AN ABSTRACT OF THE DISSERTATION OF

Soon-Chun Jeong for the degree of Doctor of Philosophy in Plant Physiology presented on September 25, 1997. Title: EVOLUTION AND ECOLOGY OF THE CEANOTHUS-FRANKIA SYMBIOSIS

Redacted for privacy

Abstract approved: _____

David D. Myrold

The evolutionary relationship between *Frankia* and actinorhizal plants was evaluated by reconstructing molecular phylogenetic trees from *nif*H, 16S rDNA, and *rbc*L nucleotide sequences. Subgroupings in *Frankia* phylogenetic trees reconstructed from *nif*H and from 16S rDNA sequences were consistent in terms of plant origins of *Frankia* strains. Although the branching order of *Frankia* 16S rDNA and plant *rbc*L trees were different, subgroupings of *Frankia* and of plants correlated well in terms of symbiotic partnership. Tree matching, estimated divergence times, and molecular clock hypothesis tests indicated that *Frankia* clades diverged more recently than plant clades and that actinorhizal symbioses originated more than three times after the plant clades diverged.

A phylogenetic tree of *Ceanothus* species, which are symbiotic partners of *Frankia*, was reconstructed using *ndh*F gene sequences. The analysis identified two main clades corresponding to two subgenera: *Ceanothus* and *Cerastes*. The analysis also suggested that three monophyletic clades within the subgenus *Ceanothus* can be delimited on the basis of vegetative characters. Based on *rbc*L sequences, the two subgenera diverged 18 - 39 million years ago whereas species within each subgenus diverged more recently. These results support the current division of *Ceanothus* into two monophyletic subgenera and agree with the postulated recent divergence of many species within each subgenus.

Specificity between *Ceanothus* species and their *Frankia* microsymbionts was evaluated by analysis of DNA in nodules collected from three copopulations of *Ceanothus* species. Sequencing of the intergenic spacer region between 16S and 23S

rRNA genes suggested that *Ceanothus*-microsymbiont *Frankia* are closely related. Nodules were further analyzed by genomic fingerprinting using repetitive sequences and PCR (rep-PCR). A newly designed, direct repeat sequence and a BOX sequence showed that *Ceanothus*-microsymbiont *Frankia* exhibited less diversity within each copopulation than among copopulations. Furthermore, geographic separation was a more important factor for divergence of *Ceanothus*-microsymbiont *Frankia* than host plant.

The population of *Ceanothus*-infective *Frankia* in soils under stands of *Ceanothus velutinus* and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), a non-host plant, were compared. The population sizes were measured using plant bioassay methods with *C. velutinus, C. sanguineus*, and *C. integerrimus* as trap plants. Population size in soil under *C. velutinus* was about 10 times higher than that under the Douglas-fir. Nodulation capacities of the three trap plants were not significantly different. The diversity of *Frankia* nodulating trap plants was examined using rep-PCR. Results suggested that infective *Frankia* is not species-specific with regard to the three *Ceanothus* species used as trap plants and that although the degree of diversity was similar in both soils, the two populations consisted of different *Frankia*.

EVOLUTION AND ECOLOGY OF THE CEANOTHUS-FRANKIA SYMBIOSIS

by

SOON-CHUN JEONG

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 25, 1997
Commencement June 1998

Doctor of Philosophy dissertation of Soon-Chun Jeong presented on September 25, 1997

APPROVED:

Redacted for privacy

Major Professor, representing Plant Physiology

Redacted for privacy

Chair of Plant Physiology Program

Redacted for privacy

Dean of Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for privacy

Soon-Chun Jeong, Author

ACKNOWLEDGMENTS

First of all, I would like to express my special appreciation to Dr. David Myrold for his support and guidance to make me a good scientist, for discussion, understanding, and thoughts leading to this dissertation. I am greatly indebted to him for his help over past two years.

Special thanks should also go to Dr. Aaron Liston for teaching and helping phylogenetic analysis. I would like to thank my committee Dr. Daniel Arp, Dr. Patrick Breen, Dr. Mark Daeschel, and Dr. Jeffrey Steiner for their scientific advices.

I would like to thank Ms. Nancy Ritchie for technical assistance and discussion and Mr. David Martin for reading rough drafts. I would like to recognize those who sent me plant and bacterial material or who help me at the critical experimental steps: Dr. Dwight Baker, Dr. Beth Mullin, Dr. Alison Berry, Dr. Gordon Johnson, Dr. C.-Y. Lee, and Dr. Arthur McKee. I would also like to thank Dr. Sang-Gu Kim and Dr. Jun-Won Choi for their encouragements during bad and good times.

I would like to thank my wife Junghwa Lee for her love and help during the last period.

Finally, my parents deserve special credit. This dissertation was possible only by their support and permanent trust.

CONTRIBUTION OF AUTHORS

Dr. Aaron Liston was involved in the design, analysis, and writing of the Chapter 3. Ms. Nancy J. Ritchie was involved in the discussion of the Chapter 2.

TABLE OF CONTENTS

	Page
Chapter 1. The *Ceanothus-Frankia* Symbiosis	1
1.1 Attributes of the actinorhizal symbiosis	4
1.2 Ecology of *Frankia* in the soil	5
1.3 Size and diversity of *Frankia* populations that nodulate *Ceanothus* species	6
1.3.1 Diversity of *Ceanothus* microsymbiont *Frankia*	6
1.3.2 Ecology of *Ceanothus-Frankia* symbiosis	7
1.4 Methods of studies of *Frankia* populations	8
1.5 Repetitive sequences and the polymerase chain reaction	9
1.6 Research plan	11
1.7 References	13
Chapter 2. Molecular phylogenies of plants and *Frankia* support multiple origins of actinorhizal symbioses	23
2.1 Abstract	23
2.2 Introduction	24
2.3 Methods	26
2.4 Results	30
2.4.1 Phylogenetic analysis of *nif*H and 16S rDNA sequences	20
2.4.2 Phylogenies of *rbc*L sequences	33
2.4.3 Comparison of phylogenetic trees for *Frankia* and actinorhizal plants.	33
2.5 Discussion	40
2.6 References	43
Chapter 3. Molecular phylogeny of the genus *Ceanothus* (Rhamnaceae) using *rbc*L and *ndh*F sequences	47
3.1 Abstract	47
3.2 Introduction	48
3.3 Materials and methods	50

TABLE OF CONTENTS (Continued) Page

3.3.1 Plant materials..... 50
3.3.2 DNA sequencing..... 47
3.3.3 Phylogenetic analysis..... 54
3.4 Results..... 54
3.5 Discussion..... 59
3.6 Acknowledgements..... 61
3.7 References..... 62

Chapter 4. Specific genomic fingerprinting of microsymbiont *Frankia* from *Ceanothus* copopulations using repetitive sequences and PCR..... 65
4.1 Abstract..... 65
4.2 Introduction..... 66
4.3 Materials and Methods..... 68
4.3.1 Bacterial strains and field-collected nodules..... 68
4.3.2 DNA extraction from bacterial cells and plant nodules..... 68
4.3.3 PCR amplification and DNA sequencing of the 16S to 23S rRNA intergenic spacer region..... 70
4.3.4 Rep-PCR..... 71
4.4 Results..... 72
4.4.1 PCR amplification and DNA sequencing of 16S to 23S rDNA intergenic spacer region..... 72
4.4.2 DR PCR fingerprinting and its comparison with BOX-PCR fingerprinting..... 75
4.4.3 BOX and DR PCR fingerprinting of *Ceanothus* nodules..... 82
4.5 Discussion..... 90
4.6 References..... 94

TABLE OF CONTENTS (Continued) Page

Chapter 5. Population size and diversity of *Frankia* in soils under *Ceanothus velutinus* and Douglas-fir stands..... 98
5.1 Abstract..... 98
5.2 Introduction..... 99
5.3 Materials and methods..... 101
5.3.1. Soil collection..... 101
5.3.2 Plant material and growth conditions..... 101
5.3.3 Inoculation..... 103
5.3.4 Nodulation and measurement of root and shoot dry weight..... 104
5.3.5 Acetylene reduction activity assay..... 104
5.3.6 DNA extraction..... 104
5.3.7 Rep-PCR..... 105
5.4 Results..... 105
5.4.1 Measurement of *Frankia* population size by plant nodulation.....
5.4.2 *Frankia* diversity determined by the rep-PCR..... 108
5.5 Discussion..... 116
5.6 References..... 120

Chpater 6. Conclusions..... 124

Bibliography..... 126

Appendices..... 140

LIST OF FIGURES

Figure		Page
2.1	Phylogenetic tree of *Frankia* strains obtained by the Neighbor-joining method of *nif*H genes	31
2.2	Phylogenetic tree of *Frankia* strains obtained by the Neighbor-joining method of 16S rDNAs	32
2.3.	Phylogenetic tree of representative actinorhizal plant taxa obtained via the Neighbor-joining method of *rbc*L sequence data	34
2.4	A reconciled trees for actinorhizal plants and their symbiotic partner, *Frankia*	37
3.1	Single most parsimonious tree constructed by using the branch-and-bound algorithm with maximum parsimony of 13 *ndh*F sequences of *Ceanothus* species plus outgroup taxa	57
4.1	Comparison of ISR sequences between 16S and 23S rRNA genes from *Ceanothus*-microsymbiont *Frankia* (HJVEL and CWVEL), *Frankia* sp. AcN14a, *Frankia* sp. ORS020606, and *Clavibacter michiganesis insidiosum* LMG 3663	73
4.2	A possible reason of generation of complex PCR amplification patterns by DR oligonucleotide-genomic DNA interactions	74
4.3	Ethidium bromide-stained agarose gel patterns of BOX- and DR-PCR products from genomic DNAs of gram-negative and gram-positive bacteria..	76
4.4	Ethidium bromide-stained agarose gel patterns of BOX- and DR-PCR products from various genomic DNA concentrations and DNA mixtures of *E. coli*, *Frankia* sp. EaI12, *C. velutinus*, and *Nicotiana tabacum*	78
4.5	Ethidium bromide-stained agarose gel patterns of BOX PCR(panel A) and DR PCR (panel B) products from six nodules collected from three *Ceanothus* copopulations	84
4.6	Ethidium bromide-stained agarose gel patterns of BOX PCR (panel A) and DR PCR (panel B) products from three nodule lobes of each nodule collected from *C. sanguineus-C. integerrimus* copopulation	86
4.7	Ethidium bromide-stained agarose gel patterns of BOX PCR (panel A) and DR PCR (panel B) products from six products collected from three *C. velutinus-C. prostratus* copopulation sites	88
4.8	Dendrogram of relatedness of rep-PCR fingerprints for *Ceanothus* nodules determined by cluster analysis and the unweighted pair-group method with arithmetic averages	90

LIST OF FIGURES (Continued)

Page

5.1 Relationship between NUs determined by the nodulation capacity and the MPN methods from *C. velutinus* and Douglas-fir stand soils using three *Ceanothus* species 109

5.2 A representative ethidium bromide-stained agarose gel patterns of BOX- (panel A) and DR-PCR (panel B) products from nodules of *C. velutinus C. sanguineus*, and *C. integerrimus Frankia*-inoculated with soils under *C. velutinus* and Douglas-fir stands 112

5.3 Two-dimensional plot of fingerprint patterns of *C. velutinus* (○), *C. sanguineus* (□), and *C. integerrimus* (◇) nodules by principal component analysis 110

LIST OF TABLES

Table		Page
1.1	Actinorhizal plant genera and *Frankia* isolates and accession numbers of *rbc*L and 16S rDNA sequences deposited in GenBank	3
2.1	Sources of *rbc*L sequences	27
2.2	Sources of 16S rDNA and *nif*H sequences	28
2.3	The estimated divergence times among the three clades of *Frankia* and among four clades of actinorhizal plants	38
3.1	Characteristics of representative *Ceanothus* species used in this study	51
3.2	Sources of taxa sampled for DNA sequences and previously published sequences	52
3.3	Oligonucleotides designed for PCR amplification and DNA sequencing	53
3.4	Pairwise distances of *ndh*F sequences among ten *Ceanothus* species	56
4.1	Sources of bacterial strains, plants, or IGS DNA sequences used in this study	69
4.2	Sources of *Ceanothus* nodules	69
5.1	Characteristics of soil samples	102
5.2	Comparison of shoot and root dry weight in terms of bioassay species, stand type, and soil dilution	107
5.3	Comparison of acetylene reduction rate of nodule lobes from three *Ceanothus* species	107
5.4	Mean NU calculated using nodulation capacity and MPN method for three *Ceanothus* species using soils from *Ceanothus velutinus* and Douglas-fir stands	110
5.5	Number of nodules analyzed by BOX- and DR- PCRs	114

LIST OF APPENDICES

Appendix		Page
1	The data matrix of nucleotide sequences used for analyzing actinorhizal symbioses	141
2	The data matrix of 13 *ndh*F nucleotide sequences used for parsimony analysis	169
3	Binary data matrix of fingerprints generated by BOX and DR PCRs of nodule DNAs from *C. velutinus, C. sanguineus*, and *C. integerrimus* inoculated with soils from the *C. velutinus* and Douglas-fir stands	179

Evolution and Ecology of the *Ceanothus-Frankia* Symbiosis

Chapter 1
The *Ceanothus-Frankia* Symbiosis

Actinorhizae are N_2-fixing symbioses between dicotyledonous plants and *Frankia*. Actinorhizal plants, including *Ceanothus*, form microsymbiotic root nodules with *Frankia* (Frankiaceae). *Frankia* are N_2-fixing, sporulating, Gram-positive, filamentous bacteria. These actinorhizal associations contribute significant quantities of N (1 - 150 kg N ha^{-1} y^{-1}) to temperate forest ecosystems (Torrey, 1978; Dawson, 1983). In addition to enhancing forest productivity by serving as a major input of N, some actinorhizal plants, such as many *Alnus* and *Casuarina* species, are commercially valuable (Diem and Dommergues, 1990; Hibbs and Cromack, 1990; Wheeler and Miller, 1990).

Actinorhizal plants occur in diverse environments and are early successional plants. In conventional taxonomic treatment actinorhizal plants are diverse, belonging to eight plant families and 24 genera (Table 1.1). All are perennial woody shrubs or trees except *Datisca*, which is herbaceous. The term "actinorhizal" plants was coined to describe their common features (Torrey and Tjepkema, 1979). In contrast to conventional taxonomy (Cronquist, 1988), recent molecular phylogenies of angiosperm plants based on *rbc*L sequences suggest a close genealogical relationship among N_2-fixing plants (Chase et al., 1993). Actinorhizal plants were found in three of the four subclades that contain N_2-fixing symbioses (Soltis et al., 1995). Soltis et al. (1995) also suggested that N_2-fixing plants, including actinorhizal plants and legumes, have a single origin in common with several non- N_2-fixing plants. An analysis of *rbc*L sequences combined with anatomical and morphological characters (Swensen, 1996) suggested that actinorhizal symbioses originated at least four times, not like the previously accepted idea that the symbioses are chaotic (Sprent, 1994; Bousquet and Lalonde, 1990).

The genus name *Frankia* was proposed in 1887 by J. Brunchorst (1987). However, it took 100 years to isolate pure cultures. Since the first successful isolation of *Frankia* from nodule tissue (Callaham et al., 1978), much has been learned about *Frankia* taxonomy, host specificity groupings, metabolism, biochemistry, and genetics (Lechevalier and Lechevalier, 1984; Normand and Lalonde, 1986; Tjepkema et al., 1986; Schwintzer and Tjepkema, 1990; Benson and Silvester, 1993). For example, studies of the isolated strains have allowed *Frankia* strains to be grouped on the basis of host specificity (Baker, 1987), DNA-DNA reassociation kinetics, restriction fragment length polymorphisms, DNA sequences, etc. Recent phylogenetic positioning of *Frankia* using full-length 16S rDNAs (Normand et al., 1996) suggests that *Frankia* are monophyletically distinct from close bacterial genera and are grouped into four subclades. Ecological studies of *Frankia* in soil are in their infancy, however, primarily because of the difficulty of isolating *Frankia* from soil and of differentiating *Frankia* strains from each other and from other soil microorganisms. Although microsymbionts have been isolated from 20 of 24 actinorhizal plant genera, typical *Frankia* isolates confirmed by host reinfection and N_2-fixing ability measurement have been reported from only 10 plant genera. *Frankia* isolates from *Ceanothus* (Lechevalier and Ruan, 1984), which is the focus of this study, are unable to reinfect *Ceanothus*, although some can nodulate members of the Elaeagnaceae and Myricaceae (Baker, 1987).

This study explored the phylogeny and ecology of the dicotyledonous plant-*Frankia* association, focusing on the *Ceanothus-Frankia* symbiosis. It builds upon the information collected on the growth and survival of *Alnus*-infective *Frankia* in soil at Myrold's laboratory (Hilger et al., 1991; Hilger and Myrold, 1992; Crannell et al., 1994; Huss-Danell and Myrold, 1994; Myrold and Huss-Danell, 1994) and a legacy of research on *Ceanothus* in the forests of Oregon by researchers associated with Oregon State University (Zavitkovski and Newton, 1968; Youngberg and Wollum, 1976; Binkley et al., 1982, McNabb et al., 1979, Kim, 1987).

Table 1.1. Actinorhizal plant genera and *Frankia* isolates and accession numbers of *rbc*L and 16S rDNA sequences deposited in GenBank.

Family and genus	Isolates[a]	*rbc*L[b]	16S rDNA[b]	Selected references
Betulaceae				
Alnus	Y, T	X56618	M88466	Baker et al., 1979
Casuarinaceae		-		
Allocasuraina	Y, T	X69527		Zhang and Torrey, 1985
Casuarina	Y, T	L01893	M55343	Diem et al., 1983
Gymnostoma	Y, T	X69531		Racette and Torrey, 1989
Ceuthostoma	N	-		
Coriariceae				
Coriaria	Y, A	L01897	L18981	Chaudhary and Mirza, 1987
Datiscaceae				
Datisca	Y, A	L21939	L18979	Chaudhary and Mirza, 1987
Elaeagnaceae				
Elaeagnus	Y, T	U17038	L40618	Baker et al, 1979
Hippophaë	Y, T	U17039		Gauthier et al., 1981
Shepherdia	Y, T	-		Baker, 1982
Myricaceae				
Comptonia	Y, T	X69529		Callaham et al., 1978
Myrica	Y, T	L01934	L40622	St.-Laurent and Lalonde, 1986
Rhamnaceae				
Ceanothus	Y, A	U06795		Lechevalier and Ruan, 1984
Colletia	Y, A	U59819		Burggraaf and Shipton, 1983
Discaria	Y, A	U59826		Longeri and Aberzua, 1989
Kentrothamnus	N	-		
Retanilla	Y, ND	-		Caru, 1993
Talguenea	N	-		
Trevoa	Y, T	U59828		Carrasco et al., 1995
Rosaceae		-		
Cercocarpus	Y, A	U06796		Baker, 1987
Chamaebatia	N	-		
Cowania	Y, ND	U59817		Baker and O'Keefe, 1984
Dryas	N	U06825	L40616	
Purshia	Y, A	U06821		Baker and O'Keefe, 1984

[a] Symbols: Y, isolates reported; N, isolates not reported; Y, typical strains which reinfect their original host plant species and fix N_2; A, atypical strains which do not reinfect their host plant species but usually fix N_2 in culture or associated with other host plant species; ND, infectivity and effectivity not determined.

[b] More than one sequence have been deposited from some plant genera or their *Frankia* isolates. The Genbank accession numbers provided are those of the longest or representative sequences.

1.1 Attributes of the actinorhizal symbiosis

Free-living *Frankia* differentiate into three different cell types: septate hyphae, multilocular sporangia, and thick-walled vesicles. Hyphae are septate, often tightly interwoven in culture, and produce either terminal or intercalary sporangia (Horrière et al., 1983; Lancelle et al., 1985). Segmentation within the enlarging sporangia produce a multilocular sporangium containing many spores (van DijK and Merkus, 1976). Sporangia are produced readily in culture by all *Frankia* strains isolated (Callaham et al., 1979). In symbiosis, nodules contain spores (sp^+) or lack spores (sp^-) (van Dijk, 1978). When suspensions from sp^+ and sp^- nodules are used to inoculate cultured plants, the plants produce sp^+ and sp^- nodules, respectively, suggesting genetic differences (van Dijk, 1984). These two types of nodules have important ecological implications, because several studies have shown that plants inoculated with sp^- cultures produce significantly more biomass than those inoculated with sp^+ cultures (Normand and Lalonde, 1982; VandenBosch and Torrey, 1984; Monz and Schwintzer, 1989; van Dijk et al., 1988; Weber, 1990). Vesicles are normally produced in culture on nitrogen-free or nitrogen-poor media and are the site of nitrogenase (Tjepkema et al., 1980). Provesicles initially develop as terminal swellings on hyphae or on short side branches (Fontaine et al., 1984). Provesicles rapidly develop into mature vesicles which are characterized by the multilayered envelope and internal septum.

In nodules, *Frankia* typically differentiate into vesicles and sometimes sporangia. However, actinorhizal nodules are morphologically diverse and the host plant obviously plays a significant role in modifying *Frankia* morphology. For example, the mature effective form of *Frankia* strains in nodules is often but not exclusively associated with symbiotic vesicle formation, although all effective *Frankia* strains described to date form vesicles in culture (Newcomb and Wood, 1987). *Alnus*, *Ceanothus*, and *Elaeagnus* nodules have large, spherical, multiseptate vesicles. Members of the Rosaceae (*Cercocarpus* and *Dryas* species) have nonseptate elliptical vesicles. *Coriaria*, *Myrica*, and *Comptonia* species have simple club-shaped hyphal endings, *Casuarina* species have an entirely filamentous mature structure.

Frankia enter root tissue by root hair infection (for *Alnus, Casuarina, Comptonia*, and *Myrica* species) or by intercellular penetration (for *Ceanothus, Elaeagnus*, and

Shepherdia species) (Benson and Silvester, 1993). Infection route is determined by the host plant. That is, one *Frankia* strain may enter either by root hair or by intercellular space, depending on host plant species, but a plant species shows only one infection route (Racette and Torrey, 1989; Miller and Baker, 1986). Root hair infection is characterized by branching and curling of the root hair and only one root hair infection leads to nodulation (Berry et al., 1986). The root epidermis is the site of intercellular penetration. In *Ceanothus* species this occurs via the intercellular spaces in the presence of root hairs, which are not infected (Liu and Berry, 1991), whereas in *Elaeagnus* and *Shepherdia* species infection occurs in the absence of root hairs (Racette and Torrey, 1989; Miller and Baker, 1986).

1.2 Ecology of *Frankia* in the soil

Frankia can proliferate in two niches: root nodules and soil. The isolation from root nodules (Callaham et al., 1978) and soil (Baker and O'Keefe, 1984) of *Frankia* that can be successfully cultured in simple media suggests that *Frankia* is a saprophyte and a facultative symbiont. However, evidence for the saprophytic condition comes from indirect observations and experiments, because of the difficulty in differentiating between spores and hyphae (Benson and Silvester, 1993).

Generally, high *Frankia* populations in soil are observed near soils of host plants. However, many observations also indicate that *Frankia* are present in soils well outside the normal geographic range of host plants, under non-host plant stands, or long after host plants have disappeared from a site. For example, in New Zealand the reliable nodulation of both *Alnus* and *Elaeagnus* species that are recent arrivals to the country was observed at every site where they grow, from sea level to 1700 m (Benecke, 1969). In a study of Finnish soils, Smolander and Sundman (1987) found higher populations of *Alnus*-infective *Frankia* in birch (not-host plant) soils than in alder (host plant) soils. As a forest stand develops and *Ceanothus* drops out of the stand, the nodulation capacity of the soil decreases, although some *Frankia* seem to persist even in 300-year-old stands (Wollum et al., 1968).

In addition to host plants, *Frankia* soil populations are controlled by soil variables. The population size of *Frankia* in soils was positively correlated with pH 3 up to pH 8.0 in a study of Finnish soils (Smolander and Sundman, 1987). In contrast, Myrold and Huss-Danell (1994) observed no significant correlation between pH and population size in a study of Swedish soils. However, they found that C level may regulate the size of infective *Frankia* populations. There is some indication that the speed of nodulation during stand development is a function of soil Ca levels. Scott (1973) observed that nodulation of *Ceanothus* is more rapid on sites with higher amounts of Ca. This was supported by observations made with *Alnus rubra* that numbers of infective *Frankia* (Hilger and Myrold, 1992) and degree of seedling nodulation (Crannell et al., 1994) increase following liming or the addition of Ca to soils.

1.3 Size and diversity of *Frankia* populations that nodulate *Ceanothus* species.

1.3.1 Diversity of Ceanothus-microsymbiont Frankia

The genus *Ceanothus* (Rhamnaceae) is endemic to North America. Of the 55 species of *Ceanothus*, almost half are found in the western U. S., with the richest diversity in California (Klemmedson, 1979). Eight *Ceanothus* species occur in Oregon, of which snowbrush (*C. velutinus*), redstem (*C. sanguineus*), and deerbrush (*C. integerrimus*) are the most prevalent (Hibbs and Cromack, 1990). *Ceanothus* species are found on early successional sites, most commonly after fire, and are well adapted to xeric sites (Conrad et al., 1985). Most *Ceanothus* species are nodulated by *Frankia* and can fix up to 108 kg N ha^{-1} y^{-1} during the first 10-25 years following colonization (Hibbs and Cromack, 1990).

Relatively few isolates of *Frankia* from *Ceanothus* species have been reported -- six in the second edition of the Catalog of *Frankia* strains (Lechevalier, 1985-86). All of these are from the eastern species *C. americanus*. Wollum et al. (1966) also reported isolating several *Streptomyces* species from *C. velutinus*, but these were clearly not *Frankia*. None of the true *Frankia* isolates from *C. americanus* are capable of reinfecting *Ceanothus* (Lechevalier and Ruan, 1984; Lechevalier, 1985-86), although some strains nodulate species of the Elaeagnaceae and Myricaceae (Baker, 1987). This result is

interesting because *Ceanothus* is infected via intercellular penetration (Liu and Berry, 1991), which is the mode of entry in *Elaeagnus* species as well (Miller and Baker, 1985).

Recently, partial 16S rDNA sequences amplified and sequenced directly from nodules have grouped *Ceanothus*-microsymbiont *Frankia* either with those that nodulate members of the Rosaceae (Benson et al., 1996) or the Elaeagnaceae (Murry et al., 1997). Both of these plant families are in the same subclade of nitrogen-fixing plants, although *Ceanothus* is much more closely related to members of the Elaeagnaceae (Soltis et al., 1995). At the intrageneric level of *Ceanothus*, Baker and Mullin (1994) detected no plant population or geographic patterns of *Frankia* within *C. americanus* nodules. Murry et al. (1997) detected some degree of diversity in a narrow geographic range containing three *Ceanothus* species using rep-PCR techniques.

1.3.2 Ecology of the Ceanothus-Frankia symbiosis

The lack of infective *Ceanothus* isolates of *Frankia* has hindered ecological studies of this particular actinorhizal symbiosis, especially those focused on the bacterial partner. There are, however, a few interesting observations about the nodulation of *Ceanothus* stands in nature. Zavitkovski and Newton (1968) found that nodulation of seedlings decreased in response to high N litter, an observation that has also been observed when inorganic N is added to soil at concentrations greater than 50 mg L^{-1} (Thomas and Berry, 1989). This may mean that nodulation and rates of N_2 fixation in the field vary inversely with soil N content. Youngberg and Wollum (1976) found slower nodulation of *C. velutinus* in a Douglas-fir clearcut, which had about 50% more soil N than a stand of similar density developing on a *Pinus ponderosa* site. Of course, there were probably differences other than soil N, because after 10 years there was 50% greater N accretion at the Douglas-fir site than the *Pinus* site. By the end of the 10-year period most *Ceanothus* at both sites were nodulated. There is some indication that the nodulation rate of *Ceanothus* during stand development is a function of soil Ca levels, with more rapid nodulation on sites with higher amounts of Ca (Scott, 1973). As a forest stand develops and *Ceanothus* drops out of the stand, the nodulation capacity of the soil decreases, although some *Frankia* seem to persist even in 300-year-old stands (Wollum et al., 1968). In a greenhouse study using several *C. velutinus* genotypes and several different soils,

Nelson and Lopez (1989) found that nodulation and N_2 fixation rely on geographic origins of soils rather than host plants. More quantitative estimates are needed to build upon these earlier observations if the ecology of *Ceanothus*-microsymbiont *Frankia* is to be better understood.

1.4 Methods of studies of *Frankia* populations

Only recently have *Frankia* populations in soil been studied quantitatively by plant bioassay using a nodulation capacity (e.g. van Dijk, 1984) and a most-probable-number (MPN) method (e.g. Huss-Danell and Myrold, 1994). All of these studies have been with *Alnus* species. The MPN method, which measures nodulation units (NUs), has been useful in surveying *Frankia* populations in various forest soils (Smolander and Sundman, 1987; van Dijk et al., 1988) and in following the survival of *Frankia* introduced into soil under laboratory conditions (Smolander et al., 1988). These studies and work in Myrold's laboratory (Hilger and Myrold, 1992; Huss-Danell and Myrold, 1994) have given similar results: (1) numbers of *Frankia* range from zero to a few thousand NU g^{-1} soil, (2) *Frankia* populations differ according to the tree species present, and (3) native and introduced populations of *Frankia* are favored by higher soil pH or liming.

Although the MPN method provides a quantitative measure of NUs, it is unclear whether these numbers reflect *Frankia* biomass in soil. A complicating factor is that spores appear to be much more infective than hyphae (van Dijk, 1984; Burleigh and Torrey, 1990). Furthermore, calculations based on DNA content extracted from soil and assayed with *Frankia*-specific probes (Hahn et al., 1990; Simonet et al., 1991) suggest that even viable counts represent only a fraction of the *Frankia* genomes present in soil (Myrold et al., 1990). More recently use of the polymerase chain reaction (PCR) has substantiated that the number of *Frankia* genomes (genomic units, GUs) in soil is significantly greater than the number of *Frankia* NUs (Hilger and Myrold, 1992; Picard et al., 1992; Myrold and Huss-Danell, 1994). It is now possible to expand research on *Frankia* ecology to include forms of *Frankia* that have so far eluded measurement. Such methods can be applied without isolating *Frankia* strains. However, PCR methods still

need to be improved because DNA primers frequently amplify DNA from closely related bacteria or nonspecifically.

Studies of *Frankia* diversity by conventional approaches have emphasized phenotypic characteristics expressed *in planta*, such as sporulation (sp^+ vs. sp^-) (Holman and Schwintzer, 1987; Weber, 1986; van Dijk et al., 1988), nitrogenase activity (Nif^+ vs. Nif^-) (Hahn et al., 1990), or hydrogenase activity (Hup^+ vs. Hup^-) (Sellstedt, 1989). Ecological studies of genetic diversity have often involved isolation of *Frankia* strains. For example, using whole-cell protein patterns, Benson and Hanna (1983) divided 43 isolates from one alder stand into six groups with one group being dominant and containing 35 of the isolates. Bloom et al. (1989) used RFLP (restriction fragment length polymorphism) analysis to divide *Frankia* strains isolated from *Myrica* nodules into five groups. More recently, Jamann et al. (1992) used RFLP analysis to examine the diversity of *Elaeagnus*-infective *Frankia* strains and found that diversity was lower in soils with lower pH.

Molecular techniques that circumvent *Frankia* isolation has innovated population studies especially for *Frankia* strains associated with plant host recalcitrant to isolation. The first study was the use of strain specific *nifH* probes to assay nodule DNA extracts (Simonet et al., 1990). RFLP analysis of nodule DNA extracts (Baker and Mullin, 1994) can be used to examine the population structure of *Frankia* without isolation, as can a newer variant of this, PCR-RFLP (Maggia et al., 1992; Jamann et al., 1993; Rouvier et al., 1996). Several reports have been published recently about the diversity of *Frankia* in actinorhizal nodules of several plant genera by direct PCR and DNA sequencing (Nick et al., 1992; Benson et al., 1996; Normand et al., 1996, Murry et al., 1997).

1.5 Repetitive sequences and the polymerase chain reaction

The best known of interspersed repetitive DNA sequence elements is the *Alu* family of sequences in mammalian species (Schmid et al., 1982; Jelinek et al., 1980; Houck et al., 1979). The interspersed distribution and conserved nature of these *Alu* repeats have been exploited to amplify unique intervening sequences by a technique known as *Alu*-PCR (Nelson et al., 1989). Prokaryotic genomes that are 0.1% the size of those of

mammalian species may have been maintained at a minimum under natural selection for rapid DNA replication and cell reproduction. Nevertheless, families of short intergenic repeated sequences have also been described in the prokaryotic genomes. Notable examples are repetitive extragenic palindromic elements (REPs) and enterobacterial repetitive intergenic consensus repeats (ERICs) found in *Escherichia coli* and *Salmonella typhimurium* (Higgins et al., 1982; Sharples and Lloyd, 1990; Hulton et al., 1991). These elements contain highly conserved, palindromic inverted repeat sequences. As additional DNA sequence information is assembled in different microbial systems, novel repetitive sequences have been found in diverse organisms. Subsequent reports of such sequences in the Gram-negative bacteria *Deinococcus* (Lennon et al., 1991), *Calothrix* (Mazel et al., 1990), *Neisseria* (Correia et al., 1986), the Gram-positive bacteria *Streptococcus* (Martin et al., 1992), *Mycobacterium* (Doran et al., 1993), and the fungi *Candida albicans* (Scherer and Stenens, 1988) and *Pneumocystis carinii* (Stringer et al., 1991), illustrate the presence of dispersed extragenic repetitive sequences in many organisms.

The actual function of these highly repeated and conserved elements remains largely unknown. Although several posibilities (e.g. mRNA stabilization, translational coupling between genes, homologous recombination, chromosome organization and binding of HU proteins, DNA gyrase and DNA polymerase I) has been suggested (e.g. Newbury et al., 1987; Gilson et al., 1990), no single function satisfactorily explains their DNA sequence conservation and/or ubiquitous distribution. Therefore, the suggestion has also been made that these repeated elements represent "selfish" DNA sequences which have evolved via gene conversion (Higgins et al., 1988).

The comparison of REP and ERIC type elements found in *E. coli* and *S. typhimurium* has allowed the derivation of REP and ERIC consensus sequences (Stern et al., 1984; Hulton et al., 1991). These consensus sequences, in turn, allowed Versalovic et al. (1991) to design REP- and ERIC-specific oligonucleotide primers and to fingerprint the genomes of a variety of eubacteria, using the polymerase chain reaction (PCR). Surprisingly they found that REP- and ERIC-like sequences could be detected in a large variety of Gram-negative bacteria. Subsequently a consensus sequence corresponding to the repetitive BOX sequence found in Gram-positive bacteria, primarily *Streptococcus pneumonia*, was applied to fingerprint the genomes of a variety of Gram-negative and Gram-positive

bacteria using the PCR technique (Versalovic et al., 1994; Sadowsky et al., 1996). Recently, the utility of these techniques was also demonstrated to generate characteristic banding patterns with genomic DNA of *Frankia* isolates and nodules (Murry et al., 1997, 1995) Thus, it was proposed that rep-PCR may constitute a useful method to fingerprint bacterial genomes at strain level. It also suggests that the other novel repeated sequences may be useful for genomic fingerprinting coupled with the PCR technique.

1.6 Research plan

A comparison of molecular phylogenetic trees of actinorhizal plants (Soltis et al., 1995; Swensen, 1996) and *Frankia* (Normand et al., 1996; Hönerlage et al., 1995) suggests some degree of congruence between *Frankia* and actinorhizal plant phylogenetic trees, although no rigorous analysis has been done. Both phylogenetic trees indicate coevolution of *Frankia*-actinorhizal symbioses above the familial level of plants. In Chapter II, I analyzed the concordance between host plant and *Frankia* phylogenies using available *rbc*L and 16S rDNA sequences in a rigorous way, although information about *Frankia* from plant families such as Rhamnaceae and Rosaceae, which are recalcitrant to *Frankia* isolation, is meager. The approach used was similar to that used for well-known cospeciation phenomena such as plant-rhizobia symbioses, gopher-lice associations, *Blepharida-Bursera* interactions, etc.

No rigorous phylogenetic analysis in *Ceanothus* has been reported using molecular markers. The objective of Chapter III was to analyze the phylogenetic relationship among representative *Ceanothus* species using the chloroplast genes, *ndh*F and *rbc*L. The results support the division of *Ceanothus* into two monophyletic subgenera and are consistent with the postulated recent divergence of many species within each subgenus.

An experimental test of the coevolutionary relationship between *Frankia* and their hosts, *Ceanothus* species, is described in Chapter IV. Nodules and soil from three sites were collected where *C. velutinus* and *C. integerrimus, C. velutinus* and *C. prostratus,* and *C. integerrimus* and *C. sanguineus* grow together. I analyzed the diversity of *Frankia* strains in plant nodules by sequencing intergenic spacer regions between 16S and 23S rRNA and repetitive sequences and the PCR (rep-PCR). The results suggested that

geographic separation is a more important factor for *Frankia* speciation than its symbiotic partner, *Ceanothus* species.

Chapter V is a comparison of population size and diversity of *Frankia* in soils under *C. velutinus* and under old-growth Douglas-fir stands. The population size of *Ceanothus*-infective *Frankia* was determined using bioassay and MPN methods. Diversity of *Ceanothus*-infective *Frankia* was assessed using plant nodules from seedlings of three *Ceanothus* species grown as "trap" plants in soil collected from both areas. The plant nodules were analyzed by rep-PCR. Acetylene reduction activity was measured to determine the N_2-fixing ability of the nodules. Both nodulation units and total *Frankia* were higher in *C. velutinus* than in the Douglas-fir stands, but diversity was the same in both stands.

Ceanothus species are an important shrub component in the early successional stage of many western U.S. ecosystems. Through their actinorhizal association with *Frankia* they fix significant quantities of N_2 and thereby contribute to the long-term productivity of these ecosystems. Despite the ecophysiological and practical importance of this N_2-fixing symbiosis, few quantitative efforts have been undertaken exploring the ecological interactions of *Ceanothus* and *Frankia*. This is primarily because there are no isolates of *Ceanothus*-infecting *Frankia* strains that are able to reinfect their host. Although this phenomenon is not uncommon in *Frankia*, it has greatly limited the types of autecological studies that can be done. Chapters IV and V, which took advantage of molecular methods, provide fundamental information about diversity of *Ceanothus*-microsymbiont *Frankia* strains on the level of *Ceanothus* species. They also provide the first information about population size of *Ceanothus*-microsymbiont *Frankia* strains. Furthermore, this research gives insight into the interactions between *Frankia* and the plant family Rhamnaceae, one of the most poorly understood groups among eight actinorhizal plant families

1.7 References

Baker, D.D. (1982) A cumulative listing of isolated frankiae, the symbiotic, nitrogen-fixing actinomycetes. Actinomycetes 17: 35-42

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

Baker D.D., and B. C. Mullin (1994). Diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus* as assessed by RFLP patterns from single nodule lobes. Soil Biol Biochem 26:547-552

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

Baker, D., J. G. Torrey, and G. H. Kidd (1979). Isolation by sucrose-density fractionation and cultivation *in vitro* of actinomycetes from nitrogen-fixing root nodules. Nature 281:76-78

Benecke, U. (1969). Symbiosis of alder nodules in New Zealand. Plant Soil 30: 145-149

Benson, D. R., and D. Hanna (1983). *Frankia* diversity in an alder stand as estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis of whole-cell proteins. Can. J. Bot. 61:2919-2923.

Benson, D. R., and W. B. Silvester (1993). Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev. 57:293-319.

Benson D. R., D. W. Stephens, M. L. Clawson, and W. B. Silvester (1996). Amplification of 16S rRNA genes from *Frankia* strains in root nodules of *Ceanothus griseus, Coriaria arborea, Coriaria plumosa, Discaria toumatou*, and *Purshia tridenta*. Appl Environ Microbiol 62:2904-2909

Berry, A. M., L. McIntyre, and M. E. McCully (1986). Fine structure of root hair infection leading to nodulation in the *Frankia-Alnus* symbiosis. Can J Bot 64:292-305

Binkley, D., K. Cromack, Jr., and R. L. Fredrickson. (1982). Nitrogen accretion and availability in some snowbrush ecosystems. For. Sci. 30:26-35.

Bloom, R. A., B. C. Mullin, and R. L. Tate III. (1989). DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (bayberry). Appl. Environ. Microbiol. 55:2155-2160.

Bousquet, J., and M. Lalonde (1990). The Genetics of Actinorhizal Betulaceae. P. 239-261*In* The Biology of *Frankia* and Actinorhizal Plants (C. R. Schwintzer and J. D. Tjepkema, eds). Academic Press. Inc., San Diego, CA.

Brunchorst, J. (1887). Die Struktur der Inhaltskürper in den Zellen einiger Wurzelanschwellungen. Bergens Mus. Aarsb. 235.

Burggraaf, A. J. P., and W. A. Shipton (1983). Studies on the growth of *Frankia* isolates in relation to infectivity and nitrogen fixation (acetylene reduction). Can. J. Bot. 61:2774-2782.

Burleigh, S., and J. G. Torrey (1990). Effectiveness of different *Frankia* cell types as inocula for the actinorhizal plant *Casuarina*. Appl. Environ. Microbiol. 56:2565-2576.

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

Callaham, D, W. Newcomb, J. G. Torrey, and R. L. Peterson (1979). Root infection in actinomycete-induced root nodule initiation in *Casuarina*, *Myrica*, and *Comptonia*. Bot. Gaz. 140 (Suppl.):S1-S9.

Carracso, A, J. R. Salyards, and A. M. Berry (1995). Studies of two *Frankia* strains isolated from *Trevoa trinervis* Miers. Plant Soil 171: 359-363.

Caru, M. (1993). Characterization of native *Frankia* strains isolated from Chilean shrubs (Rhamnaceae). Plant Soil 157: 137-145

Chase M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H.G. Hills, Y.-L. Qui, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, G.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan, V. A. Albert (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Mo Bot Gard 80:528-580.

Chaudhary, A.H., and M. S. Mirza (1987). Isolation and characterization of *Frankia* from nodules of actinorhizal plants of Pakistan. Physiol. Plant. 70: 255-258.

Conrad, S. G., A. E. Jaramillo, K. Cromack, Jr., and S. Rose (1985). The role of the genus *Ceanothus* in western forest ecosystems. U.S.D.A. For. Serv. Gen Tech. Rep. PNW-182. 72 pp.

Correia, F. F., S. Inouye, and M. Inouye (1986). A 26-base-pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. J. Bacteriol. 167:1009-1015.

Crannell, W. K., Y. Tanaka, and D. D. Myrold. (1994). Calcium and pH interaction on root nodulation of nursery grown red alder (*Alnus rubra* Bong.) seedlings by *Frankia*. Soil Biol. Biochem. 26: 607-614.

Cronquist, A. (1988). "The Evolution and Classification of Flowering Plants," 2nd ed. Allen, Lawrence, Kansas.

Dawson, J.O. (1983). Dinitrogen fixation in forest ecosystems. Can. J. Microbiol. 29:979-992.

Diem, H. G., and Y. R. Dommergues. (1990). Current and potential uses and management of Casuarinaceae in the tropics and subtropics. p. 317-342. *In* C.R. Schwintzer and J.D. Tjepkema (ed.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA

Diem, H. G., D. Gauthier, and Y. Dommergues. (1983). An effective strain of *Frankia* from *Casuarina* sp. Can. J. Bot. 61:2815-2821

Doran, T. J., A. L. M. Hodgson, J. K. Davies, and A. J. Radford (1993). Characterisation of a highly repeated DNA sequence from *Mycrobacterium bovis*. FEMS Microbiology Letters 111: 147-152

Fontaine, M.F., S. A. Lancelle, and J. G. Torrey. (1984). Initiation and ontogeny of vesicles in cultured *Frankia* sp. strain HFPArI3. J Bacteriol 160: 921-927

Gauthier, D., H. G. Diem, and Y. Dommergues (1981). Infectivite et effective de souche de *Frankia* isolaees de nodules de *Casuarina equisetifolia* et *d'Hippopha*□ *rhamnoides*. C. R. Acad. Sc. Paris Ser. III 293:489-491.

Gilson, E, D. Perrin, M. Hofnung (1990). DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization. 18:3941-3952

Hahn, D., R. Kester, M. J. C. Starrenburg, and A. D. L. Akkermans. (1990). Extraction of ribosomal RNA from soil for detection of *Frankia* with oligonucleotide probes. Arch Microbiol 154:329-335.

Hibbs, D. E., and K. Cromack, Jr. (1990). Actinorhizal plants in Pacific Northwest forests. p. 343-363. *In* C.R. Schwintzer and J.D. Tjepkema (ed.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA.

Higgins, C. F., G. F. Ames, W. M. Barnes, J. M. Clement, M. Hofnung (1982). A novel intercistronic regulatory element of prokaryotic operons. Nature 298:760-762.

Higgins, C. F., R. S. McLaren, and S. F. Newbury (1988). Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. Gene 72:3-14.

Hilger, A. B., and D. D. Myrold (1992). Quantitation of soil *Frankia* by bioassay and gene probe methods: response to host and non-host rhizospheres and liming. Acta Œcol. 13:505-506.

Hilger, A. B., Y. Tanaka, and D. D. Myrold (1991). Inoculation of fumigated nursery soil increases nodulation and yield of bare-root red alder (*Alnus rubra* Bong.). New Forests 5:35-42

Holman, R. M. and C. R. Schwintzer (1987). Distribution of spore-positive and spore-negative nodules of *Alnus incana* ssp. *rugosa* in Maine, USA. Plant Soil 104:103-111.

Hönerlage, W., D. Hahn, K. Zepp, J. Zeyer, and P. Normand (1995). A hypervariable 23S rRNA region provides a discriminating target for specific characterization of uncultured and cultured *Frankia*. System. Appl. Microbiol. 17:433-443.

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

Houck, C. M., F. P. Rinehart, C. W. Schmid (1979). A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. 132:289-306.

Hulton, C. S. J., C. F. Higgins, and P. M. Sharp (1991). ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enteric bacteria. Mol. Microbiol. 5:825-834.

Huss-Danell, K., and D. D. Myrold. (1994). Intrageneric variation in nodulation of *Alnus*: Consequences for quantifying *Frankia* infective units in soil. Soil Biol. Biochem. 26: 525-531.

Jamann, S., M. P. Fernandez, and A. Moiroud (1992). Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. Acta Œcol. 13:395-406.

Jamann, S., M. P. Fernandez, and P. Normand (1993). Typing method for N_2-fixing bacteria based on PCR-RFLP--application to the characterization of *Frankia* strains. Mol. Ecol. 2:17-26.

Jelinek, W. R., T. P. Toomey, L. Leinwand, C. H. Duncan, P. A. Biro, P. V. Choudary, S. M. Weissman, C. M. Rubin, C. M. Houck, P. L. Deininger, and C. W. Schmid (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. Proc. Natl. Acad. Sci. USA 77:1398-1402.

Kim, D.Y. (1987). Seasonal estimates of nitrogen fixation by *Alnus rubra* and *Ceanothus* species in western Oregon forest ecosystems. M.S. thesis. Oregon State Univ., Corvallis, OR. 71 pp.

Klemmedson, J.O. (1979). Ecological importance of actinomycete-nodulated plants in the western United States. Bot. Gaz. 140(suppl.):S91-S96.

Lancelle, S. A., J. G. Torrey, P. K. Hepler, and D. A. Callaham (1985). Untrastructure of freeze-substituted Frankia strain HFPCcI3, the actinomycete islated from root nodules of *Casuarina cunninghamiana*. Protoplasma 127:64-72.

Lechevalier, M. P. (1985-86). Catalog of *Frankia* strains, second edition. The Actinomycetes 19:131-162.

Lechevalier, M. P., and H. A. Lechevalier. (1984). Taxonomy of *Frankia*. p. 575-582. *In* Biological, Biochemical, and Biomedical Aspects of Actinomycetes (L. Ortiz-Ortiz, L.F. Bojalil, and V. Yakoliff, ed.). Academic Press, New York, NY.

Lechevalier, M. P., and J.-S. Ruan. (1984). Physiological and chemical diversity of *Frankia* spp. isolated from nodules of *Comptonia peregrina* (L.) Coult. and *Ceanothus americanus* L. Plant Soil 78:15-22.

Lennon, E, P. D. Gutman, H. L. Yao, K. W. Minton (1991). A highly conserved repeated chromosomal sequence in the radioresistant bacterium *Deinococcus radiodurans* SARK. J. Bacteriol. 173:2137-2140.

Liu, Q., and A. M. Berry. (1991). The infection process and nodule initiation in the *Frankia-Ceanothus* root nodule symbiosis: a structural and histochemical study. Protoplasma 163:82-92.

Longeri, L, and M. Abarzu (1989). Ultrastructure of *Frankia* isolated from three Chilean shrubs (Rhamnaceae), p 47-53. In FA Skinner (ed.), Nitrogen Fixation with non-legumes. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Maggia, L., S. Nazaret, and P. Simonet (1992). Molecular characterization of *Frankia* isolates from *Casuarina equisetifolia* root nodules harvested in West Africa (Senegal and Gambia). Acta Œcol. 13:453-462.

Martin, B, O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J. P. Claverys (1992). A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. 20: 3479-3483

Mazel, D., J. Houmard, A. M. Castets, N. Tandeau de Marsac (1990). Highly repetitive DNA sequences in cyanobacterial genomes. J. Bacteriol. 172:2755-2761

McNabb, D.H., J. M Geist, and C. T. Youngberg. (1979). Nitrogen fixation by *Ceanothus velutinus* in northeastern Oregon. p. 481-482. *In* Symbiotic Nitrogen Fixation in the Management of Temperate Forests (J.C. Gordon, C.T. Wheeler, and D.A. Perry, ed.). For. Res. Lab., Oregon State Univ., Corvallis, OR.

Miller, I. M., and D. D. Baker (1985). The initiation, development and structure of root nodules in *Elaeagnus augustifolia* L. (Elaeagnaceae). Protoplasma 128:107-119.

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

Monz C. A., and C. Schwintzer (1989). The physiology of spore-negative and spore-positive nodules of *Myrica gale.* Plant Soil 118: 75-87

Murry M. A., A. S. Konopka, S. D. Pratt, T. L. Vandergon (1997). The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus*. Physiol. Plant. (in press).

Murry, M. A., D. Zhang, M. Schneider, and F. J. de Bruijn (1995). Use of repetitive sequences and the polymerase chain reaction (rep-PCR) to fingerprint the genomes of *Frankia* isolates. Symbiosis 19: 223-240.

Myrold, D. D., A. B. Hilger, and S. H. Strauss (1990). Detecting *Frankia* in soil using PCR. p. 429. *In* P.M. Gresshoff, L.E. Roth, G. Stacey, and W.E. Newton (ed.) Nitrogen Fixation: Achievements and Objectives. Chapman and Hall, New York, NY.

Myrold, D. D., and K. Huss-Danell (1994). Population dynamics of *Alnus*-infective *Frankia* in a forest soil with and without host trees. Soil Biol. Biochem. 26: 533-540.

Nelson, D. L, S. Ledbetter, L Corbo, M. F. Victoria, R. Ramirez-Solis, T. D. Webster, D. H. Ledbeter, and C. T. Caskey (1989). *Alu* polymearse chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. USA 86:6686-6690.

Nelson, D. L., and C. F. Lopez (1989). Variation in nitrogen fixation among populations of *Frankia* sp. and *Ceanothus* sp. in actinorhizal association. Biol. Fertil. Soils 7:269-274.

Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins (1987). Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 30:297-310.

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

Nick, G., E. Paget, P. Simonet, A. Moiroud, and P. Normand (1992). The nodular endophytes of *Coriaria* spp. form a distinct lineage within the genus *Frankia*. Mol. Ecol. 1:175-182.

Normand, P., and M. Lalonde (1982). Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. Can. J. Microbiol. 28: 1133-1142.

Normand, P., and M. Lalonde (1986). The genetics of actinorhizal *Frankia*: A review. Plant Soil 90:429-453.

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

Picard, C., C. Ponsonnt, E. Paget, X. Nesme, and P. Simonet. (1992). Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. 58:2717-2711.

Racette, S., and J. G. Torrey (1989). Root nodule initiation in *Gymnostoma* (Casuarinaceae) and *Shepherdia* (Elaeagnaceae) induced by *Frankia* strain HFPGpI1. Can. J. Bot. 67: 2873-2879.

Rouvier C, Y. Prin, P. Redell, P. Normand, P. Simonet (1996). Genetic diversity among *Frankia* strains nodulating members of the family Casuarinaceae in Australia revealed by PCR and restriction fragment length polymorphism analysis with crushed root nodules. Appl. Environ. Microbiol. 62: 979-985.

Sadowsky, M. J, L. L. Kinkel, J. H. Bowers, and J. L. Schottel (1996). Use of repetitive intergenic DNA sequences to classify pathogenic and Disease-supressive *Streptomyces* strains. Appl. Environ. Microbiol. 62:3489-3493.

Scherer, S, and D. A. Stenens (1988). A *Candida anlicans* dispersed, repeated gene family and its epidemiologic applications. Proc. Natl. Acad. Sci. USA 85:1452-1456.

Schmid, C. W., and W. R. Jelinek (1982). The *Alu* family of dispersed repetitive sequences. Science 216:1065-1070

Schwintzer, C. R., and J. D. Tjepkema. (1990). The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA. 408 p.

Scott, W. (1973). Some soil factors affecting snowbrush nodulation. Ph.D. Thesis. Oregon State Univ., Corvallis, OR. (Diss. Abstr. 33:4074).

Sellstedt, A. (1989). Occurrence and activity of hydrogenase in symbiotic *Frankia* from field-collected *Alnus incana*. Physiol. Plant. 75:304-308.

Sharples, G. J., and R. G. Lloyd (1990). A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucleic Acids Res. 18:6503-6508.

Simonet, P., M.-C. Grosjean, A. K. Misra, S. Nazaret, B. Cournoyer, and P. Normand. (1991). *Frankia* genus-specific characterization by polymerase chain reaction. Appl. Environ. Microbiol. 57:3278-3286.

Simonet, P., P. Normand, A. Moiroud, and R. Bardin (1990). Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. Arch. Microbiol. 153:235-240.

Smolander, A., and V. Sundman. (1987). *Frankia* in acid soils of forests devoid of actinorhizal plants. Physiol. Plant. 70:297-303.

Smolander, A., C. van Dijk, and V. Sundman. (1988). Survival of *Frankia* strains introduced into soil. Plant Soil 106:65-72.

Soltis D. E., P. S. Soltis, D. R. Morgan, S. M. Swensen, B. C. Mullin, J. M. Dowd, and P. G. Martin (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA 92:2647-2651.

Sprent J. I. (1994). Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? Plant Soil 161: 1-10.

Stern, M. J., G. F. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell 37:1015-1026.

St.-Laurent, L., and M. Lalonde (1986). Isolation and characterization of *Frankia* strains isolated from *Myrica gale*. Can. J. Bot. 65: 1356-1363.

Stringer, S. L., S. T. Hong, D. Giuntoli, J. R. Stringer (1991). Repeated DNA in *Pneumocystis carinii*. J. Clin. Microbial. 29:1194-1201.

