clusters (CC) are indicated on the right.

Table A.2: BLAST results for unique 255 bp *nif*H sequences recovered from soil clone libraries

Query	Count	Identity %	E-value	Accession	Description	HIG
B20	6	99.6	1.297E-124	AJ545031	BMG5.12	III
B29	1	99.2	1.58E-123	AJ545031	BMG5.12	III
B55	1	99.6	1.297E-124	AJ545031	BMG5.12	III
B26	3	99.6	6.717E-122	AJ545031	BMG5.12	III
B74	1	99.6	8.183E-121	AJ545031	BMG5.12	III
B2	1	98.2	3.052E-107	FJ977330	ARgP5ag	I
B73	1	98.6	7.178E-109	FJ977330	ARgP5ag	I
B50	1	99.6	1.138E-112	JF273726	BMG5.15	III
B35	1	97.8	1.065E-106	FJ477535	Japan-nodB7	I
B1	1	98.7	1.688E-110	FJ477535	Japan-nodB7	I
B13	1	99.1	3.97E-112	FJ477535	Japan-nodB7	I
B3	46	98.7	1.688E-110	FJ477535	Japan-nodB7	I
B47	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B16	1	97.8	8.745E-108	FJ477535	Japan-nodB7	I
B41	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B15	1	97.8	8.745E-108	FJ477535	Japan-nodB7	I
B48	1	97.8	8.745E-108	FJ477535	Japan-nodB7	I
B28	1	97.8	8.745E-108	FJ477535	Japan-nodB7	I
B46	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B8	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B25	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B68	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B39	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
B69	1	98.2	2.057E-109	FJ477535	Japan-nodB7	I
C45	1	99.6	1.297E-124	JF273729	ChI7	III
C3	58	99.6	3.05E-126	JF273729	ChI7	III
C1	1	99.6	1.297E-124	JF273729	ChI7	III
C31	1	99.6	1.297E-124	JF273729	ChI7	III
C35	1	99.6	1.297E-124	JF273729	ChI7	III
C6	1	99.6	1.297E-124	JF273729	ChI7	III
C28	1	99.2	1.58E-123	JF273729	ChI7	III
C52	1	99.6	1.297E-124	JF273729	ChI7	III
C63	1	99.6	1.297E-124	JF273729	ChI7	III
C32	1	99.6	1.297E-124	JF273729	ChI7	III
C51	1	99.6	1.297E-124	JF273729	ChI7	III
C12	1	99.6	1.297E-124	JF273729	ChI7	III
C67	1	99.6	1.297E-124	JF273729	ChI7	III
C13	1	99.6	1.297E-124	JF273729	ChI7	III
C72	1	99.6	1.297E-124	JF273729	ChI7	III
C41	1	99.6	1.297E-124	JF273729	ChI7	III
C2	4	99.6	1.297E-124	JF273729	ChI7	III
S9	13	99.6	3.05E-126	JF273729	ChI7	III
S43	1	99.6	1.297E-124	JF273729	ChI7	III
S54	1	99.6	1.297E-124	JF273729	ChI7	III

Table A.2 *continued on following page...*

Table A.2 – *Continued*

Query	Count	Identity %	E-value	Accession	Description	HIG
S17	1	99.6	1.297E-124	JF273729	ChI7	III
S18	1	99.6	1.064E-125	JF273729	ChI7	III
S3	3	99.6	1.297E-124	JF273729	ChI7	III
S34	1	99.6	4.526E-124	JF273729	ChI7	III
S61	1	99.6	3.48E-119	JF273729	ChI7	III
S4	1	99.8	3.48E-119	JF273729	ChI7	III
S29	1	99.2	8.183E-121	AJ545036	BMG5.6	III
S46	1	99.6	6.717E-122	AJ545036	BMG5.6	III
S20	1	98.8	2.345E-121	AJ545031	BMG5.12	III
S66	1	98.8	2.345E-121	AJ545031	BMG5.12	III
S51	1	98.4	2.856E-120	AJ545031	BMG5.12	III
S48	1	98.4	2.856E-120	JF273726	BMG5.15	III
S19	2	99.6	1.297E-124	JF273726	BMG5.15	III
S16	1	98.4	8.183E-121	JF273726	BMG5.15	III
S2	17	99.6	1.297E-124	AJ545031	BMG5.12	III
S64	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S21	1	98.4	8.183E-121	AJ545031	BMG5.12	III
S40	1	97.6	4.239E-118	AJ545031	BMG5.12	III
S62	1	98.0	3.48E-119	AJ545031	BMG5.12	III
S56	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S37	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S39	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S15	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S30	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S57	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S58	1	99.6	1.58E-123	AJ545031	BMG5.12	III
S11	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S65	1	99.2	1.58E-123	AJ545031	BMG5.12	III
S33	1	98.8	6.717E-122	AJ545031	BMG5.12	III
S1	1	98.2	6.717E-122	AJ545031	BMG5.12	III
S5	3	99.6	1.297E-124	JF273732	CpI1	I
S7	1	98.3	1.215E-118	JF273732	CpI1	I
T19	3	99.6	3.05E-126	JF273729	ChI7	III
T6	1	99.6	1.297E-124	JF273729	ChI7	III
T33	1	99.6	1.297E-124	JF273729	ChI7	III
T50	1	99.6	1.297E-124	JF273729	ChI7	III
T66	1	99.6	1.297E-124	JF273729	ChI7	III
T61	1	99.6	6.717E-122	JF273729	ChI7	III
T3	2	99.6	1.297E-124	JF273729	ChI7	III
T47	1	99.6	1.58E-123	JF273724	BCU110345	III
T26	1	98.8	2.345E-121	EU862989	MCNod27	III
T41	2	99.6	4.526E-124	AJ545031	BMG5.12	III
T42	1	99.2	5.514E-123	AJ545031	BMG5.12	III
T62	1	98.8	2.345E-121	AJ545031	BMG5.12	III
T9	1	99.6	1.58E-123	AJ545031	BMG5.12	III
T8	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T51	1	99.6	1.297E-124	JF273726	BMG5.15	III