Swensen, S. M. (1996). The evolution of actinorhizal symbiosis: Evidence for multiple origins of the symbiotic association. Am. J. Bot. 83: 1503-1512.

Thomas, K. A., and A. M. Berry (1989). Effects of continuous nitrogen application and nitrogen preconditioning on nodulation and growth of *Ceanothus griseus* var *horizontalis*. Plant Soil 118:181-187.

Tjepkema, J. D., W. Ormerod, and J. G. Torrey (1980). Vesicle formation and acetylene reduction activity in *Frankia* sp. CpI1 cultured in defined nutrient media. Nature 287:633-635.

Tjepkema, J. D., C. R. Schwintzer, and D. R. Benson (1986). Physiology of actinorhizal nodules. Annu. Rev. Plant Physiol. 37:209-232.

Torrey, J. G. (1978). Nitrogen fixation by actinomycete-nodulated angiosperms. BioSci. 28:586-591.

Torrey J. G. and J. D. Tjepkema (1979). Preface. Bot Gaz 140(Suppl): i-ii.

VandenBosch, K. A., and J. G. Torrey (1984). Consequences of sporangial development for nodule function in root nodules of *Comptonia peregrina* and *Myrica gale*. Plant Physiol 76:556-560

van Dijk, C. (1978). Spore formation and endophyte diversity in root nodules of *Alnus glutinosa* (L.) Vill. New Phytol. 81: 601-615

van Dijk, C. (1984). Ecological Aspects of Spore Formation in the *Frankia-Alnus* Symbiosis. Ph.D. Thesis. State University, Leiden, the Netherlands. 154 p.

van Dijk, C. and E. Merkus (1976). A microscopic study of the development of a spore-like stage in the life cycle of the root nodule endophyte of *Alnus glutinosa* (L.) Vill. New Phytol 77: 73-90

van Dijk, C., A. Sluimer, and A. Weber (1988). Host range differentiation of spore-positive and spore-negative strain types of *Frankia* in stands of *Alnus glutinosa* and *Alnus incana* in Finland. Physiol. Plant. 72:349-358.

Versalovic, J., T. Koeuth, and J. R. Lupski (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. 24:6823-6831.

Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski (1994). Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. 5:25-40.

Weber, A. (1986). Distribution of spore-positive and spore-negative nodules in stands of *Alnus glutinosa* and *Alnus incana* in Finland. Plant Soil 96:205-213.

Weber, A. (1990). Host specificity and efficiency of nitrogenase activity of *Frankia* strains from *Alnus incana* and *Alnus glutinosa*. Symiosis 8: 47-60

Wheeler, C. T., and I. M. Miller (1990). Current and potential uses of actinorhizal plants in Europe. p. 365-389. *In* C.R. Schwintzer and J.D. Tjepkema (ed.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA.

Wollum, A. G. II, C. T. Youngberg, and F. W. Chichester (1968). Relation of previous timber stand age to nodulation of *Ceanothus velutinus*. For. Sci. 14:114-118.

Wollum, A. G. II, C. T. Youngberg, and C. M. Gilmour (1966). Characterization of a *Streptomyces* sp. isolated from root nodules of *Ceanothus velutinus* Dougl. Soil Sci. Soc. Am. Proc. 30:463-467.

Youngberg, C. T., and A. G. Wollum II (1976). Nitrogen accretion in developing *Ceanothus velutinus* stands. Soil Sci. Soc. Am. J. 40:109-112.

Zavitkovski, J., and M. Newton. (1968). Ecological importance of snowbrush *Ceanothus velutinus* in the Oregon Cascades. Ecology 49:1134-1145.

Zhang, Z and J. G. Torrey (1985). Studies of an effective strain of *Frankia* from *Allocasuarina lehmanniana* of the Casuarinaceae. Plant Soil 87: 1-16

Chapter 2

Molecular Phylogenies of Plants and *Frankia* Support Multiple Origins of Actinorhizal Symbioses

SC Jeong[1], DD Myrold[1,2], and NJ Ritchie[2]

[1] Plant Physiology Program and [2]Department of Crop and Soil Sciences, Oregon State University, Corvallis, OR 97331

Formatted for submission to *Molecular Phylogenetics and Evolution*

2.1 Abstract

The molecular phylogenetic trees of *Frankia* and actinorhizal plants were reconstructed using *nif*H, 16S rDNA, and *rbc*L nucleotide sequences. The comparison of *Frankia* phylogenetic trees reconstructed using *nif*H and 16S rDNA sequences indicated that subgroupings of both trees correspond with each other in terms of plant origins of *Frankia* strains. The results suggested that 16S rDNAs can be utilized for coevolution analysis of actinorhizal symbioses. *Frankia* and plant phylogenetic trees reconstructed using 16S rDNA and *rbc*L sequences were compared. The comparison indicated that although branching orders of both trees do not correspond with each other, subgroupings of *Frankia* and their host plants correspond with each other in terms of symbiotic partnership. Tree matching, estimated divergence times among *Frankia* and plant clades, and molecular clock hypothesis tests indicated that *Frankia* clades diverged more recently than plant clades. Taken together, actinorhizal symbioses originated more than three times after the four plant clades diverged.

2.2 Introduction

Actinorhizae are symbiotic associations between angiosperm roots and *Frankia* (Torrey and Tjepkema, 1979). Actinorhizal symbioses fix dinitrogen. The significance of the actinorhizal symbioses in the global nitrogen cycle is primarily ecological; actinorhizal plants are usually early successional species.

The first attempts to classify members of the genus *Frankia* using bacterial isolates were based on infectivity groups. Baker (1987) grouped *Frankia* strains into the following four infectivity groups by using pure cultures in cross-inoculation tests: strains that infect *Alnus* and *Myrica* species, strains that infect *Casuarina* and *Myrica* species, strains that infect *Elaeagnus* and *Myrica* species, and strains that infect only *Elaeagnus* species. Lalonde et al. (1988) used a more complex approach in which they relied on diverse phenotypic characteristics. However, these studies used a limited diversity of *Frankia*; most of *Frankia* strains came from only four of eight actinorhizal plant families. It is also questionable whether cross-inoculation tests reflect host-specificity under normal conditions in nature because strains are not subject to competition. The observation that autoregulation of nodules leads to arrest of symbiotic interactions also casts doubt on the reliability of inoculation tests. Nevertheless, Maggia and Bousquet (1994) suggested some degree of coevolution between the divergence of host plants and their promiscuity toward *Frankia*. The actinorhizal taxa that diverged more recently in this group of plants were shown to be susceptible to a narrower spectrum of *Frankia*, whereas earlier diverging ancestors were highly promiscuous, indicating that evolution has proceeded toward narrower promiscuity and greater specialization. Thus, host promiscuity is likely to be a key determinant in the establishment of an effective symbiosis.

Molecular phylogenetic analyses of nitrogen-fixing symbioses suggest closer relationships among these plants than have been indicated by traditional classification (Soltis et al., 1995; Swensen et al., 1994). Swensen (1996) combined morphological and anatomical data with an analysis of *rbc*L gene sequences of 17 genera of the total 24

representative actinorhizal plant genera. Her results suggested that actinorhizal symbioses have arisen at least four times in different lineages of angiosperms and experienced many loss events during the course of angiosperm evolution. However, her analysis did not consider molecular phylogeny of *Frankia*.

Many previous studies have shown that *Frankia* is monophyletic. Based on molecular phylogeny using full-length of 16S rDNA, Normand et al. (1996) suggested that *Frankia* is monophyletic and is emended to a new family. The phylogenetic analyses of 16S rDNA sequences suggest that *Frankia* form a well-defined, coherent cluster and the genus *Frankia* is further subdivided into four clusters.

Taken together, phylogenetic analyses of both actinorhizal plants and *Frankia* suggested that they are monophyletic, in other words, they have single origins. The next question is how many times have they interacted with each other to form symbiotic relationship and how closely have they interacted to maintain mutualistic associations after establishing symbioses. In a recent report (Jeong et al., 1997), we discussed the coevolutionary relationship between actinorhizal plants and *Frankia* on the basis of reported molecular phylogenetic analyses. In it, we suggested that actinorhizal plants and *Frankia* coevolved at the above familial level of plants. The objective of this study is to rigorously compare the phylogenetic trees of plant and *Frankia* using *rbc*L and 16S rDNA nucleotide sequences.

A complication in molecular phylogenetic studies is unintentional comparisons of paralogous rather than orthologous descendants, and even an occasional horizontal gene transfer. The well-understood *Rhizobium* phylogeny using 16S rDNA sequences does not show host-microbe coevolution compared to the plant phylogenetic tree (Young and Johnston, 1989; Oyaizu et al., 1993), however nitrogen-fixing genes, *nif* and *nod* genes, which are encoded on its *Sym* plasmid show some degree of coevolutionary relationship (Ueda et al., 1995). The observation supports frequent lateral gene transfer among organisms of genes encoded by plasmid.

Nitrogen-fixing genes in *Frankia* are, however, encoded on the chromosome. Therefore, we evaluated the relationship between 16S rDNA and *nif*H gene. The sequences of 16S rDNA and *nif* genes were aligned and phylogenies were reconstructed. After we made sure that 16S rDNA and nitrogen-fixing genes have evolved in a similar

evolutionary time scale, we took advantage of *Frankia* phylogenetic trees from 16S rDNA to compare with phylogenetic tree of actinorhizal plants using *rbc*L.

2.3 Methods

The analyzed sequences were selected from GenBank. All available *rbc*L sequences from actinorhizal plant genera were used to reconstruct actinorhizal plant phylogeny, as shown in Table 2.1. In cases where more than one *rbc*L sequence per genus was available, we used only one representative sequence. Several additional non-actinorhizal *rbc*L sequences that were very similar to the actinorhizal *rbc*L sequence were included in our data set; however, we minimized the use of those *rbc*L sequences while constraining our tree to the same topology as the previously reported trees (Soltis et al., 1995; Swensen, 1996). This sequence selection strategy allowed us to focus on gain events of actinorhizal symbioses rather than loss events. In fact, previous studies yielded complicated phylogenetic trees because too many sequences were used, although the phylogenies are perhaps more accurate.

Frankia strains are divided into typical and atypical strains in terms of their host infectivity and nitrogen-fixing ability (Baker, 1987; Myrold, 1994). Typical strains infect their host plant and form nitrogen-fixing nodules. Atypical strains have lost the ability to fix nitrogen but retain the ability to associate with the host plants as co-symbionts or as contaminants in the surface layers of the nodule epidermis where they escape disinfection treatments. Only *Frankia nif*H and 16S rDNA sequences from typical strains were used (Table 2.2). We also included *nif*H and 16S rDNA gene sequences amplified directly from uncultured microsymbiotic *Frankia* within root nodules. We considered that the directly amplified sequences from nodules represented natural *Frankia* populations that infect their host plants. Additionally, we sequenced 16S rDNA genes amplified directly from *Purshia tridentata, Ceanothus americanus,* and *C. thyrsiflorus* nodules.

Sequences were aligned by using ASSEMBLE CONTIGS option within the Genetic Data Environment, version 2.0, sequence analysis software (provided by Steven Smith, Millipore Corporation, Marlborough, MA). Adjustments were manually made to maximize sequence alignments. Indel-containing regions and ambiguous bases (regions

Table 2.1. Sources of *rbc*L sequences. Actinorhizal families and species in bold

Family	species	GenBank accession number
Betulaceae	***Alnus crispa***	X56618
	Betula nigra	L01889
	Corylus cornuta	X56619
Begoniaceae	*Begonia oxyloba*	U59815
Casuarinaceae	***Allocasuarina verticillata***	X69527
	Casuarina litorea	L01893
	Gymnostoma webbianum	X69531
Coriariaceae	***Coriaria myrtifolia***	L01897
Cucurbitaceae	*Cucubita pepo*	L21938
Datiscaceae	***Datisca cannabina***	L21939
	Tetrameles nudiflora	L21943
Elaeagnaceae	***Elaeagnus angustifolia***	U17038
	Hippophaë salicifolia	U17037
	Shepherdia canadensis	U17039
Fagaceae	*Castanea sativa*	M94936
Juglandaceae	*Juglans nigra*	U00437
Moringaceae	*Moringa oleifera*	L11359
Myricaceae	***Comptonia peregrina***	X69529
	Myrica gale	L01934
Nothofagaceae	*Nothofagus dombeyi*	L13350
Rhamnaceae	***Colletia ulicina***	U59819
	Discaria chacaye	U59826
	Ceanothus sanguineus	U06795
	Rhamnus cartharticus	L13189
	Trevoa trinervis	U59828
Rosaceae	***Cercocarpus ledifolius***	U06796
	Cowania stansburiana	U59817
	Dryas drummondi	U06825
	Neillia sinensis	U59828
	Purshia tridentata	U06821
	Rosa woodsii	U06824
Urticaceae	*Pilea pumila*	U00438
Ulmaceae	*Parasponia rigida*	U59820
Vitaceae	*Vitis aestiyalis*	L01960
Platanaceae	*Platanus oxidentalis*	L01943

Table 2.2. Sources of 16S rDNA and *nif*H seqences. Unless noted, all sequences are full length

Names of species or strains	GenBank accession number	
	16S rDNA	*nif*H
ACN14a (*Alnus crispa* infective)	M88466	
ArI3 (*Alnus* infective)		L41344
ArI4 (*Alnus rugosa* infective)	L11307	
AgKG4-84 (*Alnus glutinosa* infective)	L18976	
ARgP5 (*Alnus* infective)	L40612	
ACoN24d (*Alnus* infective)	L40610	
AVN17s (*Alnus* infective)	L40613	
CeD (*Casuarina eqisetifolia* infective)	M55343	
Ea1-28 (*Elaeagnus* infective)	L40618	
EuiK1 (*Elaeagnus umbellata* infective)		U53362
FaC1 (*Alnus* infective)		U53363
HR27-14 (*Hipphphaë* infective)	L40617	
HRN18a (*Alnus* infective)		X17522
SCN10a (*Sepherdia* infective)	L40619	
Alnus.ru-micro (Unisolated *Alnus rugosa* microsmybiont)	L40956	
Myrica-micro (Unisolated *Myrica* microsymbiont)	L40622	
Coriaria-micro (Unisolated *Coriaria nepalensis* microsymbiont)	L18981[a]	X76399[a]
Datisca-micro (Unisolated *Datisca cannabina* microsymbiont)	L18979[a]	X76398[a]
Dryas-micro (Unisolated Dryas microsymbiont)	L40616	
Purshia-micro (Unisolated *Purshia tridentata* microsymbiont)	This study[b]	
Ceano.ca-micro (Unisolated *Ceanothus caeruleus* microsymbiont)		U78306[a]
Ceano.th-micro (Unisolated *Ceanothus thyrsiflorus* microsymbiont)	This study[b]	
Ceano.am-micro (Unisolated *Ceanothus americanus* microsymbiont)	This study[b]	
Catenuloplanes cauruleus	X93202	
Geodermatophilus obscurus	L40621	
Alcaligenes faecalis		X96609
Rhizobium ORS571		M16709

[a] Only partial sequence available.

[b] A nodule of *P. tridentata,* was collected by Alison Berry (University of California, Davis) at Mesa Verde National Park, CO; that of *C. thyrsiforus* by Nancy Ritchie (Oregon State University) at Bear Creek Recreational Area, OR; *C. americanus* by Beth Mullin (University of Tennessee) at Oak Ridge, TN. 16S rDNAs of microsymbionts from the nodules were directly amplified using polymerase chain reaction and the PCR products were directly sequenced by an automatic sequencer.

containing an insertion in one sequence or a deletion in another sequence) were corrected. Phylogenetic trees were constructed by the Neighbor-joining method with the Kimura two-parameter model for nucleotide change and a transition-transversion ratio 2.0. We also inferred parsimony trees with the heuristic search option of PAUP 3.0s+4(beta) (Swofford, 1991). A bootstrap confidence analysis was performed with 100 replicates to determine the robustness of branches in both Neighbor-joining and parsimony trees. We edited phylogenetic trees with the program TREEVIEW (Page, 1996). The *rbc*L and 16S rDNA data were used to calculate divergence times (Li and Graur, 1991). We estimated pairwise distance between taxa using PAUP 3.0s+4(beta) (Swofford, 1991).

We compared the Neighbor-joining trees of the plant and *Franka* using Tree Mapping option of COMPONENT 2.0 (Page, 1994). For comparison, non-actinorhizal plant taxa of the plant tree were pruned except *Vitis* as an outgroup. All outgroups of the *Frankia* tree were pruned. In this study, tree mapping makes changes in the plant tree by duplicating branches until differences between the plant and *Frankia* trees are reconciled. A reconciled tree represents the most-parsimonious interpretation of the data subject to the constraint of no horizontal transmission. This method provides two measures of fit between host and associate trees. "Leaves added" is the difference between the number of nodes in the *Frankia* and reconciled tree, and "losses" is the number of absences of *Frankia* strains where they are predicted to occur on the reconciled tree. Both parameters decrease with increasing similarities of plant and *Frankia* trees. The significance of the observed parameters was evaluated by reconciling the *Frankia* tree with 1000 binary rooted trees randomly generated using a Markovian model (Harding, 1971; Savage, 1983) with the same number of taxa as the host.

We examined molecular clock hypothesis using likelihood ratio test (LRT). The LRT statistic for comparing two hypotheses (Λ) is defined (see Huelsenbeck and Rannala (1997) for a review) as

$$\Lambda = \frac{\max[L_o(\text{null hypothesis} \mid \text{data})]}{\max[L_1(\text{alternative hypothesis} \mid \text{data})]}$$

The likelihood L is maximized under both the null and alternative hypothesis. The null hypothesis for the molecular clock hypothesis assumes that the rates among lineages are equal. If $\Lambda > 1$, the null hypothesis is favored; the alternative hypothesis is favored if $\Lambda < 1$. In case of test of nested hypotheses, Λ will always be <1 and -2 log Λ is approximately χ^2 distributed under the null hypothesis, with q degrees of freedom, which is the difference in the number of free parameters between the null and alternative hypotheses.

2.4 Results

2.4.1 Phylogenetic analysis of nifH and 16S rDNA sequences

The first phase of our study entailed the evaluation of relationship between 16S rDNA and nitrogen-fixing genes of *Frankia*, because nitrogen-fixing ability is an essential feature of actinorhizal symbioses. We reconstructed a phylogenetic tree by the Neighbor-joining method using published partial *nif*H nucleotide sequences, as shown in Fig. 2.1. We have also reconstructed a phylogenetic tree by the Neighbor-joining method using 16S rDNA sequences (Fig. 2.2). The distance matrix and phylogenetic tree of 16S rDNA indicated that *Frankia* strains are divided mainly into three monophyletic clades. Clade I corresponds with Coriariceae-, Datiscaceae-, Rhamnaceae- and Rosaceae-microsymbiont *Frankia*; Clade II with Betulaceae-, Casuarinaceae-, and Myricaceae-infective *Frankia*; Clade III with Elaeagnaceae-infective *Frankia*. Subgroups of *Frankia* in the phylogenetic trees were well supported by bootstrap values higher than 75. The tree topology obtained with the Neighbor-joining method is in accord with the phylogenetic tree reported by Normand et al. (1996). Parsimony trees exhibited essentially the same topology with the exception that Clade II and Clade III sequences are grouped together with bootstraping values less than 30, while Clade II sequences are themselves grouped together with bootstrap values of 88. The 16S rDNA sequences of *Elaeagnus*-infective *Frankia* have a particular hypervariable region between *E. coli* coordinates 900 and 1200 (Brosius et al., 1987) different from the rest of *Frankia* 16S rDNA sequences (Normand

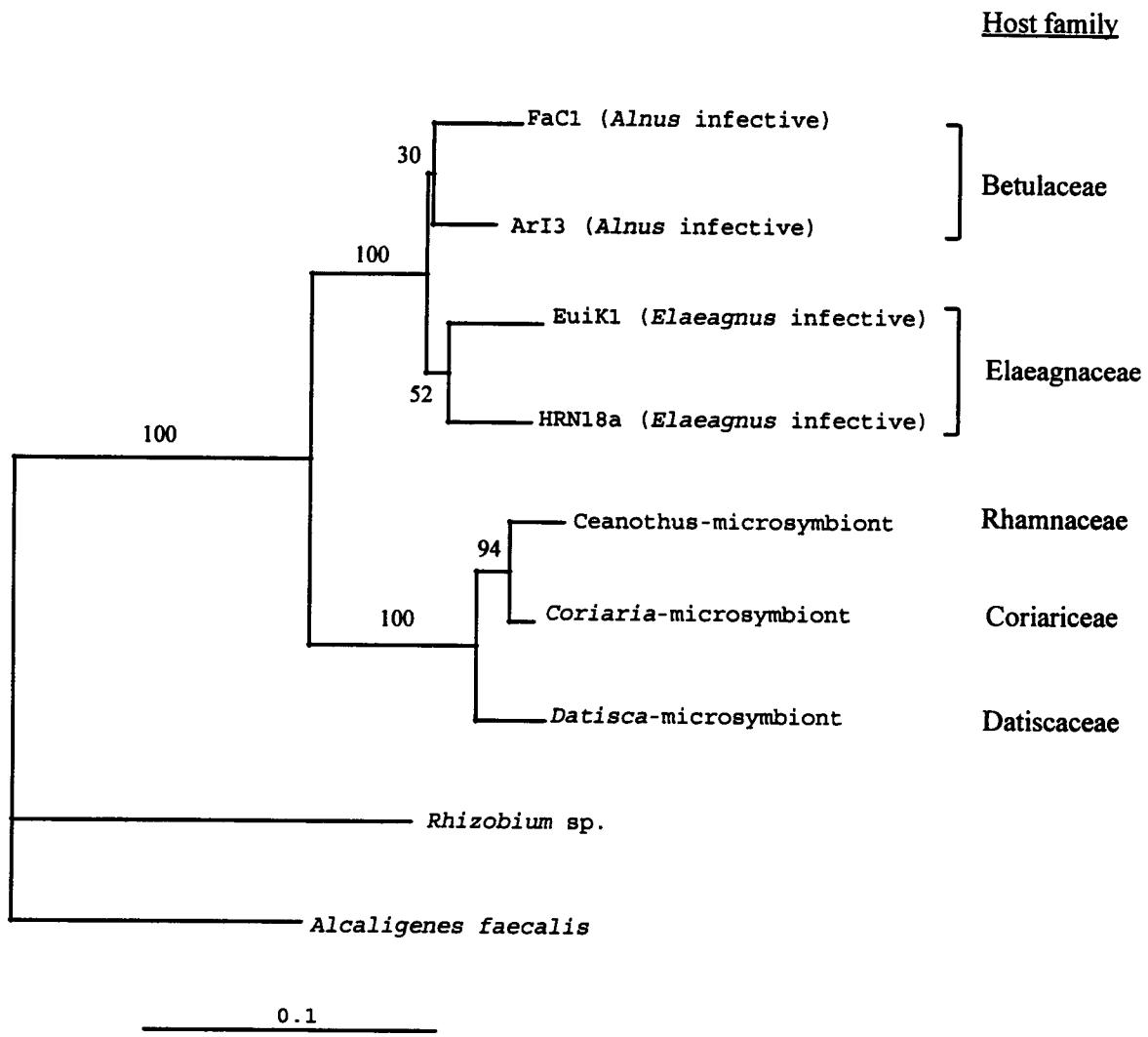


Fig. 2.1. Phylogenetic tree of *Frankia* strains obtained by the Neighbor-joining method of *nif*H genes. The bootstrap values of 100 resamplings are indicated.

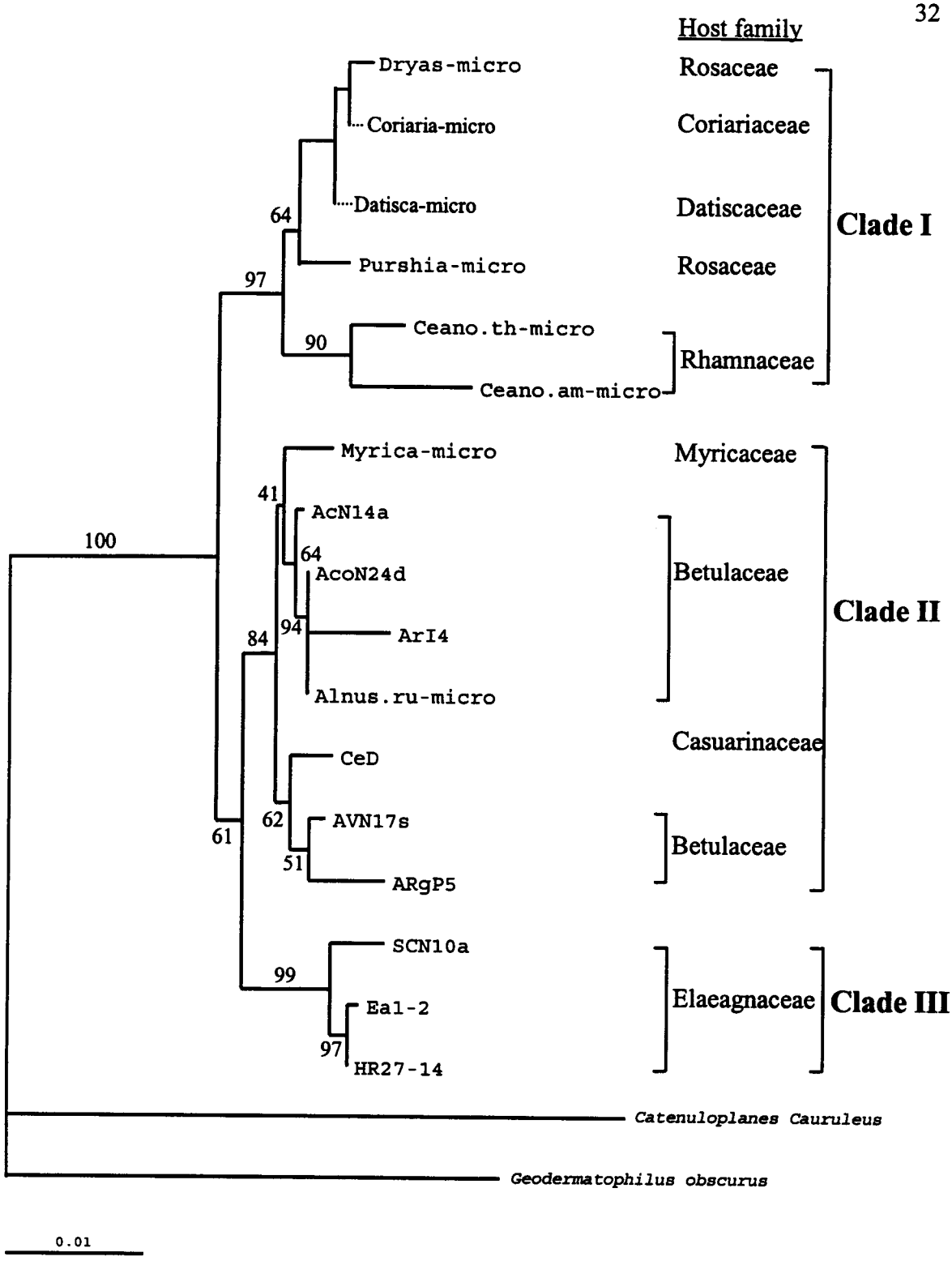


Fig. 2.2. Phylogenetic tree of *Frankia* strains obtained by the Neighbor-joining method of 16S rDNAs. The bootstrap values of 100 resamplings are indicated.

et al., 1996). Thus, we considered the clade II and III *Frankia* as separate monophyletic sister groups, as we found in the Neighbor-joining tree.

Although data for only seven *nif*H sequences were available at the moment, we found that the topology of an evolutionary tree of *nif*H sequences was quite similar to that of 16S rDNA sequences in terms of plant origins of *Frankia* strains. This analysis suggests that 16S rRNA and *nif*H genes of *Frankia* have evolved together in more tightly linked state relative to the legume-*Rhizobium* symbiosis, in which the topology of the evolutionary tree of 16S rRNA is quite dissimilar to that of nitrogen-fixing genes (Oyazu et al., 1993). Whether the lateral gene transfer of *nif* genes has occurred in *Frankia* will be revealed by more *nif*H gene sequences from diverse *Frankia*. Thus, the result also supports that 16S rDNA sequences could be used for evaluating the evolutionary relationship between *Frankia* and actinorhizal plants.

2.4.2 Phylogenetic analysis of rbcL sequences

The *rbc*L sequence alignment has one gap as the result of three ambiguous bases of the *Northofagus rbc*L sequence. A phylogenetic tree of representative actinorhizal plants obtained by the Neighbor-joining method is shown in Fig. 2.3. Actinorhizal plants are grouped into four clades in accordance with Swensen's report (1996). The four clades are well-supported by high bootstrap values. Clade 1 corresponds with actinorhizal plant families Betulaceae, Casuarinaceae, and Myricaceae; Clade 2 with Coriariceae and Datiscaceae; Clade 3 with Rhamnaceae and Elaeagnaceae; Clade 4 with Rosaceae. A parsimony tree exhibited the same topology with a few branching order differences. Subgroups of actinorhizal plants in phylogenetic trees were well supported by high bootstrap values.

2.4.3 Comparison of phylogenetic trees for Frankia and actinorhizal plants

Viewed together phylogenetic trees for *Frankia* and actinorhizal plants reconstructed using 16S rDNA and *rbc*L sequences, subgroups of both plant and *Frankia* trees show partial correspondence in terms of symbiotic partnerships. Although the concordance of tree topology is consistent with the hypothesis of cospeciation, the concordant phylogenies might instead result from dispersal, extinction, or incomplete sampling of

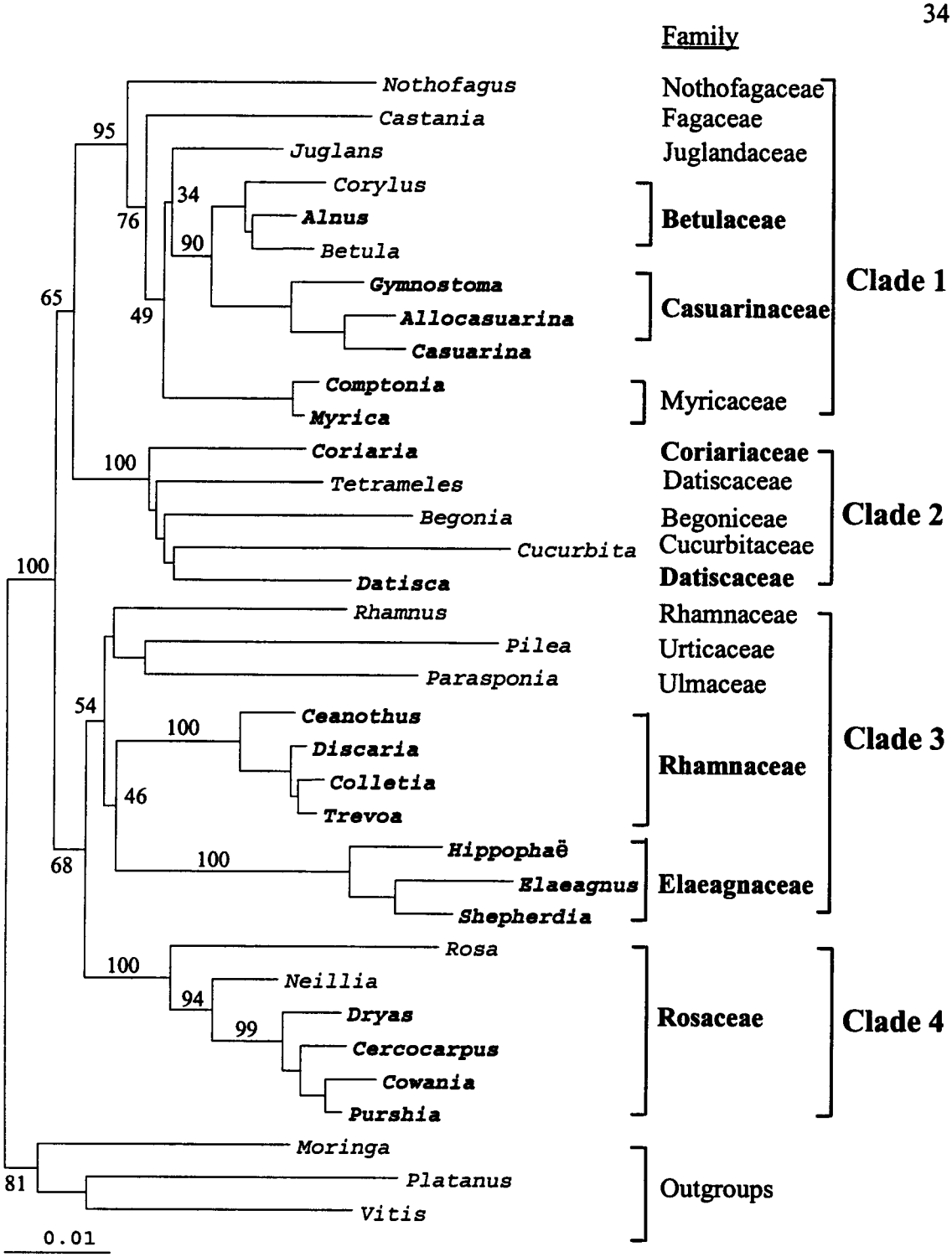


Fig. 2.3. Phylogenetic tree of representative actinorhizal plant taxa obtained via the Neighbor-joining method of *rbc*L sequence data. The bootstrap values of 100 resamplings are indicated. Actinorhizal families and genera are in bold

closely related taxa. Only the cospeciation hypothesis predicts temporal congruence of plant and *Frankia* speciation events, which would result in a significant relation between measures of molecular differentiation in the plant and *Frankia* trees. Thus, we rigorously tested our cospeciation hypothesis using tree mapping and molecular clock hypothesis.

The COMPONENT program determined that the fit between observed *Frankia* and plant trees was significantly better than the fit between the *Frankia* and plant tree drawn at random from the set of all possible plant trees (Fig. 2.4). The observed degree of fit between the *Frankia* and plant trees was significantly better ("leaves added," $P(H_0) = 0.003$; "losses," $P(H_0) = 0.004$) than the fit between the *Frankia* and 1000 randomized plant trees. These results, which are robust with respect to the method of phylogenetic inference and to the evolutionary models used, falsify the null hypothesis of chance similarity between the plant and *Frankia* trees. A duplicated node reflects a speciation event in the *Frankia* that happened prior to the differentiation of the corresponding host. However, in this study, it is unclear whether duplicated nodes reflect independent speciation of *Frankia* or codivergence of *Frankia* and plant with one exception, because plant tree is essentially phylogeny above genus levels. *Alnus*-infective *Frankia* appear to have speciated independent of plant speciation before codivergence events between *Alnus*- and *Myrica*- or *Casuarina*-infective *Frankia*. Thus, our reconciled tree show eight events of cospeciation and hypothesizes three corresponding subgroups of plants and *Frankia*.

The molecular clock hypothesis is satisfied if the mean rate of substitution has remained constant among lineages and DNA substitutions follow a Poisson process. The log likelihood calculated under the clock hypothesis is $\log L_o = -6535.67$ for the plants (*rbc*L sequence from *Hippophaë* were now considered for the test because of ambiguous sequence of 71 bp) and $\log L_o = -2584.86$ for the *Frankia* when a simple model of DNA substitution is used. A more general model assumes that each branch of the phylogenetic tree has a unique unconstrained rate of substitution. This introduces $s - 2$ additional parameters; the likelihood for this latter model is therefore higher than that under the molecular clock ($\log L_1 = -6482.06$ for the plants and $\log L_1 = -2567.26$ for the *Frankia*). Because models are nested and the phylogenetic tree is held constant, the statistic $-2 \log \Lambda$ can be compared with a χ^2 distribution with $s - 2$ degrees of freedom to

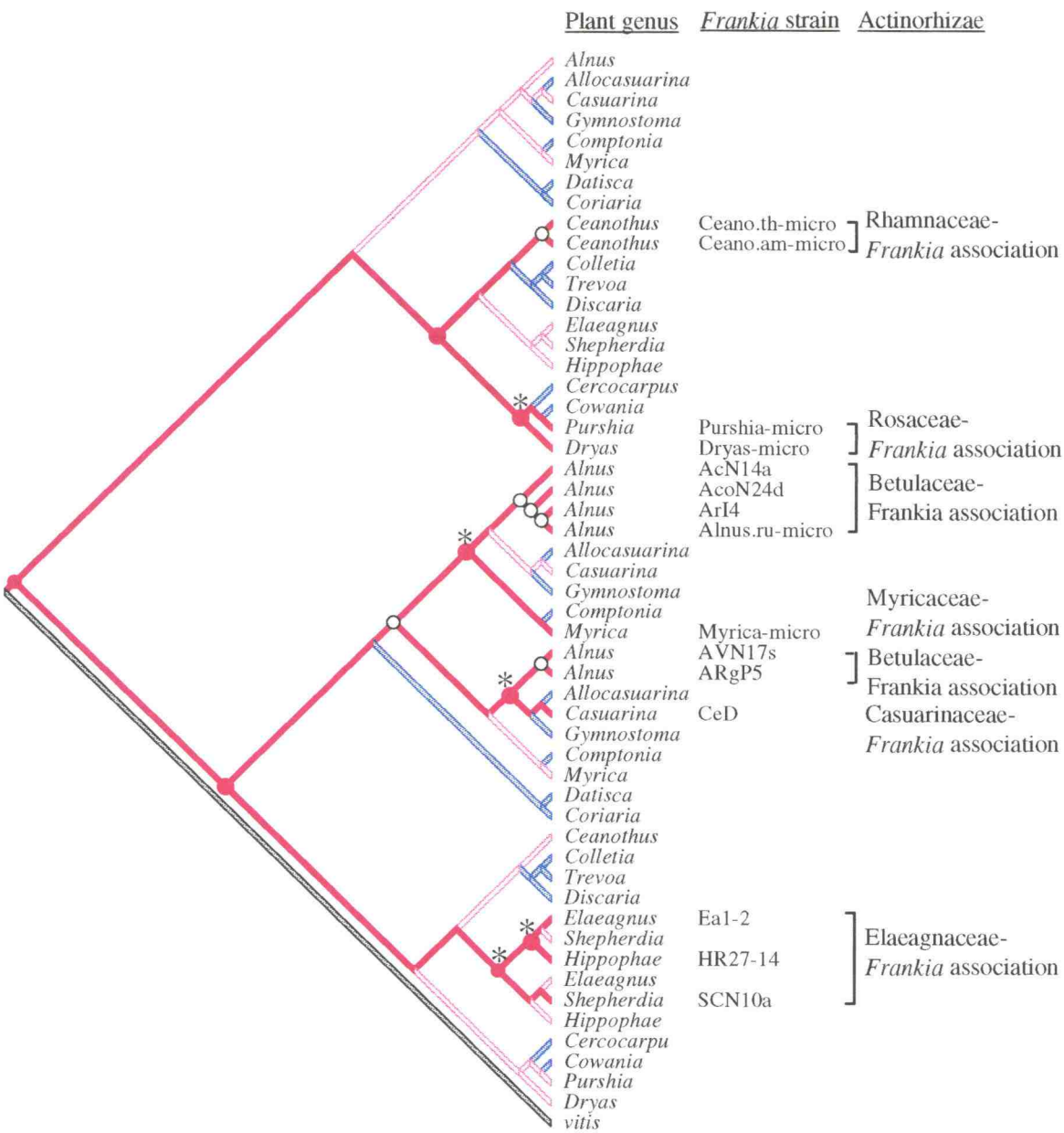


Fig. 2.4. A reconciled trees for actinorhizal plants and their symbiotic partners, *Frankia*. This reconciled tree has eight codivergent (cospeciation) events (●). ○ = duplications; red solid branches = presence of *Frankia*; pink hollow branches = inferred absence of *Frankia*; blue shaded branches = actinorhizal plants that miss 16S rDNA sequence data of *Frankia*. Asterisks indicate nodes used for comparison of divergence times within subgroups.

determine the significance of the test. In this case, the molecular clock hypothesis can be rejected for either the plants ($P < 0.001$) or the *Frankia* ($P < 0.001$), suggesting that actinorhizal symbioses do not have single origin.

Because tree mapping identified three corresponding clades of *Frankia* and plant trees, we also tested molecular clock hypothesis for each clade. For Clade I of *Frankia,* $\log L_o$ and $\log L_1$ were –2073.75 and –2069.57; for Clade II of *Frankia*, they were –2072.74 and –2060.52. For Clade 1 of the plant tree $\log L_o$ and $\log L_1$ were –3190.26 and –3169.83; for Clade 2, –2831.90 and -2827.29; for Clade 3, –3312.21 and –3300.47; for Clade 4, –2492.20 and –2487.27. When *Nothofagus* from Clade 1 and *Rhamnus, Pilea*, and *Parasponia* from Clade 3, which are not actinorhizal plant taxa but are distantly grouped with actinorhizal plant lineages within each Clade were not considered for the test, $\log L_o$ and $\log L_1$ are -2951.93 and -2943.18 for Clade 1 and -2529.84 and -2526.39 for Clade 3. Therefore, the molecular clock hypothesis cannot be rejected for both the subgroups of the actinorhizal plants and the *Frankia* ($P > 0.05$) and it appears that plants and *Frankia* have interacted several times with each other to form symbiotic relationship.

We estimated approximate divergence times (Table 2.3) among the three clades of *Frankia* and among four clades of plants using average pairwise distances estimated between taxa . The mean calculated divergence rate of *rbc*L estimated by Albert et al (1994) is approximately $2.05 \pm 0.75 \times 10^{-10}$ total substitutions per site per year for a time range potentially spanning 500 - 5 million years before present; that of 16S rDNA estimated by Moran et al (1993) is 2 to 4×10^{-10} per year. We used these mean rates to estimate the divergence time (T)

$$T = S/\lambda$$

where S is the estimated average pairwise distance among taxon groups and λ is the mean base substitution rate. The estimated divergence time among *Frankia* clades ranged from 39 to 125 million years; for plant clades it ranged from 170 to 429 million years ago. Although the maximum approximate value of plant clade divergence times is unrealistically high when we consider the origin of angiosperms 415 million years before present, these calculations indicate that plant clades diverged earlier than *Frankia* clades.

Table 2.3. The estimated divergence times[a] among the three clades of *Frankia* and among four clades of actinorhizal plants[b]

	Clade I / Clade 1	Clade II / Clade 2	Clade III / Clade 3	Clade 4
Clade I / Clade 1		181 - 389	199 - 429	191 - 413
Clade II / Clade 2	53 - 107		192 - 414	180 - 388
Clade III / Clade 3	62 - 125	39 - 78		170 - 367
Clade 4				

[a] Million years

[b] Below diagonal: divergence times of *Frankia* and above diagonal: divergence times of actinorhizal plants

Not only does the cospeciation hypothesis predict similar patterns of speciation in plant-*Frankia* assemblages, but it also predicts that the speciation events were causally related and, therefore, approximately contemporaneous. Thus, we can test the cospeciation hypothesis from another perspective if DNA substitution evolves in a roughly clock-like fashion. Hafner and Nadler (1990) proposed a framework for comparing relative timing of cospeciation events using cytochrome oxidase I DNA sequences of gophers and their lice, given molecular clocks (which may tick at different rates) in both groups. A line fitted to a plot of parasite divergence versus host divergence simultaneously describes two aspects of host-parasite divergence (Hafner and Page, 1995). The slope of the line is an estimate of the relative rate of genetic change in the two groups. The y-intercept of the line measures genetic divergence in the parasites at the time of host speciation. For example, an intercept of zero indicates synchronous cospeciation, wherein hosts and parasites speciate simultaneously. Here, we adopted this method for comparing divergence times of plant and *Frankia*. Because the plant and *Frankia* phylogenies were reconstructed using different molecular sequences, we used divergence times estimated using pairwise distances between corresponding taxa rather than nucleotide substitutions for comparisons. We used simple regression analysis to quantify the relation between plant and *Frankia* divergence times. Comparisons were restricted to *Frankia* whose phylogenetic history is topologically identical to that of their hosts. Because most of the uncertainty in the phylogenetic analyses involved branches near the base of the trees, only terminal and subterminal branches were compared between plants and *Frankia* (indicated in Fig. 2.4). The slope of the regression was 0.31 ($r = 0.73$; $R^2 = 54$; $P = 0.006$), which indicates that the rate of evolutionary change is approximately three times greater in plants than in *Frankia*. The estimated intercept (–8 million years), which is not significantly different from zero, indicated that plants and *Frankia* have cospeciated synchronously.

2.5 Discussion

The phylogenies of representative *Frankia* strains and actinorhizal plants were reconstructed using 16S rRNA and *rbc*L nucleotide sequences as shown in Fig. 2.2 and Fig 2.3 and then compared. The comparison revealed partial correspondence of grouping at the higher taxonomic level. In other words, *Frankia* strains that originated from closely related plant taxa are grouped together. For example, *Frankia* isolated from *Alnus, Myrica*, and *Casuarina,* which are grouped together in the plant phylogenetic tree, belong to the same cluster. *Frankia* isolated from *Elaeagnus,* which belong to another phylogenetic group, form a separate sister cluster of Clade II. 16S rDNA sequences directly amplified and sequenced from *Ceanothus, Coriaria, Datisca, Dryas*, and *Purshia* (Clade I), which belong to Clade 2, 3, and 4 of the plant phylogenetic tree, form yet another cluster. Grouping of *Datisca* and *Coriaria* microsymbionts was assigned using a phylogenetic tree that was reconstructed using a sequence alignment corresponding with partial 16S rDNA sequences of *Datisca* and *Coriaria*-microsymbiont. Thus, full-length 16S rDNA sequences of these taxa may show different groupings in future.

Relationship between Clade 1 plants and their microsymbiotic *Frankia* (Clade II) have been analyzed using *rbc*L gene sequences and *Frankia* cross-inoculation tests (Maggia and Bousquet, 1994). The comparison showed that plants originating early have a broader range of *Frankia* cross-inoculation than plants originating later. As expected, branching orders of Clade I plants and their microsymbiont *Frankia* do not correspond with each other. Our analysis indicates that coevolutionary analysis by Maggia and Bousquet (1994) covered only one-third of *Frankia* diversity. Our analysis also pointed out that more studies should be conducted on *Frankia* that infect Rhamnaceae and Rosaceae, which are recalcitrant to *Frankia* isolation. Comparison of the two phylogenetic trees does not show exact correspondence between *Frankia* and actinorhizal plants above the familial level. Nowhere do branching topologies of both phylogenetic trees correspond to each other. This suggests that actinorhizal symbioses originated at least three times before separation of present actinorhizal plant families. Loss of symbiotic ability occurred many times according to Swensen's (1996) plant tree.

Together, the results support multiple, rather than single, origins of actinorhizal symbioses. However, a few gain and many loss events complicate rigorous analysis of evolutionary origin using phylogenetic analysis.

Although concurrence of tree topologies is necessary to guide us to the correct answer, concurrence is not sufficient to firmly support coevolution. If hosts and symbionts are actually speciating in synchrony, then according to the molecular clock hypothesis we should expect to see roughly equivalent amounts of nucleotide change in associated host and symbiont lineages following speciation events (represented by nodes on phylogenetic trees). Therefore, to be consistent with the strict cospeciation model, host and symbiont phylogenies must agree not only in branching pattern, but also in branch lengths. Divergence time can be estimated using mean substitution rates obtained from comparison of 16S rDNA and *rbc*L nucleotide sequences on the basis of fossil record, geographic separation, etc. (Ochman and Wilson, 1987; Moran et al., 1993; Albert et al., 1994). Normand et al. (1993) considered divergence time to take into account the diversity of *Frankia*. They compared the estimated divergence times of *Frankia* subgroups with the appearance of plant families obtained from fossil records. Although their calculation shows some degree of concurrence of evolutionary timing, it was inconclusive regarding coevolutionary relationships. The plant phylogenetic tree based on *rbc*L sequences indicated that actinorhizal plants are well grouped into four clades (Fig 2.3). Thus, first we considered divergence times of the four plant clades rather than the appearance time of an individual actinorhizal plant family. We also estimated divergence times of the three *Frankia* clades. The estimated divergence times indicate that clades of *Frankia* diverged more recently than clades of plants (Table 2.5). Furthermore, corresponding subgroups identified by tree mapping not only fulfill the molecular clock hypothesis but also support cospeciation hypothesis within subgroups. These results indicate that actinorhizal symbioses originated several times long after plant clades diverged. Together with tree topology comparison, our results also indicated that once symbioses established, plants and *Frankia* were retained within certain taxonomic group with limited lateral transfer.

Thus far, most molecular phylogenetic studies of *Frankia* have used only 16S rDNA sequences. However, it is important to use other sequences at the same time, because

conflicting molecular phylogenies could arise if lateral transfer among *Frankia* occurs. Lateral transfer of nitrogen-fixing genes is well-documented in rhizobia studies. In many cases nitrogen-fixing genes in rhizobia are encoded on *Sym* plasmids. The phylogenetic tree for NodD protein does not correspond with 16S rRNA phylogeny, suggesting lateral gene transfer of *Sym* plasmids among rhizobia strains. Oyaizu et al. (1993) also suggested lateral gene transfer, because the phylogeny of 16S rRNA of rhizobia does not relate to the host specificity of rhizobia. Interestingly, Ueda et al. (1995) reported that the topology of the evolutionary tree of NodC protein is similar to that of leghemoglobin in host plants. Coevolution might have occurred in plant-plasmid interaction. Moreover, the phylogeny of 16S rRNA molecules of rhizobia does not correspond with that of host plants. Nitrogen-fixing genes in *Frankia* have been detected on the chromosome. In contrast to *Rhizobium,* however, *Bradyrhizobium* and *Azorhizobium* have nodulation genes encoded on the chromosome and tend to have a rather broad host range (Fisher and Long, 1992). *Frankia* also exhibits a broad host range. These observations may explain lower host specificity of *Frankia* and lack of correspondence of *Frankia* and plant phylogenetic trees at lower taxonomic levels. Comparison of the phylogenetic trees of 16S rDNA and *nif*H gene sequences suggests that both genes have evolved in a relatively tightly-linked state. Thus, the phylogenetic tree of 16S rDNA sequences may reveal origins of actinorhizal symbiosis. However, it would be worthwhile to reconstruct a phylogenetic tree using more extensive *nif* genes to pinpoint coevolutionary origins. A recent report dealt with *Frankia* phylogeny using the hypervariable *nif*DK intergenic spacer (Navarro et al., 1997). The report suggested that *Gymnostoma* microsymbiont are phylogenetically close to *Elaeagnus*-infective *Frankia* strains. However, they compared such a short 16S rDNA (268 bp; in general, full-length sequence of a 16S rDNA is about 1400 bp) with 33 other rDNA sequences of *Frankia* that the result may not reveal true phylogenetic tree. Thus, the possibility of lateral transfer of nitrogen-fixing genes between microsymbiotic *Frankia* of *Gymnostoma* and *Elaeagnus* is still an open question, as our study of the phylogenetic tree of *nif*H genes suggests.

In general the degree of sequence conservation is critical to resolving the phylogenetic problem at hand. It is reported that a partial sequence of 16S rRNA cannot be used for identification of rhizobia at the species level because of the high degree of

conservation of the molecule. Indeed, comparison of phylogenetic tree reconstructed by Benson et al. (1996) and Murry et al. (1997) using different regions of 16S rDNA partial sequences show different topologies. Yet *Frankia* from distant plant groups are distantly grouped together. Normand et al. (1996) demonstrated why they got different trees. They showed that the three *Elaeagnus*-infective *Frankia* strains, which form the well-defined cluster, have most of their mismatches between *E. coli* 16S rRNA coordinates 974 and 1267. Most mismatches of most other pairs of sequences occurred between coordinates 83 and 182.

Consequently the direct comparison of phylogenetic trees using representative *Frankia* and actinorhizal plants indicated some degree of coevolution. The phylogenetic analysis of additional 16S rDNA sequences and other nitrogen-fixing genes will be needed in order to elucidate the coevolution in more detail. Such an analysis should clarify coevolutionary relationship between *Frankia* and actinorhizal plants. The observation of *Frankia* phylogenetic trees using partial 16S rDNA sequences apparently does not imply coevolutionary relationship at the species level of host plants (Benson et al., 1996; Murry et al., 1997). Thus, the analysis should address the question of at which plant taxonomic level the actinorhizal symbiosis has coevolved. The evidence for multiple symbiotic origins indicates a genetic predisposition toward root nodule symbiosis in nodulating species and their close relatives. This assumption contradicts the previous widespread idea that the origin of nitrogen-fixing plant-microbe symbiosis is chaotic (Sprent, 1994; Bousquet and Lalonde, 1990). Therefore, this molecular evolutionary study of the *Frankia*-actinorhizal symbiosis may provide insight into the evolution of symbiosis. The outcome may be valuable for the effective development and function of symbiosis in land management and reforestation efficiency of disturbed site.

2.6 References

Albert V. A., Backlund, A., Bremer K., Chase M. W., Manhart, J. R., Mishler, B. D., and Nixon, K. C. (1994). Functional constraints and *rbc*L evidence for land plant phylogeny. *Ann. Mo. Bot. Gard.* 81: 534-567.

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

Benson, D. R., Stephens, D. W., Clawson, M. L., and Silvester, W. B. (1996). Amplification of 16S rRNA genes from *Frankia* strains in root nodules of *Ceanothus griseus*, *Coriaria arborea*, *Coriaria plumosa*, *Discaria toumatou*, and *Purshia tridentata*. *Appl. Environ. Microbiol.* 62: 2904-2909.

Bousquet, J., and Lanonde, M, (1990). The genetics of actinorhizal Betulaceae. *In* "The Biolgy of *Frankia* and Actinorhizal Plants" (C. R. Schwintzer, J. D. Tjepkema, Eds.), pp. 236-261. Academic Press, Inc., SanDiego.

Brosius, J,, Palmer, M. L., Kennedy, J. P., and Noller, H. F. (1987). Complete nucleotide sequence of a 16S ribosomal RNA gene from *E. coli*. *Proc. Natl. Acad. Sci. USA* **75:** 4801-4805.

Fisher, R. F., and Long, S. R. (1992). *Rhizobium*-plant signal exchange. *Nature* (London) **357:** 655-660.

Hafner, M. S., and Nadler, S. A. (1990). Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation. *Syst. Zool.* **39:** 192-204.

Hafner, M. S., and Page, R. D. M. (1995). Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Phil Trans R Soc* Lond B **349:** 77-83.

Hafner, M. S., Sudman, P. D., Villabalnca F. X., Spradling, T. A., Demastes, J. W., Nadler, S. A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265:** 1087-1090.

Harding, E. F. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Prob.* **3:** 44-77.

Huelsenbeck, J. P., and Rannala, B. (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276:** 227-232.

Jeong, S.-C., Liston, L., and Myrold, D. D. (1997). Molecular phylogeny of the genus *Ceanothus* using *ndh*F and *rbc*L sequences. *Theor. Appl. Genet.* **94:** 852-857.

Lalonde, M., Simon, L., Bousquet, J., and Seguin, A. (1988). Advances in the taxonomy of *Frankia*: recognition of species *alni* and *eleagni* and novel subspecies *pommerii* and *vandijkii*. *In* "Nitrogen Fixation: Hundred Years After" (H. Bothe, F. J. de Bruijn, W. E. Newton, Eds.),. pp. 671-680. Gustav Fisher Verlag, Stutgart.