Table A.2 *continued on following page...*

Table A.2 – *Continued*

Query	Count	Identity %	E-value	Accession	Description	HIG
T34	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T60	1	99.6	8.183E-121	AJ545031	BMG5.12	III
T1	35	99.6	1.297E-124	AJ545031	BMG5.12	III
T38	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T57	2	99.2	1.58E-123	AJ545031	BMG5.12	III
T39	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T28	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T30	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T43	1	99.2	1.58E-123	AJ545031	BMG5.12	III
T23	1	100	3.05E-126	AJ545031	BMG5.12	III
T35	1	99.6	1.297E-124	AJ545031	BMG5.12	III
T4	2	100	3.05E-126	EU862917	Cc1.17	III
T44	1	99.6	1.297E-124	EU862917	Cc1.17	III
T53	1	99.6	1.297E-124	EU862917	Cc1.17	III
W12	4	99.6	3.05E-126	JF273729	ChI7	III
W22	1	99.6	1.297E-124	JF273729	ChI7	III
W17	1	99.6	3.48E-119	JF273729	ChI7	III
W66	1	99.6	4.526E-124	EU862989	MCNod27	III
W23	1	99.6	6.717E-122	AJ545036	BMG5.6	III
W28	1	98.0	3.48E-119	AJ545031	BMG5.12	III
W51	1	99.6	4.526E-124	AJ545031	BMG5.12	III
W58	1	99.2	5.514E-123	AJ545031	BMG5.12	III
W26	1	99.6	1.297E-124	JF273726	BMG5.15	III
W4	3	100	3.05E-126	AJ545031	BMG5.12	III
W9	1	99.6	1.297E-124	AJ545031	BMG5.12	III
W21	2	99.2	1.58E-123	AJ545031	BMG5.12	III
W1	1	98.8	6.717E-122	AJ545031	BMG5.12	III
W2	1	98.8	6.717E-122	AJ545031	BMG5.12	III
W6	41	99.6	1.297E-124	AJ545031	BMG5.12	III
W5	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W11	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W55	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W61	1	98.8	6.717E-122	AJ545031	BMG5.12	III
W60	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W54	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W27	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W45	1	99.2	1.58E-123	AJ545031	BMG5.12	III
W63	1	98.0	1.215E-118	JF273732	CpI1	I
W14	1	100	3.05E-126	JF273732	CpI1	I
W71	1	99.6	1.297E-124	JF273732	CpI1	I
W15	1	98.8	6.717E-122	JF273732	CpI1	I
W3	1	96.5	8.745E-108	EU863029	MPNod32	I
Z33	1	100	3.48E-119	EU863047	SANod2	III
Z23	1	99.2	5.514E-123	EU862989	MCNod27	III
Z24	1	98.8	6.717E-122	EU862989	MCNod27	III
Z28	1	98.8	2.345E-121	EU862989	MCNod27	III
Z9	1	97.2	6.291E-116	EU862989	MCNod27	III