Li, W.-H., and Graur, D. (1991). "Fundamentals of molecular evolution," Sinauer, Sunderland.

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

Moran, N. A., Munson, M. A., Bauman P., and Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B Biol. Sci.* **253:** 167-171.

Murry, M. A., Konopka, A. S., Pratt, S. D., and Vandergon, T. L. (1997). The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus*. *Physiol. Plant.* (In press)

Myrold, D. D. (1994). *Frankia* and the actinorhizal symbiosis. *In* "Methods of Soil Analysis: Part 2. Microbiological and Biochemical properties" R. W. Weaver, V. S. Angle, P. J. Bottomley, Eds.), pp 291-329. Soil Science Society of America, Madison.

Navarro, E., Nalin, R., Gauthier, D., and Normand, P. (1997). The nodular microsymbionts of *Gymnostoma* spp. Are *Elaeagnus*-infective *Frankia* strains. *Appl. Environ. Microbiol.* **63:** 1610-1616.

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

Ochman, H., and Wilson, A. C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26:** 74-86.

Oyaizu, H., Matsumoto, S., Minamisawa, K., and Gamou, T. (1993). Dostribution of rhizobia in leguminous plants surveyed by phylogenetic identification. *J. Gen. Appl. Microbiol.* **39:** 339-354.

Page, R. D. M. (1994). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* **43:** 58-77.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12:** 357-358.

Savage, H. M. (1983). The shape of evolution: Systematic tree topology. *Biol. J. Linn. Soc.* 20: 225-244.

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowd, J. M., and Martin, P. G. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. USA* **92:** 2647-2651.

Sprent, J. I. (1994). Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? *Plant Soil* **161:** 1-10.

Swensen, S. M. (1996). The evolution of actinorhizal symbiosis: Evidence for multiple origins of the symbiotic association. *Am. J. Bot.* **83:** 1503-1512.

Swensen, S. M., Mullin, B. C., and Chase, M. W. (1994). Phylogenetic affinities of Datiscaceae based on an analysis of nucleotide sequences from the plastid *rbc*L gene. *Syst. Bot.* **19:** 157-168.

Swofford, D. L. (1991). "PAUP: phylogenetic analysis using parsimony, version 3.0" Illinois National History Survey, Champaign.

Torrey, J. G., and Tjepkema, J. D. (1979). Preface. *Bot. Gaz. (Suppl.):* i-ii.

Ueda, T., Suga, Y., Yahoo, N., and Matsuguchi, T. (1995). Phylogeny of sym plasmids of rhizobia by PCR-based sequencing of a *nod*C segment. *J. Bacteriol.* **177:** 468-472.

Young, J. P., and Johnston, A. W. B. (1989). The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol. Evol.* **4:** 341-349.

Chapter 3
Molecular Phylogeny of the Genus *Ceanothus* (Rhamnaceae) Using *rbc*L and *ndh*F Sequences

S.-C. Jeong[1], A. Liston[2], D. D. Myrold[3]
[1,3]Plant Physiology Program and [3]Department of Crop and Soil Science, Oregon State University, Corvallis, OR 97331
[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331

Published in *Theoretical and Applied Genetics* 94:852-857

3.1 Abstract

Intrageneric phylogeny among ten representative *Ceanothus* species was investigated using DNA sequences of the chloroplast encoded *ndh*F and *rbc*L genes. Parsimony analysis of the *ndh*F sequences identified two main clades corresponding to two subgenera *Ceanothus* and *Cerastes*. The phylogenetic results suggest that three monophyletic clades within the subgenus *Ceanothus* can be delimited on the basis of 1) evergreen or 2) deciduous leaves and 3) thorn presence within the evergreen clade. The estimated divergence time based on *rbc*L sequences suggests that the two subgenera diverged 18 - 39 million years ago whereas species within each subgenus diverged more recently. Taken together, the results support the division of *Ceanothus* into two monophyletic subgenera and are consistent with the postulated recent divergence of many species within each subgenus.

3.2 Introduction

Ceanothus (Rhamnaceae), a genus of 55 species of shrubs or rarely small trees (McMinn 1942), is restricted to North America with a center of diversity in California. This genus is grouped into two subgenera (Schmidt, 1993): *Ceanothus* (33 species) and *Cerastes* (22 species). Forty-five species (over 80 %) occur on the Pacific coast of North America; however, several other species occur widely in other areas of the continent (e.g., *C. americanus* and *C. fendleri*). Species of the subgenus *Ceanothus* and all of the species of the subgenus *Cerastes* are often important components of a broad leaf sclerophyllous vegetation.

Ceanothus is an important genus as cultivated ornamentals, forage plants, and plants for site amelioration and erosion prevention. Thirty-one species of the genus have been reported to be associated with nitrogen-fixing bacteria of the genus *Frankia* (Becking 1977). Host plants that form endosymbiotic nodules with *Frankia* are called actinorhizal plants (Torrey and Tjepkema 1979). The nitrogen-fixing ability of *Ceanothus* is a particularly important ecophysiological trait. *Ceanothus* species contribute substantially to the nitrogen status of soil in northwest North America (Conard et al. 1985).

The most recent comprehensive monograph of *Ceanothus* is that of McMinn (1942) and here we follow his taxonomic treatment. In the subgenus *Ceanothus,* the species have alternate leaves, hornless fruit, thin early-falling stipules, and stomata located on the lower surfaces of the leaves. Species of subgenus *Cerastes* are characterized by the presence of opposite leaves, horned fruits, persistent corky stipules, and stomata located in sunken pits on the underside of the leaves. Among these characters, the thick leaves with sunken stomatal pits and persistent stipules exist in all members of subgenus *Cerastes*. Most *Ceanothus* species have evergreen leaves with the exception of six species of the subgenus *Ceanothus*. Seven species within this subgenus have thorny twigs, which often occur in members of the Rhamnaceae.

Fossil records also support the distinctiveness of the two *Ceanothus* subgenera. *Ceanothus* has been present at least since the Oligocene (Chaney 1927; Raven and Axelrod 1978). The distinctive leaf forms of both subgenera, *Ceanothus* and *Cerastes,* are

occasionally found in Miocene fossil floras, suggesting that genetic barriers between the two sections may already have been developed at this time (Nobs 1963).

Although *Ceanothus* species can be conveniently resolved by morphological characters, karyological and hybridization studies indicate that species boundaries are not clear. Chromosome counts of 17 species and three varieties of the subgenus *Ceanothus* together with 17 species and five varieties of the subgenus *Cerastes* investigated by Nobs (1963) have a diploid number 2n = 24. Hybrid plants were found to have the same number of chromosomes as their parents, with one exception that had 12 pairs of normal-size chromosomes, one very small pair, and three fragments. All reduction divisions examined appeared regular. Therefore, he suggested that no evident cytological barrier prevented hybridization. Field observation and hybridization studies (Nobs 1963) also indicated that species within each subgenus are interfertile, but the species of the two subgenera do not hybridize.

We are interested in elucidating plant-*Frankia* symbiotic relationship. From experiences of our previous *Alnus*-infective *Frankia* studies (Crannell et al. 1994; Huss-Danell and Myrold 1994) we selected the genus *Ceanothus* as model plants because of their reticulate relationship among species, high species diversity within relatively small areas, and their ecophysiological importance. No rigorous phylogenetic analysis of *Ceanothus* has been reported using molecular markers. The objective of this study was to investigate the phylogenetic relationship among representative *Ceanothus* species (Table 3.1) using the chloroplast genes, *ndh*F and *rbc*L.

Several candidate chloroplast genes have been suggested for phylogenetic studies at lower taxonomic levels (Olmstead and Palmer 1994). *ndh*F has recently been demonstrated to be useful for phylogenetic studies at intrafamilial levels compared with the *rbc*L gene which has been widely used at higher taxonomic levels (Olmstead and Sweere 1994; Kim and Jansen 1995). The utility of sequence change in the *rbc*L gene has been recently evaluated for the estimation of divergence time between woody-taxon pairs (Albert et al. 1994). In the present study, we use *ndh*F sequences to reconstruct phylogenetic relationship and *rbc*L sequences to estimate divergence times.

3.3 Materials and methods

3.3.1 Plant materials

Ten species of *Ceanothus* (Tables 3.1 and 3.2) represent the taxonomic diversity of the plant genus, its geographical range, and a habitat gradient of environmental and soil conditions. *Rhamnus purshiana* (Rhamnaceae) was used as an outgroup (Table 3.2). *Nicotiana tabacum* (Solanaceae) and *Barnadesia caryophylla* (Asteraceae) were also included as distant outgroups. All collections by S.-C. Jeong, G. V. Johnson, B. C. Mullin, and N. J. Ritchie are deposited as vouchers in the herbarium of Oregon State University (OSC).

3.3.2 DNA Sequencing

Total DNA was isolated from leaves following the protocol of Doyle and Doyle (1987). When needed, DNA was further purified by polyethylene glycol precipitation.

We utilized the polymerase chain reaction (PCR) to amplify *ndh*F and *rbc*L nucleotide sequences from all taxa listed in Table 3.2. DNA sequences for the two chloroplast genes were determined following similar procedures. DNA sequencing was carried out using a set of internal primers. The primers described by Olmstead and Sweere (1994) were used for the amplification and internal sequencing of *ndh*F, with the modification of five internal primers specified in Table 3.3. Primers for amplification of *rbc*L and the internal sequencing primers were provided by G. Zurawski (DNAX Research Institute, Palo Alto, CA), with the exception of C331F and C1022R (Table 3.3). The Z1 and C1022R primer combinations were used to amplify *rbc*L. The *ndh*F sequences of *Ceanothus* were amplified in two overlapping segments, using primers 1F-1318R and C803F-2110R. For *R. purshiana*, two different primer combinations were used (C538F-1318R and C803F-2110R) because of problems with 1F, the 5′ forward primer.

Ten microliters of each PCR reaction was run on a 0.8% agarose gel to check the quality of amplification. To purify the resulting DNA, 40 µl of each PCR reaction was run on a 1.0% TBE agarose gel. Agarose blocks containing the DNA were excised from the gel with a scalpel over UV light, and purified with the GeneClean II gel-extraction kit

Table 3.1. Characteristics of representative *Ceanothus* species used in this study

Ceanothus species[a]	Distinguishing characteristics	
	Plant features[b]	Distribution
Subgenus *Ceanothus*		
C. americanus	Deciduous, W, leaves L and broadly ovate to oblong-ovate	Eastern U.S.
C. sanguineus	Deciduous, W, leaves L and broadly ovate but not orbicular	Pacific coast, northern Rocky Mountains
C. integerrimus	Semidecidous, W or B, leaves M/ L and elliptical but never approaching orbicular	Pacific coast, southern Rocky Mountains
C. cordulatus	Evergreen, thorny, W, leaves M and not two-times longer than wide	Pacific coast
C. fendleri	Evergreen, thorny, W, leaves M and two-times longer than wide	Southern Rocky Mountains
C. thyrsiflorus	Evergreen, B, leaves M/L, branchlets flexible	Pacific coast
C. velutinus	Evergreen, W, leaves L and glutinous above	Pacific coast, northern Rocky Mountains
Subgenus *Cerastes*		
C. cuneatus	Erect, W, leaves S and typically entire	Pacific coast
C. prostratus	Prostrate, B, leaves S and usually broader	Pacific coast
C. pumilus	Prostrate, B, leaves M and narrowly oblanceolate	Pacific coast

[a]*Ceanothus* is divided into two subgenera: *Ceanothus* and *Cerastes*

[b]W - white flower, B - blue flower, S - leaves less than 10 mm long or broad, M - leaves 10 to 25 mm long or broad, L - leaves more than 25 mm long or broad

Table 3.2. Sources of taxa sampled for DNA sequences and previously published sequences

Taxon	DNA source/voucher	Genbank accession number	
		*rbc*L	*Ndh*F
C. americanus L.	Tennessee, Anderson Co., 3 miles from Oak Ridge/ B. C. Mullin 01.		U78893
C. cordulatus Kell	Oregon, Douglas Co., 100 yards from Toketee Ranger Station/ N. J. Ritchie 004.	U78904	U78894
C. fendleri Gray	New Mexico/ G. V. Johnson 01.		U78895
C. integerrimus H. & A.	Oregon, Lane Co., 0.5 mile west of Blue River, Hwy. 126 /N. J. Ritchie 003.		U78896
C. sanguineus Pursh	Oregon, Lane Co., H.J. Andrews Experimental Forest, Unit L/ N. J. Ritchie 002.	U06795 (Morgan et al. 1994)	U78897
C. thyrsiflorus Esch.	Oregon, Douglas Co., Bear Creek Recreational area, Hwy. 42/ N. J. Ritchie 005.		U78898
C. velutinus Dougl.	Oregon, Benton Co., 5.8 miles from Hwy. 34 on Marys Peak Rd./N. J. Ritchie 001.		U78899
C. cuneatus (Hook.) Nutt.	Oregon, Benton Co., 0.6 miles north of Adair Village, Hwy. 99W/ N. J. Ritchie 008.		U78900
C. prostratus Benth.	Oregon, Douglas Co., 6.5 miles west of Diamond Lake/ N. J. Ritchie 006.		U78901
C. pumilus Greene	Oregon, Coos Co., Powers Ranger District: FS 5325, S. FS 530/ N. J. Ritchie 007.	U78905	U78902
R. purshiana DC.	Oregon, Benton Co., Jackson-Frazier Wetland of Corvallis/ S.-C. Jeong 001.		U78903
R. carthartica L.		G13189 (Chase et al. 1993)	
N. tabacum L.			L14953 (Olmstead and Sweere 1994)
B. caryophylla Blake			L39394 (Kim and Jansen 1995)

Table 3.3. Oligonucleotides designed for PCR amplification and DNA sequencing

Gene	Primer	5′ - 3′ nucleotide sequence
*ndh*F	C538F	GTAACTAATCGTGTAGGGGATT
*ndh*F	C538R	AATCCCCTACACGATTAGTTAC
*ndh*F	C803F	CTATGGTAGCAGCCGGAATTTTTC
*ndh*F	C803R	GAAAAATTCCGGCTGCTACCATAG
*ndh*F	C1601F	TATCCGCAGGAATCGGACAATACTAT
*rbc*L	C331F	TCTACGTAGTAAATCAACAAAGCCTAAA
*rbc*L	C1022R	ATCACGTAGTAATAAATCAACAAAGCCTAAA

according to the manufacturer's protocol (BIO 101, Vista, Calif.). Final pellets were dissolved in 22 µl water and DNA amounts measured using a fluorometer. The purified double-stranded DNA products were sequenced directly by automatic sequencing with dye terminator extension in the Center for Gene Research and Biotechnology, Oregon State University.

Eleven *ndh*F sequences, corresponding to ten taxa and one outgroup, and two *rbc*L sequences were generated. Sequence alignment was performed using the PILEUP program of GCG (Genetics Computer Group, Madison, Wis.). No gaps (indels) were observed in the *ndh*F sequences of *Ceanothus* species. Two gaps (indels) in the *ndh*F sequence of *R. purshiana* were observed. There were no missing cells in the data matrix except for the gaps and missing sequences caused by outgroups. In all cases where sequences were ambiguous, due to compression or otherwise, samples were sequenced in both directions. All sequences determined as part of this study have been submitted to GenBank (Table 3.2).

3.3.3 Phylogenetic Analysis

The *rbc*L data were used to calculate divergence times (Li and Graur 1991). Maximum-parsimony analyses of the *ndh*F data were conducted using PAUP 3.0s + 4 (beta) (Swofford 1991), with all changes weighted equally. Gaps caused by outgroups were treated as missing data. We used the branch-and-bound algorithm with the delayed transformation option. Analyses of the complete data set were performed rooting the tree with *R. purshiana*, *B. caryophylla*, and *N. tabacum*. To evaluate relative levels of support for individual clades, the bootstrap method was used with 100 replicates of a heuristic search.

3.4 Results

A total of 2122 bp of *ndh*F from ten *Ceanothus* species and 993 bp of *rbc*L from two *Ceanothus* species were sequenced. As an outgroup, 1616 bp of the *ndh*F of *R.*

purshiana was sequenced due to difficulty in PCR amplification of the gene. No length variation was observed for *rbc*L sequences. No length variation was observed for *ndh*F nucleotide sequences among *Ceanothus* species. In the sequenced 1616 bp of *ndh*F from *R. purshiana*, we found one deletion (9 bp) and one insertion (39 bp), relative to published sequences (Olmstead and Sweere 1994; Clark et al. 1995; Kim and Jansen 1995). Interestingly, primer pair 1F and 803R for the amplification of *ndh*F nucleotides of *R. purshiana* produced a single PCR product of approximately 800 bp. However, the sequencing of the PCR product gave a partial *ndh*H nucleotide sequence, suggesting that both primers have high homology to the *ndh*H nucleotide sequence.

The *rbc*L sequences allowed us to estimate divergence times between the *Ceanothus* subgenera *Ceanothus* and *Cerastes* and between *Rhamnus* and *Ceanothus* pairs using the mean calculated divergence rate estimated by Albert et al. (1994). Based on patristic distances between woody taxon pairs from Search II of Chase et al. (1993) and the divergence time assumption, they obtained a mean divergence rate of approximately 2.05 $\pm$ 0.72 x 10^{-10} total substitutions per site per year. Estimated divergence time of the two *Ceanothus* subgenera is 18 - 39 million years ago; for *Rhamnus* and *Ceanothus* it is 147 - 317 million years ago. The *rbc*L sequences of *C. cordulatus* and *C. sanguineus,* which belong to the same subgenus, showed 100% similarity.

Pairwise distances between taxa are shown in Table 3.4. There were 1 - 3 base differences within the subgenus *Cerastes* (0.0 - 0.1%) and 1 - 11 base differences within the subgenus *Ceanothus* (0.0 - 0.5%). A total of 20 - 25 (0.9 - 1.1%) base differences were observed between the two subgenera. In the data set 30 of the 2122 bp were variable, 11 of which represented third-codon-position substitutions. Seven were due to first-position substitutions, whereas the remaining 12 were second-position substitutions. One percent of the nucleotides were potentially phylogenetically informative.

Parsimony analyses of the 13 *ndh*F sequences by the branch-and-bound algorithm yielded one most parsimonious tree (Fig. 3.1), with a length of 567 steps, a consistency index (CI) of 0.947, homoplasy (HI) of 0.053, a CI excluding uninformative characters of 0.878, a HI excluding uninformative characters of 0.122, a retention index of 0.888, and re-scaled CI of 0.841. Bootstrap analysis of the nucleotide-sequence data yielded a topology that was exactly congruent with the branch-and-bound analysis.

Table 3.4. Pairwise distances of *ndh*F sequences among ten *Ceanothus* species[a]

	1	2	3	4	5	6	7	8	9	10
1. *C. cuneatus*	-	0.000	0.001	0.011	0.010	0.010	0.010	0.010	0.011	0.011
2. *C. pumilus*	1	-	0.001	0.010	0.009	0.009	0.010	0.010	0.010	0.011
3. *C. prostratus*	3	2	-	0.011	0.010	0.010	0.011	0.011	0.011	0.012
4. *C. americanus*	23	22	24	-	0.002	0.004	0.004	0.004	0.005	0.005
5. *C. sanguineus*	21	20	22	4	-	0.003	0.003	0.003	0.004	0.004
6. *C. integerrimus*	21	20	22	8	6	-	0.000	0.000	0.002	0.002
7. *C. thyrsiflorus*	22	21	23	9	7	1	-	0.001	0.002	0.003
8. *C. velutinus*	22	21	23	9	7	1	2	-	0.002	0.003
9. *C. fendleri*	23	22	24	10	8	4	5	5	-	0.002
10. *C. cordulatus*	24	23	25	11	9	5	6	6	5	-

[a] Numbers the above diagonal indicate percent distance and below the diagonal is the number of base differences between species

Fig. 3.1. Single most parsimonious tree constructed by using the branch-and-bound algorithm with maximum parsimony of 13 *ndh*F sequences of *Ceanothus* species plus outgroup taxa. Percentages based on 100 bootstrap replicates are shown above the branches; branch lengths are shown below. Column A-E are a key to the distribution of selected morphological characters of *Ceanothus* species: A--flower color (W-white, B-blue); B--leaf type (D-deciduous, E-evergreen); C--thorny twigs (T-thorny, N-non-thorny); D--growth form (P-prostrate, E-erect); E--leaf size (S-<10 mm long or broad, M-10 to 25 mm long or broad, L-> 25 mm long or broad).

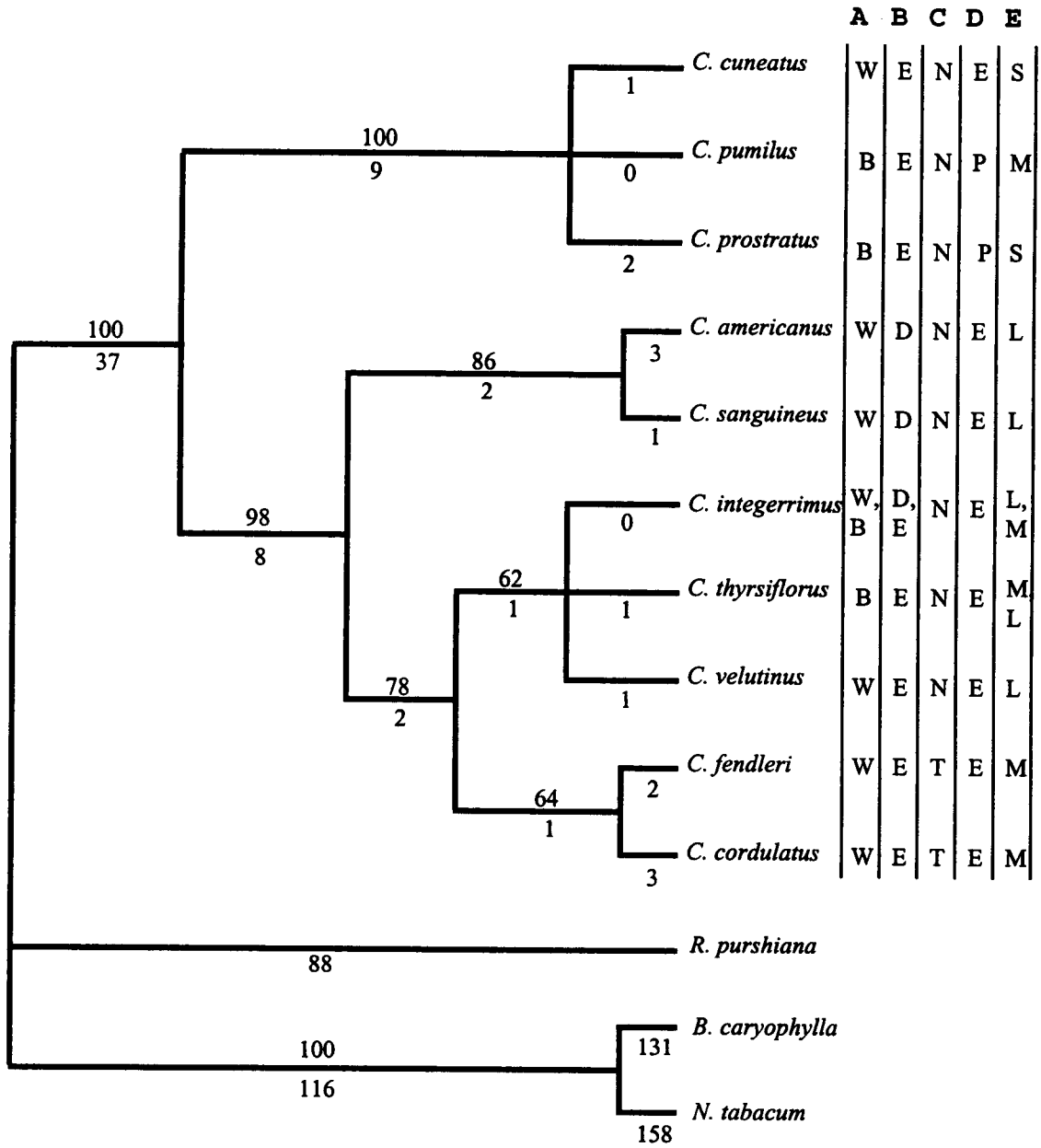

A B C D E
C. cuneatus W E N E S
C. pumilus B E N P M
C. prostratus B E N P S
C. americanus W D N E L
C. sanguineus W D N E L
C. integerrimus W, B D, E N E L, M
C. thyrsiflorus B E N E M, L
C. velutinus W E N E L
C. fendleri W E T E M
C. cordulatus W E T E M
R. purshiana
B. caryophylla
N. tabacum
100
9
1
0
2
100
37
86
2
3
1
98
8
0
62
1
1
1
78
2
2
64
1
3
88
100
116
131
158

Fig. 3.1

3.5 Discussion

The results presented here suggest that *Ceanothus* is well divided into two subgenera, in accord with the traditional classification system of the genus. McMinn (1942) divided the subgenus *Ceanothus* into two groups and the subgenus *Cerastes* into three groups on the basis of flower color. Our phylogenetic analysis indicates that flower color varies within putatively monophyletic clades. In contrast, our analysis hypothesizes three monophyletic clades corresponding to vegetative characters within the subgenus *Ceanothus* (Fig. 3.1). The three monophyletic clades correspond to (1) evergreen or (2) deciduous leaves and (3) thorn presence within the evergreen clade. An exception is *Ceanothus integerrimus*, which has semideciduous leaves (Table 3.1). The cladogram topology suggests that deciduous leaves in this species are a derived feature (Fig. 3.1). This polymorphism for leaf persistence is apparently a response to water availability, and not a result of hybridization (McMinn 1942).

The phylogenetic tree (Fig. 3.1) is consistent with low levels of genetic differentiation (Table 3.4) among representative western North America *Ceanothus* species. *C. integerrimus*, *C. thyrsiflorus*, and *C. velutinus*, which belong to the same clade, are sympatric in southern Oregon and northern California (Conard et al. 1985). *C. cuneatus*, *C. pumilus*, and *C. prostratus*, which belong to another clade, are distributed together in the southern Cascade Range and in the Sierra Nevada of southern Oregon and northern California. Although *C. cordulatus* and *C. fendleri*, which do not have a similar distribution pattern, belong to the same clade, they show substantial *ndh*F sequence divergence.

Albert et al. (1994) suggested that the *rbc*L sequence is quasi-ultrametric in terms of a clock assumption, implying that the extent of nucleotide substitution in a given taxon should roughly reflect the underlying cladogenetic time. Divergence time calculated from the *rbc*L nucleotide sequence suggested that the two *Ceanothus* subgenera separated in a comparable time to woody taxon pairs (see Albert et al. 1994), but species within a subgenus originated more recently. Low sequence divergency of *ndh*F and complete

identity of the *C. sanguineus* and *C. cordulatus* partial *rbc*L nucleotide sequences correspond with previous hybridization and karyological studies (Nobs 1963), indicating a high genetic similarity of many *Ceanothus* species.

We used the sequence information to make oligonucleotides for sequencing the *ndh*F gene. However, we had to modify several internal primers proposed by Olmstead and Sweere (1994) on the basis of the partial sequences obtained because of sequence differences between *Ceaonthus* and tobacco (Table 3.3). Therefore, our revised primer information will be useful for sequencing plant species related to *Ceanothus*. However, we succeeded in sequencing only three-fourths of the *ndh*F sequence of *R. purshiana* This difficulty may come from the lack of homology of primer 1F. The fact that the PCR product from the 1F and C803R primer pair is a part of the *ndh*H nucleotide sequence (unpublished data) points out the necessity for a new design of primer 1F for extensive use of the *ndh*F gene as phylogenetic data. Olmstead and Palmer (1994) suggested several candidate genes to study plant phylogenetic relationships on the species or genus level. One advantage of the *ndh*F gene is that it is long enough to resolve phylogenetic relationships at the species level. Recently, several groups have reported intrageneric analyses using *mat*K and ITS nucleotide sequences (e.g., Soltis et al. 1996; Yuan et al. 1996). Our report demonstrates the utility of the *ndh*F nucleotide sequence to estimate phylogenetic relationship at the intrageneric level.

Recent work on the phylogeny of nitrogen-fixing plants suggests a single evolutionary origin (Soltis et al. 1995), which is in contrast with conventional taxonomy. Actinorhizal plants are found in three of the four subclades that contain nitrogen-fixing symbioses. At this higher taxonomic level there is general congruence between host plant and *Frankia* phylogenies (Hönerlage et al. 1995; Normand et al. 1996), although *Frankia* that nodulate members of the Rhamnaceae are poorly represented in these trees. Two recent studies have grouped *Ceanothus*-infective *Frankia* either with those that nodulate members of the Rosaceae (Benson et al. 1996) or the Elaeagnaceae (Murry et al. 1997). Both of these plant families are in the same subclade of nitrogen-fixing plants; however, *Ceanothus* is much more closely related to members of the Elaeagnaceae (Soltis et al. 1995).

At lower taxonomic levels there is considerably less specificity between plant host and microbial symbiont (Myrold 1994), although the study of Rouvier et al. (1996) showed some degree of specificity between *Frankia* strains and species within the Casuarinaceae. Whether such specificity exists with *Ceanothus*-infective *Frankia* is unknown, although Baker and Mullin (1994) detected no plant population or geographic patterns of *Frankia* within *C. americanus* nodules. We have collected nodules from the same *Ceanothus* populations as the present plant-collection populations and analyzed them by PCR-RFLP (restriction fragment length polymorphism) and DNA sequencing. The preliminary data show that we have uncovered several putative new *Frankia* strains from nine *Ceanothus* species (Ritchie and Myrold, unpublished data). When these data are combined with our present results, they may give an insight into the elucidation of *Frankia*-host plant relationship. Therefore, the present molecular phylogenetic study of *Ceanothus* species provides important data to reveal how *Frankia* and *Ceanothus* have co-evolved on the intrageneric level as well as to clarify systematic relationships among *Ceanothus* species.

3.6 Acknowledgments

The authors thank Nancy J. Ritchie for collecting most plant material in Oregon and helpful technical assistance; Beth C. Mullin and Gordon V. Johnson for providing plant material; G. Zurawski for sequencing primers; David Martin for helpful comments on the manuscript. This research was supported by USDA-NRICGP grant 93-60017860-A4. Contribution from the Oregon Agricultural Experiment Station, Technical paper number 11154.

3.7 References

Albert VA, Backlund A, Bremer K, Chase MW, Manhart JR, Mishler BD, Nixon KC (1994) Functional constraints and *rbc*L evidence for land plant phylogeny. Ann Mo Bot Gard 81:534-567

Baker DD, Mullin BC (1994) Diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus* as assessed by RFLP patterns from single nodule lobes. Soil Biol Biochem 26:547-552.

Becking JH (1977) Dinitrogen-fixing associations in higher plants other than legumes. In Hardy RWF, Silver WS (eds) A treatise on dinitrogen fixation Section 3, Biology. Wiley, New York, pp 185-275

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

Chaney RW (1927) Geology and paleontology of the Crooked River Basin, with special reference to the Bridge Creek flora. Carnegie Inst Wash Publ 346:45-138

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qui Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedrén M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang G-Y, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn Jr GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Mo Bot Gard 80:528-580

Clark LG, Zhang W, Wendel JF (1995) A phylogeny of the grass family (Poaceae) based on *ndh*F sequence data. Systematic Bot 20:436-460

Conard SG, Jaramillo AE, Cromack Jr K, Rose S (1985) The role of the genus *Ceanothus* in western forest ecosystems. USDA For Serv Gen Tech Rep PNW-182

Crannell WK, Tanaka Y, Myrold DD (1994) Calcium and pH interaction on root nodulation of nursery-grown red alder (*Alnus rubra* Bong.) seedlings by *Frankia*. Soil Biol Biochem 26:607-614

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15

Hönerlage W, Hahn D, Zepp K, Zeyer J, Normand P (1995) A hypervariable 23S rRNA region provides a discriminating target for specific characterization of uncultured and cultured *Frankia*. System Appl Microbiol 17:433-443

Huss-Danell K, Myrold DD (1994) Intrageneric variation in nodulation of *Alnus*: Consequences for quantifying *Frankia* nodulation units in soil. Soil Biol Biochem 26:525-531

Kim K-J, Jansen RK (1995) *ndhF* sequence evolution and the major clades in the sunflower family. Proc Natl Acad Sci USA 92:10379-10383

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland, Massachusetts.

McMinn HE (1942) *Ceanothus*: Part II; a systematic study of the genus *Ceanothus*. Santa Barbara Botanic Garden, Santa Barbara, California pp 129-308

Morgan DR, Soltis DE, Robertson KR(1994) Systematic and evolutionary implications of *rbc*L sequence variation in Rosaceae. Am J Bot 81:890-903

Murry MA, Konopka AS, Pratt SD, Vandergon TL (1997) The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus*. Physiol Plant (in press)

Myrold DD (1994) *Frankia* and the actinorhizal symbiosis. In Weaver RW, Angle VS, Bottomley PJ (eds) Methods of soil analysis: part 2. Microbiological and biochemical properties. Soil Science Society of America, Madison, Wiscosin pp 291-328

Nobs MA (1963) Experimental studies on species relationships in *Ceanothus*. Carnegie Inst Wash Publ, Washington D. C.

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

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81:1205-1224

Olmstead RG, Sweere JA (1994) Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Syst Biol 43:467-481

Raven PH, Axelrod DI (1978) Origin and relationships of the California flora. Univ Calif Publ Bot 72:1-134

Rouvier C, Prin Y, Reddell P, Normand P, Simonet P (1996) Genetic diversity among *Frankia* strains nodulating members of the family Casuarinaceae in Australia

revealed by PCR and restriction fragment length polymorphism analysis with crushed root nodules. Appl Environ Microbiol 62:979-985

Schmidt CL (1993) *Ceanothus*. In: Hickman JC (ed) The Jepson manual: higher plants of California. University of California Press, Berkeley, California pp 932-938

Soltis DE, Kuzoff RK, Conti E, Gornall R, Ferguson K (1996) *mat*K and *rbc*L gene sequence data indicate that *Saxifraga* (Saxifragaceae) is polyphyletic. Am J Bot 83:371-382

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci USA 92:2647-2651

Swofford DL (1991) PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois National History Survey, Champaign, Illinois

Torrey JG, Tjepkema JD (1979) Preface. Bot Gaz 140(Suppl):i-ii

Yuan Y-M, Küpfer P, Doyle JJ (1996) Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. Am J Bot 83:641-652

Chapter 4

Specific Genomic Fingerprinting of Microsymbiont *Frankia* from *Ceanothus* Copopulations Using Repetitive Sequences and PCR

Soon-Chun Jeong and David D. Myrold

Formatted for submission to *Applied and Environmental Microbiology*

4.1 Abstract

Specificity between *Ceanothus* species and their microsymbionts, *Frankia,* were investigated with nodules collected from three geographically separated copopulations of *Ceanothus* species. Nodules were analyzed using DNA sequencing and rep-PCR techniques. DNA sequencing of the intergenic spacer region between 16S and 23S rRNA genes suggested that *Ceanothus*-microsymbiont *Frankia* are closely related at the intraspecific level. Diversity of the microsymbionts was further analyzed by genomic fingerprinting using repetitive sequences and PCR. A newly designed direct repeat sequence and a BOX sequence were used as PCR primers after justification that these primers can generate *Frankia*-specific fingerprints from nodule DNA. Further analysis of the nodules using BOX and DR PCR showed that *Ceanothus* microsymbiont *Frankia* exhibit less diversity within each copopulation than among copopulations. These data suggested that geographic separation plays a more important role for divergence of *Ceanothus*-microsymbiont *Frankia* than host plant.

4.2 Introduction

The genus *Frankia* is the bacterial partner of N_2-fixing actinorhizal symbioses. According to the molecular phylogeny of *Frankia* 16S rDNA, *Frankia* is monophyletic and divided into four clades (Normand et al., 1996). Based on 16S rDNA sequencing, considerable genetic diversity has been found among *Alnus*-infective *Frankia,* which are grouped into one clade. In contrast, low genetic diversity has been found among Elaeagnaceae-infective *Frankia* that have been sequenced, although previous studies using other methods and a larger number of strains showed that this group of *Frankia* are much more diverse (see Benson and Silvester (1993) for a review). The 16S rDNA data cannot be expanded to evaluate *Frankia* diversity among *Frankia* strains below the species level because of its low mutation rate, however (Fox et al., 1992). To date, only a few studies have addressed the diversity and host specificity of *Frankia* at the subspecies level (e.g. Baker and Mullin, 1994; Rouvier et al., 1996). Results have been inconclusive, mainly because *Frankia* are recalcitrant to isolation and are slow-growing in pure culture.

Relatively few isolates of *Frankia* from *Ceanothus* species have been reported--six in the second edition of the Catalog of *Frankia* strains (Lechevalier, 1985-86)--and all of these were from the eastern species, *C. americanus*. Wollum et al. (1966) did report isolating several *Streptomyces* species from *C. velutinus*, but these were clearly not *Frankia*. Of the *Frankia* isolates coming from *C. americanus*, none are capable of reinfecting *Ceanothus* (Lechevalier and Ruan, 1984; Lechevalier, 1985-86), although some strains can nodulate species of the Elaeagnaceae and Myricaceae (Baker, 1987). This is interesting because *Ceanothus* has been shown to be infected via intercellular penetration (Liu and Berry, 1991), which is the mode of entry in *Elaeagnus* species as well (Miller and Baker, 1985). The lack of infective *Frankia* isolates from *Ceanothus* has hindered ecological studies of this particular actinorhizal symbiosis, particularly on the *Frankia* side.

Recently several studies have been reported using techniques that circumvent *Frankia* isolation. One example is the use of a strain-specific *nifH* segment to probe nodule DNA extracts (Simonet et al., 1990; Navarro et al., 1997). RFLP analysis of nodule DNA extracts (Baker and Mullin, 1994) has also been used to examine the population structure of *Frankia*, as well as PCR-RFLP (Maggia et al., 1992; Jamann et al., 1993; Rouvier et al, 1996). Here, we studied the diversity of *Frankia* associated with the genus *Ceanothus,* an important actinorhizal shrub in western USA. *Ceanothus* is a typical example of plants recalcitrant to *Frankia* isolation. Two previous studies (Baker and Mullin, 1994; Murry et al., 1997) reported low diversity of *Ceanothus*-microsymbiont *Frankia* relative to *Alnus*-infective *Frankia*. However, host specificity or extent of diversity of the microsymbiont *Frankia* remains an enigma.

The intergenic spacer region (ISR) between 16S and 23S rRNA genes has a demonstrated use as a fast molecular chronometer to measure phylogeny of bacterial strains at the intraspecific level (Garcia-Martines et al., 1996; Dolzani et al., 1994). Most of *Frankia* 16S-23S ISR is presumably nonfunctional DNA (Normand et al., 1992) and should reveal a considerable degree of sequence variation by genetic drift (Bousquet et al., 1995). Rep-PCR using consensus sequence corresponding to BOX elements (Martin et al., 1992) has been utilized to differentiate Gram-positive bacterial strains at the intraspecific level (e.g. Louws et al., 1994; Versalovic et al. 1994; Murry et al., 1995). Additional repetitive sequences, such as the DR (direct repeats) found in high GC *Mycobacterium bovis* (Doran, 1993), may also be useful in fingerprinting high GC bacteria such as *Frankia*.

Conventional taxonomy and molecular phylogeny of *Ceanothus* species divided the genus *Ceanothus* into two subgenera and show that species within each subgenus have reticulate relationships (McMinn, 1942; Jeong et al., 1997). Thirty-three of the 55 *Ceanothus* species have been observed in N_2-fixing association with *Frankia* (Becking, 1977). We were interested in elucidating specificity in the symbiotic association of *Ceanothus* species and *Frankia*. The objective of this study was to elucidate the genotypic diversity of *Frankia* microsymbiotically associated with *Ceanothus* species where *Ceanothus* species grow together. *Frankia* strains in nodules collected from

copopulations of *Ceanothus* species were analyzed by sequencing DNA of the ISR between 16S and 23S rRNA, and by fingerprinting of nodule DNAs using rep-PCR.

4.3 Materials and Methods

4.3.1 Bacterial strains and field-collected nodules

Sources of bacterial strains are listed in Table 4.1. Nodules were collected from three copopulations of *Ceanothus* species: *C. veltinus-C. prostratus* copopulation, *C. velutinus-C. integerrimus* copopulation, and *C. integerrimus-C. sanguineus* copopulation (Table 4.2).

4.3.2 DNA extraction from bacterial cells and plant nodules

A modification of the cetyltrimethylammonium bromide (CTAB) procedure (Murray and Thomson, 1980) was used. For DNA extraction of bacterial cells, a broth culture was centrifuged (27000 *g*) in a microcentrifuge tube to yield a final cell volume of 100 µl. Cells were resuspended in 500 µl TNE (10 mM Tris-Cl, 100 mM NaCl, 1 mM EDTA (pH 8.0)) and centrifuged twice. Cells were homogenized with a motorized pestle grinder, 500 µl of 2× CTAB extraction buffer (CTAB 2%, 100 mM Tris (pH 8.0), 20 mM EDTA, 1.4 M NaCl) was added, and the sample was vortexed briefly.

Field-collected nodule clusters were washed with distilled water to remove soil debris, surface sterilized in 2.5% sodium hypochlorite, rinsed with sterilized distilled water, and stored at -20 °C. For DNA extraction, 1 tip of a nodule lobe was cut, quick-frozen in a dry ice/ethanol bath, homogenized to a fine powder with ceramic mortar and pestle, and 300 µl 2× CTAB extraction buffer was added directly to the mortar. The contents were transferred to a 1.7-ml tube. An additional 300 µl 2x buffer was used to rinse the mortar and added to the first extract. The sample was vortexed briefly.

All samples were then incubated at 65°C for 30-40 min and centrifuged for 10 min at 7000 *g* to remove cell debris. Supernatant was transferred to a new tube. One volume of chloroform:isoamyl alcohol (24:1) was added, mixed gently by inversion for several minutes until an emulsion formed, and centrifuged for 2 min. Using a wide-bore pipette

Table 4.1. Source of bacterial strains, plants, or IGS DNA sequences used in this study

Species or nodule	Source or reference
Frankia AvcI1	Provided by Dwight D. Baker (Panlabs, Bothell, WA)
Frankia CpI1	Provoded by Dwight D. Baker (Panlabs, Bothell, WA)
Frankia EaI12	Provided by Dwight D. Baker (Panlabs, Bothell, WA)
Escherichia clil DH5α	Life Technologies (Grand Island, NY)
Azospirillum brasilense ATCC 29710	Provided by C.-L. Li (Oregon State University)
Mocrococcus lureus	ATCC
Geodermatophilus obscurus ATCC 25078	ATCC
C. velutinus	Grown in laboratory
C. integerrimus	Grown in laboratory
C. sanguineus	Grown in laboratory
C. prostratus	Grown in laboratory
Nicotiana tobaccum	Provided as genomic DNA by David Martin (Oregon State University)

Table 4.2. Sources of *Ceanothus* nodules

Name of copopulation	Site description and nodule collection
C. velutinus-C. integerrimus copopupation	Oregon, Lane Co., H.J. Andrews Experimental Forest, Crossroad site between Road # 720 and Road # 745 (44°09′ N; 122°23′ W). Nodules were collected from side of the road.
C. sanguineus-C. integerrimus copopulation	Oregon, Douglas Co., Umpqua National Forest, NDF 4710; 10 km west of Steamboat (43°20′ N; 122°80′ W) . Nodules were collected from the clear-cut and slash-burned site.
C. velutinus-C. prostratus copopulation	Oregon, Douglas Co., Umpqua National Forest, Clearwater Falls (43°20′ N; 122°15′ W), Hwy. 138. Nodules were collected at three sites on side of Hwy. 138: 1.5 km west, 1.5 km east, and entrance of the Clearwater Falls

tip, the upper aqueous phase was transferred to a new tube. This chloroform:isoamyl extraction was repeated until no visible debris was observed. One volume ice-cold isopropyl alcohol or two volumes ethanol were added, mixed thoroughly by inversion, incubated at -70 °C for >20 min (dry ice bath), centrifuged at 0 °C for 30 min, decanted, and blotted on Kimwipes. The pellet was reuspended with 50 µl 0.1 TE (10 mM Tris (pH 8.0), 0.1 mM EDTA).

The sample was reprecipitated with 1/4 volume 10 M ammonium acetate and one volume isopropyl alcohol at -70 °C for >20 min and centrifuged as before. The pellet was dried under vacuum for more than 1 h. The pellet was resuspended in 40 µl 0.1 TE or H_2O and one volume of PEG/NaCl (20% polyethylene glycol 8000, 2.5 M NaCl) was added. The sample was incubated for 15 min at 37°C and then centrifuged for 3 min. The supernatant was pipetted off carefully and the pellet was washed two times with 80% ethanol. The pellet was dried under vacuum, resuspended in one volume (starting sample volume) 0.1 TE or H_2O, and stored at -20 °C.

DNA was quantified using Hoechst dye 33258 and a DNA fluorometer (Model TKO 100; Hoefer Scientific Instruments, San Francisco, CA) according to the manufacturer's instructions.

4.3.3 PCR amplification and DNA sequencing of the 16S to 23S rDNA intergenic spacer region

Amplification of the 16S to 23S rDNA ISR was performed with primers 16F1646 (GATTGGGACGAAGTCGT) and 23R1609 (ATCGCATGCCTACTACC). Amplifications were in a total volume of 50 µl, including 1.5 µl of template DNA and 0.2 µM each primer and were run for 35 cycles (95 °C for 45 sec, 50 °C for 60 sec, and 72 °C for 90 sec with final extension at 72 °C for 7 min). To check the efficiency of amplification, 10 µl of the PCR products were visualized by electrophoresis on a 0.8% (wt/vol) agarose gel run at 7 V/cm for 1 h and stained with ethidium bromide.

DNA was purified from agarose gels with a GeneClean kit (Bio 101, Vista, CA) as described previously (Jeong et al., 1997). Samples were sequenced by dye terminator chemistry with an ABI 373 at Center for Gene Research and Biotechnology at Oregon

State University. Alignment of sequences was made with the program Clustal (PC-Gene; IntelliGenetics). Phylogenetic tree was constructed using the Neighbor-joining method.

4.3.4 Rep-PCR

Genomic DNAs were amplified using BOX (CTACGGCAAGGCGACGCTGACG) or DR (GCGGCAACGGCGGCAACGGCGG) primers synthesized by Genosys (The Woodlands, TX). Modified PCR protocols to that described previously (Versalovic et al., 1994) were applied. PCR protocols with BOX and DR primers were referred to as BOX-PCR and DR-PCR, collectively referred to as rep-PCR. Amplification was performed in a model 110s Tempcycler II (Coy Corp., Grass Lake, MI). The PCR reactions were performed in GeneAmp® 1× buffer II (10 mM Tris-HCl (pH 8.3), 50 mM KCl; Perkin-Elmer, Roche Molecular Systems, inc., Branchburg, NJ), with 2.0 mM $MgCl_2$ and 5% (v/v) DMSO, 100 pmol BOX primer or 150 pmol DR primer, 5-10 ng (if not specified) template DNA, and 1.25 mM of each of the 4 dNTPs in a 25 µl reaction mix. Each PCR experiment included a control lacking template DNA. PCR mixtures were overlaid with 25 µl of mineral oil (M3516; Sigma, St. Louis, MO). Samples were denatured at 95 °C for 5 min before adding 2 units of AmpliTaq DNA polymerase (Perkin-Elmer) and then run by using the following 30 cycles: denaturing at 94°C for 1 min, annealing at 52 °C (BOX) or 57 °C (DR) for 1 min, and extension at 65 °C for 8 min, with a final extension at 65 °C for 15 min and a final soak at 4 °C.

A portion (7 µl) of amplified PCR product was separated by gel electrophoresis on 1.5% agarose gels, in 0.5 x TBE buffer for 3 h at 5 V/cm, stained with ethidium bromide and photographed on a UV transilluminator. Fingerprinting generated from different strains or nodule tips were compared visually and all bands were scored. The presence or absence of particular DNA fragments were converted into binary data and the resulting matrix was analyzed with the SHAN (Sequential Agglomerative Hierarchical and Nested; Sneath and Sokal, 1973) of the NYSYS software (Applied Biostatistic Inc.).

4.4 Results

4.4.1 PCR amplification and DNA sequencing of 16S to 23S rDNA intergenic spacer region

PCR using primers 16F1646 and 23R1609 produce single DNA products of the expected 2 kbp length. PCR products were sequenced using primer 16F1646 as a sequencing primer. We obtained two sequences of 592 bp from each *C. velutinus* nodule of the *C. velutinus-C. prostratus* and *C. velutinus-C. integerrimus* copopulations (Fig. 4.1). The two sequences, which contained full length of ISR, 3′ end of 16S rDNA, and 5′ end of 23S rDNA, were unambiguously readable except for apparently compressed regions resulting from sequencing chemistry. We also obtained 150-200 bp DNA sequences from four nodules (*C. prostratus* nodule of the *C. velutinus-C. prostratus* copopulation; *C. integerrimus* and *C. sanguineus* nodules of the *C. integerrimus-C. sanguineus* copopulation; *C. integerrimus* nodule of the *C. veltinus-C. integerrimus* copopulation). The four short DNA sequences apparently resulted from secondary structure in a region of high GC content, because DNA sequencing abruptly stopped at about 170 bp in all cases. The short sequences may also have resulted from contaminants, because we pooled several PCR reaction tubes to make appropriate amount of DNA for sequencing. In any case, all sequences were readable, indicating that the PCR products were homogeneous and our primer was specific to *Frankia* DNA.

Sequences of two full and four partial 16S-23S ISRs were compared with two published *Frankia* sequences and one *Clavibacter* sequence. All of the short DNA sequences were identical. We observed only two bases differences between the two 592-bp sequences. The DNA sequences of *C. velutinus* microsymbiont *Frankia* were compared with published sequences (Fig. 4.1). The alignment resulted in several indels, which were not included in the phylogenetic analysis. A phylogenetic tree, reconstructed using Neighbor-joining method, indicated that the *C. velutinus* microsymbiont *Frankia* are distantly related to *Alnus*- and *Casuarina*-infective *Frankia*.

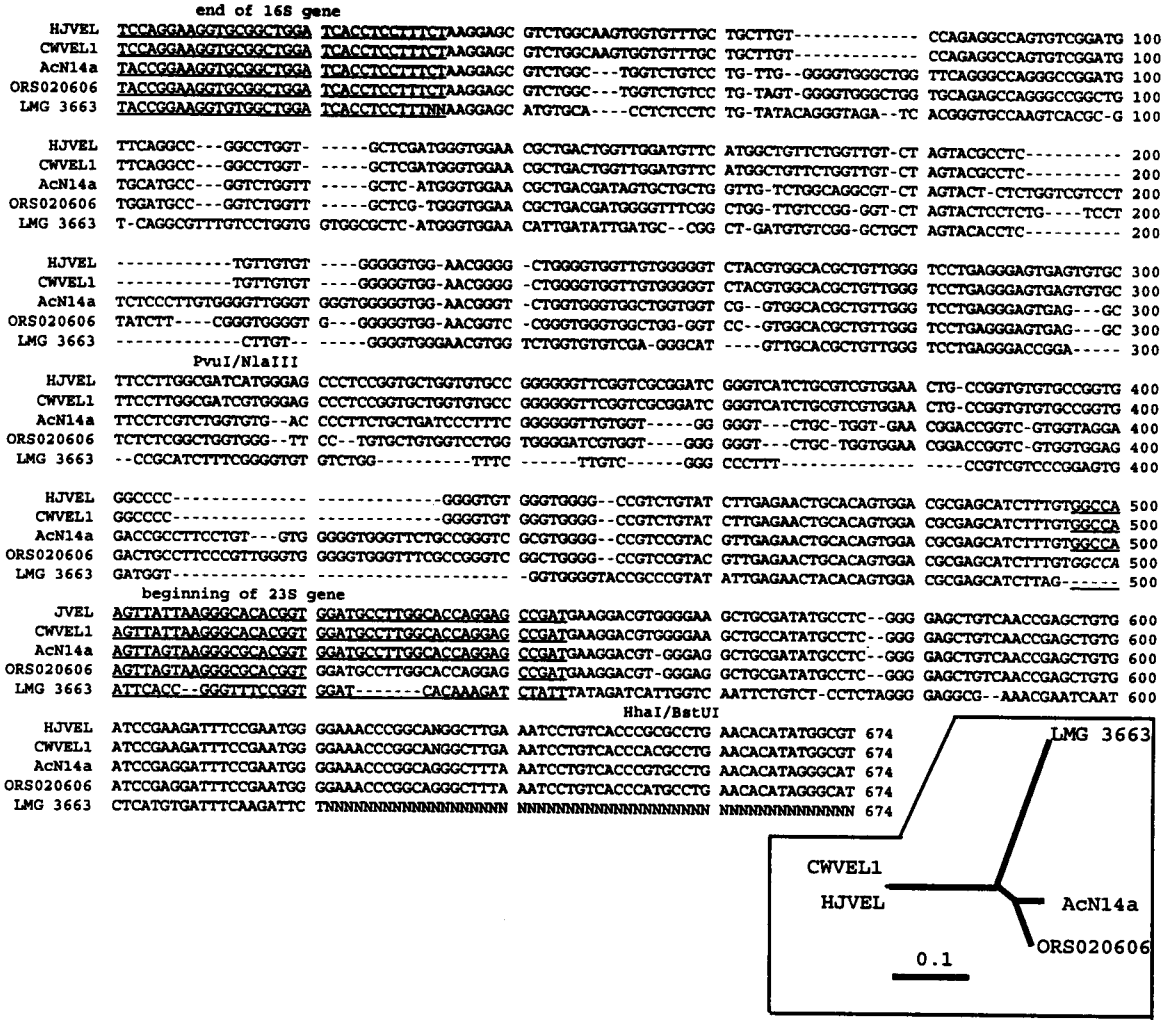


Figure 4.1. Comparison of ISR sequences between 16S and 23S rRNA genes from *Ceanothus* microsymbiont *Frankia* (HJVEL and CWVEL), *Frankia* sp. AcN14a, *Frankia* sp. ORS020606, and *Clavibacter michiganesis insidiosum* LMG 3663. Restriction sites unique to the ISR sequences of the two *Ceanothus*-microsymbiont *Frankia* are identified above the sequence line. The beginning and end of 16S and 23S rRNA genes are underlined. The insert cladogram is a phylogenetic tree of this sequence alignment reconstructed using Neighbor-joining method.

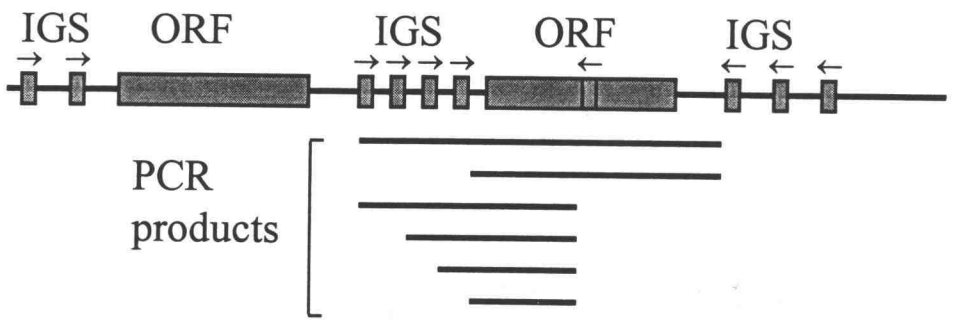


Fig. 4.2. A possible reason for generation of complex PCR amplification patterns by DR oligonucleotide-genomic DNA interactions. ORF represents open reading frame of a gene. Small boxes indicate DR regions and arrows the direction of DR.