Table A.2 *continued on following page...*

Table A.2 – *Continued*

Query	Count	Identity %	E-value	Accession	Description	HIG
Z21	1	99.6	3.48E-119	JF273734	NRRLB-16219	III
Z1	3	99.6	1.297E-124	JF273729	ChI7	III
Z17	1	99.6	1.297E-124	JF273729	ChI7	III
Z53	1	99.6	1.58E-123	JF273729	ChI7	III
Z10	4	99.6	1.58E-123	JF273729	ChI7	III
Z15	1	99.6	1.58E-123	JF273729	ChI7	III
Z30	1	99.6	6.717E-122	JF273729	ChI7	III
Z25	1	99.6	1.58E-123	JF273729	ChI7	III
Z56	1	98.8	6.717E-122	JF273729	ChI7	III
Z29	4	99.6	3.05E-126	JF273729	ChI7	III
Z4	1	99.6	1.297E-124	JF273729	ChI7	III
Z26	1	99.6	1.297E-124	JF273729	ChI7	III
Z39	1	99.6	1.297E-124	JF273729	ChI7	III
Z46	1	99.6	1.58E-123	JF273729	ChI7	III
Z44	1	99.6	1.297E-124	JF273729	ChI7	III
Z51	1	99.2	1.58E-123	AJ545031	BMG5.12	III
Z35	1	98.8	6.718E-122	AJ545031	BMG5.12	III
Z5	2	99.2	1.58E-123	JF273726	BMG5.15	III
Z43	1	98.8	6.718E-122	AJ545031	BMG5.12	III
Z66	1	98.8	6.718E-122	AJ545031	BMG5.12	III
Z42	1	99.2	1.58E-123	JF273726	BMG5.15	III
Z8	6	99.6	1.297E-124	AJ545031	BMG5.12	III
Z32	1	99.2	1.58E-123	AJ545031	BMG5.12	III
Z20	4	99.2	1.58E-123	AJ545031	BMG5.12	III
Z58	1	99.6	1.58E-123	AJ545031	BMG5.12	III
Z54	1	98.8	6.718E-122	JF273726	BMG5.15	III
Z13	1	100	3.48E-119	JF273724	BCU110345	III
Z3	10	99.6	1.297E-124	AJ545031	BMG5.12	III
Z18	5	100	3.05E-126	JF273732	CpI1	I
Z11	2	99.6	1.297E-124	JF273732	CpI1	I
Z69	1	99.2	1.58E-123	JF273732	CpI1	I
Z2	1	98.4	2.856E-120	JF273732	CpI1	I
Z49	1	97.6	4.239E-118	JF273732	CpI1	I

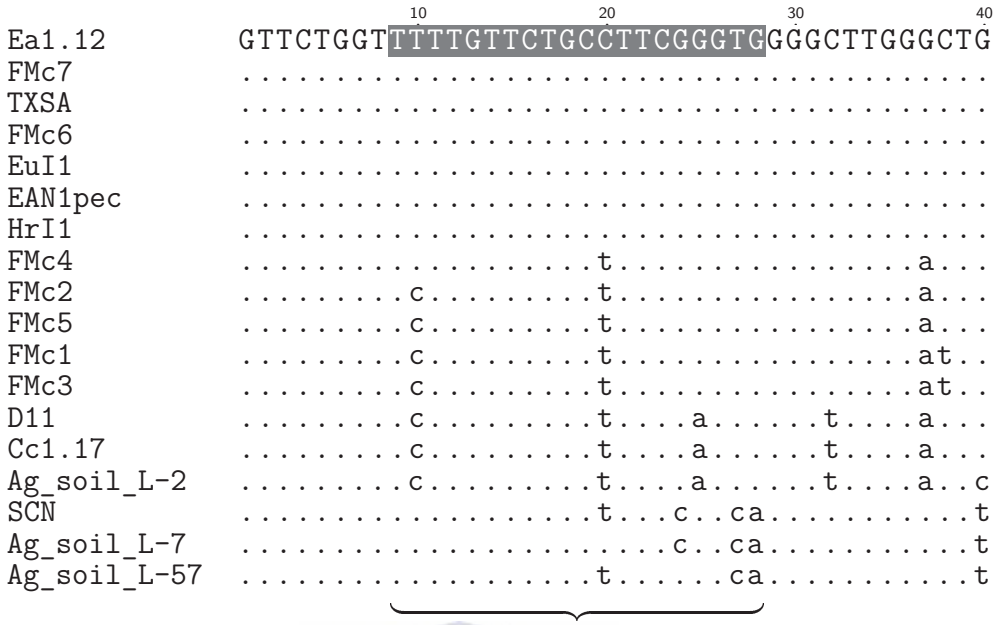
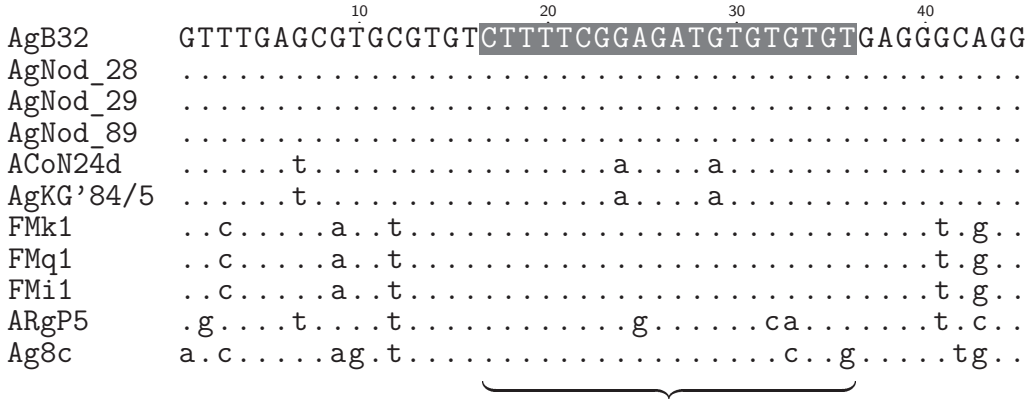


Figure A.2: Conserved sequences within the 23S rRNA insertion with potential use in discriminating *Elaeagnus*-infective *Frankia* from subgroup VI.



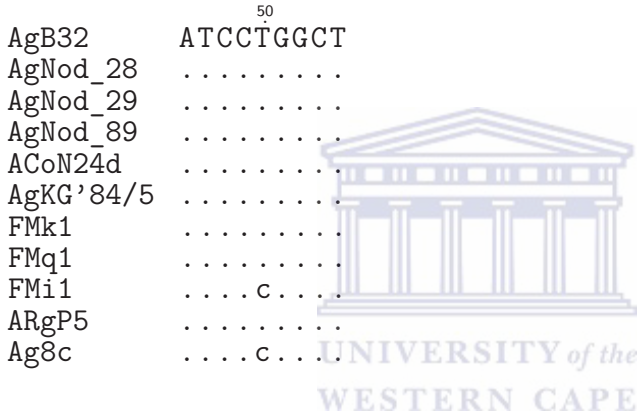


Figure A.3: Conserved sequences within the 23S rRNA insertion potentially useful for discriminating Alnus-infective *Frankia* from subgroup IIIb.

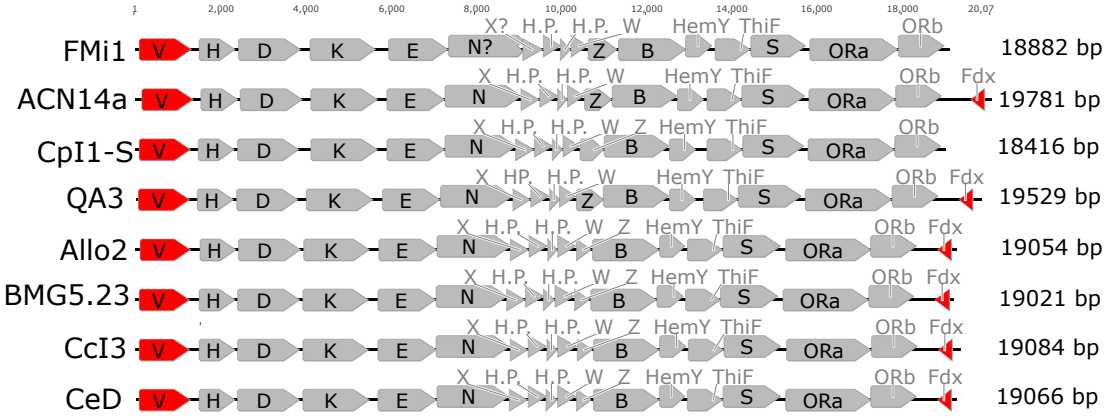


Figure A.4: Comparison of *nif* and N_2 fixation-associated gene cluster arrangement between Cape *M. integra* isolate FMi1, and cluster 1 *Frankia* strains infective on *Alnus* and *Casuarina*. In these strains *nif*V is upstream of *nif*H, and the ferredoxin gene (Fdx) is on the opposite strand downstream of ORb . The apparent absence of Fdx in FMi1 is due to termination of the contig immediately after ORb. Single capitalized letters represent *nif* genes, H.P. is hypothetical protein. Also indicated are ORa, ORb, HemY and ThiF.

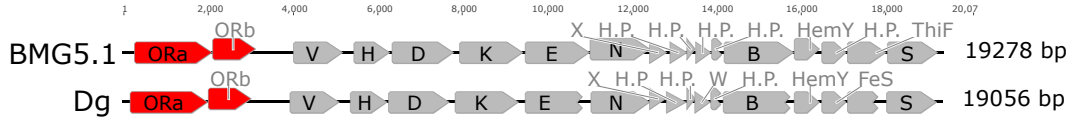


Figure A.5: Arrangement of *nif* and N_2 fixation-associated gene clusters in cluster 2 *Frankia* strains. ORa (ferredoxin oxidoreductase alpha) and ORb (ferredoxin oxidoreductase beta) are upstream of *nif*V which, as with cluster 1 strains, is upstream of *nif*H. The ferredoxin gene (Fdx) is found elsewhere on the genome. Single capitalized letters represent *nif* genes, H.P., hypothetical proteins. Also indicated are HemY and FeS.