4.4.2 DR-PCR fingerprinting and its comparison with BOX-PCR fingerprinting

Previous studies have reported that an oligonucleotide primer corresponding to BOX elements in a Gram-positive bacteria can be utilized to generate genomic fingerprints of both Gram-positive and Gram-negative bacteria (Versalovic et al., 1994). In this study we designed a novel primer corresponding to conserved DNA sequences of direct repeats in *Mycobacterium bovis* and demonstrated its utility for genomic fingerprinting. We used a single primer for DR-PCR in a similar way to that of BOX-PCR, because a BLAST search (Altschul et al., 1990) showed that DR sequences tend to be dispersed in the *M. bovis* and *M. tuberculosis* genomic DNA (Doran et al., 1993; Philipp et al., 1996) in the same orientation in a particular intergenic region but in different orientations in others. DR sequences are also found occasionally in intragenic regions. Thus, we postulated that a single DR consensus sequence could be used as a PCR primer to yield distinct patterns of DNA fingerprints varying in size when separated by agarose gel electrophoresis (Fig.4.2).

We examined systematically the distribution and utility of DR-PCR by comparing it with the well-characterized BOX-PCR. DR-PCR yielded 8 to 20 distinct PCR products, ranging in size from approximately 100 bp to over 2 kbp, from Gram-positive and Gram-negative bacteria, including three *Frankia* pure culture strains (Fig. 4.3 A). BOX-PCR also yielded distinct fingerprint profiles from diverse bacteria (Fig. 4.3A) as reported (Sadowsky et al., 1996; Murry et al., 1995). To determine the reproducibility of generating fingerprint profiles from different DNA preparations of the same strain, we analyzed DNA from cultures that had been initiated from single colonies at different times: *E. coli* DH5α cells kept for four years in a 27 °C growth chamber and freshly cultured DH5α cells yielded identical fingerprinting profiles (Fig. 4.3B, lanes 1 and 2). Likewise, four-year-old and newly grown *Frankia* EaI12 cells also showed identical fingerprinting profiles (Fig. 4.3B lanes 8 and 9).

BOX- and DR-PCR also yielded several distinct PCR products from tobacco and *Ceanothus* root DNA (Fig. 4.4 A and B) and *Ceanothus* leaf DNA (data not shown). We tested the specificity of primers by mixing two DNAs. Interestingly, BOX- and DR-PCR of a mixture of plant and bacterial DNA yielded only bacterial PCR profiles (Fig. 4.4 A, B, C, and D). A mixture of the same amount of *E. coli* and *Frankia* DNA yielded almost

Fig 4.3. Ethidium bromide-stained agarose gel pattern of BOX- and DR-PCR products from genomic DNAs of Gram-negative and Gram-positive. The DNA molcular weight marker (lane M) is a 100-bp ladder (Gibco BRL).

A. The pattern of PCR products generated by using the BOX primer. Lanes 1 to 6 show the BOX-PCR products of genomic DNA from *E. coli* DH5α (lane 1), *A. brasilense* ATCC 29710 (lane 2), *M. luteus* (lane 3), *G. obscurus* (lane 4), *Frankia* sp. AvcI1 (lane 5), and *Frankia* sp. EaI12 (lane 6); lane 7 is negative control reaction lacking template DNA.

B. The pattern of PCR products generated by using the DR primer. Lanes 1 to 9 show the DR-PCR products of genomic DNA from freshly cultured *E. coli* DH5α (lane 1), four-year old cultured *E. coli* DH5α (lane 2), *A. brasilense* ATCC 29710 (lane 3), *M. luteus* (lane 4), *G. obscurus* (lane 5), *Frankia* sp. CPI1 (lane 6), *Frankia* sp. AvcI1 (lane 7), freshly cultured *Frankia* sp. EaI12 (lane 8), and four-year old cultured *Frankia* sp. EaI12 (lane 9); lane 10 is negative control reaction lacking template DNA.

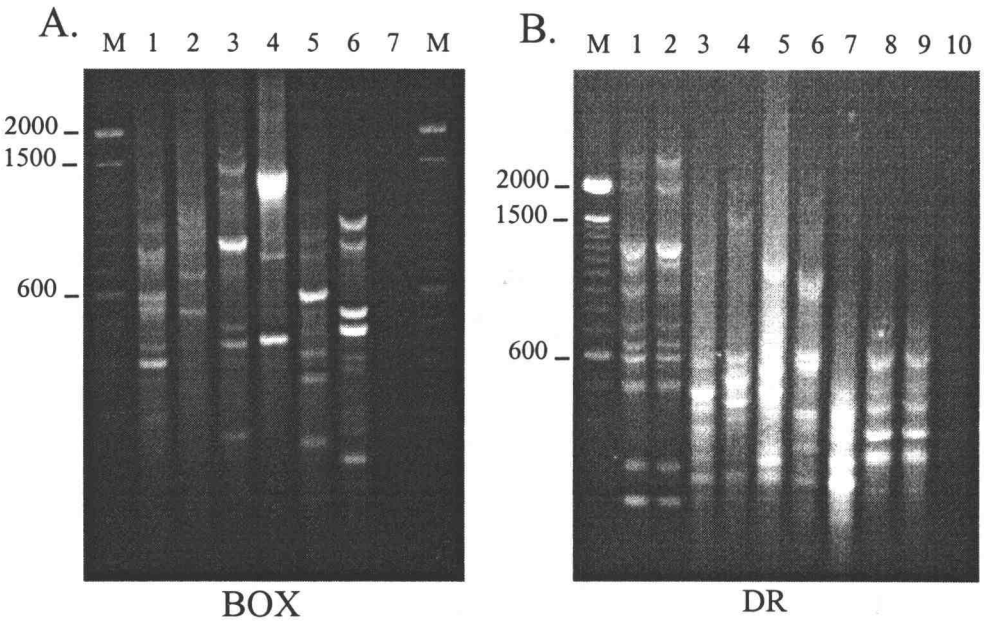

A.
M 1 2 3 4 5 6 7 M
2000
1500
600
BOX
B.
M 1 2 3 4 5 6 7 8 9 10
2000
1500
600
DR

Fig. 4.3.

Fig. 4.4. Ethidium bromide-stained agaroae gel pattern of BOX- and DR-PCR products from various genomic DNA concentrations and DNA mixtures of *E. coli*, *Frankia* sp. EaI12, *C. velutinus*, and *Nicotiana tabacum*. The DNA molecular weight marker (lane M) is a 100-bp ladder (Gibco BRL).

A. The pattern of PCR products generated by using the BOX primer: lanes 1 to 11 show the BOX-PCR products of genomic DNA from *E. coli* (15 ng, lane 1), EaI12 (15 ng, lane 2), *E.coli* (7.5 ng) +EaI12 (7.5 ng) (lane 3), *N. tobacuum* (15 ng, lane 4), *E. coli* (7.5 ng) + *N. tobacuum* (7.5 ng, lane 5), EaI12 (7.5 ng) + *N. tabacuum* (7.5 ng) (lane 6), *C. velutinus* (15 ng, lane 7), *E. coli* (7.5 ng) + *C. velutinus* (7.5 ng) (lane 8), EaI12 (5 ng) + *C. velutinus* (10 ng) (lane 9), EaI12 (7.5 ng) + *C. velutinus* (7.5 ng) (lane 10), EaI12 (10 ng) + *C. velutinus* (5 ng) (lane 11).

B. The pattern of PCR products generated by using the DR primer: lanes 1 to 11 correspond to lane 1 to 11 in panel A.

C. The pattern of PCR products generated by using the BOX primer: lanes 1 to 7 show the BOX-PCR products of genomic DNA from EaI12 (20 ng, lane 1), EaI12 (5 ng, lane 2), EaI12 (1 ng, lane 3), *C. velutinus* (5 ng) + EaI12 (20 ng) (lane 4), *C. velutinus* (5 ng) + EaI12 (5 ng) (lane 5), *C. velutinus* (5 ng) + EaI12 (1ng) (lane 6), *C. velutinus* (5 ng, lane 7), and negative control reaction lacking template DNA (lane 8).

D. The pattern of PCR products generated by using the DR primer: lanes 1 to 8 correspond to lanes 1 to 8 in panel C.

E. The pattern of PCR products generated by using the BOX primer: lanes 1 to 7 show the BOX-PCR products of genomic DNA from *E. coli* (10 ng, lane 1), *E. coli* (9.5 ng) + EaI12 (0.5 ng) (lane 2), *E. coli* (7.5 ng) + EaI12 (2.5 ng) (lane 3), *E. coli* (5 ng) + EaI1 (5 ng) (lane 4), *E. coli* (2.5 ng) + EaI12 (7.5 ng) (lane 5), *E. coli* (0.5 ng) +EaI12 (9.5 ng) (lane 6), EaI12 (10 ng) (lane 7), and negative control reaction lacking template DNA (lane 8).

F. The pattern of PCR products generated by using the DR primer: lanes 1 to 8 correspond to lanes 1 to 8 in panel E.

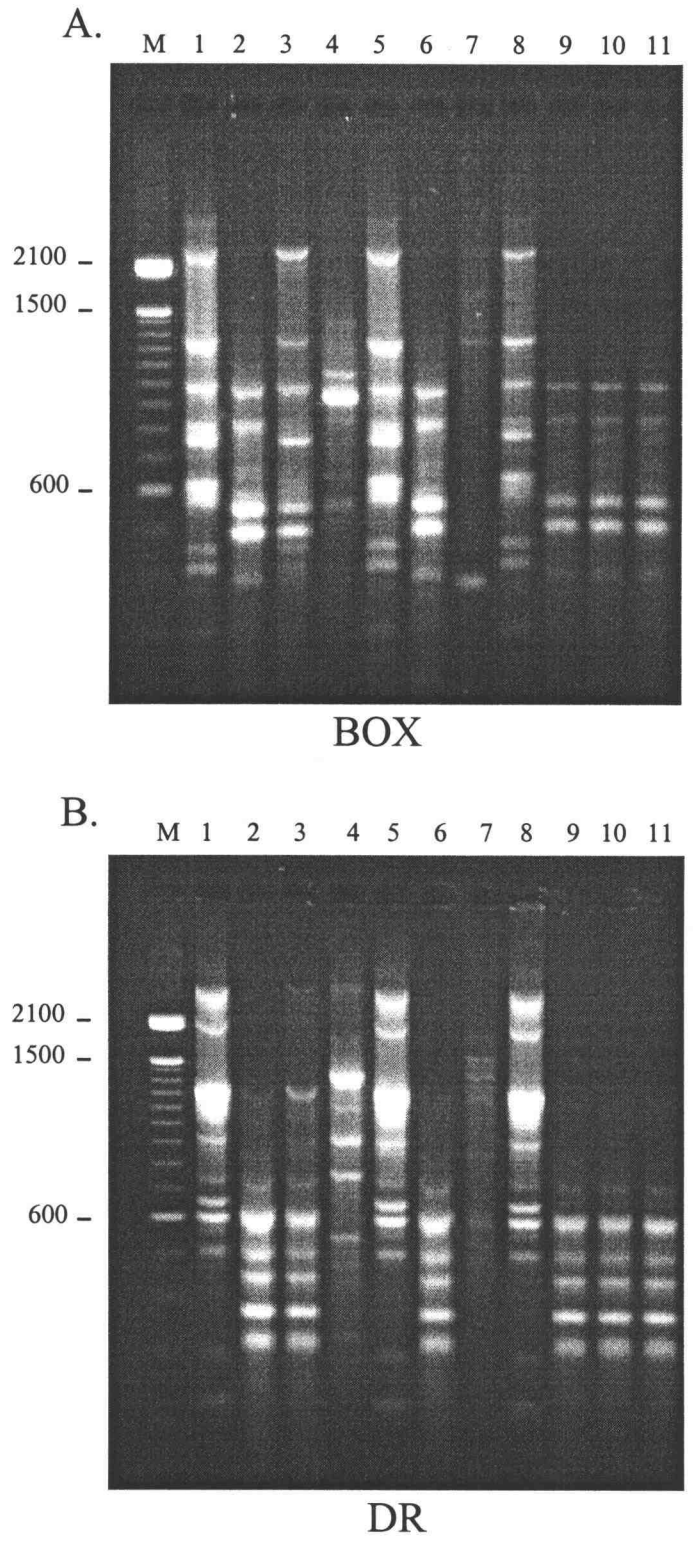
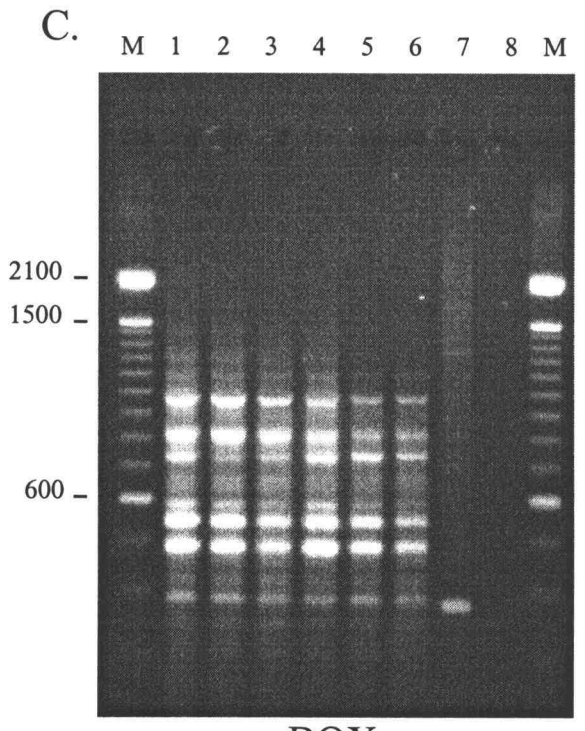

A.
M 1 2 3 4 5 6 7 8 9 10 11
2100
1500
600
BOX
B.
M 1 2 3 4 5 6 7 8 9 10 11
2100
1500
600
DR

Fig. 4.4.



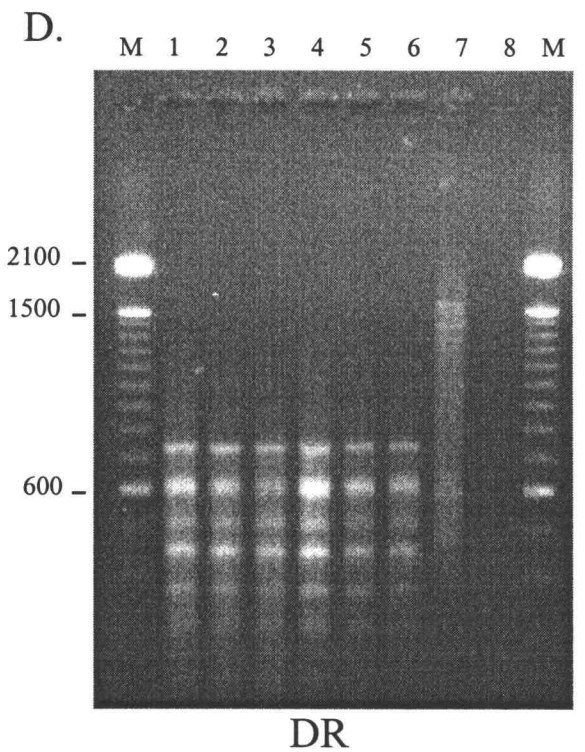
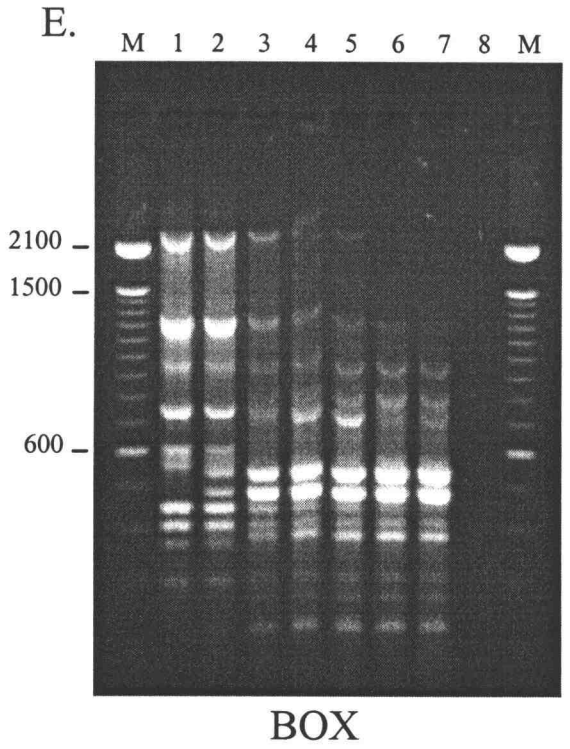


Fig. 4.4.



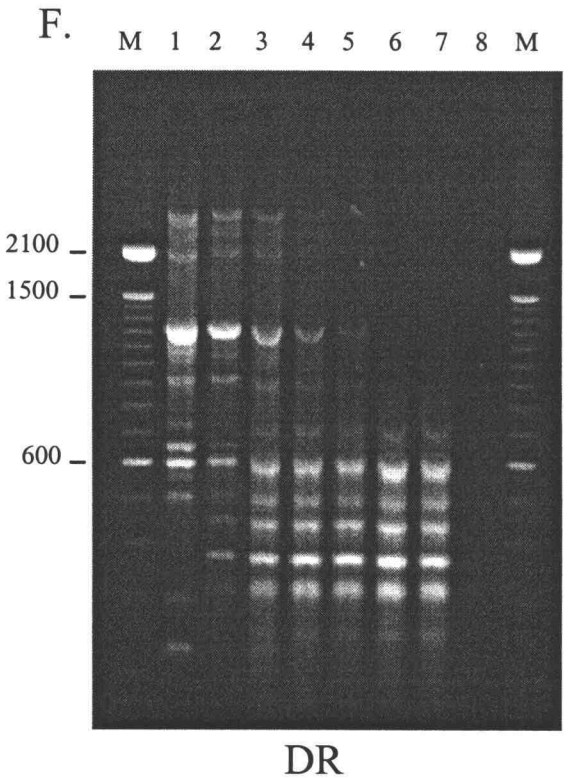


Fig. 4.4.

the combination of the two bacterial genomic fingerprint patterns (Fig. 4.4 A, B, E, and F). However, disproportionate mixtures of the two bacterial DNAs tended to retain more bands from that at the higher concentration, although the PCR profiles showed that *Frankia* EaI12 DNA was a better template than *E. coli* DNA. Furthermore, no distinct bands derived from *E. coli* DNA were observed in the fingerprint profile from the mixture of 0.5 ng *E. coli* DNA and 9.5 ng *Frankia* EaI12 DNA. Different amounts of genomic DNA, ranging from 0.5 ng to 20 ng, showed essentially the same profiles with change of an intensity of just a few bands (for example, lanes 1 and 2 of Figs. 4.4 C and D). Generally, the quantity of DNA in the extracts did not substantially affect fingerprint profiles. This result is interesting because we estimated that more than 80% of nodule DNA came from microsymbiont *Frankia*. We estimated the approximate proportion of *Frankia* DNA by subtracting the DNA contents of roots and leaves from that of fresh nodule lobes. The average DNA contents of four nodule lobes of *C. velutinus* were 105 ± 25 ng DNA mg^{-1} nodule; that of three nodule lobes of *C. sanguineus* was 69 ± 4 ng DNA mg^{-1} nodule; that of three nodule lobes of *C. integerrimus* was 115 ± 45 ng DNA mg^{-1} nodule. The average DNA contents of four root samples adjacent/near the nodule of the above three species was 8 ± 4 ng DNA mg^{-1} root (*C. velutinus*), 9 ± 3 ng DNA mg^{-1} root (*C. sanguineus*), 4 ± 3 ng DNA mg^{-1} root (*C. integerrimus*).

4.4.3 BOX- and DR-PCR fingerprinting of Ceanothus nodules

Primers corresponding to conserved DNA sequences of BOX elements and DR sequences annealed to genomic DNA and generated unique genomic fingerprints of nodule DNA (Fig. 4.5 and 4.6). Rep-PCR DNA fingerprints clearly distinguish different nodules from different *Ceanothus* plants. BOX- and DR-PCR yielded 8 to more than 15 distinct PCR products, ranging in size from approximately 100 bp to over 2 kb. We also ran PCR products from leaf or root DNA. The PCR products of root or leaf alone did not correspond with those from nodule DNAs (Data not shown) Differences among nodules were assessed visually on the basis of the migration patterns of PCR products. Fingerprinting profiles generated with each primer set were complex but quite similar among nodules with a few band differences. In total, 17 distinct bands from BOX-PCR were visualized. Although DR-PCR produced more smeared band patterns than BOX-

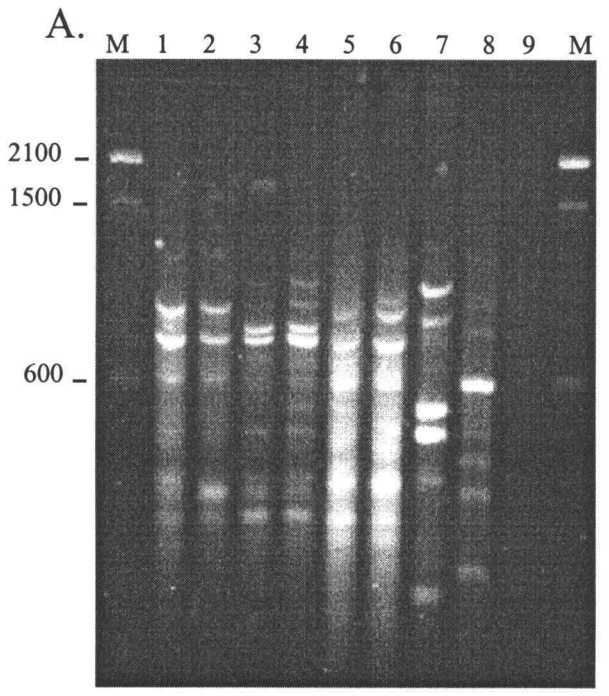
PCR, DR-PCR showed the same degree of resolution compared with BOX-PCR. Nine major PCR products, generated by BOX-PCR, appeared to comigrate among nodules; for DR-PCR, four major products comigrated. *Ceanothus*-microsymbiont *Frankia* are closely related on the basis of DNA sequences. The presence of several comigrating bands also suggested that *Frankia* are closely related. The BOX- and DR-PCR profiles of *Ceanothus*-microsymbiont *Frankia* were found to be completely different from those of both *Frankia* CPI1 and AvcI1 (Fig. 4.5 A and B).

Fig. 4.6 showed that BOX and DR primers yielded quite similar banding patterns between nodules of two *Ceanothus* species within a copopulation with one to three band differences. BOX-PCR yielded a single band difference for nodules of the *C. integerrimus-C. sanguineus* copopulation (Fig. 4.6A). DR-PCR of the nodules gave three band differences (Fig. 4.6B). No notable differences were observed between the two nodules of the *C. velutinus-C. integerrimus* copopulation or between the two nodules of the *C. velutinus-C. prostratus* copopulation (data not shown). We did not observe any difference of genomic fingerprinting patterns among nodule lobes of a nodule (Fig. 5 B and D).

We observed some degree of difference in genomic fingerprints among the three copopulations. However, the results may underestimate the diversity of *Frankia* within a copopulation, because only one nodule from each species was analyzed from each copopulation. We tested this possibility using nodules collected from three *C. velutinus-C. prostratus* copopulations about 1.5 km apart from each other. Fig. 4.7 shows that the PCR-profiles from six nodules showed greater differences among copopulation sites than between the two species.

Among BOX- and DR-PCR fingerprints from 10 nodules, 8 different profiles were identified. The binary matrix from the BOX- and DR-PCR profiles was analyzed with UPGMA option in the SHAN analysis of the NTSYS software. The dendrogram obtained showed high similarity between nodule fingerprints from the same geography rather than from the same *Ceanothus* species (Fig. 4.8).

Fig. 4.5. Ethidium bromide-stained agarose gel patterns of BOX PCR (panel A) and DR PCR (panel B) products from six nodules collected from three *Ceanothus* copopulations. Lanes 1-6 show the BOX and DR PCR products obtained from nodule lobes of *C. velutinus* (lane 1) and *C. integerrimus* (lane 2) of *C. velutinus-C. integerrimus* copopulation, *C. sanguineus* (lane 3) and *C. integerrimus* (lane 4) of *C. sanguineus-C. integerrimus* copopulation, and *C. velutinus* (lane 5) and *C. prostratus* (lane 6) *of C. velutinus-C. prostratus* copopulation. In panel A lane 7 is BOX-PCR products of *Frankia* sp. EaI12 and lane 8 *Frankia* sp. AvcI1. In panel B lane 7 is DR-PCR products of *Frankia* EaI12 and lane 8 CpI1. Lane 9 is negative control reaction lacking template DNA. The lanes labeled M show the DNA molecular size marker (100-bp ladder, Gibco-BRL.


A.
M 1 2 3 4 5 6 7 8 9 M
2100
1500
600
BOX

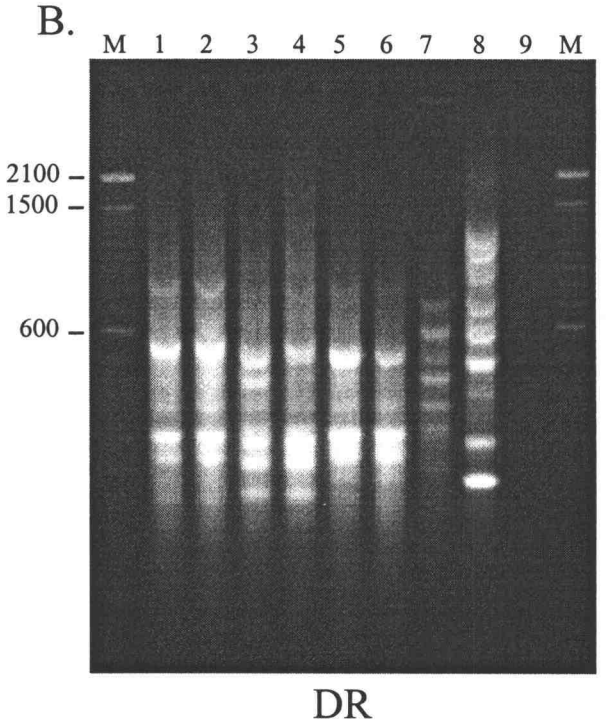

B.
M 1 2 3 4 5 6 7 8 9 M
2100
1500
600
DR

Fig. 4.5.

Fig. 4.6. Ethidium bromide-stained agarose gel patterns of BOX PCR (panel A) and DR PCR (panel B) products from three nodule lobes of each nodule collected from *C. sanguineus* and *C. integerrimus* in copopulation. Lanes 1 to 6 show the BOX and DR PCR products obtained from lobes of *C. sanguineus* nodule (lane 1-3) and of *C. integerrimus* (lane 4-6). In panel A lane 7 is BOX PCR products of *Frankia* sp. CpI1 and lane 8 is negative control reaction lacking template DNA. In panel B lane 7 is the negative control. The lanes labeled M show the DNA molecular size marker (100-bp ladder, Gibo-BRL).

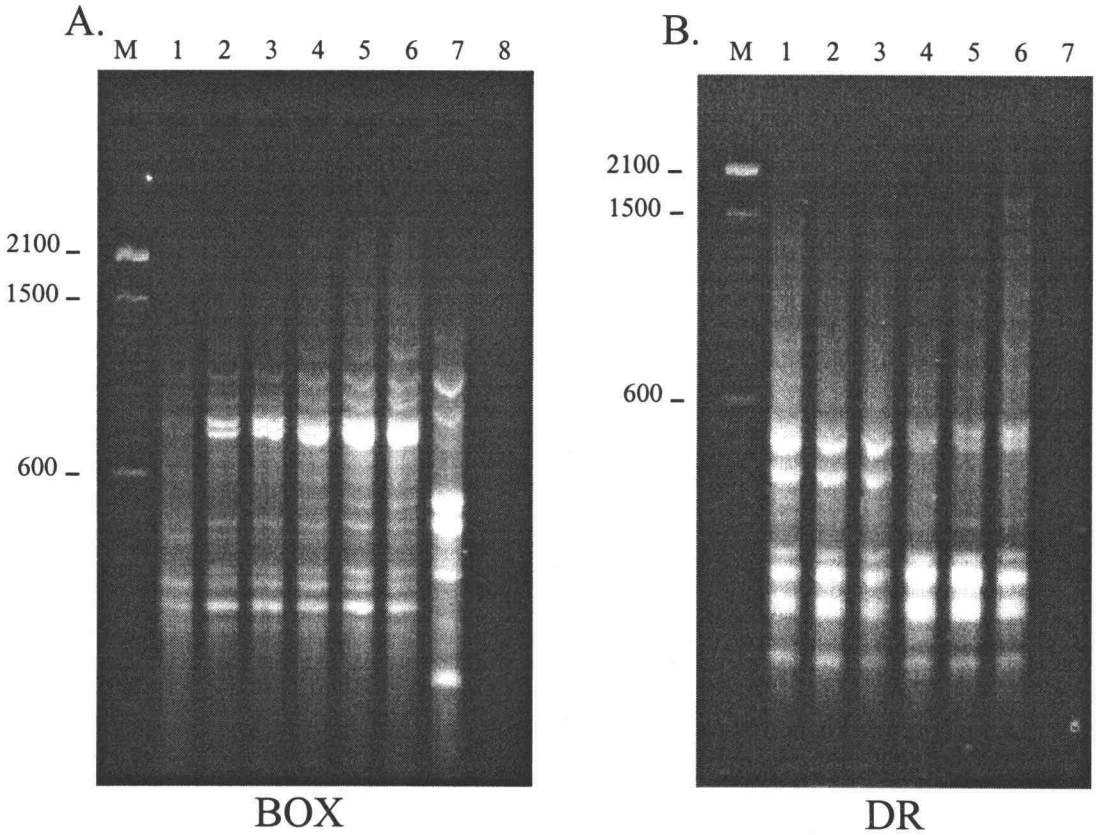


Fig. 4.6.

Fig. 4.7. Ethidium bromide-stained agarose gel patterns of BOX PCR (panel A) and DR PCR (panel B) products from six products collected from three *C. velutinus-C. prostratus* copopulation sites. Lanes 1 to 6 show the BOX and DR PCR products obtained from lobes of *C. velutinus* (site 619, lane1; site 1, lane 3; and site 2, lane 5) and of *C. prostratus* (site 619, lane 2; site 1, lane 4; and site 2, lane 6). Lane 7 is PCR products of *Frankia* sp. CpI1 and lane 8 is negative control reaction lacking template DNA.

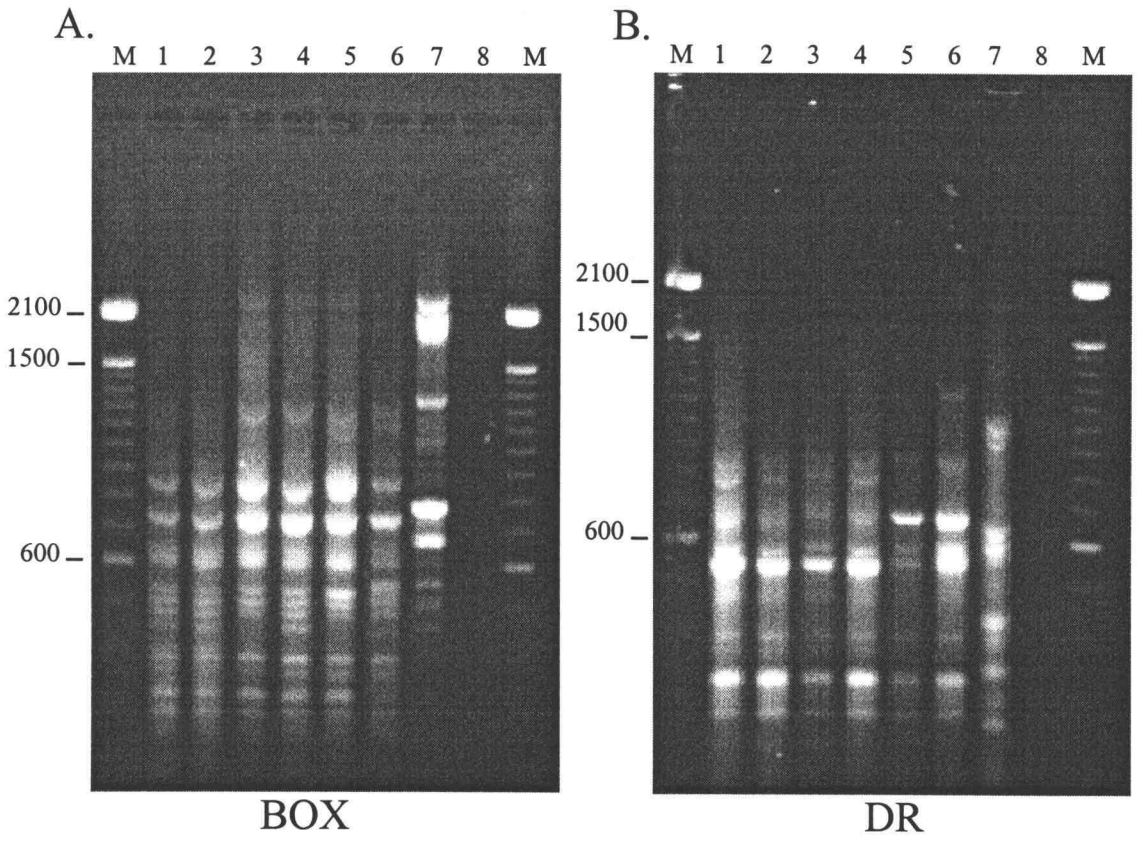


Fig. 4.7.

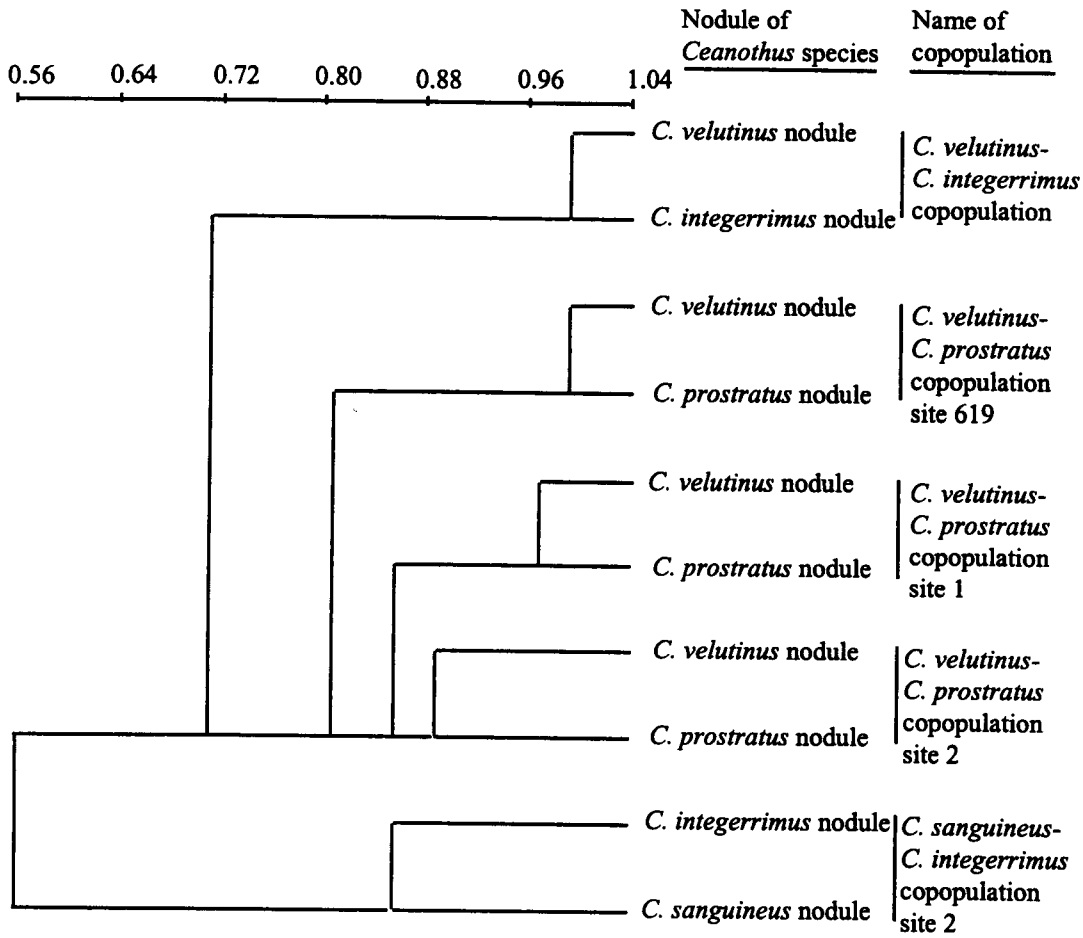


Fig. 4.8. Dendrogram of relatedness of rep-PCR fingerprints for *Ceanothus* nodules determined by cluster analysis and the unweighted pair-group method with arithmetic averages (Sneath and Sokal, 1973). Analyses were conducted with the UPGMA option in SHAN Clustering in NT-SYS program (Applied Biostatistics Inc.)

4.5 Discussion

The low sequence variation of 16S-23S ISR in *Ceanothus*-microsymbiont *Frankia* from three *Ceanothus* copopulations (Fig. 4.1) suggested that those microsymbiont *Frankia* were related at the sub-species or lower level. The phylogenetic analysis of 16S-23S ISR suggested that the two *Ceanothus*-microsymbiont *Frankia* are distantly related to *Alnus*- and *Casuarina*-infective *Frankia* strains. The result is similar to those of earlier analyses of the full or partial length 16S sequences where *Ceanothus*-microsymbiont *Frankia* cluster with those of the Elaeagnus-infective staains or Dryas microsymbiont (Murry et al., 1997; Benson et al., 1996; Chapter 2). This is interesting because both *Ceanothus*-microsymbiont and *Elaeagnus*-infective *Frankia* infect the root by direct intercellular penetration (Liu and Berry, 1991) and *Elaeagnus*-infective *Frankia* show low diversity relative to *Alnus*-infective *Frankia* (Normand et al., 1996). It also suggests that the high variability region of the ISR can be useful for designing PCR primers specific for *Ceanothus*-microsymbiont *Frankia*.

Because ISR sequences showed such a low variation, we further analyzed *Ceanothus*-microsymbiont *Frankia* using BOX and DR-PCR methods, referred to as a rep-PCR collectively. Rep-PCR methods have been known to resolve bacterial strains highly specifically at the sub-species level (Frey et al., 1997; Latour et al., 1996). Two previous reports have used the rep-PCR method to assess the genetic diversity of *Frankia* pure cultures and *Ceanothus*-microsymbiont *Frankia* (Murry et al., 1995; 1997). However, the previous nodule study by BOX-PCR assumed, without clear justification, that the resulting rep-PCR products revealed genomic fingerprinting of nodule microsymbiont. We justified in this study that the BOX primers are particularly suitable for the rapid molecular characterization of *Ceanothus*-microsymbiont *Frankia* at the subspecies level using nodules. Additionally, we demonstrated a novel DR-PCR technique, which has almost the same degree of utility as BOX-PCR.

The results shown in Fig. 4.3 suggested that the DR primer could be a useful tool for genomic fingerprinting, similar to BOX, REP, and ERIC primers. First, our results suggest that the genomic fingerprint patterns generated by DR-PCR are stable over time. Identical profiles from four-year-old cultured isolates and freshly cultured isolates

support the reproducibility of the rep-PCR protocol. Second, we have demonstrated that closely-related *Frankia* strains can be easily differentiated using DR-PCR profiles, as determined by other experimental protocols such as DNA sequencing studies and that bacteria associated with the same host can be easily differentiated by these profiles. Third, we have demonstrated that diverse arrays of bacteria ranging from Gram-negative to Gram-positive bacteria yielded a sufficient complexity of DR-PCR products.

We suggest that rep-PCR profiles generated with nodule DNA represent those of the microsymbiont *Frankia* genome, as shown in Fig. 4.4. Rep-PCR of plant-bacterial DNA mixtures kept most PCR bands observed in those of plant DNA alone from amplifying, indicating that DR and BOX primers are more specific to bacterial DNA. Rep-PCR of a mixture of equal amounts of DNA from two bacterial strains tended to retain PCR bands observed in each strain alone. However, highly disproportionate mixtures tended to reduce PCR amplification from bacterial DNA of lower concentration and in extreme disproportionate mixture (20:1), no PCR bands are amplified from that of lower concentration. Taken together, BOX- and DR-PCR yield most likely PCR products amplified only from *Frankia* DNA when fingerprinting nodule DNA, even though nodule DNA contains some degree of plant DNA and ineluctable contaminant bacterial DNA. The results correspond to a previous report by Louws et al. (1994) that plant tissue infected with a known *Xanthomonas campestris* strain yielded the rep-PCR profile identical to that of the bacterial isolate.

We systematically compared the general utility of BOX and DR primers. In general, differences between microsymbiont *Frankia* were detected by each primer with similar resolution, suggesting that the distribution of BOX and DR sequences is a true reflection of genomic structure. For detecting limited polymorphisms within a copopulation between microsymbiont *Frankia*, each primer set offered unique information. By using two different primers, a broader survey of the chromosomal structure was possible and more specific conclusions concerning diversity or similarity among bacterial strains and microsymbiont *Frankia* were achieved. More faint bands from *Frankia* DNA were observed in DR-PCR. In general, depending on the efficiency of PCR, bright bands and faint bands were not always amplified to the same extent or were not equally visible after ethidium bromide staining. Thus, it was necessary to repeat PCR reaction at least once in

order to confirm PCR patterns. More prominent bands, however, were consistently present and were almost always sufficient to distinguish among pure culture or microsymbiont *Frankia*.

Ceanothus-microsymbiont *Frankia* are closely related on the basis of DNA sequences. The presence of several comigrating bands also suggested that *Frankia* are closely related. The BOX profiles of *Frankia* were comparatively simple, and at least nine PCR products comigrated. DR-PCR also yielded multiple bands of parallel mobility. However, the overall patterns of BOX- and DR-PCR were distinct. The BOX- and DR-PCR profiles of *Ceanothus*-microsymbiont *Frankia* were found to be completely different from those of *Frankia* CPI1, AvcI1, or EaI12, confirming that previously isolated *Ceanothus*-microsymbiont *Frankia* are distantly related to *Alnus*-infective *Frankia* on the basis of sequence data of ribosomal DNA (Chapter 2, Murry et al., 1997). Each *Ceanothus* nodule tested had unique BOX- and DR-PCR fingerprint profiles, but the presence of multiple bands of apparent equal mobility suggested that microsymbiont *Frankia* among nodules from three different geographic sites had a common evolutionary heritage. The BOX- and DR-PCR protocols provided similar conclusions about the apparent relatedness among nodules that had similar fingerprint profiles. When nodules within a copopulation demonstrated polymorphisms, each primer set offered unique information, generating nodule-specific profiles.

It is interesting that we did not observe any differences in rep-PCR fingerprints among nodule lobes within a nodule, even though great diversity of *Rhizobium* has been observed within a single nodule. This may be due to difference of the infective process and the ultrastructure of nodules. For example, rhozobia infect root tissue via infection thread and are located in intracellular organelle, called bacteroids, which are like plastids. *Frankia* penetrate *Ceanothus* root tissue via intercellular space and are located outside plasma membrane of plant cells.

Ceanothus-microsymbiont *Frankia* exhibit a lower degree of diversity within each copopulation than among copopulations. The result is interesting in respect to host specificity because the same *Ceanothus* species from different geographic origins is associated with different microsymbiont *Frankia*, whereas two different *Ceanothus* species from the same geographic origins are associated with quite similar microsymbiont

Frankia. Taken together, the results suggest that *Frankia* exhibit limited host-specificity within *Ceanothus* copopulations. However, geography is the more important factor for *Frankia* divergence than host plants.

4.6 References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-10 (1990).

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

Baker, D. D, and B. C. Mullin. 1994. Diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus* as assessed by RFLP patterns from single nodule lobes Soil Biol. Biochem. 26: 547-552.

Becking, J. H. 1977. Dinitrogen-fixing associations in higher plants other than legumes. In R. W. F. Hardy. and W. S. Silver (eds) A treatise on dinitrogen fixation section 3, Biology. Wiley, New York, pp 185-275.

Benson, D. R., D. W. Stephens, M. L. Clawson, and W. B. Silvester. 1996. Amplication of 16S rRNA genes from *Frankia* strains in root nodules of *Ceanothus griseus*, *Coriaria arborea*, *Coriaria plumosa*, *Discaria toumatou*, and *Purshia tridenta*. Appl. Environ. Microbiol. 62: 2904-2909.

Bourquet, S. N., J. R. Valero, M. C. Lavoie, and R. C. Levesque. 1995. Comparative analysis of the 16S to 23S ribosomal intergenic spacer sequences of *Bacillus thuringiensis* strains and subspecies and of closely related species. Appl. Environ. Microbiol. 61: 1623-1626.

Dolzani, L, E. Tonin, C. Lagatolla, and C. Monti-Bragadin. 1994. Typing of *Staphylococcus aureus* by amplification of the 16S-23S rRNA intergenic spacer sequences. FEMS Microbiol. Lett. 119: 167-174.

Doran, T. J., A. L. M. Hodgson, J. K. Davies. and A. J. Radford. 1993. Characterisation of a highly repeated DNA sequence from *Mycrobacterium bovis*. FEMS Microbiol. Lett. 111: 147-152

Frey, P, P. Frey-Klett, J. Garbaye, O. Berge, and T. Heulin 1997. Metabolic and genotypic fingerprinting of fluorescent pseudomonads associated with the Douglas Fir-*Laccaria bicolor* mycorrhizosphere. Appl. Environ. Microbiol. 63: 1852-1860.

Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk. Jr. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. System. Bacteriol. 42: 166-170.

Garcia-Martines, J., A. Martinez-Murcia, A. I Anton, and F Rodriguez-Valera 1996. Comparison of the small 16S to 23S intergenic spacer region (ISR) of the rRNA operons of some *Escherichia coli* strains of the ECOR collection and *E. coli* K-12. J. Bacteriol. 178: 6374-6377.

Jamann, S., M. P. Fernandez, and P. Normand. 1993. Typing method for N_2-fixing bacteria based on PCR-RFLP--application to the characterization of *Frankia* strains. Mol. Ecol. 2:17-26.

Jeong, S. C., A. Liston, and D. D. Myrold. 1997. Molecular phylogeny of the genus *Ceanothus* (Rhamnaceae) using *rbc*L and *ndh*F sequences Theor.Appl. Genet. 94: 852-857.

Latour, X, T. Corberand, G. Languerre, F. Allard, and P. Lemanceau. 1996. The comparison of fluorescent pseudomonad populations associated with roots is influenced by plant and soil type. Appl. Environ. Microbiol. 62: 2449-2456.

Lechevalier, M. P. 1985-86. Catalog of *Frankia* strains, second edition. Actinomycetes 19:131-162.

Lechevalier, M.P., and J.-S. Ruan. 1984. Physiological and chemical diversity of *Frankia* spp. isolated from nodules of *Comptonia peregrina* (L.) Coult. and *Ceanothus americanus* L. Plant Soil 78:15-22.

Liu, Q., and A. M. Berry. 1991. The infection process and nodule initiation in the *Frankia-Ceanothus* root nodule symbiosis: a structural and histochemical study. Protoplasma 163:82-92.

Louws, F. J., D. W. Fulbright, C. T. Stephens, and F.J. de Bruijn. 1994. Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. Appl. Environ. Microbiol. 60: 2286-2295.

Maggia, L., S. Nazaret, and P. Simonet. 1992. Molecular characterization of *Frankia* isolates from *Casuarina equisetifolia* root nodules harvested in West Africa (Senegal and Gambia). Acta Ècol. 13:453-462.

Martin, K. J., Y. Tanaka, and D. D. Myrold. 1992. Dual inoculation of red alder (*Alnus rubra* Bong.) with effective *Frankia alni* for nursery bare-root seedling production. Acta Ècol. 13:508-509.

McMinn, H. E. 1942. A systematic study of the genus *Ceanothus*. *In Ceanothus*. Santa Barbara Botanic Garden, CA.

Miller, I. M., and D. D. Baker. 1985. The initiation, development and structure of root nodules in *Elaeagnus augustifolia* L. (Elaeagnaceae). Protoplasma 128:107-119.

Murray, M. G., and W. F. Thomson 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8: 4321-4324

Murry M. A., D. Zhang, M. Schneider, and F. J. de Bruijn. 1995. Use of repetitive sequences and the Polymerase chain reaction (rep-PCR) to fingerprinting the genomes of *Frankia* isolates. Symbiosis 19: 223-240.

Murry, M. A., A. S. Konopka, S. D. Pratt, and, T. L. Vandergon. 1997. The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophyte of the actinorhizal shrub *Ceanothus*. Physiol. Plant. (in press).

Navarro, E, R Nalin, D Gauthier, and P. Normand. 1997. The nodular microsymbionts of *Gymnostoma* spp. are *Elaeagnus*-infective *Frankia* strains. Appl. Environ. Microbiol. 63: 1610-1616.

Nazaret, S., B. Cournoyer, P. Normand, and P. Simonet. 1991. Phylogenetic relationships among *Frankia* genomic species determined by use of amplified 16S rDNA sequences. J. Bacteriol. 173:4072-4078.

Normand, P., B. Cournoyer, P. Simonet, and S. Nazaret. 1992. Analysis of a ribosomal RNA operon in the actinomycete *Frankia*. Gene 111:119-124.

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

Philipp,WJ, S Poulet, K Eiglmeier, L Pascopella, V Balasubramanian, B Heym, S Bergh, BR Bloom, W R Jacobs Jr., and ST Cole. 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. Proc. Natl. Acad. Sci. U.S.A. 93: 3132-3137.

Rouvier, C, Y Prin, P Reddell, P Normand, and P Simonet. 1996. Genetic diversity among *Frankia* strains nodulating members of the family Casuarinaceae in Australia revealed by PCR and restriction fragment length polymorphism analysis with crushed root nodules. Appl. Environ. Microbiol. 62: 979-985

Sadowsky, M. J., L. L. Kinkel, J.H. Bowers, and J. L. Schottel (1996). Use of repetitive intergenic DNA sequences to classify pathogenic and Disease-supressive *Streptomyces* strains. Appl. Environ. Microbiol. 62:3489-3493

Simonet, P., P. Normand, A. Mooroud, and R. Bardin. 1990. Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. Arch. Microbiol. 153:235-240.

Sneath, P. H. A., and R. R. Sokal (1973). Numerical Taxonomy, Freeman, San Fransisco, CA.

Versalovic, J, M. Schneider, F. J. de Bruijn, and J. R. Lupski. 1994. Genomic fingerprinting of bacteia using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. 5: 25-40.

Wollum, A. G. II, C. T. Youngberg, and C. M. Gilmour. 1966. Characterization of a *Streptomyces* sp. isolated from root nodules of *Ceanothus velutinus* Dougl. Soil Sci. Soc. Am. Proc. 30:463-467.

Chapter 5
Population size and diversity of *Frankia* in soils under *Ceanothus velutinus* and Douglas-fir stands

Soon-Chun Jeong and David D. Myrold

Formatted for submission to *Soil Biology and Biochemistry*

5.1 Abstract

The influence of host plants on *Frankia* populations were investigated using soils from *Ceanothus velutinus* (Dougl.) and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) stands. Population sizes of *Ceanothus*-infective *Frankia* in the soils were measured using a plant bioassay. *C. velutinus, C. sanguineus* (Pursh), and *C. integerrimus* (H. & A.) were used as trap plants. The *Frankia* population size in soil from the *C. velutinus* stand soil was about 10 times higher than that from the Douglas-fir stand. This result supports previous reports that although the presence of host plants increases *Frankia* population size, *Frankia* persist without host plants. Nodulation capacities of the three trap plants were not significantly different. All nodules showed N_2 fixation ability using the acetylene reduction activity assay. The diversity of *Frankia* nodulating trap plants was examined using repetitive intergenic DNA and the polymerase chain reaction (rep-PCR). A newly designed, direct repeat sequence and a BOX sequence were used as rep-PCR primers. The results showed that infective *Frankia* in the two contrasting soils do not have specificity to *Ceanothus* species and the two soils contained a common group of *Frankia* as well as different *Frankia* strains. Taken together, the results suggest that the higher population sizes in the soil from *Ceanothus* stand may be due to preferential increase in particular groups of *Frankia* in different soils.

5.2 Introduction

The genus *Frankia* (Frankiaceae) can persist in root nodules and exist in soil, as demonstrated by its isolation from nodules (Callaham et al., 1978) and from soil (Baker and O'Keefe, 1984). Generally, high *Frankia* populations are found in soils near host plants. However, significant numbers of *Frankia* have also been observed in soils well outside the normal geographic range of host plants (Benecke, 1969; Tyson and Silver, 1979), under non-host plant stands (Smolander and Sundman, 1987), or long after host plants have disappeared from a site (Wollum et al., 1968). In this study we examined soils under *C. velutinus* (Dougl.) and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) stands to further address *Frankia* ecology in soil.

Ceanothus species are host plants of *Frankia*. A rich diversity of *Ceanothus* species exists in Oregon. They are early successional species at disturbed sites (burns, landslides, clear-cuts, etc.) and can fix significant amounts of N_2 (Hibbs and Cromack, Jr., 1990). Douglas-fir, a non-host plant, is one of the dominant forest tree species in Oregon. Stands of Douglas-fir often replace *Ceanothus* stands during plant succession.

Most previous population studies of *Ceanothus*-infective *Frankia* have been done by observations of *Ceanothus* stands in the field or in greenhouse studies. Nodulation of seedlings decreased with higher soil N level (Youngberg and Wollum, 1976) and in response to soil amendments of a high N litter (Zavitkovski and Newton, 1968) or inorganic N (Thomas and Berry, 1989), indicating that nodulation and rates of N_2 fixation in the field may vary inversely with soil N content. There is some indication that the speed of *Ceanothus* nodulation during stand development is a function of soil Ca, with more rapid nodulation on sites with higher Ca concentrations (Scott, 1973). By the end of the 10-year observation period, most *Ceanothus* were nodulated, however. As a forest stand develops and *Ceanothus* drops out of the stand, the nodulation capacity of the soil decreases, although some *Frankia* seem to persist even in 300-year-old conifer stands (Wollum et al., 1968). Results of a greenhouse experiment using *C. velutinus* germinated from different populations and soils collected under those stands (Nelson and Lopez, 1989) suggested that the rate of N_2 fixation in actinorhizal associations depended on the soil source rather than the host population.