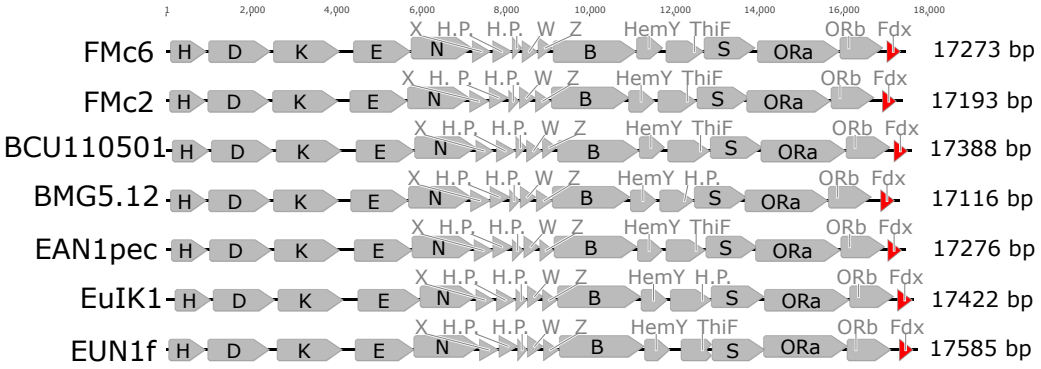


Figure A.6: Comparison of *nif* and N_2 fixation-associated gene cluster arrangement between Cape *M. cordifolia* isolates FMc6 and FMc2, and cluster 3 *Frankia* strains infective on *Elaeagnus*. In cluster 3 *nifV* is located elsewhere on the genome. The ferredoxin gene (Fdx) is downstream of ORb and has the same orientation as other genes in the cluster. Single capitalized letters represent *nif* genes, H.P., hypothetical proteins. Also indicated are ORa, ORb, HemY and FeS.

Table A.3: Similarity of individual nucleotide sequences from genes within the nitrogenase gene cluster from isolate FMc2 to those from other *Frankia* strains

	nifV	nifH	nifD	nifK	nifE	nifN	nifX	nifW	nifZ	nifB	HemY	ThiF	nifS	Ora	Orb	Fdx
BMG5.1	NA	86,4	78,5	70,6	90,5	65,4	63,5	NA	NA	64,7	61,5	49,1	62,3	66,8	64,8	NA
Dg	NA	85,3	79,7	70,4	78,5	65,9	64,1	53,1	NA	61,7	62,1	55,6	63,1	67,0	64,4	NA
BCU110501	NA	95,2	95,1	90,3	90,9	81,9	85,8	79,4	81,3	79,0	88,6	72,4	84,1	89,9	87,7	90,9
EuIK1	NA	92,9	93,3	87,7	79,3	70,9	91,8	79,7	80,7	77,3	88,4	74,0	83,7	85,9	86,1	91,2
EAN1pec	NA	94,4	95,2	90,0	90,6	82,3	82,4	81,0	80,2	79,7	86,8	78,0	84,9	90,0	87,0	88,3
EUN1f	NA	94,7	91,6	86,3	91,0	80,1	84,6	80,0	79,9	82,2	91,3	83,3	89,2	93,8	91,7	89,8
BMG5.12	NA	99,7	99,2	99,4	98,3	97,2	98,3	97,6	97,2	98,6	98,7	NA	99,2	97,7	94,1	98,5
Allo	NA	92,5	89,3	86,0	81,1	62,4	68,3	61,0	58,1	64,8	74,8	66,4	64,5	76,3	67,5	68,1
BMG5.23	NA	92,5	89,4	85,9	81,1	64,7	68,3	61,0	58,1	64,9	75,2	66,4	64,4	76,4	67,5	68,1
CcI3	NA	92,5	89,4	85,9	81,1	62,4	68,3	61,0	58,1	64,9	75,2	66,4	64,5	76,4	67,5	68,1
CeD	NA	92,5	89,3	86,0	80,8	63,0	68,3	61,0	58,1	64,9	75,1	66,4	64,4	76,3	67,4	68,1
ACN14a	NA	93,1	88,8	82,1	81,8	59,9	71,7	71,2	42,7	67,4	75,2	66,0	59,1	76,4	70,8	68,7
CpI1S	NA	93,0	89,1	82,1	81,7	64,2	72,3	71,2	47,4	66,6	74,4	66,1	60,6	75,6	71,7	NA
QA3	NA	93,3	89,0	81,7	81,8	63,4	70,8	71,7	42,1	67,3	75,2	66,2	59,1	76,3	72,0	69,9
FMc6	NA	94,5	95,5	90,1	90,5	81,9	82,3	79,9	80,2	79,2	87,8	77,5	84,4	90,4	87,3	89,8
FMi1	NA	92,1	88,8	82,5	78,6	NA	NA	66,7	39,4	66,0	75,0	63,8	65,3	73,6	69,5	NA

Table A.4: Similarity of individual peptide sequences from genes within the nitrogenase gene cluster from isolate FMc2 to those from other *Frankia* strains

Strain	nifV	nifH	nifD	nifK	nifE	nifN	nifX	nifW	nifZ	nifB	HemY	ThiF	nifS	Ora	Orb	Fdx
BMG5.1	NA	85,5	NA	63,5	74,8	53,2	60,2	NA	NA	67,0	50,0	61,7	57,3	59,7	52,6	NA
Dg	NA	85,5	74,7	62,7	75,8	53,6	60,2	52,5	NA	70,1	50,5	NA	58,1	59,9	52,9	NA
BCU110501	NA	96,2	96,7	90,2	93,2	79,8	83,1	75,4	78,0	74,6	86,9	71,7	84,8	90,0	85,5	94,6
EuIK1	NA	92,3	94,2	86,7	92,8	51,0	74,8	73,8	78,0	75,0	87,4	72,3	81,7	55,4	82,6	94,6
EANipec	NA	94,8	96,7	90,2	94,5	79,8	79,1	75,4	78,0	77,6	86,4	74,3	85,3	90,2	85,8	94,6
EUN1f	NA	94,1	92,6	83,2	79,3	78,6	91,5	81,6	81,9	83,6	90,3	NA	90,6	94,6	91,6	88,4
BMG5.12	NA	99,7	99,6	99,6	98,7	96,3	98,0	97,6	94,5	98,4	99,0	NA	99,5	98,2	93,6	100
Allo	NA	93,4	90,9	86,3	81,0	58,4	64,8	62,2	48,9	73,2	66,1	61,1	59,1	73,0	63,1	60,4
BMG5.23	NA	93,4	90,9	86,3	81,2	60,6	64,8	62,2	49,6	73,2	67,5	61,1	59,1	73,3	63,1	60,4
CcI3	NA	93,4	90,9	86,3	81,2	58,4	64,8	62,2	49,6	73,2	67,5	61,1	59,1	73,3	63,1	60,4
CeD	NA	93,4	90,9	78,7	79,9	59,0	64,8	61,4	49,6	73,7	67,5	61,1	59,1	72,9	63,1	60,4
ACN14a	NA	95,1	89,7	79,2	80,5	52,8	69,2	61,4	51,2	75,9	64,6	60,4	57,8	73,0	63,6	66,7
CpI1S	NA	95,5	89,7	78,5	80,5	57,6	69,2	59,8	48,8	75,7	64,1	61,8	56,2	72,0	64,2	NA
QA3	NA	95,1	89,7	86,3	80,7	56,4	67,8	62,2	52,8	75,9	65,0	61,4	55,6	72,7	64,2	65,8
FMc6		94,8	96,7	89,6	92,8	80,8	78,4	75,4	77,2	76,8	87,4	73,0	85,0	91,9	86,4	94,6
FMi1		92,7	89,1	79,6	80,1	NA	NA	62,2	46,9	71,6	67,0	58,9	59,7	72,2	64,5	NA

Table A.5: Similarity of individual nucleotide sequences from genes within the nitrogenase gene cluster from isolate FMc6 to those from other *Frankia* strains

	nifV	nifH	nifD	nifK	nifE	nifN	nifX	nifW	nifZ	nifB	HemY	ThiF	nifS	Ora	Orb	Fdx
BMG5.1	NA	87,9	79,2	70,4	71,8	63,0	62,9	NA	NA	60,2	61,8	49,5	63,1	67,1	65,8	NA
Dg	NA	87,3	80,6	70,4	79,7	63,1	63,5	57,0	NA	58,8	62,3	57,3	64,1	67,4	65,4	NA
BCU110501	NA	97,6	98,4	97,1	96,9	90,9	90,2	93,4	93,0	87,2	97,6	86,3	93,6	96,9	95,8	98,2
EuIK1	NA	96,6	96,4	95,3	83,6	78,5	81,0	92,9	92,7	87,8	97,1	88,7	92,9	92,2	92,2	98,0
EAN1pec	NA	99,2	99,3	98,8	99,4	96,5	86,8	98,2	99,7	98,1	98,6	94,0	97,9	97,2	95,6	96,2
EUN1f	NA	94,9	91,9	84,7	89,1	76,2	92,7	77,8	79,1	78,9	87,4	82,0	82,5	89,8	87,1	87,4
BMG5.12	NA	94,6	95,1	89,9	90,5	82,6	82,9	80,1	79,3	79,1	87,8	NA	84,6	90,0	86,8	89,8
Allo	NA	91,1	90,0	86,7	80,7	62,2	69,2	61,5	61,2	60,2	74,8	67,9	65,4	77,8	69,2	69,5
BMG5.23	NA	91,1	90,1	86,6	80,8	64,5	69,2	61,5	61,4	60,3	75,2	67,9	65,4	77,9	69,2	69,8
CcI3	NA	91,1	90,1	86,6	80,7	62,2	69,2	61,5	61,4	60,3	75,2	67,9	65,4	77,9	69,2	69,5
CeD	NA	91,1	90,0	86,7	80,4	62,8	69,2	61,5	61,4	60,2	75,1	67,9	65,3	77,8	69,2	69,8
ACN14a	NA	92,2	89,5	83,2	82,8	59,1	71,1	71,1	40,6	62,6	74,1	68,8	59,4	77,5	71,8	71,0
CpI1S	NA	92,2	89,6	83,4	82,4	63,4	71,6	71,1	45,6	61,8	74,0	68,8	61,2	76,8	72,1	NA
QA3	NA	92,2	89,7	83,4	82,9	62,7	71,3	71,1	39,8	62,3	73,4	68,6	59,8	77,3	72,5	70,4
FMc6	NA	94,5	95,5	90,1	90,5	81,9	82,3	79,9	80,2	79,2	87,8	77,5	84,4	90,4	87,3	89,8
FMi1	NA	92,3	89,4	83,9	78,5	NA	NA	68,6	37,4	61,0	74,7	67,1	66,9	75,1	70,8	NA