The diversity of *Ceanothus* microsymbiont *Frankia* has been studied only recently because the microsymbiont is recalcitrant to isolation. Genomic fingerprints of microsymbiont *Frankia* as assessed by RFLPs (Baker and Mullin, 1994) and rep-PCR (Murry et al., 1997) have revealed some degree of diversity, which is not related to geographic origins or *Ceanothus* populations. Phylogenetic studies using DNA sequences PCR-amplified directly from nodules showed that the *Ceanothus*-microsymbiont is closely related to *Elaeagnus*-infective or *Phurshia* microsymbiont *Frankia* and distantly related to *Alnus*-infective *Frankia* (Murry et al., 1997; Bensen et al., 1996; Chapter 2). However, these diversity studies still raise questions such as how many *Ceanothus*-infective *Frankia* strains exist in a handful of soil. More quantitative estimates are needed if the ecology of *Ceanothus*-microsymbiont *Frankia* is to be better understood.

The objective of this study was to elucidate how host plants affect *Frankia* population size and diversity. No study has been reported regarding population dynamics of *Frankia* in soil using actinorhizal plants other than *Alnus* species. We compared soils under a host plant, *C. velutinus,* and under a non-host plant, Douglas-fir. Population sizes of *Ceanothus*-infective *Frankia* in soils were measured using plant bioassays. Diversity of *Ceanothus*-microsymbiont *Frankia* was examined using rep-PCR.

Previous studies have used plant bioassay and PCR-MPN (most-probable number) methods to measure the population size of *Frankia* (e.g. van Dijk, Ph.D. Thesis, State University, Leiden, the Netherlands, 1984; Picard et al., 1990; Myrold and Huss-Danell, 1994). PCR-MPN methods use *Frankia*-specific primers and allow one to measure population size directly from soils or plant nodules. However, PCR-MPN methods still need improvement because PCR primers may amplify DNA of closely related bacteria. A widely used plant bioassay based on counting nodules of trap plants offers an indirect measure of *Frankia* population size in soils. So far, the plant bioassay is the only method for measuring population size of host-infective *Frankia.*

PCR-based methods can eliminate the need to isolate microorganisms in pure culture for studying the diversity and structure of microbial communities. An example is rep-PCR, a PCR technique with primers that correspond to ubiquitous repetitive DNA sequences (Versalovic et al. 1991). Rep-PCR is a highly specific genomic fingerprinting

technique. Recently, rep-PCR has been utilized to identify cultured *Frankia* strains with resolution at the strain level (Murry et al., 1995) and to assess the diversity of the *Frankia* microsymbiont in nodules from *Ceanothus* species (Murry et al., 1997). Here we used rep-PCR techniques to assess the diversity of *Ceanothus*-infective *Frankia*.

5.3 Materials and Methods

5.3.1 Soil collection

Soil samples were collected from a 20-year-old *C. velutinus* stand and a nearby Douglas-fir stand more than 100 years old at the H.J. Andrews Experimental Forest (44°09' N; 122°22' W) in Oregon. Three 3-4 kg soil cores (0 - 20 cm depth) taken near the roots of two plants were collected 10 m apart within the *C. velutinus* stand. Nodules were also collected to inoculate positive control plants. At the Douglas-fir stand, three 3-4 kg soil cores were collected from three sites 10 m apart. Different collection tools were used between stands to avoid cross-contamination. Samples were transported in an ice box and stored at -20°C. The soil chemical properties were analyzed by the Central Analytical Laboratory at Oregon State University (Corvallis, OR) (Table 5.1).

5.3.2 Plant material and growth conditions

Seeds of *C. velutinus*, *C. sanguineus* (Pursh), and *C. integerrimus* (H. & A.) were soaked for 1 hr in 2.5 % sodium hypochlorite solution and then scarified by placing approximately 500 seeds in 200 ml boiling deionized water for 1 min, removing them from the heat and slowly cooling until room temperature. The seeds were then stratified by wrapping them in several layers of moistened, sterile cheese cloth and holding at 4 °C for 60 days. Seeds were then transferred to germinate in sterilized, moistened Perlite at 21 °C under 25 $\mu mol\ m^{-2}s^{-1}$ fluorescent lighting. Seedlings were grown until one true leaf developed and then transferred to a pot containing Perlite:sand (3:1).

Pots were designed to avoid cross-contamination in growth chamber or greenhouse. A pot consisted of two Square Grower Containers (57 mm x 60 mm, S-225, McConkey Co., Sumner, WA), one upon another so there was a 12-mm space between the two

Table 5.1. Characteristics of soil samples.

Chemical property	*C. velutinus* stand	Douglas-fir stand
pH	6.0	6.3
P (mg kg^{-1} soil)	7	7
K (mg kg^{-1} soil)	234	343
Ca (cmol kg^{-1} soil)	3.5	6.8
Mg (cmol kg^{-1} soil)	0.6	1.0
Na (cmol kg^{-1} soil)	0.04	0.04
C (%)	4.4	4.1
N (%)	0.16	0.17

containers to prevent cross-contamination by diffusion of dripping solution while watering. Two plastic nets (1.25-mm mesh) were inserted between the two containers to exclude Perlite or sand from flowing down. The pots were supported on a lattice.

Pots were left for 3 days in the laboratory before inoculation so that the transferred plants became accustomed to the pot environment. Then *C. velutinus* pots were moved to a greenhouse; *C. sanguineus* and *C integerrimus* pots were moved to growth chambers. The greenhouse was held at a 21°C - 16°C day-night regime. Plants were grown from December to March, therefore the day-night period was adjusted to 16 - 8 h using *ca* 65 μmol $m^{-2}s^{-1}$ sodium lamps. Growth conditions for *C. sanguineus* were 26°C - 16°C on a 16 - 8 h day-night regime. The photosynthetic photon flux density was 145 μmol $m^{-2}s^{-1}$ for 3 h at the beginning and end of the day and 270 μmol $m^{-2}s^{-1}$ for 10 h in the middle of the day. Growth conditions for *C. integerrimus* were the same as for *C. sanguineus* except that the temperature regime was 21°C - 16°C. Plants were watered with sterilized tap water everyday and a quarter-strength, modified Hoagland's solution (Huss-Danell and Myrold, 1994) once a week.

5.3.3 Inoculation

Three soil samples from each stand were mixed thoroughly by shaking in a plastic bag. The soils were sieved to remove gravel and plant debris (2 mm-mesh). Soil corresponding to 20 g of dry weight was suspended in quarter-strength Hoagland's solution at a 1:5 dilution. From this suspension, serial dilutions corresponding to 1:5, 1:50, 1:500, and 1:5000 were prepared. At each dilution step, the soil suspension was shaken vigorously by hand. The soil suspensions were added to the root system of test seedlings. One dilution series consisted of 16 seedlings, 4 for each dilution. There were three replicate dilution series for each species.

Negative controls were left uninoculated and served to detect any contaminants. Positive controls were inoculated with a crushed nodule suspension to verify that the growth conditions used would allow nodulation to take place. Additional pots, which contained 4 g and 0.4 g soil, were made to check nodulation over time.

5.3.4 Nodulation and measurement of root and shoot dry weight.

Every week three of the extra pots were destroyed and the roots of plants inspected to determine when nodulation leveled off. *Frankia* populations, or nodulation units (NU) g^{-1} soil, were calculated from plants that had been scored as nodulated according to the most probable number (MPN) method (Koch, 1981). *Frankia* populations were also calculated from the number of nodules per plant according to a nodulation capacity method based on Van Dijk (*loc. cit.*). For measurement of root and shoot weight of seedlings, nodules were first excised from roots. Nodule lobes of some nodules were excised and tested for acetylene reduction activity. Remaining nodules were stored at -20 °C for later DNA extraction. Roots and shoots were separated approximately at the Casparian strip, dried in an oven at 70 °C for 2 days and weighed separately. Statistical analyses were performed with multifactor ANOVA in Statgraphics plus 2.0 (Statistical Graphics corp. Princeton, NJ).

5.3.5 Acetylene reduction activity assay

Commercially purified C_2H_2 (0.3ml) was injected with a plastic syringe into a 3.0-ml tube containing one nodule lobe. After 2 h a 0.1-ml gaseous sample from each tube was removed and analyzed for C_2H_4 and C_2H_2 with a Hewlett-Packard 5830A gas chromatograph (Walnut Park, CA) fitted with a 2 m x 2.1 mm, 80-100 mesh, Porapak R column; oven temperature 70°C. Injection temperature and flame ionization detector temperature were adjusted to 100 °C. Flow rate of the N_2 carrier gas was adjusted to 40 ml min^{-1}.

5.3.6 DNA Extraction

For DNA extraction, bacterial cells from liquid culture of *Frankia* were concentrated centrifuging and washed using TNE (10 mM Tris-Cl, 100 mM NaCl, 1 mM EDTA (pH 8.0) and a nodule lobe from a nodule was excised and surface-sterilized using 2.5% sodium hypochlorite solution. Genomic DNAs were extracted following a modified protocol of the cetyltrimethylammonium bromide (CTAB) procedure (Murray and Thomson, 1980) as described (Chapter 4). Basically, nodule lobes were frozen, ground, extracted with CTAB and chloroform:phenol and precipitated twice with isopropyl

alcohol and once with PEG/NaCl. This method yields sufficient DNA for PCR amplification.

DNA was quantified using Hoechst dye 33258 and a DNA fluorometer (Model TKO 100; Hoefer Scientific Instruments, San Francisco, CA) according to the manufacturers instructions.

5.3.7 Rep-PCR.

Total nodule DNA was amplified using the BOX A1R primer and the DR1R primer as described (Chapter 4). Each PCR experiment included a control lacking template DNA. For each nodule sample amplified, a control utilizing DNA isolated from root sections adjacent to that nodule was included. The amplification products were separated by gel electrophoresis and stained with ethidium bromide. The band patterns were converted into a binary matrix and analyzed using NTSYS software (Applied Biostatistic Inc.).

5.4 Results

5.4.1 Measurement of Frankia population size by plant nodulation

It was necessary to do preliminary experiments to design appropriate nodulation systems, because no previous bioassay had been reported using *Ceanothus* species. For example, we grew *C. velutinus* in a hydroponic culture system for 12 months as in bioassay system used for *Alnus* species (Huss-Danell and Myrold, 1994). However, we did not observe any nodulation. On the basis of a previous inoculation study with *Alnus rubra* Bong., in which almost no nodules were observed on plants in negative control plot (Crannell et al., 1994), we evaluated the growth of *Ceanothus* plants in Perlite:sand (3:1) pots in a greenhouse and a growth chamber. Sixteen plants of each *Ceanothus* species were grown in pots, eight were inoculated with 4 g of soil and eight with 20 g of soil from the *C. velutinus* stand. Half of the plants were grown in the greenhouse and half in the growth chamber. The pots were maintained for 11 weeks. In the greenhouse, all three species grew well. Only *C. velutinus* plants were nodulated in both 4 g and 20 g soil

treatment plants, however, and no *C. integerrimus* and *C. sanguineus* plants were nodulated. In the growth chamber, all *C. sanguineus* plants grew well, whereas all *C. velutinus* plants and half of the *C. integerrimus* plants died. All *C. sanguineus* plants and some healthy *C. integerrimus* plants were nodulated. These results may have been caused by different moisture and temperature requirements of the three *Ceanothus* species. Thus, we used three different environmental systems for the three *Ceanothus* species in subsequent plant bioassay experiments.

Nodules of three *Ceanothus* species were never observed on negative controls and the most dilute inoculum but always on positive controls. Nodulation occurred on at least one dilution within every bioassay replicate. In this study, the highest frequency of nodulated plants and the highest number of nodules per plant were found with the least dilute inoculum. Thus, our experiments reflect the number of NU in these soils.

The shoots of all nodulated plants were larger and darker green than those of non-nodulated plants. Root and shoot weights of all plants were measured (Table 5.2). A completely randomized three-way analysis of variance was then done. On this basis there were some significant main effects ($p < 0.05$), but none of the interactions was significant ($p > 0.05$). There was no significant difference in shoot weight between the soils, however, there were significant difference in shoot weight among *Ceanothus* species and among soil dilution treatments. There was a significant difference in root weight between the soils, however, there were no significant differences among *Ceanothus* species or among soil dilution treatments in root weight. Taken together, the results suggest that plant nodulation increases shoot weight rather than root weight.

We estimated N_2-fixation ability of individual nodule lobes separated from nodules of selected plants by acetylene reduction assay (Hardy et al., 1968). Some level of acetylene reduction activity was detected in 34 out of 37 nodule lobes tested, indicating that the *Ceanothus* nodules obtained from our bioassay system could fix N_2. Mean levels of acetylene reduction rate of 11 nodule lobes of *C. integerrimus*, 19 nodule lobes of *C. sanguineus*, and 7 nodule lobes of *C. velutinus* ranged from 2.8 to 6.6 $\mu mol\ h^{-1}\ g^{-1}$ (Table 5.3). Because of high variability (coefficients of variation from 23% to 36%), no significant difference among *Ceanothus* species or between two soils was found in two-way ANOVA (Table 5.3).

Table 5.2. Comparison of shoot and root dry weight in terms of bioassay species, stand type, and soil dilution. Within a factor and a column, weights followed by different letters are significantly different based on LSD (p = 0.05)

Factor	Shoot dry weight	Root dry weight
	----------g----------	
Species		
C. velutinus	0.013[a]	0.014[a]
C. sanguineus	0.036[c]	0.028[b]
C. integerrimus	0.026[b]	0.028[b]
Soil		
C.velutinus stand	0.025[a]	0.021[a]
Douglas-fir stand	0.025[a]	0.025[b]
Soil dilution		
1:5	0.031[a]	0.024[a]
1:50	0.026[b]	0.023[a]
1:500	0.021[c]	0.022[a]
1:5000	0.022[c]	0.024[a]

Table 5.3. Comparison of acetylene reduction rate of nodule lobes from three *Ceanothus* species.

Factor	Acetylene reduction rate[a]
	-------μmol $h^{-1}g^{-1}$ soil-------
Species	
C. velutinus	6.4 ± 1.7
C. sanguineus	2.8 ± 1.0
C. integerrimus	6.6 ± 1.5
Soil	
C. velutinus stand	4.8 ± 1.1
Douglas-fir stand	5.7 ± 1.3

[a] Mean ± standard deviation

The first nodules appeared on the three species at different times. Eight weeks after inoculation nodules were observed on *C. sanguineus* and *C. integerrimus* in the growth chambers and after 10 weeks on *C. velutinus* in the greenhouse. Nodulation leveled off 3-4 weeks after observing the first nodulation. Thus we counted nodule numbers of *C. sanguineus* and *C. integerrimus* 12 weeks after inoculation and those of *C. velutinus* 14 weeks after inoculation. We observed nodulation from all positive control pots after 14 weeks. However, we did not observe any nodulation from negative controls or the highest diluted soil pots for over 6 months.

Mean numbers of *Frankia* NU ranged from 0.2 to 5.2 g^{-1} soil when calculated by the nodulation capacity method. Using three replicates of each soil sample resulted in adequate precision, with coefficients of variation ranging from 2 to 21%. Mean numbers of *Frankia* NU ranged from 3.6 to 5.2 among the three bioassay species for soil under *C. velutinus* stand; 0.2 to 0.4 for soil under Douglas-fir stand. Within each bioassay species, the mean numbers of *Frankia* NU for soil under the *C. velutinus* stand was 8- to 19-fold higher than those for soil under the Douglas-fir stand. Mean numbers of *Frankia* NU calculated by the MPN method ranged from 0.9 to 2.4 for soil under the *C. velutinus* stand; 0.2 to 0.3 for soil under the Douglas-fir stand. Within each bioassay species, the mean *Frankia* NU for soil under the *C. velutinus* stand was 5- to 11-fold higher than those for soil under the Douglas-fir stand. The overall *Frankia* NU calculated by nodulation capacity were significantly higher than those by MPN methods, as shown in Fig. 5.1. A two-way analysis of variance of *Frankia* NU for the species and soil were then carried out (Table 5.4). On this basis the *Frankia* NU was significantly different between the two soils, but not significantly different among the three species and the interaction between the soil and plant species was not significant. The MPN method and the nodulation capacity method led to the same conclusion: The population size of *Ceanothus*-infective *Frankia* of soil under *C. velutinus* stand is several-fold higher than that of *Ceanothus*-infective *Frankia* of soil under Douglas-fir stand.

5.4.2 Frankia diversity determined by the Rep-PCR

The fingerprint patterns of nodule and bacterial DNAs generated by BOX and DR primers are shown in Fig. 5.2. BOX- and DR-PCR yielded 8 to more than 15 distinct

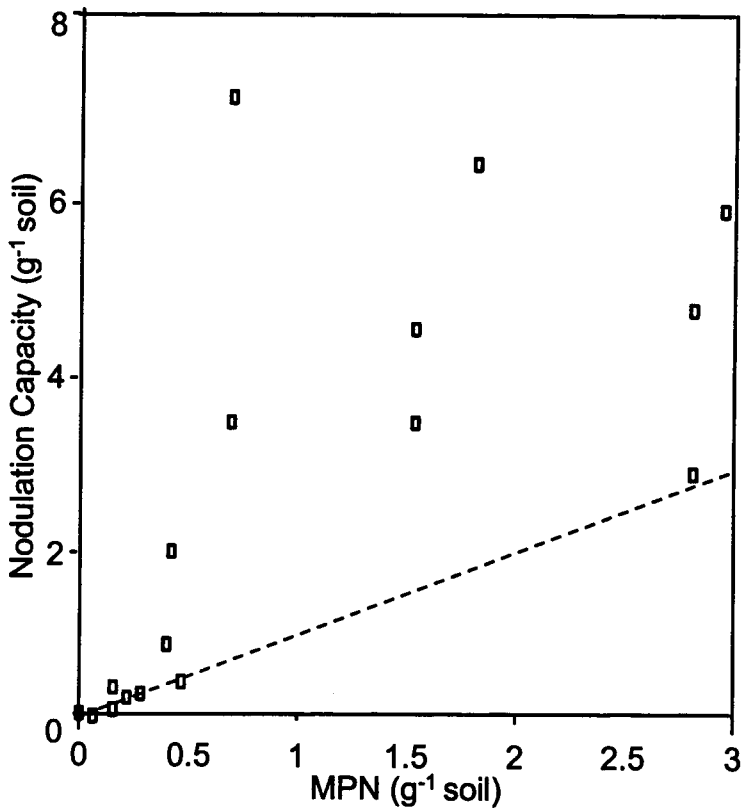


Fig. 5.1. Relationship between NUs determined by the nodulation capacity and the MPN methods from *C. velutinus* and Douglas-fir stand soils using three *Ceanothus* species. Data are from all replicates (n = 18) of the three *Ceanothus* species treated with soil from *C. velutinus* and Douglas-fir stands.

Table 5.4. Mean NU calculated using nodulation capacity and MPN method for three *Ceanothus* species using soil from *Ceanothus velutinus* and Douglas-fir stands. Within a factor and a column, populations followed by different letters are significantly different based on LSD (p = 0.05).

Factor	Nodulation capacity	MPN
	----------NU g^{-1} soil----------	
Species		
C. velutinus	2.0[a]	1.1[a]
C. sanguineus	2.9[a]	1.2[a]
C. integerrimus	2.2[a]	0.5[a]
Soil		
C. velutinus stand	4.4[a]	1.7[a]
Douglas-fir stand	0.3[b]	0.2[b]

products, ranging in size from approximately 100 bp to over 2 kbp. Differences among nodules were assessed visually on the basis of the banding patterns.

The BOX- and DR-PCR profiles of *Ceanothus* nodule DNAs were found to be completely different from those of two reference strains, *Frankia* AvcI1 and EaI12, corresponding with the previous reports that *Ceanothus* microsymbiont *Frankia* are distantly related on the basis of DNA sequences to *Frankia* isolated to date (Chapter 2, Chpater 4, Murry et al., 1997). The presence of several comigrating bands of nodule DNA suggested that *Ceanothus* microsymbionts are closely related. At least four BOX-PCR products from nodule DNA comigrated on the agarose gel. DR-PCR also gave multiple bands of equal mobility. However, the overall patterns from many nodule DNAs were distinct. In total over 49 distinct bands were visualized. Thus, the relationship among *Ceanothus*-infective *Frankia* could be surmised from the BOX and DR-PCR fingerprint patterns.

To determine the host specificity of *Ceanothus*-infective *Frankia* and to compare their diversity between *C. velutinus* stand and Douglas-fir soils, we examined fingerprint patterns of nodule DNAs collected from trap plants of three different *Ceanothus* species (Table 5.5).

According to the principal component analysis, the projection of the fingerprints of the 69 nodules on the plot defined by the first two principal components, which accounts for 11% and 10% of the total variance, respectively, reveals the presence of three major groups, named A, B, and C (Fig. 5.3). Fingerprints from the Douglas-fir soil were found mainly in A (61%) and B (32%) with only a few in C; fingerprints from the *C. velutinus* soil only in B (73%) and C (27%). Most fingerprints of cluster C were generated from nodule DNAs of *C. velutinus* stand soils. However, only fingerprints from nodule DNAs of Douglas-fir stand soil gathered in cluster A. In terms of trap plant species, most fingerprints of nodule DNAs from *C. velutinus* were observed in clusters A (31%) and B (54%) and a few were observed in group C (8%); those from C. *sanguineus* only in clusters A (35%) and B (65%); and those of *C. integerrimus* in clusters B (50%) and C (45%). Distribution of fingerprints by three groups within the two stand soils or three plant species were tabulated using two-way contingency tables and associations determined by analysis of frequencies using chi-square statistic (Snedecor and Cochran,

Fig. 5.2. A representive ethidium bromide-stained agarose gel patterns of BOX- (panel A)and DR-PCR (panel B) products from nodules of *C. velutinus C. sanguineus*, and *C. integerrimus Frankia*-inoculated with soils under *C. velutinus* and Douglas-fir stands. Panel A: Lanes 1, 3, and 5 show the BOX-PCR products obtained from nodule of *C. velutinus* (lane 1), *C. sanguineus* (lane 3) and *C. integerrimus* (lane 5) inoculated with soils under the *C. valutinus* stand; Lanes 2, 4, and 6 show the BOX-PCR products obtained from nodule of *C. velutinus* (lane 2), *C. sanguineus* (lane 4) and *C. integerrimus* (lane 6) obtained with soils under the Douglas-fir stand; lane 7 is PCR products of EaI12 and lane 8 CPI1. Panel B: Lanes 1, 3, and 5 show the DR-PCR products obtained from nodule of *C. velutinus* (lane1), *C. sanguineus* (lane 3) and *C. integerrimus* (lane 5) inoculated with soils under the Douglas-fir stand; Lane 2, 4, and 6 show the DR-PCR products obtained from nodule of *C. velutinus* (lane 2), *C. sanguineus* (lane 4) and *C. integerrimus* (lane 6) inoculated with soils under the *C. valutinus* stand; lane 7 show the PCR products of EaI12 and lane 8 *E. coli* DNA. In both panel A and B, lane M show the DNA molecular size marker (100 bp-ladder, Gibco-BRL) and lane 9 is a negative control lacking template DNA

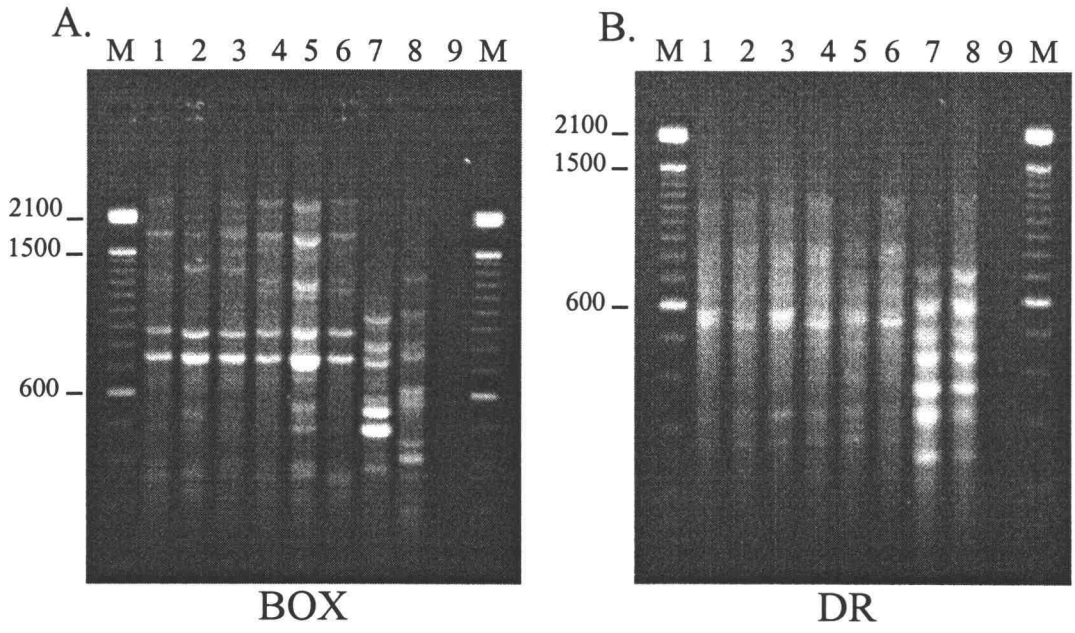

A.
M 1 2 3 4 5 6 7 8 9 M
2100
1500
600
BOX
B.
M 1 2 3 4 5 6 7 8 9 M
2100
1500
600
DR

Fig. 5.2

Table 5.5. Number of nodules analyzed by BOX- and DR-PCRs.

Species	Number of nodule	
	C. velutinus stand soil treatment	Douglas-fir stand soil treatment
C. velutinus	13	13
C. sanguineus	11	12
C. integerrimus	14	6

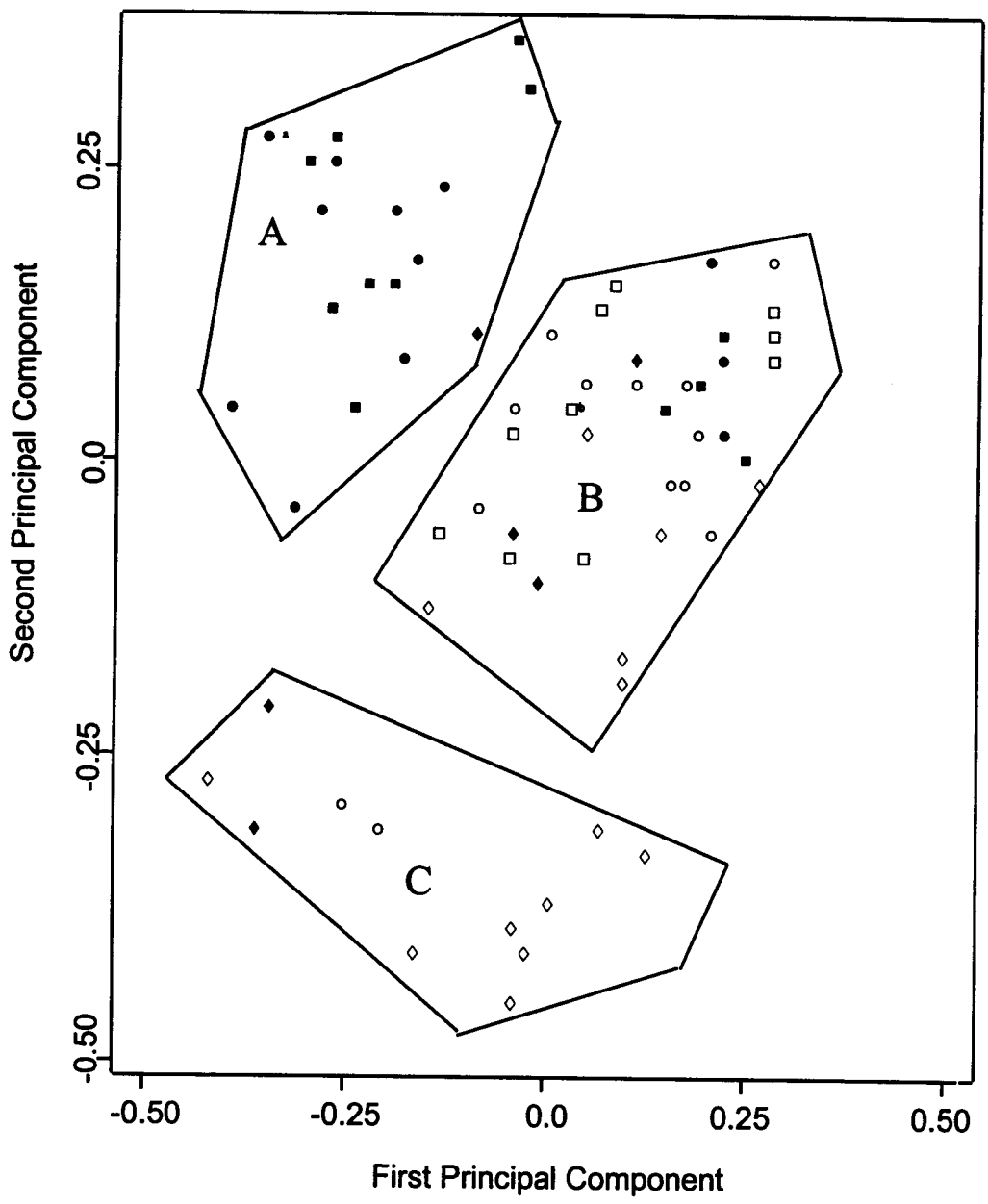


Fig. 5.3. Two-dimensional plot of fingerprint patterns of *C. velutinus* (○), *C. sanguineus* (□), and *C. integerrimus* (◇) nodules by principal component analysys. Open symbols correspond to the fingerprint patterns obtained from nodules of the plants treated with soils under *C. velutinus* stand and solid symbols Douglas-fir stand. Superscripts a and b indicate two identical patterns obtained from two *C. velutinus* nodules and two *C. sanguineus* nodules, respeciviely. The first and second principal component account for 11.3% and 10.3%, respectively, of total variance.

1980). On this basis, there was a significant interaction between fingerprint groups regarding the stand soil ($P < 0.01$) with group A different from groups B or C. There was no significant interaction between fingerprint groups A and B regarding *Ceanothus* species. However, there was a significant interaction between groups C and A or B ($P < 0.01$) probably because of fewer number of fingerprints from *C. integerrimus* nodules treated with the Douglas-fir stand soil and group C has no fingerprints from *C. sanguineus*.

5.5 Discussion

Previous studies showed that *Frankia* NUs were not affected significantly by host plant, leading to a thought that *Frankia* persist without a host plant as a saprophyte. For example, in a study of Swedish soils, Myrold and Huss-Danell (1994) found that alder (a host plant species) and lupine soils were lower in infective *Frankia* population size than the no-tree control. In fact, they found that the greatest change in *Frankia* population size occurred seasonally. Smolander and Sundman (1987) found that there was higher *Frankia* population size in birch soils than in alder soils on Finnish forest soils. In this study, the *Frankia* population size of the *C. velutinus* stand soil was about 10 times higher than that of the Douglas-fir stand soil, indicating that *C. velutinus* may have increased the size of soil *Frankia* population. This result supports an observation by Wollum et al. (1968) that the nodulation capacity of the soil decreases over 300 years, after *Ceanothus* dropped out of the stand after development of a timber stand.

The principal component analysis of BOX- and DR-PCR fingerprints indicated that *Frankia* from our soil samples do not show host specificity to three *Ceanothus* species. It also suggested that two different soils show similar degree of diversity, however the soils contain different strains of *Frankia*. In other words, we found that two contrasting soils contained a common *Frankia* group, however the two soils also contained unique *Frankia* strains. The result is consistent to a conclusion of Baker and Mullin (1994) and Murry et al. (1997) that diversity of *Ceanothus*-microsymbiont *Frankia* is not related to

differences of *Ceanothus* species or population. However, this study indicated that different soils contain different *Frankia* strains, in contrast to the previous reports (Baker and Mullin, 1994; Murry et al., 1997) that *Ceanothus*-microsymbiont *Frankia* is not related to geographic origins. The discrepancy may come from greater sampling intensity because we analyzed many more nodules than in the previous studies.

Several lines of evidence indicate that there are no specific associations between three *Ceanothus* species and their infective *Frankia*. First, we found no significant intrageneric variation in the quantification of NU from a soil sample and between the two soil samples. This is interesting because a significant intrageneric variation in nodulation of *Alnus* was found (Huss-Danell and Myrold, 1994). Second, we found no significant difference between *Ceanothus* species and between two soils in acetylene reduction activity. Dillon and Baker (1982) found that both microsymbiont and host influenced the acetylene reduction rate, but they compared host:microsymbiont sources from different species and genera. Nelson and Lopez (1989) found that variability in the N_2 fixation rate is more a function of the soil source than the *C. velutinus* source population in actinorhizal associations. Third, fingerprinting of nodules using BOX- and DR-PCR did not show host-specific grouping, although fingerprints showed some degree of difference between the two soils.

Nodulation units estimated by the nodulation capacity method were consistently higher than those estimated by MPN plant bioassay methods even though they gave a similar trend in NU between *C. velutinus* stand and Douglas-fir stand. A study by Huss-Danell and Myrold (1994) showed that agreement between the two methods was especially good between 30 and 300 NU g^{-1} soil although they are based on different assumptions. Huss-Dannell and Myrold (1994) also suggest that *Frankia* NU may be underestimated above this population level by the nodulation capacity method perhaps because of saturation of nodulation sites and, inversely, at lower population levels by the MPN method perhaps because of difficulty of dilution of potentially infective *Frankia*. Our results showed that *Frankia* NU values estimated by the nodulation capacity method were consistently higher than those estimated by the MPN method. Thus, in this study, the discrepancy between the two estimates may come from low population levels in our soil samples, ranging from 0.2 to 5.2 mean NUs g^{-1} soil.

Both *Frankia* NU values calculated by nodulation capacity and MPN methods in this study were at the low end of the range obtained by previous studies in which alders were used as trap plants. For example, the NUs varied from 4 to 400 NU g^{-1} soil for the soil from the no-tree control plot (Huss-Danell and Myrold, 1994) or 0-900 NU cm^{-3} of soil for the soil under some coniferous species (Smolander, 1990). However, low NUs in this study are most likely due to low population levels. For example, we consistently observed nodulation on the least dilute inoculum but never on the most dilute inoculum or negative control. NUs calculated by the MPN method is consistently lower than those by the nodulation capacity (Fig. 5.1), as previously noticed by Myrold and Huss-Danell (1994). Interestingly, Markham and Chanway (1996) observed a similar range of NUs, which varied from 0 to 18.9 NU cm^{-3} of soil, in a study of soils from harvested forest sites apparently because of the low population levels in the soils.

Many actinorhizal plants have been used for inoculation with isolated *Frankia* or for trapping soil *Frankia* (e.g. Baker, 1987). However, all studies on *Frankia* populations in soils have used *Alnus* species as a trap plant to quantify population size (e.g. Smolander and Sundman, 1987; Huss-Danell and Myrold, 1994). *Alnus* species grow well in hydroponic culture so that nodulation incidence can be easily observed. Here we demonstrated that *Ceanothus* species can be used as trap plants to assay plant nodulation using a pot system with solid media.

The difference of *Frankia* population size between two soils are unlikely due to differences in soil chemical properties, because the soil properties from the soil chemical analysis were quite similar between the two soils except Ca and K levels (Table 5.1). It has been suggested that nodulation rate or numbers of infective *Frankia* is significantly correlated to soil Ca levels or increase in response to the addition of Ca to soils (Scott, 1973; Hilger and Myrold, 1992; Crannell et al., 1994). However, in this study host plant seems to play more important role in regulating *Frankia* population size, because the soil from the Douglas-fir stand had two-fold higher in Ca levels but was about 10 times lower in *Frankia* NU than that from the *C. velutinus* stand. Although previous studies have suggested possible regulators of *Frankia* population in soil, there is no general agreement about what soil chemical factor is the major important regulator. For example, Smolander and Sundman (1987) found that soil pH was the major regulator of the size of infective

Frankia populations on Finnish forest soils with a range in pH 3.5-4.4. However, Myrold and Huss-Danell (1994) found that C, rather than pH, was regulating the size of the infective *Frankia* populations on Swedish soils with a narrower range of pH (4.66 - 4.93).

Most previous ecological studies of actinorhizal symbiosis were based on analysis of *Frankia* isolates from *Alnus* (Benson and Hanna, 1983), *Myrica* (Bloom et al., 1989), and *Elaeagnus* (Jamann, 1992), from which *Frankia* can be easily isolated. The lack of infective *Ceanothus* isolates of *Frankia* has hindered ecological studies of this particular actinorhizal symbiosis, particularly those focused on *Frankia.* Recent applications of molecular techniques have allowed to circumvent *Frankia* isolation. The first study was the use of a strain-specific *nif*H sequences to probe nodule DNA extracts (Simonet et al., 1990). The use of RFLP analysis of nodule DNA extracts (Baker and Mullin, 1994) was used to examine the population structure of *Frankia* without isolation, as can a newer variant of this, PCR-RFLP (Maggia et al., 1992; Jamann et al., 1993; Rouvier et al., 1996). The polymerase chain reaction (PCR) has been utilized to amplify specific regions of the *Frankia* genome for phylogenetic analysis based on nucleic acid sequence comparisons (e.g. Simonet et al., 1991; Nazaret et al., 1991; Cournoyer et al., 1993). Here we demonstrated that without *Frankia* isolation BOX- and DR-PCR can be utilized to quantify the diversity of *Ceanothus*-infective *Frankia* trapped from soils, supporting and expanding the previous report by Murry et al. (1997) that there was considerable diversity in the BOX-PCR patterns among nodules sampled between species, individuals of the same species and between sites.

DNA sequences of the intergenic spacer region between 16S rRNA and 23S rRNA genes suggested that *Ceanothus*-microsymbiont *Frankia* from several sites in Oregon are related at the intraspecific level (Chapter 4). Although several comigrating bands among rep-PCR patterns indicated close relationship among *Ceanothus*-infective *Frankia* studied here, we observed 49 distinct bands from rep-PCR fingerprints of 69 nodules. This observation is not surprising according to previous studies that rep-PCRs revealed high genetic polymorphisms at the pathovar or genotype level (e.g. Louws et al., 1994; Latour et al., 1996; Frey et al., 1997). For example, Frey et al. (1997) found 16 different rep-PCR fingerprints from one 16S rDNA type identified by restriction fragment length polymorphism

In conclusion, the plant bioassay results suggest that *Ceanothus*, a host plant of *Frankia,* supports *Frankia* persistence better than Douglas-fir, a non-host plant. Despite the large differences in *Frankia* populations between the two sites, the overall diversity of *Frankia* at each site was similar, although each site harbored some unique rep-PCR types. It is interesting that *Frankia* strains from both soils can nodulate three *Ceanothus* species equally well, as evidenced by the quantification of *Frankia* population sizes, acetylene reduction activity, genomic fingerprinting using rep-PCRs.

5.6 References

Baker D.D. and Mullin B.C. (1994) Diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus* as assessed by RFLP patterns from single nodule lobes. *Soil Biology and Biochemistry* **26,** 547-552.

Baker, D. and O'Keefe D. (1984) A modfied sucrose fractionation procedure for the isolation of frankiae from actinorhizal root nodules and soil samples. *Plant and Soil* **78,** 23-28.

Benecke U. (1969) Symbiots of alder nodules in New Zealand. *Plant and Soil* **30,** 145-149.

Benson D. R. and Hanna D. (1983) *Frankia* diversity in an alder stand as estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis of whole-cell proteins. *Canadian Journal of Botany* **61,** 2919-2923.

Bensen D. R., Stephens D. W., Clawson M. L. and Silvester W. B. (1996). Amplication of 16S rRNA genes from Frankia strains in root nodules of *Ceanothus griseus, Coriaria arborea, Coriaria plumosa, Discaria toumatou,* and *Purshia tridenta.* Applied and Environmental Microbiology **62,** 2904-2909.

Bloom R. A., Mullin B. C. and Tate R. L. III. (1989) DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (bayberry). *Applied and Environmental Microbiology* **55,** 2155-2160.

Callaham D., DelTredici P. and Torrey J. G. (1978) Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia. Science* **199,** 899-902.

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

Crannell W. K., Tanaka Y. and Myrold D. D. (1994) Calcium and pH interaction on root nodulation of nursery grown red alder (*Alnus rubra* Bong.) seedlings by *Frankia*. *Soil Biology and. Biochemistry* **26,** 607-614.

Dillon J. T. and Baker D. (1992) Variation in nitrogenase activity among pure-cultured *Frankia* strains tested on actinorhizal plants as an indication of symbiotic compatibility. *New phytologists* **92,** 215-219.

Frey P., Frey-Klett P., Garbaye J., Berge O. and Heulin T. (1997) Metabolic and genotypic fingerprinting of fluorescent pseudomonads associated with the Douglar fir-*Laccaria bicolor* mycorrizosphere. *Applied and Environmental Microbiology* **63,** 1852-1860.

Hardy R. W. F., Holsten R. D., Jackson E. K. and Burns RC (1968) The acetylene-ethylene assay for N_2-fixation: Laboratory and field evaluation. *Plant Physiology* **43,** 1185-1207.

Hibbs D. E. and Cromack K. Jr. (1990) Actinorhizal plants in Pacific Northwest forests. In *The biology of Frankia and Actinorhizal Plants* (C. R. Scrwhitzer and J. D. Tjepkema, Ed), pp 343-363. Academic Press Inc., San Diego, CA.

Hilger, A. B., and Myrold D. D. (1992) Quantitation of soil *Frankia* by bioassay and gene probe methods: response to host and non-host rhizospheres and liming. *Acta Œcologia* **13,** 505-506.

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

Jamann S., Fernandez M. P. and Moiroud A. (1992) Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. *Acta Œcologia 13,* 395-406.

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

Latour X., Corberand T., Laguerre G., Allard F. and Lemanceau P. (1996) The comparison of fluorescent pseudomonad populations associated with roots is influenced by plant and soil type. *Applied and Environmental Microbiology* **62,** 2449-2456.

Louws F. J., Fulbright D. W., Stephens C. T., and de Bruijn F. J (1994) Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. *Applied and Environmental Microbiology* **60,** 2286-2295.

Maggia L., Nazaret S. and Simonet P. (1992) Molecular characterization of *Frankia* isolates from *Casuarina equisetifilia* root nodules harvested in West Africa (Senegal and Gambia). Acta Œ*cologia* **13,** 453-462.

Markham, J. H. and Chanway C. P.(1996) *Alnus rubra* nodulation capacity of soil under five species from harvested forest sites in coastal British Columbia. *Plant and Soil* **178,** 283-286.

Murray M. G. and Thomson W. F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **8,** 4321-4324.

Murry M. A., Zhang D., Schneider M. and de Bruijn F. J. (1995) Use of repetitive sequences and the polymerase chain reaction (rep-PCR) to fingerprint the genomes of *Frankia* isolates. *Symbiosis.*

Murry M. A., Konopka A. S., Pratt S. D. and Vandergon T. L. (1997). The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophyte of the actinorhizal shrub *Ceanothus. Physiologia Plantarum* (in press).

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

Nazaret S., Cournoyer B., Normand P. and Simonet P. (1991) Phylogenetic relationships among *Frankia* genome species determined by use of amplified 16S rDNA sequences. *Journal of Bacteriology* **173,** 4072-4078.

Nelson D. L. and Lopez C. F. (1989). Variation in nitrogen fixation among populations of *Frankia* sp. and *Ceanothus* sp. in actinorhizal association. *Biology and. Fertility of Soils* **7,** 269-274.

Picard C., Ponsonnet C., Paget E., Nesme X. and Simonet P. (1990) Detection and enumeration of bacteria in soil direct DNA extraction and polymerase chain reaction. *Applied and Environmental Microbiology* **58,** 2717-2722.

Rouvier C., Prin Y., Reddell P., Normand P. and Simonet P. (1996) Genetic diversity among *Frankia* strains nodulating members of the family Casuarinaceae in Australia revealed by PCR and restriction fragment length polymerphism analysis with crushed root nodules. *Applied and Environmental Microbiology* **62,** 979-985.

Scott W. (1973) Some soil factors affecting snowbrush nodulation. Ph.D. Thesis. Oregon State Univ., Corvallis, Oregon (Dissertation Abstract **33,** 4074).

Simonet P.,. Grosjean M.-C, Misra A.K., Nazaret S., Cournoyer B. and Normand P.. 1991. *Frankia* genus-specific characterization by polymerase chain reaction. *Applied. and Environmental Microbiology* **57,** 3278-3286.

Simonet P., Normand P., Moiroud A., and Bardin R.. 1990. Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. *Archives in Microbiology* **153,** 235-240.

Smolander A. and Sundman V. (1987) *Frankia* in acid soils of forests devoid of actinorhizal plants. *Physiologia Plantarum* **70,** 297-303.

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

Snedecor G. W. and Cochran W. G. (1980) *Statistical Methods.* Iowa State Univ. Press, Ames, IA.

Thomas K.A. and Berry A.M. (1989) Effects of continuous nitrogen application and nitrogen preconditioning on nodulation and growth of *Ceanothus griseus* var *horizontalis. Plant and Soil* **118,** 181-187.

Tyson J. H. and Silver W. S. (1979). Relationship of ultrastructure to acetylene reduction (nitrogen fixation) in root nodules of *Csuarina. Botanical Gazette* (Suppliment) S44-S48.

Versalovic J., Koeuth T. and Lupski J. R. (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Research* **19,** 6823-6831.

Wollum A. G. II, Youngberg C. T. and Chichester F. W. (1968) Relation of previous timber stand age to nodulation of *Ceanothus velutinus. Forest Science* **14,** 114-118.

Youngberg C. T. and Wollum A. G. II. (1976) Nitrogen accretion in developing *Ceanothus velutinus* stands. *Soil Society of American Journal* **40,** 109-112.

Zavitkovski J. and Newton M. (1968) Ecological importance of snowbrush *Ceanothus velutinus* in the Oregon Cascades. *Ecology* **49,** 1134-1145.

6. Conclusions

The comparison of *Frankia* phylogenetic trees reconstructed using *nif*H and 16S rDNA sequences showed that subgroupings of both trees correspond with each other in terms of plant origins of *Frankia* strains, except for a *Ceanothus* microsymbiont *Frankia*. The comparison of *Frankia* and plant phylogenetic trees reconstructed using 16S rDNA and *rbc*L sequences indicated that subgroupings of *Frankia* and actinorhizal plants correspond with each other in terms of symbiotic partnership. According to tree matching, estimated divergence times, and molecular clock hypothesis tests, *Frankia* clades diverged more recently than plant clades and actinorhizal symbioses originated more than three times after the four plant clades diverged.

Parsimony analysis of the *ndh*F sequences identified two main clades corresponding to two subgenera *Ceanothus* and *Cerastes*. Phylogenetic results suggest that three monophyletic clades within subgenus *Ceanothus* can be delimited on the basis of 1) evergreen or 2) deciduous leaves and 3) thorn presence within the evergreen clade. The divergence time estimated by using *rbc*L sequences suggests that the two subgenera diverged 18 - 39 million years ago whereas species within each subgenus diverged more recently. Taken together, the results support the division of *Ceanothus* into two monophyletic subgenera and are consistent with the postulated recent divergence of many species within each subgenus.

The intergenic spacer region between 16S and 23S rRNA gene sequences suggested that *Ceanothus*-microsymbiont *Frankia* are closely related at the intraspecific level. A newly designed, direct repeat sequence and a BOX sequence were used to generate *Frankia*-specific fingerprints from nodule DNA. Fingerprints of nodule DNAs showed that *Ceanothus*-microsymbiont *Frankia* exhibit less diversity within copopulations of host plants than among copopulations, suggesting that geographic separation plays a more important role for divergence of *Ceanothus*-microsymbiont *Frankia* than host plant.

The *Frankia* population size in soil under *C. velutinus* was about 10 times higher than that under Douglas-fir, suggesting that host plants have a positive influence on *Frankia* population size. Nodulation capacities among three *Ceanothus* species used as

trap plants showed no significant difference. Most nodules could fix N_2 according to the acetylene reduction activity assay. Fingerprints of nodule DNA suggested that there was no host specificity among *Ceanothus*-infective *Frankia* in the two soils. Results also suggest that the two soil shared a large, common group of *Frankia* but there were also *Frankia* strains unique to the Douglas-fir soil.

Bibliography

Akimov, V. N., and S. V. Dobritsa (1992). Grouping of *Frankia* strains on the basis of DNA relatedness. System. Appl. Microbiol. 15:372-379

Albert, V. A., A. Backlund, K. Bremer, M. W. Chase, J. R. Manhart, B. D. Mishler, K. C. Nixon (1994). Functional constraints and *rbc*L evidence for land plant phylogeny. Ann. Mo. Bot. Gard. 81:534-567

Baker, D. D. (1982). A cumulative listing of isolated frankiae, the symbiotic, nitrogen-fixing actinomycetes. Actinomycetes 17:35-42

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

Baker, D. D., and B. C. Mullin (1994). Diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus* as assessed by RFLP patterns from single nodule lobes. Soil Biol. Biochem. 26:547-552

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

Baker, D., J. G. Torrey, and G. H. Kidd (1979). Isolation by sucrose-density fractionation and cultivation *in vitro* of actinomycetes from notrogen-fixing root nodules. Nature 281:76-78

Becking J. H. (1977). Dinitrogen-fixing associations in higher plants other than legumes. In Hardy RWF, and Silver WS (eds.) A treatise on Dinitrogen Fixation Section 3, Biology. Wiley, New York, pp 185-275

Benecke, U. (1969). Symbiots of alder nodules in New Zealand. Plant Soil 30: 145-149

Benson, D .R., and D. Hanna (1983). *Frankia* diversity in an alder stand as estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis of whole-cell proteins. Can. J. Bot. 61:2919-2923

Benson, D. R., and W. B. Silvester (1993). Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev. 57:293-319

Benson, D. R., Stephens, D. W., Clawson, M. L., and Silvester, W. B. (1996). Amplification of 16S rRNA genes from *Frankia* strains in root nodules of *Ceanothus griseus*, *Coriaria arborea*, *Coriaria plumosa*, *Discaria toumatou*, and *Purshia tridentata*. Appl. Environ. Microbiol. 62:2904-2909

Berry, A. M., L. McIntyre, and M. E. McCully (1986). Fine structure of root hair infection leading to nodulation in the *Frankia-Alnus* symbiosis. Can. J. Bot. 64:292-305

Binkley, D., K. Cromack, Jr., and R. L. Fredrickson (1982). Nitrogen accretion and availability in some snowbrush ecosystems. For. Sci. 30:26-35

Bloom, R. A., B. C. Mullin, and R. L. Tate III (1989). DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (bayberry). Appl. Environ. Microbiol. 55:2155-2160

Bourquet, S. N., J. R. Valero, M. C. Lavoie, and R. C. Levesque (1995). Comparative analysis of the 16S and 23S ribosomal intergenic spacer sequences of *Bacillus thuringiensis* strains and subspecies and of closely related species. Appl. Environ. Microbiol. 61:1623-1626

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

Brosius, J., M. L. Palmer, J. P. Kennedy, and H. F. Noller (1987). Complete nucleotide sequence of a 16S ribosomal RNA gene from *E. coli*. Proc. Natl. Acad. Sci. USA 75:4801-4805

Brunchorst, J. (1887). Die Struktur der Inhaltskürper in den Zellen einiger Wurzelanschwellungen. Bergens Mus. Aarsb. 235

Burggraaf, A. J. P., and W. A. Shipton (1983). Studies on the growth of *Frankia* isolates in relation to infectivity and nitrogen fixation (acetylene reduction). Can. J. Bot. 61:2774-2782

Burleigh, S., and J. G. Torrey (1990). Effectiveness of different *Frankia* cell types as inocula for the actinorhizal plant *Casuarina*. Appl. Environ. Microbiol. 56:2565-2576

Callaham, D. (1985). Ultrastructure of freese-substituted *Frankia* strain HFPCcI3, the actinomycete isolated from root nodules of *Casuarina cunninghamiana*. Protoplasma 127: 64-72

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

Callaham, D, W. Newcomb, J. G. Torrey, and R. L. Peterson (1979). Root infection in actinomycete-induced root nodule initiation in *Casuarina*, *Myrica*, and *Comptonia*. Bot. Gaz. 140 (Suppl.):S1-S9

Carrasco, A., J. R. Salyards, A. M. Berry (1995). Studies of two *Frankia* strains isolated from *Trevoa trinervis* Miers. Plant Soil 171: 359-363

Caru, M. (1993). Characterization of native *Frankia* strains isolated from Chilean shrubs (Rhamnaceae). Plant Soil 157: 137-145

Chaney R. W. (1927). Geology and paleontology of the Crooked River Basin, with special reference to the Bridge Creek flora. Carnegie Inst. Wash. Publ. 346:45-138

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H .G. Hills, Y.-L. Qui, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, G.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan, V. A. Albert (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. 80:528-580

Chaudhary, A.H., and M. S. Mirza (1987). Isolation and characterization of *Frankia* from nodules of actinorhizal plants of Pakistan. Physiol. Plant. 70: 255-258

Clark L. G., W. Zhang, J. F. Wendel (1995). A phylogeny of the grass family (Poaceae) based on *ndh*F sequence data. Syst. Bot. 20:436-460

Conrad, S. G., A .E. Jaramillo, K. Cromack, Jr., and S. Rose (1985). The role of the genus *Ceanothus* in western forest ecosystems. U.S.D.A. For. Serv. Gen Tech. Rep. PNW-182. 72 p.