Table A.6: Similarity of individual peptide sequences from genes within the nitrogenase gene cluster from isolate FMc6 to those from other *Frankia* strains

	nifV	nifH	nifD	nifK	nifE	nifN	nifX	nifW	nifZ	nifB	HemY	ThiF	nifS	Ora	Orb	Fdx
BMG5.1	NA	87,9	NA	64,3	76,1	52,3	63,2	NA	NA	67,2	51,0	64,4	58,2	60,9	51,3	NA
Dg	NA	88,9	74,3	63,9	76,9	52,3	63,2	54,2	NA	70,7	51,5	NA	59,4	61,1	51,6	NA
BCU110501	NA	96,9	99,6	96,7	98,5	89,8	87,3	92,1	92,4	83,5	98,1	81,9	94,4	97,4	96,0	100
EuIK1	NA	94,8	96,7	94,4	99,3	57,3	80,4	90,5	93,2	84,4	98,5	81,6	90,5	59,9	89,6	100
EANipec	NA	100	100	98,1	91,9	95,3	91,9	96,8	99,1	97,9	99,0	93,4	97,8	97,8	95,9	100
EUN1f	NA	95,8	93,0	81,9	84,1	74,2	75,7	76,0	81,1	75,7	90,3	NA	83,1	90,2	86,1	86,7
BMG5.12	NA	95,2	96,7	89,8	91,9	81,3	77,7	75,2	78,1	76,3	87,4	NA	85,0	91,5	84,7	94,6
Allo	NA	94,1	91,8	87,5	80,8	57,2	66,9	60,6	55,1	73,9	67,1	64,9	59,9	73,6	63,7	63,1
BMG5.23	NA	94,1	91,8	87,5	81,0	59,4	66,9	60,6	56,4	73,9	68,4	64,9	59,9	73,9	63,7	63,1
CcI3	NA	94,1	91,8	87,5	81,0	57,2	66,9	60,6	56,4	73,9	68,4	64,9	59,9	73,9	63,7	63,1
CeD	NA	94,1	91,8	81,2	79,7	57,8	66,9	65,4	56,4	74,1	68,4	64,9	59,9	73,4	63,7	63,1
ACN14a	NA	93,4	91,2	82,3	80,7	51,8	67,8	65,4	55,1	76,2	63,6	62,4	58,0	73,4	65,1	70,3
CpI1S	NA	93,7	91,2	82,1	80,7	56,5	67,8	63,8	56,0	75,8	63,1	63,1	57,3	72,7	65,3	NA
QA3	NA	93,4	90,7	87,5	81,0	55,1	67,8	60,6	58,5	76,0	64,1	63,4	56,6	73,1	65,9	66,7
FMc2	NA	94,8	96,7	89,6	92,8	80,8	78,4	75,4	77,2	76,8	87,4	73,0	85,0	91,9	86,4	94,6
FMi1	NA	93,7	90,3	82,9	80,1	NA	NA	62,2	49,2	71,5	66,5	62,0	61,1	72,0	66,2	NA

Table A.7: Similarity of individual nucleotide sequences from genes within the nitrogenase gene cluster from isolate FMi1 to those of other *Frankia* strains