Correia, F. F., S. Inouye, and M. Inouye (1986). A 26-base-pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. J. Bacteriol. 167:1009-1015

Cournoyer, B., M. Gouy, and P. Normand (1993). Molecular phylogeny of the symbiotic actinomycetes of the genus *Frankia* matches host-plant infection processes. Mole. Biol. Evol. 10:1303-1316

Crannell, W .K., Y. Tanaka, and D .D. Myrold (1994). Calcium and pH interaction on root nodulation of nursery grown red alder (*Alnus rubra* Bong.) seedlings by *Frankia*. Soil Biol. Biochem. 26:607-614

Cronquist, A. (1988). "The Evolution and Classification of Flowering Plants," 2nd ed. Allen, Lawrence, Kansas

Dawson, J. O. (1983). Dinitrogen fixation in forest ecosystems. Can. J. Microbiol. 29:979-992

Diem, H .G., and Y. R. Dommergues (1990). Current and potential uses and management of Casuarinaceae in the tropics and subtropics. *In* Schwintzer CR, and Tjepkema JD (ed.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA. pp. 317-342

Diem, H. G., D. Gauthier, and Y. Dommergues. (1983). An effective strain of *Frankia* from *Casuarina* sp. Can. J. Bot. 61:2815-2821

Dillon, J. T., and D. Baker (1992). Variation in nitrogenase activity among pure-cultured *Frankia* strains tested on actinorhizal plants as an indication of symbiotic compatibility. New Phytol. 92:607-614

Dolzani, L., E. Tonin, C. Lagatolla, and C. Monti-Bragadin (1994). Typing of *Straphylococcus aureus* by amplification of the 16S-23S rRNA intergenic spacer sequences. FEMS Microbiol. Lett. 119:167-174

Doran, T. J., A. L .M. Hodgson, J. K Davies, and A. J. Radford (1993). Characterisation of a highlyu repeated DNA sequence from *Mycrobacterium bovis.* FEMS Microbiol. Lett. 111: 147-152

Doyle J. J., J. L. Doyle (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11-15

Fisher, R. F., S. R Long (1992). *Rhizobium*-plant signal exchange. Nature 357:655-660

Fontaine, M. F., S. A. Lancelle, and J. G. Torrey. (1984). Initiation and ontogeny of vesicles in cultured *Frankia* sp. strain HFPArI3. J. Bacteriol. 160: 921-927

Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk. Jr. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. System. Bacteriol. 42:166-170

Frey, P., P. Frey-Klett, J. Garbaye, O. Berge, and T. Heulin (1997). Metabolic and genotypic fingerpringing of fluorescent pseudomonads associated with the Douglas Fir-*Laccaria bicolor* micorrhizosphere. Appl. Environ. Microbiol. 63:1852-1860

Garcia-Martines, J., A. Martinez-Murcia, A. I. Anton, and F. Rodriguez-Valera (1996). Comparison of the small 16S to 23S integenic spacer region (ISR) of the rRNA operons of some *Escherichia coli* strains of the ECOR collection and *E. coli* K-12. J. Bacteriol. 178:6374-6377

Gauthier, D., H. G. Diem, and Y. Dommergues (1981). Infectivite et effective de souche de *Frankia* isolaees de nodules de *Casuarina equisetifolia* et *d'Hippophaë rhamnoides*. C. R. Acad. Sc. Paris Ser. III 293:489-491

Gilson, E., D. Perrin, and M. Hofnung (1990). DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization. 18:3941-3952

Hafner, M. S., and S. A. Nadler (1990). Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation. Syst. Zool. 39:192-204

Hafner, M. S., R. D M. Page (1995). Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. Phil. Trans. R. Soc. Lond. B 349:77-83

Hafner, M. S., P. D. Sudman, F. X. Villabalnca, T. A. Spradling, J. W. Demastes, and S. A. Nadler (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265:1087-1090

Hahn, D., R. Kester, M. J. C. Starrenburg, and A. D. L. Akkermans (1990). Extraction of ribosomal RNA from soil for detection of *Frankia* with oligonucleotide probes. Arch. Microbiol. 154:329-335

Harding, E. F. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. Adv. Appl. Prob. 3:44-77

Hardy, R. W. F., R. D. Holsten, E. K. Jackson, and R. C. Burns (1968). The acetylene-ethylene assay for N_2-fixation: Laboratory and field evaluation. Plant Physiol. 43:1185-1207

Hibbs, D. E., and K. Cromack, Jr. (1990). Actinorhizal plants in Pacific Northwest forests. *In* Schwintzer CR, and Tjepkema JD (eds.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA. pp. 343-363

Higgins, C. F., G. F. Ames, W. M. Barnes, J. M. Clement, M. Hofnung (1982). A novel intercistronic regulatory element of prokaryotic operons. Nature 298:760-762

Higgins, C. F., R. S. McLaren, and S. F. Newbury (1988). Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. Gene 72:3-14

Hilger, A. B., and D. D. Myrold. (1992). Quantitation of soil *Frankia* by bioassay and gene probe methods: response to host and non-host rhizospheres and liming. Acta Œcol. 13:505-506

Hilger, A. B., Y. Tanaka, and D. D. Myrold. (1991). Inoculation of fumigated nursery soil increases nodulation and yield of bare-root red alder (*Alnus rubra* Bong.). New Forests 5:35-42

Holman, R. M., and C. R. Schwintzer. (1987). Distribution of spore-positive and spore-negative nodules of *Alnus incana* ssp. *rugosa* in Maine, USA. Plant Soil 104:103-111

Hönerlage W., D. Hahn, K. Zepp, J. Zeyer, P. Normand (1995). A hypervariable 23S rRNA region provides a discriminating target for specific characterization of uncultured and cultured *Frankia*. System. Appl. Microbiol. 17:433-443

Horrière, F., M. P. Lechevalier, and H. A. Lechavalier (1983). *In vitro* morphogenesis and ultrastructure of a *Frankia* sp. ArI3 (Actinomycetales) from *Alnus rubra* and a

morphologically similar isolate (AirI2) from *Alnus incana* subsp. *rugosa*. Can. J. Bot. 61: 2843-2854

Houck, C. M., F. P. Rinehart, C. W. Schmid (1979). A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. 132:289-306

Huelsenbeck, J. P., Rannala, B. (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science 276:227-232

Hulton, C. S. J., C. F. Higgins, and P. M. Sharp (1991). ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enteric bacteria. Mol. Microbiol. 5:825-834

Huss-Danell, K., and D. D. Myrold. (1994). Intrageneric variation in nodulation of *Alnus*: Consequences for quantifying *Frankia* infective units in soil. Soil Biol. Biochem. 26:525-531

Jamann, S., M. P. Fernandez, and A. Moiroud. (1992). Genetic diversity of Elaeagnaceae-infective *Frankia* strains isolated from various soils. Acta Œcol. 13:395-406

Jamann, S., M. P. Fernandez, and P. Normand. (1993). Typing method for N_2-fixing bacteria based on PCR-RFLP--application to the characterization of *Frankia* strains. Mol. Ecol. 2:17-26

Jelinek, W. R., T. P. Toomey, L. Leinwand, C. H. Duncan, P. A. Biro, P. V. Choudary, S. M. Weissman, C. M. Rubin, C. M Houck, P. L. Deininger, and C. W. Schmid (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. Proc. Natl. Acad. Sci. USA 77:1398-1402

Jeong, S-C, Liston, L. and Myrold D. D. (1997). Molecular phylogeny of the genus *Ceanothus* using *ndh*F and *rbc*L sequences. Theor. Appl. Genet. 94:852-857

Kim, D. Y. (1987). Seasonal estimates of nitrogen fixation by *Alnus rubra* and *Ceanothus* species in western Oregon forest ecosystems. M.S. thesis. Oregon State Univ., Corvallis, OR. 71 p.

Kim K.-J., and R. K. Jansen (1995). *ndhF* sequence evolution and the major clades in the sunflower family. Proc. Natl. Acad. Sci. USA 92:10379-10383

Klemmedson, J. O. (1979). Ecological importance of actinomycete-nodulated plants in the western United States. Bot. Gaz. 140(suppl.):S91-S96

Lalonde, M., L. Simon, J. Bousquet, and A. Seguin (1988). Advances in the taxonomy of *Frankia*: recognition of species *alni* and *eleagni* and novel subspecies *pommerii* and *vandijkii*. *In* Bothe H, de Bruijn FJ, and Newton WE (eds.), Nitrogen Fixation: Hundred Years After. Gustav Fisher Verlag, Stutgart, Germany. pp. 671-680

Lancelle, S. A., J. G. Torrey, P. K. Hepler, and D. A. Callaham (1985). Untrastructure of freeze-substituted *Frankia* strain HFPCcI3, the actinomycete islated from root nodules of *Casuarina cunninghamiana*. Protoplasma 127:64-72

Latour, X., T. Corberand, G. Languerre, F. Allard, and P. Lemanceau (1996). The comparison of fluorescent pseudomonad populations associated with roots is influenced by plant and soil type. Appl. Environ. Microbiol. 62:2449-2456

Lechevalier, M .P. (1985-86). Catalog of *Frankia* strains, second edition. The Actinomycetes 19:131-162

Lechevalier, M. P., and H. A. Lechevalier (1990). Systematics, isolation, and culture of *Frankia*. *In* Schwintzer CR, and Tjepkema JD (eds.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA. pp. 35-60

Lechevalier, M. P., and H. A. Lechevalier (1984). Taxonomy of *Frankia*. *In* Ortiz-Ortiz L, Bojalil LF, and Yakoliff V (eds.) Biological, Biochemical, and Biomedical Aspects of Actinomycetes. Academic Press, New York, NY. pp. 575-582

Lechevalier, M. P., and J.-S. Ruan (1984). Physiological and chemical diversity of *Frankia* spp. isolated from nodules of *Comptonia peregrina* (L.) Coult. and *Ceanothus americanus* L. Plant Soil 78:15-22

Lennon, E., P. D. Gutman, H. L. Yao, K. W. Minton (1991). A highly conserved repeated chromosomal sequence in the radioresistant bacterium *Deinococcus radiodurans* SARK. J. Bacteriol. 173:2137-2140

Li W.-H., and D. Graur (1991) Fundamentals of molecular evolution. Sinauer, Sunderland, Massachusetts

Liu, Q., and A. M. Berry (1991). The infection process and nodule initiation in the *Frankia-Ceanothus* root nodule symbiosis: a structural and histochemical study. Protoplasma 163:82-92

Longeri, L, and M. Abarzu (1989). Ultrastructure of *Frankia* isolated from three Chilean shrubs (Rhamnaceae). *In* Skinner FA (ed.), Nitrogen Fixation with Non-legumes. Kluwer Academic Publishers, Dordrecht, The Netherlands. Pp. 47-53

Louws, F. J., D. W. Fulbright, C. T. Stephens, and F. J. de Brujn (1994). Specific genomic fingerprinting of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. Appl. Environ. Microbiol. 60:2286-2295

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

Maggia, L., S. Nazaret, and P. Simonet (1992). Molecular characterization of *Frankia* isolates from *Casuarina equisetifolia* root nodules harvested in West Africa (Senegal and Gambia). Acta Œcol. 13:453-462

Markham, J. H. and C. P. Chanway (1996). *Alnus rubra* nodulation capacity of soil under five species from harvested forest sites in coastal British Columbia. Plant Soil 178:283-286

Martin, K. J., Y. Tanaka, and D. D. Myrold (1992). Dual inoculation of red alder (*Alnus rubra* Bong.) with effective *Frankia alni* for nursery bare-root seedling production. Acta Écol. 13:508-509

Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J. P. Claverys (1992). A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. 20:3479-3483

Mazel, D., J. Houmard, A. M. Castets, N. Tandeau de Marsac (1990). Highly repetitive DNA sequences in cyanobacterial genomes. J. Bacteriol. 172:2755-2761

McMinn H. E. (1942). *Ceanothus*: Part II; a systematic study of the genus *Ceanothus*. Santa Barbara Botanic Garden, Santa Barbara, California pp 129-308

McNabb, D. H., Geist, J. M., and C. T. Youngberg (1979). Nitrogen fixation by *Ceanothus velutinus* in northeastern Oregon. *In* Gordon JC, Wheeler CT, and Perry DA (eds.) Symbiotic Nitrogen Fixation in the Management of Temperate Forests For. Res. Lab., Oregon State Univ., Corvallis, OR. pp. 481-482

Miller, I. M., and D. D. Baker (1985). The initiation, development and structure of root nodules in *Elaeagnus augustifolia* L. (Elaeagnaceae). Protoplasma 128:107-119

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

Monz, C. A., C. Schwintzer (1989). The physiology of spore-negative and spore-positive nodules of *Myrica gale*. Plant Soil 118: 75-87

Moran, N. A., Munson, M. A., Bauman, P. Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B Biol. Sci. 253:167-171

Morgan D. R., D. E. Soltis, and K. R. Robertson (1994). Systematic and evolutionary implications of *rbc*L sequence variation in Rosaceae. Am. J. Bot. 81:890-903

Murray, M. G., and W. F. Thomson (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8: 4321-4324

Murry, M. A., D. Zhang, M. Schneider, and F. J. de Bruijn (1995). Use of repetitive sequences and the polymerase chain reaction (rep-PCR) to fingerprint the genomes of *Frankia* isolates. Symbiosis 19:223-240

Murry MA, A. S. Konopka, S. D. Pratt, T. L. Vandergon (1997). The use of PCR-based typing methods to assess the diversity of *Frankia* nodule endophytes of the actinorhizal shrub *Ceanothus*. Physiol. Plant. (in press)

Myrold D. D. (1994). *Frankia* and the actinorhizal symbiosis. *In* Weaver RW, Angle VS, and Bottomley PJ (eds) Methods of Soil Analysis: part 2. Microbiological and biochemical properties. Soil Science Society of America, Madison, WI. pp. 291-328

Myrold, D. D., A. B. Hilger, and S. H. Strauss (1990). Detecting Frankia in soil using PCR. *In* Gresshoff PM, Roth LE, Stacey G, and Newton WE. (eds.) Notrogen Fixation: Achievements and Objectives. Chapman and Hall, New York, NY. pp. 429

Myrold, D. D., and K. Huss-Danell (1994). Population dynamics of *Alnus*-infective *Frankia* in a forest soil with and without host trees. Soil Biol. Biochem. 26:533-540

Navarro, E., R. Nalin, D. Gauthier, and P. Normand (1997). The nodular microsymbionts of *Gymnostoma* spp. Are *Elaeagnus*-infective *Frankia* strains. Appl. Environ. Microbiol. 63:1610-1616

Nazaret, S., B. Cournoyer, P. Normand, and P. Simonet (1991). Phylogenetic relationships among *Frankia* genomic species determined by use of amplified 16S rDNA sequences. J. Bacteriol. 173:4072-4078

Nazaret, S., P. Simonet, P. Normand, and R. Bardin (1989). Genetic diversity among *Frankia* isolated from *Casuarina* nodules. Plant Soil 118:241-247

Nelson, D. L., and C. F. Lopez (1989). Variation in nitrogen fixation among populations of *Frankia* sp. and *Ceanothus* sp. in actinorhizal association. Biol. Fertil. Soils 7:269-274

Nelson, D. L., S. Ledbetter, L. Corbo, M. F. Victoria, R. Ramirez-Solis, T. D. Webster, D. H. Ledbeter, and C. T. Caskey (1989). *Alu* polymearse chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. USA 86:6686-6690

Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins (1987). Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 30:297-310

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

Nick, G., E. Paget, P. Simonet, A. Moiroud, and P. Normand (1992). The nodular endophytes of *Coriaria* spp. form a distinct lineage within the genus *Frankia*. Mol. Ecol. 1:175-182

Nobs, M. A. (1963). Experimental studies on species relationships in *Ceanothus*. Carnegie Inst Wash Publ, Washington D. C.

Normand, P., and M. Lalonde (1982). Evaluation of Frankia strains isolated from provenances of two *Alnus* species. Can. J. Microbiol. 28:1133-1142

Normand, P., and M. Lalonde (1986). The genetics of actinorhizal *Frankia*: A review. Plant Soil 90:429-453

Normand, P., B. Cournoyer, P. Simonet, and S. Nazaret (1992). Analysis of a ribosomal RNA operon in the actinomycete *Frankia*. Gene 111:119-124

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

Ochman, H., and A. C. Wilson (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26: 74-86

Olmstead, R. G., and J. D. Palmer (1994). Chloroplast DNA systematics: a review of methods and data analysis. Am. J. Bot. 81:1205-1224

Olmstead, R. G., and J. A. Sweere (1994). Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Sys. Biol. 43:467-481

Oyaizu, H., S. Matsumoto, K. Minamisawa, and T. Gamou. (1993). Dostribution of rhizobia in leguminous plants surveyed by phylogenetic identification. J. Gen. Appl. Microbiol. 39:339-354

Page, R. D. M. (1994). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. Syst. Biol. 43: 58-77

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Appl. Biosci. 12: 357-358

Philipp, W. J., S. Puolet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B. R. Bloom, W. R. Jacobs, Jr., and S. T. Cole (1996). An integrated map of the genome of the tubercle bacillus, *Myrobacterium tuberculosis* H37Rv , and comparison with *Mycobacterium leprae*. Proc. Natl. Acad. Sci. USA 93:3132-3137

Picard, C., C. Ponsonnet, E. Paget, X. Nesme, and P. Simonet (1992). Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. 58:2717-2711

Racette, S., and J. G. Torrey (1989). Root nodule initiation in *Gymnostoma* (Casuarinaceae) and *Shepherdia* (Elaeagnaceae) induced by *Frankia* strain HFPGpI1. Can. J. Bot. 67: 2873-2879

Raven, P. H., and D. I. Axelrod (1978). Origin and relationships of the California flora. Univ. Calif. Publ. Bot. 72:1-134

Rouvier, C., Y. Prin, P. Redell, P. Normand, and P. Simonet (1996). Genetic diversity among *Frankia* strains nodulating members of the family Casuarinaceae in Australia revealed by PCR and restriction fragment length polymorphism analysis with crushed root nodules. Appl. Environ. Microbiol. 62: 979-985

Sadowsky, M. J., L. L. Kinkel, J.H. Bowers, and J. L. Schottel (1996). Use of repetitive intergenic DNA sequences to classify pathogenic and Disease-supressive *Streptomyces* strains. Appl. Environ. Microbiol. 62:3489-3493

Savage, H. M. (1983). The shape of evolution: Systematic tree topology. Bil. J. Lin. Soc. 20:225-244

Scherer, S., and D. A. Stenens (1988). A *candida anlicans* dispersed, repeated gene family and its epidemiological applications. Proc. Natl. Acad. Sci. USA 85:1452-1456

Schmid, C. W., and W. R. Jelinek (1982). The *Alu* family of dispersed repetitive sequences. Science 216:1065-1070

Schmidt, C. L. (1993). *Ceanothus. In* Hickman JC (ed.) The Jepson Manual: Higher Plants of California. University of California Press, Berkeley, California. pp. 932-938

Schwintzer, C. R., and J. D. Tjepkema. (1990). The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA. 408 p.

Scott, W. (1973). Some soil factors affecting snowbrush nodulation. Ph.D. Thesis. Oregon State Univ., Corvallis, OR. (Diss. Abstr. 33:4074)

Sellstedt, A. (1989). Occurrence and activity of hydrogenase in symbiotic *Frankia* from field-collected *Alnus incana*. Physiol. Plant. 75:304-308

Sharples, G. J. and R G. Lloyd (1990). A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucleic Acids Res. 18:6503-6508

Simonet, P., M.-C. Grosjean, A. K. Misra, S. Nazaret, B. Cournoyer, and P. Normand (1991). *Frankia* genus-specific characterization by polymerase chain reaction. Appl. Environ. Microbiol. 57:3278-3286

Simonet, P., P. Normand, A. Moiroud, and R. Bardin (1990). Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. Arch. Microbiol. 153:235-240

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

Smolander, A., and V. Sundman (1987). *Frankia* in acid soils of forests devoid of actinorhizal plants. Physiol. Plant. 70:297-303

Smolander, A., C. van Dijk, and V. Sundman (1988). Survival of *Frankia* strains introduced into soil. Plant Soil 106:65-72

Sneath, P. H. A., and R. R. Sokal (1973). Numerical Taxonomy, Freeman, San Fransisco, CA

Snedecor, G. W., and W. G. Cochran (1980). Statistical Methods. Iowa State Univ. Press, Ames, IA

Soltis, D. E., R. K. Kuzoff, E. Conti, R. Gornall, and K. Ferguson (1996). *mat*K and *rbc*L gene sequence data indicate that *Saxifraga* (Saxifragaceae) is polyphyletic. Am J. Bot. 83:371-382

Soltis D. E., P. S. Soltis, D. R. Morgan, S. M. Swensen, B. C. Mullin, J. M. Dowd, and P. G. Martin (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA 92:2647-2651

Sprent, J. I. (1994). Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? Plant Soil 161: 1-10

Stern, M. J., G. F.-L Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell 37:1015-1026

Stringer, S. L., S. T. Hong, D. Giuntoli, and J. R. Stringer (1991). Repeated DNA in *Pneumocystis carinii*. J. Clin. Microbiol. 29:1194-1201

St.-Laurent, L. and M. Lalonde (1986). Isolation and characterization of *Frankia* strains isolated from *Myrica gale*. Can. J. Bot. 65: 1356-1363

Swensen, S. M. (1996). The evolution of actinorhizal symbiosis: Evidence for multiple origins of the symbiotic association. Am. J. Bot. 83: 1503-1512

Swensen, S. M., B. C. Mullin, and M. W. Chase (1994). Phylogenetic affinities of Datiscaceae based on an analysis of nucleotide sequences from the plastid *rbc*L gene. Syst. Bot. 19:157-168

Swofford D. L. (1991). PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois National History Survey, Champaign, Illinois

Thomas, K. A., and A. M. Berry (1989). Effects of continuous nitrogen application and nitrogen preconditioning on nodulation and growth of *Ceanothus griseus* var *horizontalis*. Plant Soil 118:181-187

Tjepkema, J. D., W. Ormerod, and J. G. Torrey (1980). Vesicle formation and acetylene reduction acitvity in *Frankia* sp. CpI1 cultured in defined nutrient media. Nature 287:633-635

Tjepkema, J. D., C. R. Schwintzer, and D. R. Benson (1986). Physiology of actinorhizal nodules. Annu. Rev. Plant Physiol. 37:209-232

Torrey, J. G. (1978). Nitrogen fixation by actinomycete-nodulated angiosperms. BioSci. 28:586-591

Torrey, J. G., and Tjepkema, J. D. (1979). Preface. Bot. Gaz. (Suppl): i-ii

Tyson, J. H., and W. S. Silver (1979). Relationship of ultrastructure to acetylene reduction (nitrogen fixation) in root nodules of *Casuarina*. Bot. Gaz. (Suppl.): S44-S48

Ueda T., Y. Suga, N. Yahoo, and T. Matsuguchi (1995). Phylogeny of sym plasmids of rhizobia by PCR-based sequencing of a *nod*C segment. J. Bacteriol. 177: 468-472

VandenBosch, K. A., and J. G. Torrey (1984). Consequences of sporangial development for nodule function in root nodules of *Comptonia peregrina* and *Myrica gale*. Plant Physiol. 76:556-560

van Dijk, C. (1978). Spore formation and endophyte diversity in root nodules of Alnus glutinosa (L.) Vill. New Phytol. 81: 601-615

van Dijk, C. (1984). Ecological Aspects of Spore Formation in the *Frankia-Alnus* Symbiosis. Ph.D. Thesis. State University, Leiden, the Netherlands. 154 p.

van Dijk, C., and E. Merkus (1976). A microscopic study of the development of a spore-like stage in the life cycle of the root nodule endophyte of *Alnus glutinosa* (L.) Vill. New Phytol. 77: 73-90

van Dijk, C., A. Sluimer, and A. Weber. (1988). Host range differentiation of spore-positive and spore-negative strain types of *Frankia* in stands of *Alnus glutinosa* and *Alnus incana* in Finland. Physiol. Plant. 72:349-358

Versalovic, J., T. Koeuth, and J. R. Lupski (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. 19:6823-6831

Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski (1994). Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. 5:25-40

Weber, A. (1986). Distribution of spore-positive and spore-negative nodules in stands of *Alnus glutinosa* and *Alnus incana* in Finland. Plant Soil 96:205-213

Weber, A. (1990). Host specificity and effieiency of notrogenase activity of Frankia strains from *Alnus incana* and *Alnus glutinosa*. Symiosis 8: 47-60

Wheeler, C. T., and I. M. Miller (1990). Current and potential uses of actinorhizal plants in Europe. *In* Schwintzer CR, and Tjepkema JD (eds.) The Biology of *Frankia* and Actinorhizal Plants. Academic Press, Inc., San Diego, CA. pp. 365-389

Wollum, A. G. II., C. T. Youngberg, and F. W. Chichester (1968). Relation of previous timber stand age to nodulation of *Ceanothus velutinus*. For. Sci. 14:114-118

Wollum, A. G. II., C. T. Youngberg, and C. M. Gilmour (1966). Characterization of a *Streptomyces* sp. isolated from root nodules of *Ceanothus velutinus* Dougl. Soil Sci. Soc. Am. Proc. 30:463-467

Young, J. P., and Johnston, A. W. B. (1989). The evolution of specificity in the legume-rhizobium symbiosis. Trends Ecol. Evol. 4:341-349

Youngberg, C. T., and A. G. Wollum II. (1976). Nitrogen accretion in developing *Ceanothus velutinus* stands. Soil Sci. Soc. Am. J. 40:109-112

Yuan Y-M, P. Küpfer, J. J. Doyle (1996) Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. Am. J. Bot. 83:641-652

Zavitkovski, J., and M. Newton.(1968). Ecological importance of snowbrush *Ceanothus velutinus* in the Oregon Cascades. Ecology 49:1134-1145

Zhang, Z.,. and J. G. Torrey (1985). Studies of an effective strain of *Frankia* from *Allocasuarina lehmanniana* of the Casuarinaceae. Plant Soil 87: 1-16

Appendices

Appendices

Appendix 1. The data matrixes of nucleotide sequences used for analyzing actinorhizal symbioses.

Appendix 1.1. The data matrix of 9 *nif*H nucleotide sequences used for reconstructing phylogeny of *Frankia*. Species or strain abbreviation: nif-hrn18 (*Frankia* sp. HRN18a), nif-euik1 (*Frankia* sp. EuiK1), nifwho-ar (*Frankia* sp. ArI3), nifhdk1-f (*Frankia* sp. FaC1), nifh-Alfa (*Alcalgenes faecalis*), coriaria- microsymbiont (*Frankia* of *Coriaria nepalensis*), datisca-f microsymbiont (*Frankia* of *Datisca cannabina*), nif-cea microsymbiont (*Frankia* of *Ceanothus caeruleus*). Indels and missing sequences were corrected using Mask.

*nif*H data matrix

```
                          1111111111222222222233333333334444444444555555555566666666
                 12345678901234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------------------
nif-hrn18        atg---------cgccagatcgcgttctacggcaagggtggtatcggcaagtccaccacccagcaga
nif-euik1        atg---------cgccagatcgcgttctacggaaagggtggtattggcaagtccaccacccagcaga
nifwho-ar        atg---------cgccagatcgccttctacggcaagggtggtattggcaagtccaccacccagcaga
nifhdk1-f        atg---------cgccagatcgccttctacggcaagggtggtatcggcaagtccaccacccagcaga
nifh-Alfa        atg---gcaatgcgtcaatgcgctatttacgggaagggtggaatcggaaaatccaccacgacccaga
nifh-Rh          atgtcgtcactgagacagatcgcattttacggcaaggggggcatcggcaagtcgacgacgtcccaga
coriaria-        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNcagcaga
datisca-f        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNcagcaga
nif-cea          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Mask             00000000000000000000000000000000000000000000000000000000000000000000000
```

*nif*H data matrix (continued)

```
                                                          1111111111111111111111111111111111
                 66777777777788888888889999999990000000001111111112222222222333333
                 89012345678901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------------
nif-hrn18        acaccatggctgccatggccgagatgggccgtcgggtcatgatcgtcggctgcgaccccaaggctga
nif-euik1        acaccatggctgccatggccgagatgggccgtcgcgtcatgatcgtcggctgcgaccccaaggctga
nifwho-ar        acaccatggcggccatggccgagatgggccagcgggtcatgatcgtcgggtgtgacccgaaggctga
nifhdk1-f        acaccatggcggccatggccgagatgggcaagaaggtcatgatcgtcggctgtgacccgaaggccga
nifh-Alfa        acctcgtggcggccctggccgaactcggcaaaaaggttatgatcgtcggctgcgaccccaaggccga
nifh-Rh          acacgctggcggcgctggcggagatgggccatcgcatcctcatcgtcggctgcgaccccaaggccga
coriaria-        acaccatggccgcaatggctgagcgtggcaaccgggtcatgatcgtcggttgcgacccgaaggcgga
datisca-f        acaccatggccgcactggctgagcggggcaaccgggtcatgatcgtcggttgtgacccgaaggctga
nif-cea          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Mask             00000000000000000000000000000000000000000000000000000000000000000000000
```

*nif*H data matrix (continued)

```
                 11111111111111111111111111111111111111111111111111111111111111122
                 33333444444444455555555556666666666777777777788888888899999999900
                 56789012345678901234567890123456789012345678901234567890123456789012345678901
-------------------------------------------------------------------------------------------
nif-hrn18        ctcgacccgcctcatcctgcactcgaaggcccagacctcggtgatccagctcgctgccg---agaag
```

```
nif-euik1          ctcgacccgcctcatcctgcactcgaaggcccagacctccgtcatcaagctcgctgccg---agaag
nifwho-ar          ctcgacccgcctgatcctgcactcgaaggcgcagacctccgtcatccagctcgcggccg---agaag
nifhdk1-f          ctcgacccgcctgatcctgcactcgaaggcgcagacctccgtcatccagctggcggccg---agaag
nifh-Alfa          ctccactcgcctgatcctgcactccaaggcgcagaacaccatcatggaaatggccgccg---aggcc
nifh-Rh            ctcaacccgcctcatcctgcacgccaaggcgcaggacaccatcctgtcgctggccgccg---ccgcc
coriaria-          ctcgacccgtctcatcctgcactccaaggcccagcacaccgtcatcgagaaggcggccg---agaag
datisca-f          ctcgacccgtctgatcctgcattccaaggcccagcacaccgtcatcgagaaggccgccg---agaag
nif-cea            NNNNNNNNNNctcatcNtgcactccaaggcccagcacaccgtcatcgagaaggcggccg---agaag
Mask               00000000001111111111111111111111111111111111111111111111111100011111
```

*nif*H data matrix (continued)

```
                   22222222222222222222222222222222222222222222222222222222222222222222
                   00000000111111111122222222223333333333444444444455555555556666666666
                   23456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------------
nif-hrn18          gggtccgtcgaggacctggagctcgacgaggtgctcgtcgagggccagtggggcatcaagtgcgtcg
nif-euik1          ggatcggtcgaggacctggagctgaacgaggtcctggtcgagggccagtggggcatcaagtgcgtcg
nifwho-ar          ggttcggtcgaggacctggagctcgacgaggtgctcgtcgagggccagtggggcatcaagtgcgtcg
nifhdk1-f          gggtcggtcgaggacctggagctcgacgaggtcctggtcgagggccagtggggcatcaagtgcgtgg
nifh-Alfa          ggtaccgtggaagacctggaactcgaggacgtgctcaagaccggctacggcgacatcaagtgcgtcg
nifh-Rh            ggcagcgtcgaagacctcgagctcgaagaggtcatgaagatcggctaccgcgacatccgctgcgtgg
coriaria-          ggctcggtggaggacctcgagctggaagaggttctgctcgagggcgcatggaacatccggtgtgtcg
datisca-f          ggctcggtggaggacctcgagctcgaggaggtcctgctcgagggcgcgtggaacattcggtgtgtcg
nif-cea            ggctcggtggaggacctcgagctggaagaggttctgctcgagggcgcatggaacgtccggtgcgtcg
Mask               11111111111111111111111111111111111111111111111111111111111111111111
```

*nif*H data matrix (continued)

```
                   22222222222222222222222222222233333333333333333333333333333333333333
                   67777777777888888888899999999900000000001111111111222222222233333333
                   90123456789012345678901234567890123456789012345678901234567890123456
-------------------------------------------------------------------------------------
nif-hrn18          agtccggtggcccggagccgggcgtcggctgcgccggccgcggtgtcatcacctccatcacgtacct
nif-euik1          agtccggtggcccggagccgggcgtgggctgcgccggccgtggcgtcatcacctccatcacctacct
nifwho-ar          agtccggtggcccggagccgggcgtcggctgcgccggccgtggcgtcatcacctccatcacctacct
nifhdk1-f          agtccggtggcccggagccggcggtgggctgcgccggccgtggcgtcatcacctccatcaactacct
nifh-Alfa          agtcgggcggtccggagccgggcgtgggctgcgccggtcgcggcgtgatcaccgcgatcaacttcct
nifh-Rh            aatccggcggcccggagccgggcgtcggctgcgcggggcgcggcgtcatcacctccatcaatttcct
coriaria-          agtcgggtggtccggagccgggtgtcggctgcgccggccgaggtgtcatcaccgccatcaccttcct
datisca-f          agtccggtggcccggagccgggtgtcggctgcgccggccgaggtgtcatcaccgccatcaccttcct
nif-cea            agtcgggtggtcccgagccgggtgtcggctgcgccggccgaggcgtcatcaccgccatcaccttcct
Mask               11111111111111111111111111111111111111111111111111111111111111111111
```

*nif*H data matrix (continued)

```
                   33333333333333333333333333333333333333333333333333333333333333333444
                   33334444444444555555555566666666667777777777888888888899999999999000
                   67890123456789012345678901234567890123456789012345678901234567890123
-------------------------------------------------------------------------------------
nif-hrn18          ggaggaggccggcgcctacgagaacc---tcgacttcgtcacctacgacgtcctcggtgacgttgtc
nif-euik1          ggaggaggccggcgcctacgagaacc---tggacttcgtcacctacgacgtcctcggtgacgtggtc
nifwho-ar          ggaggaggccggcgcgtacgagaacc---tcgacttcgtgacctacgacgtcctcggtgacgttgtg
nifhdk1-f          ggaggaggccggcgcctacgaggacc---tcgacttcgtgacctacgacgtccttggcgacgtcgtc
nifh-Alfa          cgaagagaaggccgcctacgaggatgacctggacttcgtcttctacgacgtgctcggcgacgtggtc
nifh-Rh            ggaggagaacggcgcctatgaggaca---tcgactacgtctcctacgacgtgctcggcgacgtggtg
coriaria-          ggaggagaaccgcgcctacgagaacc---tcgacttcgtcacctacgacgtcctcggtgacgtcgtg
datisca-f          gggggagaacggcgcctacgaggacc---tcgacttcgtcacctacgacgtcctcggtgacctcgtc
nif-cea            ggaggagaacggcgcctacgaggatc---tcgacttcgtcacctacgacgtcctcggtgacgtcgtg
mask               11111111111111111111111111100011111111111111111111111111111111111111
```

*nif*H data matrix (continued)

```
                   44444444444444444444444444444444444444444444444444444444444444444444
                   00000001111111111222222222233333333334444444444555555555566666666666
                   34567890123456789012345678901234567890123456789012345678901234567890
-------------------------------------------------------------------------------------
nif-hrn18          tgcggtggcttcgcgatgccgatccgccggggcaaggcccaggagatctacatcgtgacctccggcg
```

```
nif-euik1          tgcggtggcttcgcgatgccgatccgccagggcaaggcccaggagatctacatcgtgacgtccggcg
nifwho-ar          tgtggtggcttcgcgatgccgatccgccagggcaaggcgcaggagatctacatcgtcacctccggcg
nifhdk1-f          tgcggtggcttcgcgatgccgatccgccagggcaaggcgcaggagatctacatcgtcacctccggcg
nifh-Alfa          tgtggcggcttcgccatgcccatccgcgagaacaaggcccaggagatctacgtggtctgctccggcg
nifh-Rh            tgcggcggcttcgccatgcccatccgcgagaacaaggcgcaggaaatctacatcgtcatgtccggcg
coriaria-          tgcggtggtttcgccatgccgatccgccagggcaaggcccaggagatctacatcgtcacctccggtg
datisca-f          tgtggcgggttcgccatgccgatccgccagggcaaggcccaggagatctacattgtcacctccggtg
nif-cea            tgcggtggtttcgccgtgccgatccgccagggcaaggcccaggagatNtacatcgtcacctccggtg
Mask               11111111111111111111111111111111111111111111111111011111111111111111
```

*nif*H data matrix (continued)

```
                   44444444444444444444444444444455555555555555555555555555555555555
                   77777777778888888888999999999900000000001111111111222222222233333333
                   01234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------------
nif-hrn18          agatgatggcgatgtacgcggcgaacaacatcgcccgcggcatcctcaagtacgcgcactccggcgg
nif-euik1          agatgatggcgatgtacgcggcgaacaacatcgcccgcggtgtcctcaagtacgggcactccggtgg
nifwho-ar          agatgatggcgatgtacgcggcgaacaacatcgcccgaggcatcctgaagtacgcgcactccggcgg
nifhdk1-f          agatgatggcgatgtacgcggcgaacaacatcgcccgcggagtcctcaagtacgcgcactccggtgg
nifh-Alfa          agatgatggcgatgtatgccgccaacaacatctccaagggcatcgtgaagtacgccaactccggcag
nifh-Rh            agatgatggccatgtatgcggccaacaacatctccaagggcattctcaagtacgccaattccggcgg
coriaria-          agatgatggccatgtacgcggcgaacaacattgcccgaggcatcctcaagtacgcccacagcggtgg
datisca-f          agatgatggccatgtacgcggcgaacaacattgcccgaggcatcctcaagtacgcccacagcggtgg
nif-cea            agatgatggccatgtacgcggcgaacaacattgcccgaggcatcctcaagtacgcccacagcggtgg
Mask               11111111111111111111111111111111111111111111111111111111111111111111
```

*nif*H data matrix (continued)

```
                   55555555555555555555555555555555555555555555555555555555556666666
                   33344444444445555555555666666666667777777777888888888899999999900000
                   78901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------
nif-hrn18          cgtccgcctcggtggcctcatctgcaacagccgcaagaccgaccgtgaggacgagctgatcatggag
nif-euik1          cgtgcgtctcggtgggctcatctgcaacagccgtaacaccgaccgcgaggacgagctgatcatcgag
nifwho-ar          cgtccggctcggcgggctcatctgcaacagccgcaagaccgaccgcgaggacgagctgatcatggag
nifhdk1-f          tgtccggctcggcgggctcatctgcaacagccgtaacaccgaccgcgaggacgagctgatcatggag
nifh-Alfa          cgtgcggctcggcgggctgatctgcaacagccgcaacaccgaccgcgaggacgagctgatcatggcc
nifh-Rh            cgtgcgcctgggcgggctcgtctgcaacgagcgccagacggacaaggagctggaactggccgagaac
coriaria-          tgtccgcctcggtgggctgatctgcaacagccgtaacaccgaccgtgaggacgagctcatcatcgag
datisca-f          tgtccgcctcggtgggctgatctgcaacagccgtaacaccgaccgtgaggacgagctcatcatcgag
nif-cea            tgtccgcctcggtgggctgatctgcaacagccgtcagaccgaccgtgaggacgagctcatcatggag
Mask               11111111111111111111111111111111111111111111111111111111111111111111
```

*nif*H data matrix (continued)

```
                   66666666666666666666666666666666666666666666666666666666666666666666
                   00000011111111112222222222333333333344444444445555555556666666667
                   45678901234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------------
nif-hrn18          ctcgcccgccgtctcaacacccagatgatccacttcatcccgcgtaacaacgtcgtgcagcacgccg
nif-euik1          ctcgcccgtcgtctcaacacccagatgatccacttcatcccgcgtaacaacgtcgtgcagcacgccg
nifwho-ar          ctcgcccgccgcctcaacacccagatgatccacttcatcccgcgtaacaacgtcgtgcagcacgccg
nifhdk1-f          ctcgcccgccgcctcaacacccagatgatccacttcatcccgcgcaacaacgtcgtgcagcacgccg
nifh-Alfa          ctggccgacaagctgggctcgcagatgatccacttcgtcccgcgcgacaacgtggtgcagcgcgccg
nifh-Rh            ctcgccaagaagctcggcacccagctcatctatttcgtgccgcgcgacaacatcgtgcagcacgccg
coriaria-          ctcgcccgccgcctgaacacccagatgatccacttcgttccgcgtgacaacatcgtccagcacgccg
datisca-f          ctcgcccgccgcctgaacacccagatgatccacttcgttccgcgtgacaacatcgtccagcacgccg
nif-cea            ctcgcgcggcgcctgaacacccagatgatccacttcgttccgcgtgacaacatcgtccagcacgccg
Mask               11111111111111111111111111111111111111111111111111111111111111111111
```

*nif*H data matrix (continued)

```
                   66
                   77
                   12
--------------------
nif-hrn18          ag
```

```
nif-euikl       ag
nifwho-ar       ag
nifhdkl-f       ag
nifh-Alfa       aa
nifh-Rh         ag
coriaria-       ag
datisca-f       ag
nif-cea         ag
Mask            11
```

Appendix 1.2. The data matrix of twenty 16S rDNA nucleotide sequences used for reconstructing phylogeny of *Frankia*. Species or strain abbreviation: AVN17s (*Frankia* sp. AVN17s), AcN14a (*Frankia* sp. AcN14a), AcoN24d (*Frankia* sp. AcoN24d), AgKG4 (*Frankia* sp. AgKG4-84), argp5 (*Frankia* sp. ARgP5), ArI4, (*Frankia* sp. ArI4), Arugo.en microsymiont (*Frankia* of *Alnus rugosa*), CeD (*Frankia* sp. CeD), Cnep.en (*Coriaria nepalensis*), Dcan.en (microsymbiont *Frankia* of *Datisca cannabina*), Dryas.en (microsymbiont *Frankia* of *Dryas*) Purshia16 (microsymbiont *Frankia* of *Purshia tridendata*), Ceano-thy (microsymbiont *Frankia* of *Ceanothus thyrsiflorus*), Ceano-ame (mocrosymbiont *Frankia* of *Ceanothus americanus*), EaI1-2 (*Frankia* sp. EaI1-28), SCN10a (*Frankia* sp. SCN10a), HR27-14 (*Frankia* sp. HR27-14), mnagaen (microsymbiont *Frankia* of *Myrica* spp.), Catenu (*Catenuloplanes cauruleus*), Geo-ob-ob (*Geodermatophilus obscurus*).

16S rDNA data matrix

```
                        1111111111222222222233333333334444444444555555555566666666
                   12345678901234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------------------------
AVN17s             cgctggcggcgtgcttaacacatgcaagtcgagcgagggg----cttcggc-tcct------agcgg
AcN14a             cgctggcggcgtgcttaacacatgcaagtcgagcgagggg----cttcggc-tcct------agcgg
AcoN24d            cgctggcggcgtgcttaacacatgcaagtcgagcgagggg----cttcggc-tcct------agcgg
AgKG4              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5              cgctggcggcgtgcttaacacatgcaagtcgagcggggag----cttcggc-ttct------agcgg
ArI4               cgctggcggcgtgcttaacacatgcaagtcgagcgagggg----cttcggc-ttct------agcgg
Arugo.en           cgctggcggcgtgcttaacacatgcaagtcgagcgagggg----c------------------gg
CeD                cgctggcggcgtgcttaacacatgcaagtcgagcggggag----cttcggc-tctc------agcgg
Cnep.en            NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcan.en            NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en           cgctggcggcgtgcttaacacatgcaagtcgagcgaggaa----cttcggt-tcct------agcgg
purshia16          cgctggcggcgtgcttaacacatgcaagtcgagcgggga----cttcggt-CTTC------agcgg
Cean-thy           cgctggcggcgtgcttaacacatgcaagccgaacgaggag----cttcggt-tcct------agcgg
 Cean-ame          cgctgcggcggtgcttaacacatgcaagccgaacgaggag----cttcggt-tcct------agcgg
Ea1-2              cgctggcggcgtgcttaacacatgcaagtcgagcgaggag----ctt-gc--tcct------agcgg
SCN10a             cgctggcggcgtgcttaacacatgcaagtcgagcgaggag----ctt-gc--tcct------agcgg
HR27-14            cgctggcggcgtgcttaacacatgcaagtcgagcgaggag----ctt-gc--tcct------agcgg
mnagaen            cgctggcggcgtgcCtaacacatgcaagtcgagcggggag----cttcggc-tcct------agcgg
Catenu             cgctggcggcgtgcttaacacatgcaagtcgagcggaaaggccccttcgggggtactcg---agcgg
Geo-ob-ob          cgctggcggcgtgcttaacacatgcaagtcgaacggtgaa-cctgcttcgg--tgggggatcagtgg
Mask1              11111111111111111111111111111111110000000000000000000000000000011
Mask2              00000000000000000000000000000000000000000000000000000000000000000
```

16S rDNA data matrix (continued)

```
                                                  1111111111111111111111111111111111
                   66777777777788888888889999999990000000000111111111122222222233333
                   89012345678901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------------------
AVN17s             cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataa-cctcgggaaaccggggct
AcN14a             cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataa-cctcgggaaaccggggct
AcoN24d            cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataa-cttcgggaaaccggggct
AgKG4              NNNNNNNNNNNNNNNNNNNNNNNNNaacctgccccgagctctggaataa-cctcgggaaaccggggct
argp5              cgaacgggtgagtaacacgtgggcaacctgccccgatctctggaataa-cctcgggaaaccggggct
ArI4               cgaacgggtgagtaacacgtggNcaacctgccccgagctctggaataa-cttcgggaaaccggggct
Arugo.en           cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataa-cttcgggaaaccggggct
CeD                cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataa-cctcgggaaaccggggct
Cnep.en            NNNNNNNNNNNNNNNNNNNNNNNNNaacctgccccgagctctgggataa-cttcgggaaaccggggct
Dcan.en            NNNNNNNNNNNNNNNNNNNNNNNNNaacctgccccgagctctgggataa-cttcgggaaaccggggct
Dryas.en           cgaacgggtgagtaacacgtgggcaacctgccccgagctctgggataa-cttcgggaaaccggggct
purshia16          cgaacgggtgagtaacacgtgggcaacctgccccgagctctgggataa-cttcgggaaaccggggct
Cean-thy           cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataaCctcgggaaaccggggct
Cean-ame           cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataaCctcgggaaaccggggct
Ea1-2              cgaacgggtgagtaacacgtgggcaacctgccccaagctctggaataa-ctccgggaaaccggggct
SCN10a             cgaacgggtgagtaacacgtgggcaacctgccccaagctctggaataa-ctccgggaaaccggggct
```

```
HR27-14             cgaacgggtgagtaacacgtgggcaacctgccccaagctctggaataa-ctccgggaaaccggggct
mnagaen             cgaacgggtgagtaacacgtgggcaacctgccccgagctctggaataa-cctcgggaaaccggggct
Catenu              cgaacgggtgagtaacacgtgagcaacctgccctggactttgggataa-ccctcggaaacggggct
Geo-ob-ob           cgaacgggtgagtaacacgtgggcaacctgcccccggctctgggataa-ctccaagaaattggggct
Mask1               11111111111111111111111101111111111111111111111111101111111111111111
Mask2               00000000000000000000000011111111111111111111111111101111111111111111
```

16S rDNA data matrix (continued)

```
                    11111111111111111111111111111111111111111111111111111111111111111122
                    33333444444444455555555556666666666777777777788888888899999999900
                    56789012345678901234567890123456789012345678901234567890123456789012345678901
-------------------------------------------------------------------------------------
AVN17s              aatgccggatat-gacacctccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
AcN14a              aatgccggatat-gacgctaccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
AcoN24d             aatgccggatat-gacactgctgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
AgKG4               aatgccggatatNgacatcactgggcatctggtggtg-tggaaagatttatcggcttgggatgggcc
argp5               aatgccggatat-gacattgccgggcatctggtggtg-tggaaagatttatcggatcgggatgggcc
ArI4                aatgccggatat-gacactgccgggcatctggtggtg-tggaaagatttatcggctcggNatgg-c-
Arugo.en            aatgccggatat-gacactgccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
CeD                 aatgccggatat-gacattgccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
Cnep.en             aataccggatatNgacactactggacatctggtggtg-tggaaagatttatcggcttgggatgggcc
Dcan.en             aataccggatatNgacattgccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
Dryas.en            aataccggatat-gacactactggacatctggtggtg-tggaaagatttatcggcttgggatgggcc
purshia16           aataccggatat-gacattgccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
Cean-thy.           aatgccggatat-gacactaccgggcatctggtggtg-tggaaagatttatcggcttgggatgggcc
Cean-ame.           aatgccggatat-gacactaccgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
Ea1-2               aatgccggatat-gacatcatcgggcatctggtggtg-tggaaagatttatcggcttgggatgggcc
SCN10a              aatgccggatat-gacatcgtcgggcatctggtggtg-tggaaagatttatcggcttgggatgggcc
HR27-14             aatgccggatat-gacatcatcgggcatctggtggtg-tggaaagatttatcggcttgggatgggcc
mnagaen             aatgccggatat-gacatcactgggcatctggtggtg-tggaaagatttatcggctcgggatgggcc
Catenu              aataccggatac-gacctgcgcccgcatgggtgtggg-tggaaagtttttcggtctgggatgggct
Geo-ob-ob           aataccggatgt-gaccgctgGccgcatggtctggtggtggaaagggtttccggctggggatgggcc
Mask1               11111111111100000000001111111111111101111111111111111111111101111010
Mask2               11111111111100000000001111111111111101111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                    22222222222222222222222222222222222222222222222222222222222222222222
                    00000000111111111122222222223333333333444444444455555555556666666666
                    23456789012345678901234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------------
AVN17s              cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
AcN14a              cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
AcoN24d             cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
AgKG4               cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
argp5               cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
ArI4                cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacggggagccggcctgagag
Arugo.en            cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
CeD                 cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
Cnep.en             cgcggcctatcagcttgttggtggggtaacggcctaccaaggcgacgacgggtagccggcctgagag
Dcan.en             cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
Dryas.en            cgcggcctatcagcttgttggtggggtaacggcctaccaaggcgacgacgggtagccggcctgagag
purshia16           cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
Cean-thy            cgcggcctatcagcttgttggtggggtaacggcctaccaaggcgacgacgggtagccggcctgagag
Cean-ame            cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
Ea1-2               cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
SCN10a              cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
HR27-14             cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
mnagaen             cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
Catenu              cgcggcctatcagcttgttggtggggtgatggcctaccaaggcgacgacgggtagccggcctgagag
Geo-ob-ob           cgcggcctatcagcttgttggtggggtagtggcctaccaaggcgacgacgggtagccggcctgagag
Mask1               11111111111111111111111111111111111111111111111111111111111111111111
Mask2               11111111111111111111111111111111111111111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                    22222222222222222222222222222233333333333333333333333333333333333333
                    67777777777888888888899999999990000000000111111111122222222233333333
```

```
                  90123456789012345678901234567890123456789012345678901234567890123456789012345
-------------------------------------------------------------------------------------------------
AVN17s            ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
AcN14a            ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
AcoN24d           ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
AgKG4             ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
argp5             ggcgaccggccacactgggactgagacacggcccagactcctaccggaggcagcagtggggaatatt
ArI4              ggcgaccggcNacactgggactgagacacggcccNgacacctacgggaggcagcagtggggaatatt
Arugo.en          ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
CeD               ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Cnep.en           ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Dcan.en           ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Dryas.en          ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
purshia16         ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Cean-thy          ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Cean-ame          ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Eal-2             ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
SCN10a            ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
HR27-14           ggcgatcggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
mnagaen           ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Catenu            ggcgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Geo-ob-ob         ggTgaccggccacactgggactgagacacggcccagactcctacgggaggcagcagtggggaatatt
Mask1             11111111110111111111111111111111111101111111111111111111111111111111
Mask2             11111111110111111111111111111111111101111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                  33333333333333333333333333333333333333333333333333333333333333333333444
                  33334444444444555555555566666666667777777777888888888899999999990000
                  67890123456789012345678901234567890123456789012345678901234567890123456789012
-------------------------------------------------------------------------------------------------
AVN17s            gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
AcN14a            gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
AcoN24d           gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
AgKG4             gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
argp5             gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
ArI4              gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgggggatgacgggcttcgggttgtaaacct
Arugo.en          gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
CeD               gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
Cnep.en           gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgaaggccttcgggttgtaaacct
Dcan.en           gcgcaatgggcggaagcctgacacagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
Dryas.en          gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgaaggccttcgggttgtaaacct
purshia16         gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
Cean-thy          gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgaaggccttcgggttgtaaacct
Cean-ame          gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
Eal-2             gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
SCN10a            gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
HR27-14           gcgcaatgggcgaaagcctgacgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
mnagaen            gcgcaatgggcggaagcctgacgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
Catenu            gcacaatgggcggaagcctgatgcagcgacgccgcgtgagggatgacggccttcgggttgtaaacct
Geo-ob-ob         gcgcaatgggcggaagcctgacgcagcgacgccgcgtgggggatgacggccttcgggttgtaaacct
Mask1             11111111111111111111111111111111111111111111111111111111111111111111
Mask2             11111111111111111111111111111111111111111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                  44444444444444444444444444444444444444444444444444444444444444444444444
                  00000001111111111222222222233333333334444444444555555555566666666666
                  34567890123456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------------------------
AVN17s            ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
AcN14a            ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
AcoN24d           ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
AgKG4             ctttcagcagggacgaagcga-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
argp5             ctttcagcagggacgaagcg-tga-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
ArI4              ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Arugo.en          ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
CeD               ctttcagcagggacgaagcga-ga-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Cnep.en           ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
```

```
Dcan.en          ctttcagcagggacgaagcga-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Dryas.en         ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
purshia16        ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Cean-thy         ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Cean-ame         ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Ea1-2            ctttcagcagggacgaagcga-ga-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
SCN10a           ctttcagcagggacgaagcga-ga-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
HR27-14          ctttcagcagggacgaagcga-ga-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
mnagaen           ctttcagcagggacgaagcgc-aa-gtgacggtacctgcagaagaagcaccggccaactacgtgcca
Catenu           ctttcagcagggacgaagcg-taa-gtgacggtacctgcagaagaagcgccggccaactacgtgcca
Geo-ob-ob        ctttcagcagggacgaagccgtgaggtgacggtacctgcagaagaagcaccgcccaactacgtgcca
Mask1            11111111111111111111001101111111111111111111111111111111111111111111
Mask2            11111111111111111111001101111111111111111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                 44444444444444444444444444444455555555555555555555555555555555555555
                 77777777778888888888999999999900000000001111111111222222222233333333
                 01234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------------
AVN17s           gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
AcN14a           gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
AcoN24d          gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
AgKG4            gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
argp5            gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
ArI4             gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Arugo.en         gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
CeD              gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Cnep.en          gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Dcan.en          gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Dryas.en         gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
purshia16        gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Cean-thy         gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Cean-ame         gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Ea1-2            gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
SCN10a           gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
HR27-14          gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
mnagaen           gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg

Catenu           gcagNcgcggtaagacgtagggcgcgagcgttgtccggatttattgggcgtaaagagctcgtaggcg
Geo-ob-ob        gcagccgcggtaatacgtagggtgcaagcgttgtccggaattattgggcgtaaagagctcgtaggcg
Mask1            11110111111111111111111111111111111111111111111111111111111111111111
Mask2            11110111111111111111111111111111111111111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                 55555555555555555555555555555555555555555555555555555555555556666
                 33344444444445555555555666666666677777777778888888888999999999990000
                 78901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------
AVN17s           gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
AcN14a           gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
AcoN24d          gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
AgKG4            gcctgtcgcgtcggctgtgaaatatcggggctcaaccccgggcctgcagtcgatacgggcaggctag
argp5            gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
ArI4             gcctgtcgcgtcggctgtgaaaac--ggggctcaaccccgggcctgcagtcgatacgggcaggctag
Arugo.en         gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
CeD              gcctgttgcgtcggctgtgaaaacccggggctcaactccgggcctgcagtcgatacgggcaggctag
Cnep.en          gcttgtcacgtcggctgtgaaatcccggggctcaactccgggcgtgcagtcgatacgggcaggctag
Dcan.en          gcttgttacgtcggctgtgaaatcccggggctcaactccgggcgtgcagtcgatacgggcaggctag
Dryas.en         gcttgttacgtcggctgtgaaatcccggggctcaactccgggcgtgcagtcgatacgggcaggctag
purshia16        gcttgtcgcgtcggctgtgaaatcccggggctcaactccgggcgtgcagtcgatacgggcaggctag
Cean-thy         gcttgtcgcgtcggctgtgaaatcccggggctcaactccgggcgtgcagtcgatacgggcaggctag
Cean-ame         gcttgtcgcgtcCgctgtgaaatcccggggctcaactccgggcgtgcagtcgatacgggcaggctag
Ea1-2            gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
SCN10a           gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
HR27-14          gcctgtcgcgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
mnagaen           gcttgtcacgtcggctgtgaaaacccggggctcaaccccgggcctgcagtcgatacgggcaggctag
Catenu           gcttgtcgcgtcgactgtgaaaacccgcggctcaaccgcgggcctgcagccgatacgggcaggctag
```

```
Geo-ob-ob          gttggtcgcgtcggctgtgaaaacccggagctcaactccgggcctgcagtcgatacgggccgacttg
Mask1              11111111111111111111111100111111111111111111111111111111111111111
Mask2              11111111111111111111111100111111111111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                   66666666666666666666666666666666666666666666666666666666666666666
                   00000011111111112222222222333333333344444444445555555556666666667
                   45678901234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------------
AVN17s             agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
AcN14a             agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
AcoN24d            agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
AgKG4              agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgcNagatatcaggaggaacacc
argp5              agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
ArI4               agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Arugo.en           agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
CeD                agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Cnep.en            agtccggtaggggagactggaattcctggtgtagcggtgaaatgcgcTagatatcaggaggaacacc
Dcan.en            agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgcNagatatcaggaggaacacc
Dryas.en           agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
purshia16          agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaagaacacc
Cean-thy           agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Cean-ame           agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Ea1-2              agtccggtaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
SCN10a             agtccggtaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
HR27-14            agtccggtaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
mnagaen            agtccggcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Catenu             agttcggtaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Geo-ob-ob          agtgttgcaggggagactggaattcctggtgtagcggtgaaatgcgc-agatatcaggaggaacacc
Mask1              11111111111111111111111111111111111111111111111101111111111111111
Mask2              11111111111111111111111111111111111111111111111101111111111111111
```

16S rDNA data matrix (continued)

```
                   66666666666666666666666666667777777777777777777777777777777777777
                   77777777788888888889999999990000000000111111111122222222233333333
                   12345678901234567890123456789012345678901234567890123456789012345
-------------------------------------------------------------------------------------
AVN17s             ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
AcN14a             ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
AcoN24d            ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
AgKG4              ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
argp5              ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
ArI4               ggtggcgaaggcgggtctctgggccgggactgacgctaaggagcgaaagcgtggggagcgaacagga
Arugo.en           ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
CeD                ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
Cnep.en            ggtggcgaaggcgggtctctgggccggtactgacgctaaggagcgaaagcgtggggagcgaacagga
Dcan.en            ggtggcgaaggcgggtctctgggccggtactgacgctaaggagcgaaagcgtggggagcgaacagga
Dryas.en           ggtggcgaaggcgggtctctgggccggtactgacgctaaggagcgaaagcgtggggagcgaacagga
purshia16          ggtggcgaaggcgggtctctgggccggtactgacgcttaggagcgaaagcgtggggagcgaacagga
Cean-thy           ggtggcgaaagcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
Cean-ame           ggtggcgaaagcgggtctctgggccggaactgacgTtaaggagcgaaagcgtggggagcgaacagga
Ea1-2              ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
SCN10a             ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
HR27-14            ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
mnagaen            ggtggcgaaggcgggtctctgggccggaactgacgctaaggagcgaaagcgtggggagcgaacagga
Catenu             ggtggcgaaggcgggtctctgggccgatactgacgctgaggagcgaaagcgtggggagcgaacagga
Geo-ob-ob          ggtggcgaaggcgggtctctgggcaacaactgacgctgaggagcgaaagcgtggggagcgaacagga
Mask1              11111111111111111111111111111111111111111111111111111111111111111
Mask2              11111111111111111111111111111111111111111111111111111111111111111
```

16S rDNA data matrix (continued)

```
                   77777777777777777777777777777777777777777777777777777777788888
                   33444444444455555555556666666666777777777788888888889999999900000
                   89012345678901234567890123456789012345678901234567890123456789012
-------------------------------------------------------------------------------------
AVN17s             ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
```

```
AcN14a          ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
AcoN24d         ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
AgKG4           ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
argp5           ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
ArI4            ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttcca-ggcctcc
Arugo.en        ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
CeD             ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
Cnep.en         ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
Dcan.en         ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
Dryas.en        ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
purshia16       ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggactttccacggcctcc
Cean-thy        ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
Cean-ame        ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
Ea1-2           ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
SCN10a          ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
HR27-14         ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
mnagaen          ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggaccttccacggcctcc
Catenu          ttagataccctggtagtccacgctgtaaacgttgggcgctaggtgtgggggcctctccggttctct
Geo-ob-ob       ttagataccctggtagtccacgccgtaaacgttgggcgctaggtgtgggggccattccacggtctcc
Mask1           11111111111111111111111111111111111111111111111111111111111101111111
Mask2           11111111111111111111111111111111111111111111111111111111111101111111
```

16S rDNA data matrix (continued)

```
                888888888888888888888888888888888888888888888888888888888888888888
                000001111111111222222222233333333334444444444555555555566666666677
                567890123456789012345678901234567890123456789012345678901234567890
-------------------------------------------------------------------------------
AVN17s          gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
AcN14a          gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
AcoN24d         gtgccgcagctaacgcattaagcNccccgcctggggagtacggccgcaaggNtaaaactcaaaggaa
AgKG4           gtgccgcNgctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
argp5           gtgccgcagctaacgcattaagcgccccgtctggggagtacggacgcaaggctaaaactcaaaggaa
ArI4            gtgccgcacgtaacgcattaagcgccN-gcctggggagtacggccgcaaggctaaaactcaaaggaa
Arugo.en        gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
CeD             gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Cnep.en         gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Dcan.en         gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Dryas.en        gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
purshia16       gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaaGtcaaaggaa
Cean-thy        gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Cean-ame        gtgccTcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Ea1-2           gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
SCN10a          gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
HR27-14         gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
mnagaen          gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Catenu          gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Geo-ob-ob       gtgccgcagctaacgcattaagcgccccgcctggggagtacggccgcaaggctaaaactcaaaggaa
Mask1           11111110111111111111111101100111111111111111111111111101111111111111
Mask2           11111110111111111111111011001111111111111111111111111101111111111111
```

16S rDNA data matrix (continued)

```
                888888888888888888888888888999999999999999999999999999999999999999
                777777778888888888999999999900000000001111111111222222222333333333
                234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------
AVN17s          ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
AcN14a          ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
AcoN24d         ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
AgKG4           ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
argp5           ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
ArI4            ttgacgggggcc-gcacaagcggcggagcatgtggcttaattcgatgNaacgcgaagaaccttacca
Arugo.en        ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
CeD             ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
Cnep.en         ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
Dcan.en         ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
Dryas.en        ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
purshia16       ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
```

```
Cean-thy          ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
Cean-ame          ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
Ea1-2             ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
SCN10a            ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
HR27-14           ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
mnagaen           ttgacgggggcccgcacaagcggcggagcatgtggcttaattcgatgcaacgcgaagaaccttacca
Catenu            ttgacgggggcccgcacaagcggcggagcatgcggattaattcgatgcaacgcgaagaaccttacct
Geo-ob-ob         ttgacgggggcccgcacaagcggcggagcatgttgcttaattcgatgcaacgcgaagaaccttacct
Mask1             11111111111101111111111111111111111111111111111101111111111111111111
Mask2             11111111111101111111111111111111111111111111111101111111111111111111
```

16S rDNA data matrix (continued)

```
                                                                                111111
                  999999999999999999999999999999999999999999999999999999999999000000
                  344444444445555555555666666666677777777778888888888999999999000000
                  901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------
AVN17s            gggtttgacatgcagggaaatcctccagagatgggggg-tccgtaa--gggtcctgcacaggtggtg
AcN14a            gggcttgacatgcagggaaatctcgtagagatacgggg-tccgtaa--gggtcctgcacaggtggtg
AcoN24d           gggcttgacatgcagggaaatctcgtagagatacgggg-tccgtaa--gggtcctgcacaggtggtg
AgKG4             gggtttgacatgcagggaaatccatcagagatggtggg-tccgcNaa-gggtcctgcacaggtggtg
argp5             gggtttgacatgcagggaaatcctccagagatgggggg-tccgtaa--gggtcctgcacaggtggtg
ArI4              gggcttgacatgcagggaaatctcgtagagatacgggg-tccgtaa--gggtcctgcacaggtggtg
Arugo.en          gggcttgacatgcagggaaatctcgtagagatacgggg-tccgtaa--gggtcctgcacaggtggtg
CeD               gggcttgacatgcagggaaatctcgtagagatacgggg-tccgtaa--gggtcctgcacaggtggtg
Cnep.en           aggcttgacatgcagggaaatctcgtagagatacgggg-tccgtNaa-gggctctgcacaggtggtg
Dcan.en           aggcttgacatgcagggaaatctcgtagagatacgggg-tccgtNaa-gggctctgcacaggtggtg
Dryas.en          aggcttgacatgcagggaaatctcgtagagatacgggg-tccgt-aa-gggctctgcacaggtggtg
purshia16         aggcttgacatgcagggaaatctcgtagagatacgggg-tccgt-aa-gggctctgcacaggtggtg
Cean-thy          aggcttgacatgcagggaaatctcgtaaaaatacgggg-tccgt-aa-gggctctgcacaggtggtg
Cean-ame          aggcttgacatgcagggaaatctcgtaaaaatacgggg-tccgt-aa-gggctctgcacaggtggtg
Ea1-2             gggcctgacatgcagagaaatcctgtagagatatgggg-tcctt-ag-gggctctgcacaggtggtg
SCN10a            gggcttgacatgcagagaaatcctgtagagatatgggg-tcctt-ag-gggctctgcacaggtggtg
HR27-14           gggcttgacatgcagagaaatcctgtagagatatgggg-tcctt-ag-gggctctgcacaggtggtg
mnagaen           gggcttgacatgcagggaaatctcgtagagatacgggg-tccgtaa--gggtcctgcacaggtggtg
Catenu            gggtttgacatcactcgaaaactcgcagagatgcgggg-tcctt-cg-gggcgggtgacaggtggtg
Geo-ob-ob         aggcttgacatgcacggaaatctcgcagagatgcgggg-tccgtaa--gggccgtgcacaggtggtg
Mask1             11111111111111111111110000111111111111101111100001111111111111111111
Mask2             11111111111111111111110000111111111111101111100001111111111111111111
```

16S rDNA data matrix (continued)

```
                  11111111111111111111111111111111111111111111111111111111111111111111
                  00000000000000000000000000000000000000000000000000000000000000000000
                  00001111111111222222222233333333334444444444555555555566666666777
                  67890123456789012345678901234567890123456789012345678901234567890123
-------------------------------------------------------------------------------------
AVN17s            catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
AcN14a            catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
AcoN24d           catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
AgKG4             catggctgtcgtcagctcgtgtNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5             catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
ArI4              catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
Arugo.en          catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
CeD               catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
Cnep.en           catggctgtcgtcagctcgtgtNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcan.en           catggctgtcgtcagctcgtgtNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en          catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgttc
purshia16         catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgttc
Cean-thy          catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgttc
Cean-ame          catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgttc
Ea1-2             catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccccttgtcc
SCN10a            catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccccttatcc
HR27-14           catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccccttgtcc
mnagaen           catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgtcc
Catenu            catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgttc
Geo-ob-ob         catggttgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctcgttc
Mask1             11111111111111111111111111111111111111111111111111111111111111111111
```

```
Mask2           11111111111111111111100000000000000000000000000000000000000000000000
```

16S rDNA data matrix (continued)

```
                1111111111111111111111111111111111111111111111111111111111111111111111
                0000000000000000000000000011111111111111111111111111111111111111111111
                7777777888888888899999999990000000000011111111112222222223333333333
                3456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------
AVN17s          tatgttgccagcg-------agtaatgtcggggactcataggagactgccggggtcaactcggagga
AcN14a          tatgttgccagcg-------agttatgtcggggactcataggagactgccggggtcaactcggagga
AcoN24d         tatgttgccagcg-------agttatgtcggggactcataggagactgccggggtcaactcggagga
AgKG4           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5           tatgttgccagcg-------agtcgtgtcggggactcataggagactgccggggtcaactcggagga
ArI4            tatgttgccagcg-------agttatgtcggggactcataggagactgccggggccaactcggagga
Arugo.en        tatgttgccagcg-------agttatgtcggggactcataggagactgccggggtcaactcggagga
CeD             tatgttgccagcg-------agtcgtgtcggggactcataggagactgccggggtcaactcggagga
Cnep.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcan.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en        tatgttgccagcg-------agtcatgtcggggactcataggagactgccggggtcaactcggagga
purshia16       tatgttgccagcg-------agtaatgtcggggactcataggagactgccggggtcaactcggagga
Cean-thy        tatgttgccagcg-------agtaatgtcggggactcataggagactgccggggtcaactcggagga
Cean-ame        tatgttgccagcg-------agtaatgtcggggactcataAgagactgccggggtcaactcggagga
Ea1-2           tatgttgccagcg-------agtcatgtcggggactcataggagactgccggggtcaactcggagga
SCN10a          tatgttgccagcg-------cgtcatggcggggactcataggagactgccggggtcaactcggagga
HR27-14         tatgttgccagcg-------agtcatgtcggggactcataggagactgccggggtcaactcggagga
mnagaen         tatgttgccagcg-------agtaatgtcggggactcataggagactgccggggtcaactcggagga
Catenu          gatgttgccagcg-------cgttatggcggggactcatcgaagactgccggggtcaactcggagga
Geo-ob-ob       catgttgccagca-------cgtGatggtggggactcatgggagactgccggggtcaactcggagga
Mask1           11111111111110000000111111111111111111111111111111111111111111111111
Mask2           00000000000000000000000000000000000000000000000000000000000000000000
```

16S rDNA data matrix (continued)

```
                1111111111111111111111111111111111111111111111111111111111111111111111
                1111111111111111111111111111111111111111111111111111111111112222222
                4444444445555555556666666666777777777788888888899999999990000000
                0123456789012345678901234567890123456789012345678901234567890123456
-------------------------------------------------------------------------------
AVN17s          aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
AcN14a          aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
AcoN24d         aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
AgKG4           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5           aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
ArI4            aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
Arugo.en        aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
CeD             aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
Cnep.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcan.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en        aggtggggatgacgtcaagtcatcatgccccttacgtcttgggctgcacacatgctacaatggccgg
purshia16       aggtggggatgacgtcaagtcatcatgccccttacgtcttgggctgcacacatgctacaatggccgg
Cean-thy        aggtggggatgacgtcaagtcatcatgccccttacgtcTtgggctgcacacatgctacaatggccgg
Cean-ame        aggtggggatgacgtcaagtcatcatgccccttacgtcTtgggctgcacacatgctacaatggccgg
Ea1-2           aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
SCN10a          aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
HR27-14         aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
mnagaen         aggtggggatgacgtcaagtcatcatgccccttacgtcctgggctgcacacatgctacaatggccgg
Catenu          aggtggggatgacgtcaagtcatcatgccccttatgtccagggcttcacgcatgctacaatggccgg
Geo-ob-ob       aggtggggatgacgtcaaAtcatcatgccccttatgtctagggctgcaAacatgctacaatggccgg
Mask1           11111111111111111111111111111111111111111111111111111111111111111111
Nask2           00000000000000000000000000000000000000000000000000000000000000000000
```

16S rDNA data matrix (continued)

```
                1111111111111111111111111111111111111111111111111111111111111111111111
                2222222222222222222222222222222222222222222222222222222222222222222222
                0001111111111222222222233333333334444444444555555555566666666667777
                7890123456789012345678901234567890123456789012345678901234567890123
-------------------------------------------------------------------------------
```

```
AVN17s          tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
AcN14a          tacaaagggctgcgataccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
AcoN24d         tacaaagggctgcgataccgcgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
AgKG4           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5           tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
ArI4            tacaaagggctgcgataccgcgaggtggagcgaatcccaaaaagccgggctcagttcggatcggggt
Arugo.en        tacaaagggctgcgataccgcgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
CeD             tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
Cnep.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcan.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en        tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
purshia16       tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
Cean-thy        tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
Cean-ame        tacaaagggctgcgatgccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
Ea1-2           tacaatgggctgcgataccgcgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
SCN10a          tacaaagggctgcgGtaccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
HR27-14         tacaatgggctgcgataccgcgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
mnagaen          tacaaagggctgcgataccgcgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
Catenu          tacaaagggTtgcgataccgtgaggtggagcgaatcccaaaaagccggtctcagttcggatcggggt
Geo-ob-ob       tacaaagggctgcgataccgcAaggtggagcgaatcccaaaaagccggtctcagttcggattggggt
Mask1           11111111111111111111111111111111111111111111111111111111111111111
Mask2           00000000000000000000000000000000000000000000000000000000000000000
```

16S rDNA data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111
                222222222222222222222222233333333333333333333333333333333333333333
                777777888888888899999999900000000001111111111222222222233333333334
                456789012345678901234567890123456789012345678901234567890123456789
--------------------------------------------------------------------------------
AVN17s          ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
AcN14a          ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
AcoN24d         ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
AgKG4           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5           ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
ArI4            ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
Arugo.en        ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
CeD             ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
Cnep.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcan.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en        ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
purshia16       ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
Cean-thy        ctgcaactcgaccccgtgaaatcggagtcgctagtaatcgcagatcaacaatgctgcggtgaatacg
Cean-ame        ctgcaactcgaccccgtgaaatcggaAtcgctagtaatcgcagatcagcaatgctgcggtgaatacc
Ea1-2           ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
SCN10a          ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcattgctgcggtgaatacg
HR27-14         ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
mnagaen          ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaatgctgcggtgaatacg
Catenu          ctgcaactcgaccccgtgaagtcggagtcgctagtaatcgcagatcagcaacgctgcggtgaatacg
Geo-ob-ob       ctgcaactcgaccccatgaagtTggagtcgctagtaatcgcagatcagcaacgctgcggtgaatacg
Mask1           11111111111111111111111111111111111111111111111101111111111111111
Mask2           00000000000000000000000000000000000000000000000000000000000000000
```

16S rDNA data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111
                333333333333333333333333333333333333333333333333333333333344444444
                444444444555555555566666666667777777777888888888999999999900000000
                123456789012345678901234567890123456789012345678901234567890123456
--------------------------------------------------------------------------------
AVN17s          ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccca
AcN14a          ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
AcoN24d         ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
AgKG4           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
argp5           ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
ArI4            ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacNcccgaagc-ggtggccta
Arugo.en        ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
CeD             ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
Cnep.en         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

Dcan.en	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dryas.en	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
purshia16	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccca
Cean-thy	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccNa
Cean-ame	ttcccCggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccNa
Ea1-2	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccca
SCN10a	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccca
HR27-14	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccca
mnagaen	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacacccgaagccggtggccta
Catenu	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggCaacacccgaagccggtggccta
Geo-ob-ob	ttcccgggccttgtacacaccgcccgtcacgtcacgaaagtcggtaacgcccgaagccggtggccca
Mask1	1111111111111111111111111111111111111111111111110111111110111111101
Mask2	0000000000000000000000000000000000000000000000000000000000000000000

16S rDNA data matrix (continued)

	1111
	4444
	0011
	8901
AVN17s	accc
AcN14a	accc
AcoN24d	accc
AgKG4	NNNN
argp5	accc
ArI4	accc
Arugo.en	accc
CeD	accc
Cnep.en	NNNN
Dcan.en	NNNN
Dryas.en	accc
purshia16	accc
Ceano-thy	accc
Ceano-ame	accc
Ea1-2	accc
SCN10a	accc
HR27-14	accc
mnagaen	accc
cea16s1.3	accc
cea16s5.1	accc
Catenu	accc
Geo-ob-ob	accc
Mask1	1111
Mask2	0000

Appendix 1.3. The data matrix of 34 rbcL nucleotide sequences used for reconstructing phylogeny of actinorhizal plants

rbcL data matrix

```
                    1111111111222222222233333333334444444444555555555566666666
                   1234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------------
Alnus              gttggattcaaagctggtgttaaagattataaattgacttatcatactcctgactatgaaaccaaag
Allocasuarina      gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Betula             gttgcattcaaagctggtgttaaagagtataaattgacttattatactcctgactatgaaaccaaag
Castania           gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatcaaaccaaag
Casuarina          gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Comptonia          gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaagg
Corylus            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Gymnostoma         gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Juglans            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgaatatcaaaccaaag
Myrica             gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaagg
Nothofagus         gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactacgaaaccaaag
Pilea              gttggattcaaagctggtgttaaagattacaaattgacttattacactcctgaatatgaaaccaaag
Begonia            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgaatatgaaaccaaag
Coriaria           gttggattcaaagctggtgttaaagattataaattgacttattatactcctgaatatgaaaccaaag
Cucurbita          gttggattcaaagctggtgttaaagattataaattgacttattatactcctgaatatgaaaccaaag
Datisca            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Tetrameles         gttggattcaaagctggtgttaaagattataaattgacttattatactcctgaatatcaaaccaaag
Cercocarpus        gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Cowania            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Dryas              gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Neillia            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Purshia            gttggattcaaagccggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Rosa               gttggattcaaagctggtgttaaagattataaattgacttattatactccggattatgaaaccaaag
Ceanothus          gttggattcaaagccggtgttaaagattataaattgacttattacactcctgactatgaaacaaaag
Colletia           gttggattcaaagccggtgttaaagattataaattgacttattacactcctgactatgaaaccaaag
Discaria           gttggattcaaagccggtgttaaagattataaattgacttattacactcctgactatgaaaccaaag
Elaeagnus          gttggattcaaagcgggtgttaaagattataaattgacttattacactcctgactatcaaaccaaag
Hippophaë          gttggattcaaagcgggtgttaaagattataaattgacttattacactcctgactatcaaaccaaag
Parasponia         gttggattcaaagccggtgttaaagattataaattgacttattacactcctgactatcaaaccaaag
Rhamnus            gttggattcaaggctggtgttaaagattataaattgacttattatactcctgaatatgaaaccaaag
Shepherdia         gttggattcaaagNgggtgttaaagattataaattgaattattacactcctgactatcaaaccaaag
Trevoa             gttggattcaaagccggtgttaaagattataaattgacttattacactcctgactatgaaaccaaag
Moringa            gttggattcaaagctggtgttaaagattataaattgacttattatactcctgactatgaaaccaaag
Platanus           gttggattcaaagctggtgttaaagattacaaattgacttattatactcctgaatataaaaccaaag
Vitis              gttggattcaaagccggtgttaaagattacaaattgacttattatactcctgaatatgagaccaaac
Mask               11111111111110111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                                                    1111111111111111111111111111111111
                   6677777777778888888888999999999900000000001111111111222222222233333
                   8901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------
Alnus              atactgatatcttggcagcgttccgagtaactcctcaacctggagtgccgcctgaggaagcaggggc
Allocasuarina      atactgatatcttggcagcgttccgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Betula             atactgatatcttggcagcgttccgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Castania           atactgatatcttggcagccttccgagtaactcctcaacctggagttccgccggaggaagcaggggc
Casuarina          atactgatatcttggcagcgttccgagtaactcctcaacctggagttccacctgaggaagcaggggc
Comptonia          atactgatatcttggcagcgttccgagtaactcctcaacctggagtgccgcctgaggaagcaggggc
Corylus            atactgatatcttggcagcgttccgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Gymnostoma         atactgatatcttggcagcgttccgagtaactcctcaacctggagttccacctgaggaagcaggggc
Juglans            atactgatatcttggcagcgttccgagtaagccctcaacctggagttccgcctgaggaagcaggggc
Myrica             atactgatatcttggcagcgttccgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Nothofagus         atactgatatattggcagcattccgagtaactcctcaacccggagttccgcctgaggaagcgggggc
Pilea              atactgatatcttagcagcatttcgagtaactcctcaacctggagttccccctgaagaagcaggggc
Begonia            atactgatatcttggcagcattccgagtaagtcctcaaccgggagttccacctgaggaagccggggc
Coriaria           atactgatatcttggcagcattccgagtaactcctcaacccggagttccacctgaggaagcaggggc
Cucurbita          atactgatatcttggcagcattccgagtaactcctcaaccgggagttccacccgaggaagcgggggc
```

```
Datisca          atacggatatcttggcagcattccgagtaactcctcaacctggagttccacctgaggaagcaggtgc
Tetrameles       atactgatatcttggcagcattccgagtaactcctcaacccggagttccgcctgaggaacgaggggc
Cercocarpus      atactgatatcttggcagcatttcgagtaactcctcaacctggagttcctcctgaggaagcaggggc
Cowania          atactgatatcttggcagcatttcgagtaactcctcaacctggagttcctcctgaggaagcaggggc
Dryas            atactgatatcttggcagcatttcgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Neillia          atactgatatcttggcagcatttcgagtaactcctcaacctggagttccacctgaggaagcaggggc
Purshia          atactgatatcttggcagcatttcgagtaactcctcaacctggagttcctcctgaggaagcaggggc
Rosa             atactgatatcttggcagcatttcgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Ceanothus        atactgatatcttggcagcgtttcgagtaactcctcaacctggagttccgcctgaggaagcaggggc
Colletia         atactgatatcttggcagcgtttcgagtaactcctcaacctggcgttccgcctgaggaagcaggggc
Discaria         atactgatatcttggcagcgtttcgagtaactcctcaacctggcgttccacctgaggaagcaggggc
Elaeagnus        atactgatatcttagcagcatttcgagtaactcctcaaccaggagttccacccgaagaagcaggggc
Hippophaë        atactgatatcttagcagcatttcgagtaactcctcaacctggagttccacctgaagaagcaggggc
Parasponia       atacggatatcttggcagcatttcgagtaactcctcaacctggagttcccccctgaagaagcaggggc
Rhamnus          ataccgatatcttggcagcatttcgagtaactcctcaacccggagttccacctgaggaagcaggggc
Shepherdia       atactgatatcttagcagcatttcgagtaactcctcaacctggagttccacctgaagaagcaggggc
Trevoa           atactgatatcttggcagcgtttcgagtaactcctcaacctggcgttccgcctgaggaagcaggggc
Moringa          atactgatatcttggcagcattccgagtaactcctcaacccggagttccacctgaggaagccggggc
Platanus         atactgatatcttagcagcattccgagtaactcctcaaccaggagttccacctgaggaagcaggggc
Vitis            ctactgatatcttggcagcattccgagtaactcctcaacctggagttccacccgaagaagcaggggc
Mask             11111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                 11111111111111111111111111111111111111111111111111111111111111122
                 33333444444444455555555556666666666777777777788888888899999999900
                 56789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------
Alnus            agcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatggacttactagtctt
Allocasuarina    ggcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttactagtctt
Betula           agcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatggacttactagtctt
Castania         cgcggtagctgctgaatcttccactgggacatggagaactgtgtggactgacgggcttaccagtctt
Casuarina        ggcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatggagttactagtctt
Comptonia        agcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttactagtctt
Corylus          agcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatggacttactagtctt
Gymnostoma       ggcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttactagtctt
Juglans          agcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttactagtctt
Myrica           agcagtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttactagtctt
Nothofagus       tgcggtagctgcggaatcctctaccggtacatggacaacggtgtggaccgatggacttaccagtctt
Pilea            tgcagtagctgctgaatcttctactggtacatggacaactgtatggactgatgggcttaccactctt
Begonia          cgctgtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttaccagtctt
Coriaria         cgctgtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttaccagtctt
Cucurbita        cgccgtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttaccagtctt
Datisca          cgctgtagctgctgaatcttctactggtacatggacaactgtgtggaccgatggtcttaccagtctt
Tetrameles       cgctgtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttaccagtctt
Cercocarpus      tgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Cowania          tgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Dryas            tgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Neillia          agcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Purshia          tgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Rosa             agcggtagctgcggaatcttctactggtacatggacaactgtatggactgatgggcttaccagtctt
Ceanothus        cgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Colletia         cgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggctgaccagtctt
Discaria         cgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggctgaccagtctt
Elaeagnus        agcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Hippophaë        agcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Parasponia       agcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagcctt
Rhamnus          cgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Shepherdia       agcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggcttaccagtctt
Trevoa           cgcggtagctgctgaatcttctactggtacatggacaactgtatggactgacgggctgaccagtctt
Moringa          cgcggtagctgctgaatcttctactggtacatggacaactgtgtggaccgatgggcttaccagcctt
Platanus         cgcggtagctgcggaatcttctactggtacatggacaactgtgtggaccgatggacttaccagcctt
Vitis            cgcggtagctgctgaatcttctactggtacatggacaactgtgtggactgatggacttaccagcctt
Mask             11111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                 22222222222222222222222222222222222222222222222222222222222222222
```

```
                  0000000011111111112222222222333333333344444444445555555555666666666
                  2345678901234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------------
Alnus             gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Allocasuarina     gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Betula            gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Castania          gatcgttacaaaggacgatgctaccacatcgagccggttgctggagaagaaaatcaatttattgctt
Casuarina         gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Comptonia         gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Corylus           gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Gymnostoma        gatcgttacaaaggacgatgctaccacatcgagccagttgctggcgaagaaaatcaatatattgctt
Juglans           gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaaatcaatttattgctt
Myrica            gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaagtcaatttattgctt
Nothofagus        gatcgttacaaaggacgatgctaccacatcgagccagttgctggagaagaaactcaatttattgctt
Pilea             gatcgctacaaaggtcgatgctaccacatcgagcctgttcctggcgaagaaagtcaatttattgctt
Begonia           gatcgttacaaaggacgatgctattacatcgagcctgttgctggagaagaaaatcaatatattgctt
Coriaria          gatcgttacaaaggccgatgctaccaaatcgagcctgttgctggagaagaaaatcaatatattgctt
Cucurbita         gatcgttacaaaggacgatgctatggaatcgagcctgttcctggagaagaaaatcaatatattgctt
Datisca           gatcgttacaaaggacgatgctacggcatcgagcctgttgctggaggagaaaatcaatatattgctt
Tetrameles        gatcgttacaaaggacgatgctatcacatcgagcctgttgctggagaagaaagtcaatttattgctt
Cercocarpus       gatcgttacaaagggcgatgctaccacattgagcccgttgctggagaagaaagtcaatttattgctt
Cowania           gatcgttacaaagggcgatgctaccacatcgagcccgttgctggagaagaaaatcaatttattgctt
Dryas             gatcgttacaaagggcgatgctaccacatcgagcccgttcctggagaagaaagtcaatttattgctt
Neillia           gatcgttacaaaggtcgatgctaccacatcgagcccgttgctggagaagaaagtcaatttattgctt
Purshia           gatcgttacaaagggcgatgctaccacatcgagcccgttgctggagaagaaactcaatttattgctt
Rosa              gatcgttacaaagggcgatgctaccacattgaacctgttgctggagaagaaagtcaatttattgctt
Ceanothus         gatcgttacaaaggtcgatgctaccacatcgagcccgttgctggagaagaaactcaatttattgctt
Colletia          gatcgttacaaaggtcgatgctaccacctcgagcccgttgctggagaagaaaatcaatatattgctt
Discaria          gatcgttacaaaggtcgatgctaccacctcgagcccgttgctggagaagaaaatcaatatattgctt
Elaeagnus         gatcgttacaagggtcgatgctacggccttgagcccgttgctggagaagaaaatcaatatattgctt
Hippophaë         gatcgttacaaaggtcgatgctaccaccttgagcccgttgctggagaagaaaatcaatatattgctt
Parasponia        gatcgatacaaaggtcgatgctaccacatcgagccggttgctggagaagaaaatcaatatattgctt
Rhamnus           gatcgttacaaaggtcgatgctaccacatcgagcccgttgctggagaagaaagtcaatttattgctt
Shepherdia        gatcgttacaaaggtcgatgctaccacattgagcccgttgctggagatgaaaatcaatatattgctt
Trevoa            gatcgttacaaaggtcgatgctaccacctcgagcccgttgctggagaagaaaatcaatatattgctt
Moringa           gatcgttacaaaggacgatgctaccacatcgagcccattgcgggagaagaaaatcaatttattgctt
Platanus          gatcgttacaaaggacgatgctaccacatcgaaactgttactggagaagaaaatcaatttattgttt
Vitis             gatcgttacaaaggacgatgctaccacatcgagcccgttgctggagaagaaagtcaatttattgctt
Mask              11111111111111111111111111111111111111111111111111111111111111111111

rbcL data matrix (continued)

                  2222222222222222222222222222223333333333333333333333333333333333333
                  6777777777788888888889999999990000000000111111111122222222223333333
                  9012345678901234567890123456789012345678901234567890123456789012345
-------------------------------------------------------------------------------------
Alnus             atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Allocasuarina     atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Betula            atgtagcttacccttagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Castania          atgtagcttacccctagacctctttgaagaaggttctgttactaacatgtttacttccattgtggg
Casuarina         atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Comptonia         atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Corylus           atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Gymnostoma        atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Juglans           atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Myrica            atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Nothofagus        atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Pilea             atgtagcttatccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Begonia           atgtagcttatccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Coriaria          atgtagcttatccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Cucurbita         atgtagcttatcccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Datisca           atgtagcttatcccttagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Tetrameles        atgtagcttatcccttagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
Cercocarpus       atgtagcttacccttacaccttttgaagaaggttctgttactaacatgtttacttccattgtagg
Cowania           atgtagcttaccacttagaccttttgaagaaggttctgttactaacatgtttacttacattgtagg
Dryas             atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtagg
Neillia           atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtagg
Purshia           atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtagg
Rosa              atgtagcttacccctagaccttttgaagagggttcggttactaacatgtttacttccattgtagg
Ceanothus         atgtagcttacccctagaccttttgaagaaggttctgttactaacatgtttacttccattgtggg
```

```
Colletia            atgtagcttacccctagacctttttgaagaaggttctgttactaacatgtttacttccattgtggg
Discaria            atgtagcttacccctttagacctttttgaagaaggttctgttactaacatgtttacttccattgtggg
Elaeagnus           atgtagcttacccctagacctttttgaagaaggctctgttactaacatgtttacttccattgtggg
Hippophaë           atgtagcttacccctagacctttttgaagaaggctctgttactaacatgtttacttccattgtggg
Parasponia          atgtagcttacccctagacctttttgaggaaggttctgttactaacatgtttacttccattgtggg
Rhamnus             atgtagcttacccctagacctttttgaagaaggttctgttactaacatgtttacttccattgtggg
Shepherdia          atgtagcttacccctagacctttttgaagaaggctctgttactaacatgtttacttccattgtggg
Trevoa              atgtagcttacccctagacctttttgaagaaggttctgttactaacatgtttacttccattgtggg
Moringa             atgtagcttacccctagacctttttgaagaaggttctgttactaacatgtttacttctattgtggg
Platanus            atgtagcttacccctttagacctttttgaagaaggttctgttactaacatgtttacttccattgtggg
Vitis               atgtagcttacccctttagacctttttgaagaaggctctgttactaacatgtttacttccattgtggg
Mask                11111111111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                    33333333333333333333333333333333333333333333333333333333333333333333333444
                    33334444444444555555555566666666667777777777888888888899999999999000
                    67890123456789012345678901234567890123456789012345678901234567890123456789012
------------------------------------------------------------------------------------------
Alnus               taatgtatttggattcaaggccctgcgtgctctacgtctggaggatttgcgaatccctcctgcttat
Allocasuarina       taatgtatttggattcaaggccctgcgtgctctacgtctggaggatttgcgaatccctactgcttat
Betula              taatgtatttggattcaaggccctgcgtgctctacgtctggaggatttgcgaatccctcctgcttat
Castania            taatgtatttggattcaaggccctgcgcgctctacgtctggaggatttgcgaatccctacttcttat
Casuarina           taatgtatttggattcaaggccctgcgtgctctacgtctggaggatttgcgaatccctactgcttat
Comptonia           taatgtatttggattcaaggccctgcgcgctctacgtctggaggatttgcgaatccctacttcttat
Corylus             taatgtatttggattcaaggccctgcgtgctctacgtctggaggatttgcgaatccctcccgcttat
Gymnostoma          taatgtatttggattcaaggccctgcgtgctctacgtctggaggatttgcgaatccctacttcttat
Juglans             taatgtatttggattcaaggccctgcgcgctctacgtctggaggatttgcgaatccctcctgcttat
Myrica              taatgtatttggattcaaggccctgcgcgctctacgtctggaggatttgcgaatccctacttcttat
Nothofagus          taatgtatttggattcaaggccctgcgcgctctacgtctggaggatttgcgaatccctcctgcttat
Pilea               taatgtatttgggttcaaggccctgcgagctctacgtttggaggatttgcgaattcccactgcttac
Begonia             taatgtatttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccctactgcttat
Coriaria            taatgtatttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccctcctgcttat
Cucurbita           taatgtatttggattcaaggctctgcgtgctctacgtctggaggatttgcgaatccctactgcttat
Datisca             taatgtatttgggttcaaggccctgcgcgctctacgtctggaagatttgcgaatccctcctgcttat
Tetrameles          taatgtatttgggttcaaggccctgcgcgctctacgtctggaggatttgagaatccctactgcttat
Cercocarpus         taatgtgtttgggttcaaggccctgcgcgctctacgtctggaggatttacgaatccctcctgcttat
Cowania             taatgtgtttgggttcaaggccctgcgcgctctacgtctggaggatttacgaatccctcctgcttat
Dryas               taatgtgtttgggttcaaggccctgcgcgctctacgtctggaggatttacgaatccctactgcttat
Neillia             taatgtatttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccctactgcttat
Purshia             taatgtgtttgggttcaaggccctgcgctctacgtctggaggatttacgaatccctcctgcttat
Rosa                taatgtgtttgggttcaaggccttgcgcgctctacgtctggaggatttacgaatccctactgcttat
Ceanothus           taatgtttttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccctcctgcttat
Colletia            taatgtttttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccccccctgcttat
Discaria            taatgtttttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccccccctgcttat
Elaeagnus           taatgtgtttggattcaaggccctgcgcgctctacgtctggaagatttgcgaatccctactgcttat
Hippophaë           taatgtgtttggattcaaggcccttcgcgctctacgtctggaggatttgcgaatccctactgcttat
Parasponia          taacgtatttgggttcaaggctctgcgcgctctacgtctggaggatttgagaatccctcctgcttat
Rhamnus             taatgtatttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccccccctgcttat
Shepherdia          taatgtgtttggattcaaggccctgcgcgctctacgtctggaggatttgcgaatccctactgcttat
Trevoa              taatgtttttgggttcaaggccctgcgcgctctacgtctggaggatttgcgaatccccccctgcttat
Moringa             taatgtatttgggttcaaagccctgcgcgctctacgtctagaggatctgcgaatccctcctgcttat
Platanus            tgatgtatttgggttcaaagccctacgcgctctacgtctggaggatctacgaatccctcctgcttat
Vitis               taatgtgtttgggttcaaagctctgcgcgctctacgtctagaggatctgcgaatccccccctgcttat
Mask                11111111111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                    44444444444444444444444444444444444444444444444444444444444444444444444444
                    00000001111111111222222222233333333334444444444555555555566666666666
                    34567890123456789012345678901234567890123456789012345678901234567890123456789
------------------------------------------------------------------------------------------
Alnus               tctaaaactttccaaggcccgcctcacggcatccaagttgagagagataaattaaacaagtatggcc
Allocasuarina       tctaaaactttccaaggcccgcctcacggcatccaagttgagagagataaattaaacaagtatggcc
Betula              tctaaaactttccaaggcccgcctcacggcatccaagttgagagagataaattaaacaaatatggcc
Castania            tctaaaactttccaaggtccgcctcatggcatccaagttgagagggataaattaaacaagtatggcc
Casuarina           tctaaaactttccaaggtccgcctcatggcatccaagttgagagagataaattaaacaagtatggcc
Comptonia           gttaaaactttccaaggcccgccgcatggcatccaagttgagagagataaattaaacaagtatggcc
Corylus             tctaagactttccaaggcccgcctcacggcatccaagttgagagagataaattaaacaagtatggcc
```

```
Gymnostoma      tctaaaactttccaaggtccgcctcatggcatccaagttgagagagataaattaaacaagtatggcc
Juglans         tctaaaactttccaaggcccgcctcatggaatccaagttgagagagataaattaaacaagtatggcc
Myrica          gttaaaactttccaaggcccgccgcatggcatccaagttgagagagataaattaaacaagtatggcc
Nothofagus      gctaaaactttccaaggtccgcctcatggcatccaagttgagagagataaattaaacaagtatggcc
Pilea           actaaaactttccaaggcccacccccacggcatccaagttgagagagataaattgaataagtatggcc
Begonia         gttaaaactttttcaaggccctcctcatggtatccaggttgagagagataaattgaacaagtatggcc
Coriaria        gttaaaactttccaaggccctcctcatggtatccaagttgagagagataaattgaacaagtatggcc
Cucurbita       attaaaactttccaaggcccgcctcatggtatccaggttgaaagagataaattgaacaagtatggtc
Datisca         gttaaaactttccaagggccgcctcatggtatccaagttgagagagataaattgaacaagtatggcc
Tetrameles      attaaaactttccaaggcccgcctcatggtatccaagttgagagagataaattgaacaagtatggcc
Cercocarpus     gttaaaactttccaaggcccgcctcatgggatccaagttgagagagataaattgaacaagtacggcc
Cowania         gttaaaactttccaaggcccgcctcatgggatccaagttgagagagataaattgaacaagtacggcc
Dryas           gttaaaactttccaaggcccgcctcatgggatccaagttgagagagataaattgaacaagtacggcc
Neillia         gttaaaactttccaaggcccgcctcatgggatccaagttgagagagataaattgaacaagtacggcc
Purshia         gttaaaactttccaaggcccgcctcatgggatccaagttgagagagataaattgaacaagtatggcc
Rosa            gttaaaactttccaaggcccgcctcacgggatccaagttgaaagagataaattgaacaagtatggcc
Ceanothus       tctaaaactttccaaggaccgcctcacggcatccaagttgaaagagataaattgaacaagtatggcc
Colletia        tctaaaactttccaaggaccgcctcacggcatccaagttgaaagagataaattgaacaagtatggcc
Discaria        tctaaaactttccaaggaccgcctcacggcatccaagttgaaagagataaattgaacaagtatggcc
Elaeagnus       actaaaactttccaaggcccgcctcatggtatccaagttgagagagataaattgaacaagtatggcc
Hippophaë       actaaaactttccaaggcccgcctcatggtatccaagttgagagagataaattgaacaagtatggcc
Parasponia      tctaaaactttccaaggcccacctcatggcatccaagttgagagagataaattgaacaagtatggcc
Rhamnus         actaaaactttccaaggcccgcctcatggcatccaagttgagagagataagttgaacaagtatggcc
Shepherdia      actaaaactttccaaggcccgcctcatggtatccaagttgagagagataaattgaacaagtacggcc
Trevoa          tctaaaactttccaaggaccgcctcacggcatccaagttgaaagagataaattgaacaagtatggcc
Moringa         tctaaaactttccagggaccacctcatggtatccaagttgaaagagataaattgaacaagtatggcc
Platanus        tctaaaactttccaaggcccgcctcatggcatccaagttgagagagataaattgaacaagtatggtc
Vitis           actaaaactttccaaggcccgcctcatggcatccaagttgagagagataaattgaacaagtatggtc
Mask            11111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                444444444444444444444444444444555555555555555555555555555555555555
                777777777788888888889999999999000000000011111111112222222223333333
                012345678901234567890123456789012345678901234567890123456789012345678
--------------------------------------------------------------------------------
Alnus           gccccctattgggatgtactattaagcctaaattgggattatccgctaagaattacggtagagcagt
Allocasuarina   gaccccctattaggatgtactattaaacctaaattgggattatctgctaagaattacggcagagcagt
Betula          gccccctattaggatgtactattaaacctaaattgggattatccgctaagaattacggtagagcagt
Castania        gccccctattaggatgtactattaaacctaaattgggattatccgctaagaattacggtagagcagt
Casuarina       gaccccctattaggatgtactattaaacctaaattgggattatctgctaagaattacggtagagcagt
Comptonia       gccccctattgggatgtactattaaacccaaattgggattatccgctaagaattacggtagagcagt
Corylus         gccccctattgggatgtactattaaacctaaattgggattatccgctaagaattacggtagagcggt
Gymnostoma      gaccccctattaggatgtactattaaacctaaattgggattatccgctaagaattacggtagagcagt
Juglans         gccctctattgggatgtactattaaacctaaattgggattatccgctaagaattacggtagagcagt
Myrica          gccccctattgggatgtactattaaacccaaattgggattatccgctaagaattacggtagagcagt
Nothofagus      gccccctattgggatgtactattaaaccgaaattggggttatccgctaagaattacggtagagcagt
Pilea           gccctctattgggatgtactattaaacctaaattggggctatccgctaagaattatggtagagctgt
Begonia         gccccctattgggatgtactattaaaccgaaattgggattatccgctaagaattacggtagagctgt
Coriaria        gccccctattgggatgtactattaaacctaaattgggattatctgctaagaattacggtagagcagt
Cucurbita       gccctctattgggatgtactattaaaccaaaattgggattatccgctaagaattatggtagagcagt
Datisca         gccctctattgggatgtactattaaacctaaattgggattatccgctaagaattacggtagagctgt
Tetrameles      gccctctattgggatgtactattaaacctaaattgggattatccgctaagaattacggtagagctgt
Cercocarpus     gccccctattgggatgtactattaaacctaaattggggttatccgctaagaattatggtagagcagt
Cowania         gccccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagcagt
Dryas           gccccctattgggatgtactattaaacctaaattggggttatctgctaagaattacggtagagcagt
Neillia         gccctctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagcagt
Purshia         gccccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagcagt
Rosa            gccccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagcagt
Ceanothus       gccccctattgggatgtactattaaacctaaactggggttatccgctaagaattacggtagagcagt
Colletia        gtcccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagcagt
Discaria        gtcccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagcagt
Elaeagnus       gtcccctattaggatgtactattaaacctaaattggggttatccgctaagaattatggtagagctgt
Hippophaë       gtcccctattaggatgtactattaaacctaaattggggttatccgctaaaaattacggtagagctgt
Parasponia      gccccctattgggatgtactattaaaccgaaattgggattatccgctaagaattacggtagagcagt
Rhamnus         gccccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtagagccgt
Shepherdia      gtcccctattaggatgtactattaaacctaaattggggttatccgctaaaaattacggtagagctgt
Trevoa          gtcccctattgggatgtactattaaacctaaattggggttatccgctaagaattacggtcgagcagt
Moringa         gtcccctattaggatgtactattaaacctaaattgggtttatccgctaagaactacggtagagcggt
Platanus        gtcccctattgggatgtactattaaaccaaaattggggttatctgctaagaactatggtagagcagt
```

```
Vitis           gtcccctattgggatgtactattaaacctaaattggggttatccgctaagaactatggtagagctgt
Mask            11111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                555555555555555555555555555555555555555555555555555555555556666
                333444444444455555555556666666666777777777788888888899999999990000
                789012345678901234567890123456789012345678901234567890123456789123
----------------------------------------------------------------------------------
Alnus           ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Allocasuarina   ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgaaaatgtgaattcccaaccattt
Betula          ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Castania        ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgagaacgttaattcccaaccattt
Casuarina       ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgaaaatgtgaattcccaaccattt
Comptonia       ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccgttt
Corylus         ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Gymnostoma      ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgagaatgtgaattcccaaccattt
Juglans         ttatgaatgtctccgcggtgggcttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Myrica          ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccgttt
Nothofagus      ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaacgtaaattcccaaccattt
Pilea           ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Begonia         ttatgaatgtctacgcggtggacttgattttaccaaagatgatgaaaacgtgaattcccaaccattt
Coriaria        ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgaaaacgtgaattctcaaccattt
Cucurbita       ttatgaatgtctacgcggtggacttgattttaccaaagatgatgaaaacgtgaattcccaaccattt
Datisca         ttatgaatgtctacgtggtggacttgattttaccaaagatgatgaaaacgtgaattcccaaccattt
Tetrameles      ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgaaaacgttaattcccaaccattt
Cercocarpus     ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaatgtgaattcccaaccattt
Cowania         ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaatgtgaattcccaaccattt
Dryas           ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaatgtgaattcccaaccattt
Neillia         ttatgaatgtctacgcggtggacttgattttaccaaagatgatgagaatgttaattcccaaccattt
Purshia         ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaatgtgaattcccaaccattt
Rosa            ttatgaatgtctccgcggtggacttgattttaccaaagatgatgagaatgttaattcccaaccattt
Ceanothus       ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Colletia        ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Discaria        ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Elaeagnus       ttatgaatgtcttcgcggtgggcttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Hippophaë       ttatgaatgtcttcgtggtggacttgattttaccaaggatgatgagaacgtgaattcccaaccattt
Parasponia      ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Rhamnus         ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccgttt
Shepherdia      ttatgaatgtcttcgcggtggacttgattttaccaaggatgatgagaacgtgaattcccaaccattt
Trevoa          ttatgaatgtcttcgcggtggacttgattttaccaaagatgatgagaacgtgaattcccaaccattt
Moringa         ttatgaatgtctacgcggtggacttgattttaccaaagatgatgagaatgtgaactcccaaccattt
Platanus        ttatgaatgtctccgcggtggacttgattttaccaaggatgatgagaacgtgaactcccaaccattt
Vitis           ttatgaatgtctccgcggtggacttgattttactaaagatgatgagaacgtgaactcccaaccattt
Mask            11111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                666666666666666666666666666666666666666666666666666666666666666
                000000111111111122222222223333333333444444444455555555556666666667
                456789012345678901234567890123456789012345678901234567890123456789
----------------------------------------------------------------------------------
Alnus           atgcgttggagagaccgtttcctattttgtgccgaagcaatttataaagcgcaggctgaaacaggtg
Allocasuarina   atgcgttggagagaccgtttcctattttgtgccgaagcactttataaagcgcaggccgaaacaggtg
Betula          atgcgttggagagaccgtttcctattttgtgccgaagcaatttataaagcgcaggctgaaacaggtg
Castania        atgcgttggagagaccgtttcctattttgtgccgaagcaatttataaagcgcaggctgaaacaggtg
Casuarina       atgcgttggagagaccgtttcctattttgtgccgaagcactttataaagcgcaggccgaaacaggtg
Comptonia       atgcgttggagagatcgtttcctattttgtgccgaagctctttataaagcgcaggctgaaacaggtg
Corylus         atgcgttggagagatcgtttcctattttgtgccgaagcaatttataaagcgcaggctgaaacaggtg
Gymnostoma      atgcgttggagagaccgtttcctattttgtgccgaagcactttataaagcgcaggctgaaacaggtg
Juglans         atgcgttggagagaccgtttcctattttgtgccgaagcaatttataaagcgcaggctgaaaccggtg
Myrica          atgcgttggagagatcgtttcctattttgtgccgaagctctttataaagcgcaggctgaaacaggtg
Nothofagus      atgcgttggagagaccgtttcctattttgtgccgaagcaatttataaagcacaggctgaaacaggtg
Pilea           atgcgttggagagaccgtttcttattttgtgccgaagcaatttataaatcacaggctgaaacaggtg
Begonia         atgcgttggagagaccgtttcctattttgtgcggaagcactttataaagcacaggctgaaacaggtg
Coriaria        atgcgttggagagaccgtttcctattttgtgcggaagcaatttataaatcacaggctgaaacaggtg
Cucurbita       atgcgttggagagaccgtttcctattttgtgcggaagccatttataaatcacaggctgaaacaggtg
```

```
Datisca          atgcgttggagagaccgtttcctattttgtgcggaacgaatttataaatcacaggctgaaacaggtg
Tetrameles       atgcgttggagagaccgtttcctattttgtgcggaagcaatttataaagcacaggctgaaacaggtg
Cercocarpus      atgcgttggagagaccggttcttattttgtgccgaagcactttataaagcacaggctgaaacaggtg
Cowania          atgcgttggagagaccggttcttattttgtgccgaagcaatttataaagcacaggccgaaacaggtg
Dryas            atgcgttggagagaccggttcttattttgtgccgaagcactttataaagcacaggctgaaacaggtg
Neillia          atgcgttggagagaccgtttcttattttgtgccgaagcaatttataaagcacaggctgaaacaggtg
Purshia          atgcgttggagagaccggttcttattttgtgccgaagcaatttataaagcacaggccgaaacaggtg
Rosa             atgcgttggagagaccgtttcttattttgtgccgaagcaatttataaatcgcaggctgaaacaggtg
Ceanothus        atgcgttggagagaccgtttcttattttgtgccgaagccctttataaagcacaggctgaaacaggtg
Colletia         atgcgttggagagaccgtttcttattttgtgccgaagccctttataaagcacagactgaaacaggtg
Discaria         atgcgttggagagaccgtttcttattttgtgccgaagccctttataaagcacaggctgaaacaggtg
Elaeagnus        atgcgttggagagaccgtttcttattttgtgccgaagccctttttaaagcacaggctgaaacaggtg
Hippophaë        atgcgttggagagaccgtttcttattttgtgccgaagccctttttaaagcacaggctgaaacaggtg
Parasponia       atgcgttggagagatcgtttcttattttgtgccgaagctatttataaagcacaggctgaaacaggcg
Rhamnus          atgcgttggagagaccggttcttattttgtgccgaagcaatttataaagcacaggccgaaactggtg
Shepherdia       atgcgttggagagaccgtttcttattttgtgccgaagccctttttaaagcacaggctgaaacaggtg
Trevoa           atgcgttggagagaccgtttcttattttgtgccgaagccctttataaagcacaggctgaaacaggtg
Moringa          atgcgttggagagaccgtttcttattttgtgccgaagctatttataaagcacaggctgaaacaggtg
Platanus         atgcgttggagagaccgtttcttattttgtgccgaagccctttataaagcacaggctgagacaggtg
Vitis            atgcgttggagagaccgtttcttattttgtgccgaagccatttttaaatcacaggctgaaacaggtg
Mask             11111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                 66666666666666666666666666667777777777777777777777777777777777777
                 77777777788888888889999999990000000000111111111122222222233333333
                 12345678901234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------------
Alnus            aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggctgtatt
Allocasuarina    aaatcaaagggcattacttgaatgctaccgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Betula           aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggctatatt
Castania         aaatcaaggggcattacttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Casuarina        aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggctatatt
Comptonia        aaatcaaaggacattacttgaatgctactgcaggtacatgcgaagaaatgattaaaagggctgtatt
Corylus          aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Gymnostoma       aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgataaaaagggctgtatt
Juglans          aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Myrica           aaatcaaaggacattacttgaatgctactgcaggtacatgcgaagaaatgattaaaagggctgtatt
Nothofagus       aaatcaaagggcattacttgaatgctactgcgggtacatgcgaagaaatgatcaaaagggctgtatt
Pilea            aaatcaaaggacattatttgaatgctactgcaggtacatgtgaagaaatgatgaaaagggcaatctt
Begonia          aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgaataaaagggctgtatt
Coriaria         aaatcaaggggcattacttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Cucurbita        aaatcaagggacattacttgaatgctactgcgggtacatgcgaagaaatgatgaaaagggctatatt
Datisca          aaatcaaggggcattacttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Tetrameles       aaatcaaggggcattacttgaatgctactgcgggtacatgcgaagaaatgatgaaaagggctgtatt
Cercocarpus      aaatcaaagggcattacttgaacgctactgcaggtacatgtgaagagatgatcaaaagagctgtatt
Cowania          aaatcaaagggcattacttgaacgctactgcaggtacatgtgaagagatgatgaaaagagctgtatt
Dryas            aaatcaaagggcattacttgaacgctactgcaggtacatgtgaagagatgatcaaaagagctgtatt
Neillia          aaatcaaagggcattacttgaacgctactgcaggtacatgcgaagagatgatcaaaagagctgtatt
Purshia          aaatcaaagggcattacttgaacgctactgcaggtacatgtgaagagatgatcaaaagagctgtatt
Rosa             aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagatatgatgaaaagagccgtatt
Ceanothus        aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgattaaaagggctgtatt
Colletia         aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggctatatt
Discaria         aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgattaaaagggctgtatt
Elaeagnus        aaatcaaagggcattatttgaatgctacggcaggtacatgcgaagaaatgatgaaaagggctatatt
Hippophaë        aaatcaaagggcattatttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggcagtatt
Parasponia       aaatcaaaggacattacttgaatgctaccgcaggtacatgggaagaaatgatcaaaagggctgtatt
Rhamnus          aaatcaaagggcattacttgaatgctacggcaggtacatgcgaagaaatgatgaaaagggctgcatg
Shepherdia       aaatcaaagggcattatttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggcagNatt
Trevoa           aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgattaaaagggctgtatt
Moringa          aaatcaaagggcattatttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Platanus         aaatcaaaggacattacttgaatgctactgcaggtacatgcgaagaaatgatcaaaagggctgtatt
Vitis            aaatcaaagggcattacttgaatgctactgcaggtacatgcgaagaaatgatgaaaagggctatatt
Mask             11111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                 77777777777777777777777777777777777777777777777777777777777788888
                 33444444444455555555556666666666777777777788888888899999999900000
```