Strain	nifV	nifH	nifD	nifK	nifE	nifN	nifX	nifW	nifZ	nifB	HemY	ThiF	nifS	Ora	Orb	Fdx
BMG5.1	64,0	86,5	79,6	70,5	70,0	NA	NA	NA	NA	67,7	61,7	51,9	62,0	63,7	62,5	NA
Dg	63,8	85,5	80,9	70,0	76,5	NA	NA	54,9	NA	66,9	62,3	62,9	63,1	63,6	61,9	NA
BCU110501	69,8	91,7	88,9	83,7	78,5	NA	NA	69,9	37,5	62,0	75,5	61,5	66,8	74,9	70,0	NA
EuIK1	NA	91,1	88,6	82,5	71,3	NA	NA	69,4	37,5	61,0	75,4	62,7	65,6	73,8	69,5	NA
EAN1pec	69,2	92,5	89,5	83,7	78,6	NA	NA	69,1	37,5	61,4	74,1	65,6	66,9	74,7	70,6	NA
EUN1f	68,8	91,1	89,2	80,9	78,7	NA	NA	69,2	38,4	60,1	75,4	72,8	65,0	73,1	69,6	NA
BMG5.12	68,7	92,2	89,0	82,7	78,0	NA	NA	65,7	39,7	65,8	74,7	NA	65,4	73,5	69,5	NA
Allo	81,0	94,8	89,5	84,0	83,0	NA	NA	63,7	38,4	76,7	81,5	78,1	72,8	80,9	78,0	NA
BMG5.23	81,0	94,8	89,5	84,0	83,1	NA	NA	63,7	38,4	76,8	82,1	78,1	72,8	80,9	78,0	NA
CcI3	81,0	94,8	89,5	84,0	83,0	NA	NA	63,7	38,4	76,8	82,1	78,1	72,8	81,0	78,0	NA
CeD	81,0	94,8	89,4	84,0	82,7	NA	NA	63,7	38,4	76,8	82,0	78,0	72,8	80,9	77,9	NA
ACN14a	90,5	94,7	94,5	93,1	87,5	NA	NA	77,0	57,3	81,0	85,7	86,7	72,7	86,4	85,4	NA
CpI1S	90,7	94,6	94,6	93,0	87,2	NA	NA	77,7	53,3	80,8	85,7	86,1	75,2	86,4	85,6	NA
QA3	90,6	95,1	94,5	93,3	87,3	NA	NA	77,5	55,7	81,1	86,3	87,4	74,3	86,3	86,3	NA
FMc6	NA	92,3	89,4	83,9	78,5	NA	NA	68,6	37,4	61,0	74,7	67,1	66,9	75,1	70,8	NA
FMc2	NA	92,1	88,8	82,5	78,6	NA	NA	66,7	39,4	66,0	75,0	63,8	65,3	73,6	69,5	NA

Table A.8: Similarity of individual peptide sequences from genes within the nitrogenase gene cluster from isolate FMi1 to those from other *Frankia* strains

Strain	nifV	nifH	nifD	nifK	nifE	nifN	nifX	nifW	nifZ	nifB	HemY	ThiF	nifS	Ora	Orb	Fdx
BMG5.1	62,6	86,4	NA	61,0	71,2	NA	NA	NA	NA	68,2	50,0	65,2	57,5	62,0	49,9	NA
Dg	63,1	86,8	75,8	60,4	73,2	NA	NA	50,0	NA	71,3	50,5	NA	58,9	62,0	50,1	NA
BCU110501	69,5	93,0	90,1	82,5	79,9	NA	NA	59,8	49,2	69,0	66,5	61,8	61,8	71,4	64,8	NA
EuIK1	NA	92,3	88,9	80,6	80,1	NA	NA	59,1	49,2	70,1	66,5	61,8	58,6	43,1	62,5	NA
EANipec	67,6	93,7	90,3	82,9	78,7	NA	NA	63,0	49,2	71,5	66,5	62,9	61,3	71,6	64,8	NA
EUN1f	68,8	90,9	89,1	78,3	70,2	NA	NA	63,6	48,8	73,2	67,5	NA	59,1	72,1	65,1	NA
BMG5.12	69,5	93,0	89,3	80,0	78,8	NA	NA	61,1	46,9	71,7	67,0	NA	59,7	71,8	63,6	NA
Allo2	81,3	97,9	90,1	81,0	86,6	NA	NA	59,1	54,1	80,4	79,7	76,6	66,9	83,3	79,5	NA
BMG5.23	81,3	97,9	90,1	81,2	86,8	NA	NA	59,1	54,8	80,4	81,1	76,6	66,9	83,6	79,5	NA
CcI3	81,3	97,9	90,1	81,2	86,8	NA	NA	59,1	54,8	80,4	81,1	76,6	66,9	83,6	79,5	NA
CeD	81,3	97,9	90,1	93,1	85,4	NA	NA	68,2	54,8	80,6	81,1	76,3	67,9	83,2	79,5	NA
ACN14a	91,4	95,5	97,1	92,7	90,4	NA	NA	65,4	38,0	85,8	84,5	84,8	73,2	84,3	85,8	NA
CpI1S	91,4	95,1	97,5	93,1	90,8	NA	NA	66,7	38,8	85,6	84,5	86,6	75,1	85,3	86,1	NA
QA3	90,9	95,5	96,9	81,2	90,4	NA	NA	59,1	34,5	85,2	85,4	87,4	74,0	85,2	86,4	NA
FMc6	NA	93,7	90,3	82,9	80,1	NA	NA	62,2	49,2	71,5	66,5	62,0	61,1	72,0	66,2	NA
FMc2	NA	92,7	89,1	79,6	80,1	NA	NA	62,2	46,9	71,6	67,0	58,9	59,7	72,2	64,5	NA