```
                    8901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------------
Alnus               tgccagagaattgggagttcctatcgtaatgcatgactacttaaccgggggattcactgcaaatact
Allocasuarina       tgccagagaattgggagttcctatcgtaatgcatgactacttaaccgggggattcactgcaaatact
Betula              tgccagagaattgggagttcctatcgtaatgcatgactacttaaccgggggattcactgcaaatact
Castania            tgccagagaactgggagttcctatcgtaatgcatgactacttaaccgggggattcactgcaaatact
Casuarina           tgccagagaattgggagttcctatcgtaatgcatgactacttaaccgggggttcactgcaaatact
Comptonia           tgccagagaattgggagttcctatcgtaatgcacgactacttaaccgggggattcacggcaaatact
Corylus             tgccagagaattgggagttcctatcgtaatgcatgactacttaaccgggggattcactgcaaatact
Gymnostoma          tgccagagaattgggagttcctatcgtaatgcacgactacttaaccgggggattcactgcaaatact
Juglans             tgccagagaattgggagttcctatcgtaatgcatgactacttaaccgggggatttactgcaaatact
Myrica              tgccagagaattgggagttcctatcgtaatgcacgactacttaaccgggggattcacggcaaatact
Nothofagus          tgccagagaattgggagctcctatcgtaatgcacgactacttaacaggggattcactgcaaatact
Pilea               tgccagagaattgggagctcctatcgtaatgcacgattacttaacaggaggattcactgcaaatact
Begonia             tgccagagaattggcagttcctatcgtaatgcatgactacttaacaggggattcacggcaaatact
Coriaria            tgccagagaattgggagttcctatcgtaatgcatgactacttaacgggggattcactgcaaatact
Cucurbita           tgctcgagaattgggagctcctatcgtaatgcatgactacttaacaggtggattcactgcaaatact
Datisca             tgccagagaattgggagttcctatcgtaatgcatgactacttaacaggggcttcactgcaaatact
Tetrameles          tgccagagaattgggagttcctatcataatgcatgactacttaacaggggattcactgcaaatact
Cercocarpus         tgccagagaattgggagttcctatcgtaatgcatgattacttaacaggggattcactgcaaatacc
Cowania             tgccagagaattgggagttcctatcgtaatgcatgattacttaacaggggattcactgcaaatacc
Dryas               tgccagagaattgggagttcctatcgtaatgcatgattacttaacaggggattcactgcaaatacc
Neillia             tgccagagaattgggagttcctatcgtaatgcatgattacttaacaggggattcactgcaaatact
Purshia             tgccagagaattgggagttcctatcgtaatgcatgattacttaacaggggattcactgcaaatacc
Rosa                tgccagagaattgggagttcctatcgtaatgcacgattacttaacgggggattcactgcaaatact
Ceanothus           tgccagagaattgggagttcctattgtaatgcatgattacttaacaggggattcactgcaaatact
Colletia            tgccagagaattgggagttcctattgtaatgcatgattacttaacaggggattcactgcaaatact
Discaria            tgccagagaattgggagttcctattgtaatgcatgattacttaacaggggattcactgcaaatact
Elaeagnus           tgccagagaattgggagctcctatcgtaatgcacgattacttaactgggggattcactgcaaacact
Hippophaë           tgccagagaattgggagttcctatcgtaatgcacgattacttaactgggggattcactgcaaacact
Parasponia          tgccagagaattgggagttcctatcgtaatgcatgattacttaacaggaggattcactgcaaatact
Rhamnus             tgccagagaattgggagttcctatcgtaatgcacgattacttaacgggggattcactgcaaatact
Shepherdia          tgccagagaattgggagttcctatcgtaatgcacgattacttaactgggggattcactgcaaacact
Trevoa              tgccagagaattgggagttcctattgtaatgcatgattacttaacaggggattcactgcaaatact
Moringa             tgccagagaattgggagctcctatcataatgcatgactacttaacaggggattcaccgcaaatact
Platanus            tgccagagaattgggagttcctatcgtaatgcatgactacttaacaggggattcactgcaaatact
Vitis               tgccagagaattgggagttcctatcgtaatgcatgactacttaacaggggattcaccgcaaatact
Mask                11111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                    8888888888888888888888888888888888888888888888888888888888888888888
                    0000011111111112222222222333333333344444444445555555555666666666677
                    5678901234567890123456789012345678901234567890123456789012345678901
-------------------------------------------------------------------------------------------
Alnus               agcttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Allocasuarina       agtttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Betula              agcttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Castania            agcttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Casuarina           agtttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Comptonia           agcttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Corylus             agcttggctcattattgccgggataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Gymnostoma          agtttggctcattattgccgggataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Juglans             agcttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Myrica              agcttggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Nothofagus          agcttggctcattattgccgagataatggtctacttcttcacattcatcgtgcaatgcatgcagtta
Pilea               accctggctcattattgccgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Begonia             agcttggctcattattgccgagataatggtctacttcttcacattcatcgtgcaatgcatgcagtta
Coriaria            agcttggctcattattgccgagacaatggtctacttcttcacattcaccgtgcaatgcatgcagtta
Cucurbita           agcttggctcattattgccgagataatggtctacttcttcacattcaccgtgcaatgcatgccgtta
Datisca             agcttggctcattattgccgagataatggtctacttcttcacattcaccgtgcaatgcatgcagtta
Tetrameles          agcttggctcattattgccgagataatggtctacttcttcacattcaccgtgcaatgcatgcagtta
Cercocarpus         agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Cowania             agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Dryas               agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Neillia             agcttggctcattattgccgagataatggtctacttcttcatattcaccgtgcaatgcatgcagtta
Purshia             agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Rosa                accttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgctgtta
Ceanothus           agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Colletia            agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
```

```
Discaria          agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Elaeagnus         agcttggctcattattgccgagataatggtttacttcttcacatccaccgtgcaatgcatgcagtta
Hippophaë         accttggctcattattgccgagataatggtctacttcttcacattcaccgtgcaatgcatgcagtta
Parasponia        agcctggctcattattgtcgagataatggtctacttcttcacatccatcgtgcaatgcatgcagtta
Rhamnus           accttggctcattattgccgagataatggtctgcttcttcacatccaccgtgcaatgcatgcggtta
Shepherdia        accttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Trevoa            agcttggctcattattgccgagataatggtctacttcttcacatccaccgtgcaatgcatgcagtta
Moringa           agcttggctcattattgccgagataatggtctacttcttcacatccatcgcgcaatgcatgcagtta
Platanus          accttgtcttattattgccgagataatggcctccttcttcacatccatcgcgcaatgcatgcagtta
Vitis             agcttggctcagtattgccgagataatggcctacttcttcacatccatcgtgcaatgcatgcagtta
Mask              11111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                  88888888888888888888888888899999999999999999999999999999999999999999
                  77777777888888888899999999990000000000111111111122222222223333333333
                  23456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------------------
Alnus             ttgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Allocasuarina     ttgatagacagaagaatcatggtatgcactttcgtgtattagctaaagcgttacgtatgtccggtgg
Betula            ttgatagacagaagaatcatggtatgcactttcgtgtactagctaaagcgttacgtatgtctggtgg
Castania          ttgatcgacagaagaatcatggtatacactttcgtgtactagctaaagcattacgtatgtctggtgg
Casuarina         ttgatagacagaagaatcatggtatgcactttcgtgtattagctaaagcgttacgtctgtccggtgg
Comptonia         ttgatagacagaaaaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Corylus           ttgatagacagaagaatcatggtatgcactttcgtgtactagctaaagcgttacgcatgtctggtgg
Gymnostoma        ttgatagacagaagaatcatggtatgcactttcgtgtattagctaaagcgttacgtctgtctggtgg
Juglans           ttgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Myrica            ttgatagacagaaaaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Nothofagus        ttgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Pilea             ttgatagacagaagaatcatggtatacactttcgcgtgctagctaaagccttacgtatgtctggtgg
Begonia           ttgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Coriaria          ttgatagacagaagaatcatggtatccattttcgtgtactagctaaagcgttacgtatgtctggtgg
Cucurbita         ttgatagacagaagaatcatggtatgcacttccgtgtactagctaaagcgttacgtctgtctggtgg
Datisca           ttgatagacagaagaatcacggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Tetrameles        ttgatagacagaagaatcatggtatgcactttcgtgtactagctaaagcgttacgtctgtctggtgg
Cercocarpus       tcgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Cowania           tcgatagacagaagaatcatggtatgcacttccgagtactagctaaagcgttacgtatgtctggtgg
Dryas             tcgatagacagaagaatcatggcatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Neillia           ttgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Purshia           tcgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Rosa              ttgatagacagaagaatcatggtatgcactttcgtgtactagctaaagcattacgtatgtctggtgg
Ceanothus         ttgatagacagaaaaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Colletia          ttgatagacagaaaaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Discaria          ttgatagacagaaaaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Elaeagnus         ttgatagacagaaaaatcatggtatgcacttccgtgtactggctaaagggttacgtatgtctggtgg
Hippophaë         ttgatagacagaaaaatcatggtatgcactttcgNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Parasponia        ttgatagacaaaagaatcatggtatgcactttcgtgtactagctaaagcgttacgcctgtctggtgg
Rhamnus           ttgatagacagaaaaatcatggtatgcactttcgtgtactagctaaagcgttacgtatgtctggtgg
Shepherdia        ttgatagacagaaaaatcatggtatgcacttccgtgtactggctaaagcgttacgtatgtctggtgg
Trevoa            ttgatagacagaaaaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggtgg
Moringa           ttgatagacagaagaatcatggtatgcactttcgtgtactagctaaagccttacgtttgtctggtgg
Platanus          ttgatagacagaagaatcatggtatacactttcgtgtactagctaaagcgttacgtatgtctggcgg
Vitis             ttgatagacagaagaatcatggtatgcactttcgtgtactagctaaagccttacgtatgtctggagg
Mask              11111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                                                                                  111111
                  99999999999999999999999999999999999999999999999999999999999999000000
                  34444444444555555555566666666667777777777888888888999999999000000
                  90123456789012345678901234567890123456789012345678901234567890123456
-------------------------------------------------------------------------------------
Alnus             agatcatattcacgcgggtaccgtagtaggtaaacttgaaggggaaagagagatcactttaggcttt
Allocasuarina     agatcatattcacgccggtaccgtagtaggtaaacttgaaggggaaagagatatcactttaggcttt
Betula            agatcatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagagatcactttaggcttt
Castania          agatcatattcatgccggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Casuarina         agatcatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagatatcactttaggcttt
```

```
Comptonia       agatcatattcactctggtaccgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Corylus         agatcatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagagatcactttaggcttt
Gymnostoma      agatcatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagagatcactttaggcttt
Juglans         agatcatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Myrica          agatcatattcactctggtaccgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Nothofagus      agatcatattcacgccgggaccgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Pilea           agatcatattcactctggtactgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Begonia         agaccatattcactctggtactgtagtaggtaaacttgaaggggaaagagaaatcactttagggttt
Coriaria        agatcatattcacgcaggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Cucurbita       agaccatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Datisca         agatcatattcactctggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Tetrameles      agatcatattcacgctggtaccgtagtaggtaaacttgaaggggaaagagagatcactttaggtttt
Cercocarpus     agatcatatacactctggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Cowania         agatcatatacacgctggtaccgtagtaggaaaacttgaaggggaaagagaaatcactttaggcttt
Dryas           agatcatatacactctggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Neillia         agatcatatacactctggtaccgtagtaggtaaacttgaaggggaaagagaaataactttaggcttt
Purshia         agatcatatacacgctggtaccgtagtaggtaaacttgaaggggaaagagagatcactttaggcttt
Rosa            agatcatatacacgctggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Ceanothus       agatcatatccacgctggtaccgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Colletia        agatcatattcactctggtactgtagtaggtaaacttgaaggtgaaagagaaataactttaggcttt
Discaria        agatcatattcacgctggtactgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Elaeagnus       agatcatattcactctggtactgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Hippophaë       agatcatattcacgctcgtactgtagtaggtaagcttgaaggggaaagagaaattactttaggcttt
Parasponia      NNNNNNNNNNNNNNgctggtactgtagtaggtaaacttgaaggggcaagagacataactttaggcttt
Rhamnus         agatcatattcacgctggtactgtagtaggtaaacttgaaggggaaagagaaatcactttaggcttt
Shepherdia      agatcatattcacgctggtactgtagtgggtaaacttgaaggggaaagagagatcactttaggcttt
Trevoa          agatcatattcacgctggtactgtagtaggtaagcttgaaggggaaagagaaattactttaggcttt
Moringa         agatcatattcactctggtactgtagtaggtaaacttgaaggggaaagagacatcactttaggcttt
Platanus        agatcatattcacgctggtacagtagtaggtaaacttgaaggagaaagagaaattactttgggcttt
Vitis           agatcatattcactcaggtatcgtagtaggtaaacttgaaggggaaagagacataactttgggcttt
Mask            agatcatattcacgccggtaccgtagtaggtaaacttgaaggagaaagagagatcactttgggcttt
                11111111111111111111111111111111111111111111111111111111111111111111
```

Input data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111111
                000000000000000000000000000000000000000000000000000000000000000000000
                000011111111112222222222333333333344444444445555555555666666666777
                678901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------------
Alnus           gttgatttactacgtgatgattata ttgaaaaagatcgaagccgc ggtatttatttcactcaaga
Allocasuarina   gttgatttactacgtgatgattatattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Betula          gttgatttactacgtgatgattatattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Castania        gttgatttactacgtgatgattatattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Casuarina       gttgatttactgcgtgatgattattttgaaaaagatcgaagccgcggtatttatttcactcaagatt
Comptonia       gttgatttactacgcgatgattttattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Corylus         gttgatttactgcgtgatgattatattgaaaaagatcgaagccgcggtatttatttttactcaagatt
Gymnostoma      gttgatttactacgtgatgattatattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Juglans         gttgatttactacgtgatgattttgttgaaaaagatcgaagccgcggtatttatttcactcaagatt
Myrica          gttgatttactacgcgatgattttattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Nothofagus      gttgatttactacgcgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Pilea           gttgatttactacgtgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Begonia         gttgatctactacgtgatgattatattgaaaaagatcgcgaccgcggtatttatttcactcaagatt
Coriaria        gttgatttactacgtgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Cucurbita       gttgatttactacgtgatgattttgttgaaaaagatcgaagccgcggtatttatttcactcaagatt
Datisca         gttgatttactacgtgatgattttattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Tetrameles      gttgatttactacgcgatgattatgttgaaaaagatcgaagccgcggtatttatttcactcaagatt
Cercocarpus     gttgatttactacgtgatgattttgttgaaaaagatcgaagccgcggtatttatttcactcaggatt
Cowania         gttgatttactacgtgatgattttgttgaaaaagatcgaagccgcggtatttatttcactcaggatt
Dryas           gttgatttactacgtgatgattttattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Neillia         gttgatttactacgtgatgattttattgaaaaagatcgaagccgcggtatttatttcactcaagatt
Purshia         gttgatttactacgtgatgattttgttgaaaaagatcgaagccgcggtatttatttcactcaggatt
Rosa            gttgatttactacgtgatgattatattgaaaaagaccgaagccgcggtatttatttttactcaagatt
Ceanothus       gttgatttactacgtgatgattttattgacaaagatcgaagccgtggtatttatttcactcaagatt
Colletia        gttgatttactacgtgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Discaria        gttgatttactacgtgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Elaeagnus       gttgatttactacgtgatgattatattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Hippophaë       NNNNNNNNNNNNNNNNNNNNNNNNNNNgaaaaagatagaagccgtggtatttatttcacccaagatt
Parasponia      gttgatttactacgtgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Rhamnus         gttgatttactacgtgatgattttgttgataaagatcgaagccgtggtatttatttcactcaagatt
Shepherdia      gttgatttactacgtgatgattatattgaaaaagatcgaagccgtggtatttatttcacccaagatt
```

```
Trevoa          gttgatttactacgtgatgattttattgaaaaagatcgaagccgtggtatttatttcactcaagatt
Moringa         gttgatttactacgtgatgattttgttgaaaaagatcgaagccgcggtatttttttcactcaagatt
Platanus        gttgatttactacgtgatgatttggttgaagaagaccgaagtcgtggtatttatttcactcaagatt
Vitis           gttgatttattacgtgatgattttgttgaaaaagaccgaagtcgcggtatttatttcactcaagatt
Mask            11111111111111111111111111111111111111111111111111111111111111111
```

*rbc*L data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111
                000000000000000000000000001111111111111111111111111111111111111111
                777777788888888889999999990000000000111111111122222222223333333333
                345678901234567890123456789012345678901234567890123456789012345678
----------------------------------------------------------------------------------
Alnus           gggtctctctacctggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Allocasuarina   gggtctctctaccaggtgttctgcccgtggcttcaggaggtattcatgtttggcatatgcccgctct
Betula          gggtctctctacctggcgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Castania        gggtctctttaccaggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Casuarina       gggtctctctaccaggtgttctgcccgtggcttcaggaggtattcatgtttggcatatgcccgctct
Comptonia       gggtctctctacctggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Corylus         gggtctctctaccaggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Gymnostom       gggtctctctacctggtgttctgcccgtggcttcaggaggtattcatgtttggcatatgcctgctct
Juglans         gggtctctctacctggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Myrica          gggtctctctacctggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Nothofagus      gggtttctctaccaggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Pilea           gggtctctctaccaggtgttctgcccgtggcttcaggggtattcatgtttggcatatgcctgcttt
Begonia         gggtctctctaccaggtgttctgcctgtagcttcaggggtattcatgtttggcatatgcccgctct
Coriaria        gggtctctttaccaggtgttctgccggtggcttcgggggtattcacgtttggcatatgcccgctct
Cucurbita       gggtctctttaccaggtgttctgccagtggcttccggtggtattcacgtttggcatatgcctgctct
Datisca         gggtctctctaccaggtgttctgcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Tetrameles      gggtctctctaccaggtgttttgcccgtggcttcaggggtattcacgtttggcatatgcccgctct
Cercocarpus     gggtctctctaccaggtgttttgcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Cowania         gggtctctctaccaggtgttttgcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Dryas           gggtctctctaccaggtgttttgcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Neillia         gggtctctctaccaggtgttttgcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Purshia         gggtctctctaccaggtgttttgcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Rosa            gggtttctttaccgggtgttttacctgtggcttccgggggtattcacgtttggcatatgcctgctct
Ceanothus       gggtctctctaccaggtgttctgcctgtggcttcaggggtattcatgtttggcatatgcctgctct
Colletia        gggtctctctaccaggtgttctgcctgtggcttcaggggtattcatgtttggcatatgcctgctct
Discaria        gggtctctctaccaggtgttctgcctgtggcttcaggggtattcatgtttggcatatgcctgctct
Elaeagnus       gggtctctctaccgggtgttcttcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Hippophaë       gggtctctctaccgggtgttcttcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Parasponia      gggtctctttaccaggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Rhamnus         gggtctctttaccgggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Shepherdia      gggtctctctaccgggtgttcttcctgtggcttcaggggtattcacgtttggcatatgcctgctct
Trevoa          gggtctctctaccaggtgttctgcctgtggcttcaggggtattcatgtttggcatatgcctgctct
Moringa         gggtctctctaccaggtgttctgcccgtggcttcgggggtattcacgtttggcatatgcctgctct
Platanus        gggtctctctaccaggtgttctgcccgtggcttcaggggtattcacgtttggcatatgcctgctct
Vitis           gggtctctctaccaggtgtcctgccagtggcttctgggggtattcacgtttggcatatgcctgctct
Mask            11111111111111111111111111111111111111111111111111111111111111111
```

*rbc*L data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111
                111111111111111111111111111111111111111111111111111111111111112222
                444444444455555555556666666666777777777788888888899999999990000000
                012345678901234567890123456789012345678901234567890123456789012345
----------------------------------------------------------------------------------
Alnus           gaccgaaatctttggagatgattccgtactacaattcggtggaggaactttagggcacccttgggga
Allocasuarina   gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcatccttgggga
Betula          gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcacccttgggga
Castania        gaccgaaatctttggagatgattccgtactacaatttggcggaggaactttggggcacccttgggga
Casuarina       gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcatccttgggga
Comptonia       gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcacccttgggga
Corylus         gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcacccttgggga
Gymnostoma      gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcatccttgggga
Juglans         gaccgaaatctttggagatgattccgtactacaattcggtggaggaactttagggcacccttgggga
Myrica          gaccgaaatctttggagatgattccgtactacaattcggcggaggaactttagggcacccttgggga
```

```
Nothofagus      gaccgaaatctttggggatgattccgtattacaatttggcggaggaactttagggcacccttgggga
Pilea           gaccgagatctttggagatgattccgtactacaattcggtggaggaactttagggcatccttgggga
Begonia         gaccgagatctttggagatgattccgtactacaatttggtggagggactttaggacacccatgggga
Coriaria        gaccgagatctttggagatgattccgtactacaattcggtggaggaactttagggcacccttgggga
Cucurbita       aaccgagatttttggagatgattctgtactacaattcggcggaggaactttggggcacccttggggt
Datisca         gaccgagatctttggagatgattccgtactacaattcggtggaggaactttagggcacccttggggt
Tetrameles      gaccgagatctttggagatgattccgtactacaattcggtggaggaactttaggacacccttgggga
Cercocarpus     gaccgagatctttggagatgattctgtactacaattcggcggcggaactttaggacacccttgggga
Cowania         gaccgagatctttggagatgattctgtactacaattcggcggcggaactttaggacacccttgggga
Dryas           gaccgagatctttggagatgattctgtactacaattcggcggcggaactttaggacacccatgggga
Neillia         gaccgagatctttggagatgattctgtactacaatttggcggcggaactttaggacacccttgggga
Purshia         gaccgagatctttggagatgattctgtactacaattcggcggcggaactttaggacacccttgggga
Rosa            gaccgagatctttggagatgattctgtactacaattcggtggaggaactttaggacacccttgggga
Ceanothus       gaccgagatctttggagatgattccgtactacagtttggcggaggaactttaggacacccttgggga
Colletia        gaccgagatctttggagatgattccgtactacagtttggcggaggaacgttaggacacccttgggga
Discaria        gaccgagatctttggagatgattccgtactacagtttggcggaggaactttaggacacccttgggga
Elaeagnus       gaccgagatctttggagatgatgccgtactacaattcggcggaggaactttagggcacccttgggga
Hippophaë       gaccgagatctttggagatgattccgtactacaattcggcggaggaactttagggcacccttgggga
Parasponia      gactgagatctttggcgatgattccgtactacaattcggcggaggaactttaggacatccttgggga
Rhamnus         gaccgagatctttggagacgattccgtactacaattcggcggaggaactatagggcacccttgggga
Shepherdia      gaccgagatctttggagatgatgccgtactacaattcggcggaggaactttagggcacccttgggga
Trevoa          gaccgagatctttggagatgattccgtactacagtttggcggaggaacgttaggacacccttgggga
Moringa         gaccgagatctttggagatgattccgtactacaattcggtggaggaactttaggacacccttgggga
Platanus        gaccgagatctttggggatgattccgtactacagtttggtggaggaactttaggacacccttgggga
Vitis           gaccgagatctttggagatgattccgtactacagttcggtggaggaactttaggacacccttgggga
Mask            11111111111111111111111111111111111111111111111111111111111111111111
```

rbcL data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111111
                222222222222222222222222222222222222222222222222222222222222222222222
                000111111111122222222223333333333444444444455555555556666666666777 7
                789012345678901234567890123456789012345678901234567890123456789012345
-------------------------------------------------------------------------------------
Alnus           aatgcacctggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Allocasuarina   aatgcacccggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgaggggc
Betula          aatgcacccggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Castania        aatgcacctggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Casuarina       aatgcacccggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgaggggc
Comptonia       aatgcacctggtgccgtagctaatcgagtagctctcgaagcatgtgtacaagctcgtaatgagggac
Corylus         aatgcaccgggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Gymnostoma      aatgcacctggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Juglans         aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Myrica          aatgcacctggtgctgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Nothofagus      aatgcacctggcgccgtagctaatcgagtagctctagaagcatgtgtacaagcccgtaatgagggac
Pilea           aatgcaccaggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Begonia         aatgcacccggtgccgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Coriaria        aatgcacctggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Cucurbita       aatgcaccaggtgccgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Datisca         aatgcacccggtgccgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Tetrameles      aatgcacctggtgccgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Cercocarpud     aatgcacctggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Cowania         aatgcacctggtgccgtagctaatcgagtagctctagaagcatgtgtacaagcccgtaatgaggggc
Dryas           aatgcacctggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Neillia         aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Purshia         aatgcacctggtgccgtagctaatcgagtagctctagaagcatgtgtacaagcccgtaatgaggggc
Rosa            aatgcaccaggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Ceanothus       aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Colletia        aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Discaria        aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Elaeagnus       aatgcacctggtgctgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgaggggc
Hippophaë       aatgcacctggtgctgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgaggggc
Parasponia      aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Rhamnus         aatgcacccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Shepherdia      aatgcacctggtgctgtagctaaccgagtagctctagaagcatgtgtacaagctcgtaatgaggggc
Trevoa          aatgcaccggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Moringa         aatgcaccgggtgccgtagctaatcgagtagctctagaagcatgtgtacaagctcgtaatgagggac
Platanus        aatgcaccgggtgccgtagctaatcgagtagctctagaagcatgtgtacaggctcgtaatgagggac
Vitis           aatgcaccgggtgccgtagctaatcgagtagctttagaagcatgtgtacaagctcgtaatgagggac
Mask            11111111111111111111111111111111111111111111111111111111111111111111
```

*rbc*L data matrix (continued)

```
                  1111111111111111111111111111111111111111111111111111111111111111111
                  2222222222222222222222222233333333333333333333333333333333333333333
                  7777778888888888999999999900000000001111111111222222222233333333334
                  4567890123456789012345678901234567890123456789012345678901234567890
-------------------------------------------------------------------------------------
Alnus             gtgatcttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgagctagctgc
Allocasuarina     gcgatcttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgagttagctgc
Betula            gtgatcttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgagctagctgc
Castania          gtgatcttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgagctagctgc
Casuarina         gcgatcttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgagttagctgc
Comptonia         gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgagctagctgc
Corylus           gtgatcttgctcgtgagggtaatgaaattattcgtgcggctgctaaatggagtcctgagctagctgc
Gymnostoma        gtgatcttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgagttagctgc
Juglans           gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgagctagcggc
Myrica            gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgagctagctgc
Nothofagus        gcgaccttgctcgtgagggtaatgaaattatacgcgaggcttgtaaatggagtcctgagctagctgc
Pilea             gtgaccttgctcgtgagggtaatgaaattattcgtgaggctgctaaatggagtcctgaactagcagc
Begonia           gtgatcttgctcgtgagggtaatgaaattatccgtgaggcgagtaaatggagtcccgaactagctgc
Coriaria          gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagtaaatggagtcctgaactagctgc
Cucurbita         gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagtaaatggagtcctgaactagctgc
Datisca           gtgatcttgctcgtgagggtaatgaaattaaccgtgaggctactaaatggagtccagaactagctgc
Tetrameles        gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagtaaatggagtccggaactagctgc
Cercocarpus       gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgaactagctgc
Cowania           gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgaactagctgc
Dryas             gtgatcttgctcgtgagggtaatgaaattattcgtgaggctactaaatggagtcctgaactagctgc
Neillia           gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgaactagctgc
Purshia           gtgatcttgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgaactagctgc
Rosa              gtgatctcgctcgtgagggtaatgaaattattcgtgaggctagtaaatggagtcctgaactagctgc
Ceanothus         gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagtaaatggagtcctgaactagctgc
Colletia          gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagcNNNtggagtgctgaactagctgc
Discaria          gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagtaaatggagtgctgaactagctgc
Elaeagnus         gtgatcttgctcgtgagggtaatgYaattattcgtgaggcttgtaaatggagtcctgaactagctgc
Hippophaë         gtgatcttgctcgtgagggtaatgacattattcgtgaggcttgtaaatggagtccggaactagctgc
Parasponia        gtgatcttgctcgagaaggtaatgaaattattcgcgaggctagtaaatggagtcctgaactagctgc
Rhamnus           gtgatcttgctcgtgaaggtaatgaaattattcgtgaggctagtaaatggagtcctgaactagctgc
Shepherdia        gtgatcttgctcgtgagggtaatgaaattattcgtgaggcttgtaaatggagtcctgaactagctgc
Trevoa            gtgatcttgctcgtgagggtaatgaaattatccgtgaggctagtaaatggagtgctgaactagctgc
Moringa           gtgatcttgctcgcgagggtaatgaaattatacgtaaggctagcaaatggagtcctgaactagctgc
Platanus          gtgatcttgctcgtgaaggtaatgaaattatccgtgaggctagcaaatggagtcctgaactagctgc
Vitis             gtgatcttgctcgtgagggtaatgaaattatccgtgcagctagcaaatggagtcctgaactagctgc
Mask              1111111111111111111111111111111111111111111111110001111111111111111111
```

*rbc*L data matrix (continued)

```
                  1111111111111111111111111111111111111111111111111111111111111111111
                  3333333333333333333333333333333333333333333333333333333333333333344
                  4444444445555555555666666666677777777778888888888999999999900
                  1234567890123456789012345678901234567890123456789012345678901
-------------------------------------------------------------------------------
Alnus             cgcttgtgaagtatggaaggagatcaaatttgaattc---ccagcaatggatactttgtaa
Allocasuarina     tgcttgtgaagtatggaaggcaatcaaatttgaattc---cccgcaatggatactttgtaa
Betula            cgcttgtgaagtatggaaggagatcaaatttgaattc---ccagcaatggatactttgtaa
Castania          tgcttgtgaagtatggaaagagatcaaatttgaattc---ccagcaatggatactttgtaa
Casuarina         tgcttgtgaagtatggaaggaaatcaaatttgaattc---ccggcaatggatactttgtaa
Comptonia         tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Corylus           cgcttgtgaagtatggaaggagatcaaatttgaattc---ccagcaatggatactttgtaa
Gymnostoma        tgcttgtgaagtatggaaggaaatcaaatttgaattc---ccggcaatggatactttgtaa
Juglans           tgcttgtgaagtatggaaggagatcaaatttgaattc---ccagcaatggatactttgtaa
Myrica            tgcttgtgaagtatggaaggagatcaaatttgaattc---ccagcaatggatactttgtaa
Nothofagus        tgcttgtgaagtatggaaggagatcaaatttgaattcNNNgaagcaatggatactttgtaa
Pilea             cgcttgtgaagtatggaaggagatcaaatttgaattt---ccagcaatggatactttgtaa
Begonia           tgcttgtgaagtatggaaggagatcaaatttgaattt---ccggcaatggatactttgtaa
Coriaria          tgcttgtgaagtatggaaagaaatcaaatttgaattt---gaagcaatggatactttgtaa
Cucurbita         tgcttgtgaagtatggaaggcgatcaaattcgaattt---gaggcagtagatactttgtaa
Datisca           tgcttgtgaagtatggaaggagatcaaatttgaattt---gaagcaatggatactttgtaa
Tetrameles        tgcttgtgaagtatggaaggaaatcaaatttgaattt---ccagcaatggatactttgtaa
Cercocarpus       tgcttgtgaagtctggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
```

```
Cowania      tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Dryas        tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Neillia      tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Purshia      tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Rosa         cgcttgtgaggtatggaaagagatcaaatttgaattc---gaagcaatggatactttgtaa
Ceanothus    tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Colletia     tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Discaria     tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Elaeagnus    tgcttgtgaagtatggaaggagatcaaatttgaattt---gaagctatggatacaatttga
Hippophaë    tgcttgtgaagtatggaaggagatcaaatttgaattt---caagctatggatactttgtaa
Parasponia   tgcttgtgaagtgtggaaggaaatcaaatttgaattc---ccagcaatggatactttgtaa
Rhamnus      tgcttgtgaagtatggaaggagattaaatttgaattc---ccagcaatggatactttgtaa
Shepherdi    tgcttgtgaagtatggaaggagatcaaatttgaattt---gaagctatggatacaatttga
Trevoa       tgcttgtgaagtatggaaggagatcaaatttgaattc---gaagcaatggatactttgtaa
Moringa      tgcttgtgaagtatggaaggagatcaaatttgaattc---ccagcagtggatactttgtaa
Platanus     tgcttgtgaggtatggaaggagatcaaatttgaattt---gaagcaatggatactttataa
Vitis        tgcttgtgaagtatggaaggaaatcaaatttgaattc---ccagcaatggatactttgtaa
Mask         11111111111111111111111111111111111111110001111111111111111111111
```

Appendix 2. The data matrix of 13 *ndh*F nucleotide sequences used for parsimony analysis. Species abbreviation: ccun (*C. cuneatus*), cpu (*C. pumilus*), cpro (*C. prostratus*), cam (*C. americanus*), csan (*C. sanguineus*), cin (*C. integerrimus*), cth (*C. thyrsiflorus*), cve (*C. velutinus*), cfen (*C. fendleri*), cco (*C. cordulatus*), rham (*R. purshiana*), baca (*B. caryophylla*), and ntab (*N. tabacum*)

*ndh*F data matrix

```
                    1111111111222222222233333333334444444444555555555566666666
           1234567890123456789012345678901234567890123456789012345678901234567
-----------------------------------------------------------------------------
ccun       ATGGATCATACCTTTCATTCCCCTTCCAATTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cpu        ATGGATCATACCTTTCATTCCCCTTCCAATTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cpro       ATGGATCATACCTTTCATTCCCCTTCCAATTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cam        ATGGATCATACCTTTCCTTCCCCTTCCAACTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
csan       ATGGATCATACCTTTCATTCCCCTTCCAACTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cin        ATGGATCATACCTTTCATTCCCCTTCCAACTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cth        ATGGTTCATACCTTTCATTCCCCTTCCAACTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cve        ATGGATCATACCTTTCATTCCCCTTCCAACTCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cfen       ATGGATCATACCTTTCATTCCCCTTCCAACCCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
cco        ATGGATCATACCTTTCATTCCCCTTCCAACCCCTATGTTAATAGGAGTAGGACTTCTCCTTTTTCCG
rham       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca       GTGGATAATACCTTTCCTTCCACTTCCAGTTCCTATGTTAATAGGGTTGGGACTTCTTCTTTTTCCG
ntab       GTGGATCATACCTTTCATTCCACTTCCAGTCCCTATGTTAATAGGAGCGGGACTTTTTCTTTTTCCA
```

*ndh*F data matrix (continued)

```
                                                  1111111111111111111111111111111111
           6677777777778888888888999999999900000000001111111111222222222233333
           8901234567890123456789012345678901234567890123456789012345678901234
-----------------------------------------------------------------------------
ccun       ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCGAGTGTTTTACTGTTAAGTATAGTTATGG
cpu        ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCGAGTGTTTTACTGTTAAGTATAGTTATGG
cpro       ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCGAGTGTTTTACTGTTAAGTATAGTTATGG
cam        ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTGTTAAGTATAGTTATGG
csan       ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTTTTAAGTATAGTTATGG
cin        ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTGTTAAGTATAGTTATGG
cth        ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTGTTAAGTATAGTTATGG
cve        ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTGTTAAGTATAGTTATGG
cfen       ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTGTTAAGTATAGTTATGG
cco        ACGGCAATAAAAAATCTTCGCCGTATGTGGGTTTTTCCTAGTGTTTTACTGTTAAGTATAGTTATGG
rham       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca       ACGGCAACAAAAAGTCTTCGTCGTATGTGGGCTTTTCAAAGCGTTTTATTGTTAAGTATAGTCATGA
ntab       ACGGCAACAAAAAGTTTTCGCCGTATGTGGGCTTTTCAGAGTGTTTTATTGTTAAGCATAGTCATGG
```

*ndh*F data matrix (continued)

```
           1111111111111111111111111111111111111111111111111111111111111111122
           3333444444444455555555556666666666777777777788888888899999999900
           5678901234567890123456789012345678901234567890123456789012345678901
-----------------------------------------------------------------------------
ccun       TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cpu        TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cpro       TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cam        TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTTTGGTCTTG
csan       TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cin        TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cth        TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cve        TTTTTTCAGCCCATATATCTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cfen       TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
cco        TTTTTTCAGCCCATATATTTATTCAACAAATAAATAACAGTTCTATCTATCTATCTGTATGGTCTTG
rham       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca       TTTTTTCGATGAATCTGTCTATTCAGCAAATAAATAGCAGTTCTGTCTATCAATATGTATGGTCTTG
ntab       TTTTTTCAATCTACCTGTCTATTCAGCAAATAAATAGCAGTTCTTTTTATCAATATGTATGGTCTTG
```

*ndh*F data matrix (continued)

```
           2222222222222222222222222222222222222222222222222222222222222222222
           0000000011111111112222222222333333333344444444445555555555666666666
```

```
           234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------
ccun       GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cpu        GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cpro       GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cam        GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
csan       GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cin        GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cth        GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cve        GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cfen       GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
cco        GACCATCAATAATGATTTTTCTTTAGAATTCGGCTACTTAATTGATCCACTTACTTCTCTTATGTTA
rham       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca       GATTATCAATAATGATTTTTCTTTAGAATTCGGATACTTGATCGATCCACTTACTTCTATTATGTCA
ntab       GATCATCAATAATGATTTTTCTTTAGACTTCGGATACTTGATCGACCCACTTACTTCTATTATGTCA
```

ndhF data matrix (continued)

```
           222222222222222222222222222223333333333333333333333333333333333
           677777777778888888888999999999000000000011111111112222222222333333
           901234567890123456789012345678901234567890123456789012345678901234
-------------------------------------------------------------------------------
ccun       ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCATGATCAGG
cpu        ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCATGATCAGG
cpro       ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCATGATCAGG
cam        ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCATGATCAGG
csan       ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCATGATCAGG
cin        ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCACGATCAGG
cth        ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCACGATCAGG
cve        ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCACGATCAGG
cfen       ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCACGATCAGG
cco        ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCACGATCAGG
rham       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca       ATATTAATCACTACTGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGTCTCATGATCAGG
ntab       ATATTAATCACTACGGTTGGAATTATGGTTCTTATTTATAGTGATAATTATATGGCTCATGATCAAG
```

ndhF data matrix (continued)

```
           333333333333333333333333333333333333333333333333333333333333333444
           333344444444445555555555666666666677777777778888888888999999999000
           678901234567890123456789012345678901234567890123456789012345678901
-------------------------------------------------------------------------------
ccun       GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cpu        GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cpro       GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cam        GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
csan       GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cin        GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cth        GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cve        GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cfen       GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
cco        GATATTTGAGATTTTTTGCTTATATGAGTTTTTCTAATACTTCAATGTTAGGATTAGTTACTAGTTC
rham       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca       GATATTTGAGATTTTTTGCTTATATGAGTTTTTTCAGTACTTCCATGTTGGGATTAGTTACTAGTTC
ntab       GCTATTTGAGATTTTTTGCTTATATGAGTTTTTTCAGTACTTCCATGTTGGGATTAGTTACTAGTTC
```

ndhF data matrix (continued)

```
           444444444444444444444444444444444444444444444444444444444444444444
           000000011111111112222222222333333333344444444445555555556666666666
           345678901234567890123456789012345678901234567890123456789012345678
-------------------------------------------------------------------------------
ccun       AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cpu        AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cpro       AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cam        AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
csan       AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cin        AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cth        AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cve        AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
cfen       AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
```

```
cco         AAATTTGATACAAATTTATTTTTTTTGGGAATTAGTTGGAATGTGTTCTTATCTATTAATAGGTTTT
rham        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca        AAATTTGATACAAATTTATATTTTTTGGGAATTAGTTGGAATGTGTTCGTATCTATTAATAGGATTT
ntab        GAATTTGATACAAATTTATATTTTTTGGGAATTGGTTGGGCTGTGTTCCTATCTATTAATAGGATTT
```

ndhF data matrix (continued)

```
            4444444444444444444444444444455555555555555555555555555555555555
            7777777777888888888899999999900000000001111111111222222222233333333
            0123456789012345678901234567890123456789012345678901234567890123456
------------------------------------------------------------------------------
ccun        TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cpu         TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cpro        TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cam         TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
csan        TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cin         TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cth         TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cve         TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cfen        TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
cco         TGGTTCACCCGACCTATTGCGGCGAATGCTTGTCAAAAGGCGTTTGTAACTAATCGTGTAGGGGATT
rham        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
baca        TGGTTCACACGACCTGTTGCAGCAAAGGCTTGTCAAAAAGCGTTTGTAACTAATCGGATTGGGGATT
ntab        TGGTTTACACGACCTGTTGCGGCAAATGCTTGTCAAAAAGCGTTTGTAACTAATCGTGTAGGGGATT
```

ndhF data matrix (continued)

```
            5555555555555555555555555555555555555555555555555555555555556666
            3334444444444555555555566666666667777777777888888888899999999990000
            7890123456789012345678901234567890123456789012345678901234567890123
------------------------------------------------------------------------------
ccun        TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cpu         TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cpro        TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cam         TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
csan        TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cin         TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cth         TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cve         TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cfen        TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
cco         TTGGGTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTAGGGATTTGTT
rham        TTGGTTTATTATTAGGAATTTTAGGTTTTTATTGGATAACGGGTAGTTTAGAATTTCGGGATTTGTT
baca        TTGGTTTATTATTAGGAATTTTAGGGTTTTATTGGATAACGGGGAGTTTCGAATTTCGTGATTTATT
ntab        TTGGTTTATTATTAGGAATTTTAGGTTTTTATTGGATAACAGGGAGTTTCGAATTTAGGGATTTATT
```

ndhF data matrix (continued)

```
            6666666666666666666666666666666666666666666666666666666666666666666
            0000001111111111222222222233333333334444444444555555555566666666667
            4567890123456789012345678901234567890123456789012345678901234567890
------------------------------------------------------------------------------
ccun        TGAAATATTCAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cpu         TGAAATATTCAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cpro        TGAAATATTCAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cam         TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
csan        TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cin         TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cth         TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cve         TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cfen        TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
cco         TGAAATAGTAAATAACGTGGTTGATAATAATGAAGTTGATTTTTTATTTGTTACTTTGTGTGCCTGT
rham        TGAAATATTCAATAATGTGGTTTATAATAATGAAGTTAATTTTTTATTTGTTACTTTGTGTGCCTTT
baca        CCAAATATTCAATAACTTGATTTCTAATAATGAGGTCAATTTTGTATTTGTTACTTTGTGCGCAGTT
ntab        CGAAATATTCAATAACTTGATTTATAATAATGAAGTCGATTTTTTATTTGTTACTTTATGTGCTGTT
```

ndhF data matrix (continued)

```
            6666666666666666666666666666677777777777777777777777777777777777777
            7777777778888888888999999999900000000001111111111222222222233333333
            1234567890123456789012345678901234567890123456789012345678901234567
------------------------------------------------------------------------------
```

```
ccun        CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cpu         CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cpro        CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cam         CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
csan        CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cin         CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cth         CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cve         CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cfen        CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
cco         CTATTATTTGCCGGTGCAGTTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
rham        CTATTATTTGCCGGTGCAATTGCTAAATCTGCCCAATTTCCCCTTCATGTATGGTTACCCGATGCTA
baca        CTATTATTTGCCGGTGCGATTGCTAAATCTGCACAATTTCCCCTTCATGTATGGTTACCTGATGCAA
ntab        CTCTTATTTGCCGGTGCAGTTGCTAAATCCGCCCAATTCCCCCTTCATGTATGGTTACCTGATGCCA
```

ndhF data matrix (continued)

```
            7777777777777777777777777777777777777777777777777777777777777788888
            3344444444445555555555666666666677777777778888888888999999999900000
            8901234567890123456789012345678901234567890123456789012345678901234
------------------------------------------------------------------------------
ccun        TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cpu         TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cpro        TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cam         TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
csan        TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cin         TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cth         TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cve         TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cfen        TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
cco         TGGAAGGACCTACCCCTATTTCGGCTCTTATACATGCTGCTACTATGGTAGCAGCCGGAATTTTTCT
rham        TGGAAGGACCTACCCCTATTTCGGCTCTTATACACGCTGCTACTATGGTAGCGGCGGGAATTTTTCT
baca        TGGAAGGGCCGACTCCTATTTCGGCCCTTATACATGCTGCTACGATGGTAGCAGCGGGAATTTTTCT
ntab        TGGAGGGGCCTACTCCCATTTCGGCTCTTATACATGCTGCTACTATGGTAGCGGCGGGAATTTTTCT
```

ndhF data matrix (continued)

```
            8888888888888888888888888888888888888888888888888888888888888888888
            0000011111111112222222222333333333344444444445555555555666666666677
            5678901234567890123456789012345678901234567890123456789012345678901
------------------------------------------------------------------------------
ccun        TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTCTTGGG
cpu         TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTCTTGGG
cpro        TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTCTTGGG
cam         TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
csan        TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
cin         TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
cth         TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
cve         TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
cfen        TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
cco         TGTAGCTCGGCTTCTTCCGCTTTTTATAGTTATACCTTTCATAATGAATCTAATAGCTTTTATTGGG
rham        TGTAGCTCGGCTTCTTCCGCTTTTCATAGTTATACCTTTCATAATGAATATAATAGCTTTGATTGGT
baca        TGTAGCTCGACTTCTGCCTCTTTTCATAGTCATACCTCACATAATGAATTTTATCTCTTTGATAGGG
ntab        TGTAGCTCGGCTTCTTCCTCTTTTCAGAGTTATACCTTACATAATGTATTTGATCTCGGTTATAGGA
```

ndhF data matrix (continued)

```
            8888888888888888888888888888999999999999999999999999999999999999999
            7777777788888888889999999990000000000111111111122222222233333333 3
            2345678901234567890123456789012345678901234567890123456789012345678
------------------------------------------------------------------------------
ccun        ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cpu         ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cpro        ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cam         ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAAGGGTTTAGCTT
csan        ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAAGGGTTTAGCTT
cin         ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cth         ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cve         ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cfen        ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
cco         ATAATAACATTATTTTTAGGAGCCACTTTAGCTCTTGCTCAAAAAGATATTAAGAGGGGTTTAGCTT
rham        ATAATAACATTATTTTTAGGAGCCACTTTTGCTCTTGGTCAAAAAGATATTAAGAGAGGTCTAGCTT
baca        ATAATAACAGTATTTTTAGGAGCTACTTTAGCTCTTGCTCAAAAAGACATTAAGAGGGGTTTAGCCT
```

```
ntab       ATAATAACAGTATTATTAGGAGCTACTTTAGCTCTTGCTCAAAAAGACATTAAGAGAGGTTTAGCCT
```

*ndh*F data matrix (continued)

```
                                                                        111111
           9999999999999999999999999999999999999999999999999999999999999000000
           3444444444455555555556666666666777777777788888888889999999999000000
           9012345678901234567890123456789012345678901234567890123456789012345
-----------------------------------------------------------------------------
ccun       ATTCGACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cpu        ATTCGACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cpro       ATTCGACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cam        ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
csan       ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cin        ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cth        ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cve        ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cfen       ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
cco        ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
rham       ATTCTACAATGTCTCAATTGGGTTATATGATGTTGGCTCTAGGTATGGGGTCTTATCGAGCTGCTTT
baca       ATTCCACAATGTCTCAATTGGGTTATATGATGTTAGCTCTAGGTATGGGGTCTTATCGCAGTGCTTT
ntab       ATTCCACAATGTCTCAATTGGGTTATATGATGTTAGCTCTTGGTATGGGGTCTTATCGAAGCGCTTT
```

*ndh*F data matrix (continued)

```
           1111111111111111111111111111111111111111111111111111111111111111111
           0000000000000000000000000000000000000000000000000000000000000000000
           0000111111111122222222223333333333444444444455555555556666666666777
           6789012345678901234567890123456789012345678901234567890123456789012
-----------------------------------------------------------------------------
ccun       ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCAATTATTCAT
cpu        ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCAATTATTCAT
pro       ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCAATTATTCAT
cam        ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
csan       ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
cin        ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
cth        ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
cve        ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
cfen       ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
cco        ATTTCATTTGATTACTCACGCTTATTCGAAAGCATTGTTGTTTTTAGGATCCGGATCCATTATTCAT
rham       ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTGTTATTTTTAGGGTCCGGATCAATTATTCAT
baca       ATTTCATTTGATTACTCATGCTTATTCGAAAGCATTATTGTTTTTAGGATCGGGATCCGTTATTCAT
ntab       ATTTCATTTGATTACTCATGCTTATTCCAAAGCATTATTATTTTTAGGATCCGGATCCATTATTCAT
```

*ndh*F data matrix (continued)

```
           1111111111111111111111111111111111111111111111111111111111111111111
           0000000000000000000000000001111111111111111111111111111111111111111
           7777777888888888899999999990000000000111111111122222222223333333333
           3456789012345678901234567890123456789012345678901234567890123456789
-----------------------------------------------------------------------------
ccun       TCCATGGAAACAATTGTTGGATATTCTCCGGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cpu        TCCATGGAAACAATTGTTGGATATTCTCCGGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cpro       TCCATGGAAACAATTGTTGGATATTCTCCGGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cam        TCCATGGAAACAATTGTTGGATATTCTCCAGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
csan       TCCATGGAAACAATTGTTGGATATTCTCCAGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cin        TCCATGGAAACAATTGTTGGATATTCTCCAGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cth        TCCATGGAAACAATTGTTGGATATTCTCCAGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cve        TCCATGGAAACAATTGTTGGATATTCTCCAGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cfen       TCCATGGAAACAATTGTTGGATATTCTCCAGATAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
cco        TCCATGGAAACAATTGTTGGATATTCTCCAGCTAAAAGCCAGAATATGGTTCTTATGGGCGGTTTAA
rham       TCAATGGAAACTATTGTTGGATATTCTCCAAATAAAAGTCAGAATATGGTTCTTATGGGCGGTTTAA
baca       TCAATGGAAACTATTGTTGGATATTGTCCAAAAAAAAGTCAGAATATGGTGCTTATGGGGGGTTTAA
ntab       TCAATGGAAACTATTGTTGGATATTCTCCAGCTAAAAGTCAGAATATGGGTCTTATGGGAGGTTTAA
```

*ndh*F data matrix (continued)

```
           1111111111111111111111111111111111111111111111111111111111111111111
           1111111111111111111111111111111111111111111111111111111111112222222
           4444444444555555555566666666667777777777888888888899999999990000000
           0123456789012345678901234567890123456789012345678901234567890123456
```

```
                                                                                
ccun            GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cpu             GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cpro            GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cam             GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
csan            GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cin             GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cth             GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cve             GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cfen            GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
cco             GAAAACATGTACCAATTACAAAAACCGCGTTTTTATTAGGTACACTTTCTCTGTGTGGCATTCCACC
rham            GAAAACATGTACCAATTACAAAAACTGCGTTTTTATTAGGTACACTTTCTCTTTGTGGTATTCCACC
baca            CAAAACATGTACCAGTTACTAAAACTGCTTTTTTATTAGGTACACTTTCTCTTTGCGGTATTCCACC
ntab            GAAAACATGTACCAATTAGCAAAATCACATTTTTATTAGGTACACTTTCTCTTTGTGGTATTCCACC
```

*ndh*F data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111111111
                222222222222222222222222222222222222222222222222222222222222222222222222
                000111111111122222222233333333334444444444555555555566666666667777
                789012345678901234567890123456789012345678901234567890123456789012345678
----------------------------------------------------------------------------------
ccun            CCTTGCCTGTTTTTGGTCCAAAGATGCAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cpu             CCTTGCCTGTTTTTGGTCCAAAGATGCAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cpro            CCTTGCCTGTTTTTGGTCCAAAGATGCAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cam             CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
csan            CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cin             CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cth             CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cve             CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cfen            CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
cco             CCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAGTGATAGTTGGTTATATTCACCGATTTTTGCA
rham            TCTTGCCTGTTTTTGGTCCAAAGATGAAATTCTTAATGATAGTTGGTTATATTCACCTATTTTTGCA
baca            CCTTGCTTGTTTTTGGTCCAAAGATGAAATTCTTAATGATAGTTGGTTGTATTCACCTATTTTTGCA
ntab            TCTTGCTTGTTTTTGGTCCAAAGATGAAATTCTTAATGATAGTTGGTTGTATTCGCCAATTTTCGCA
```

*ndh*F data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111111111
                222222222222222222222222223333333333333333333333333333333333333333333333
                777777888888888899999999990000000000111111111122222222223333333333 4
                456789012345678901234567890123456789012345678901234567890123456789012345678890
----------------------------------------------------------------------------------
ccun            ATAATAGCTTTTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cpu             ATAATAGCTTTTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cpro            ATAATAGCTTTTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cam             ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
csan            ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cin             ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cth             ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cve             ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cfen            ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
cco             ATAATAGCTTGTTCCACAGCTGGATTAACTGCATTTTATATGTTTAGGATCTATTTACTTACTTTTG
rham            ATAATCGCTTGTTACACAGCAGGATTAACTGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
baca            ATAATAGCTTGGTCTACGGCGGGATTAACCGCATTTTATATGTGTCGGATCTATTTACTTACTTTTG
ntab            ATAATAGCTTGGGCCACGGCGGGATTAACCGCATTTTATATGTTTCGGATCTATTTACTTACTTTTG
```

*ndh*F data matrix (continued)

```
                111111111111111111111111111111111111111111111111111111111111111111111111
                333333333333333333333333333333333333333333333333333333333333333333333333
                444444444555555555566666666667777777777888888888999999999900000000
                123456789012345678901234567890123456789012345678901234567890123456789012345678
----------------------------------------------------------------------------------
ccun            AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cpu             AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cpro            AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cam             AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
san            AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cin             AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cth             AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
```

```
cve         AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cfen        AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
cco         AAGGACATTTAAACGTTCAGTTTCAAATTTACAGTGGAAAAAAAAATAGCTCGTTCTATTCAATATC
rham        AAGGACACTTAAATGTTCATTTTCAAACTTACAGCGGCAAAAAAAAAAGCCCGTTCTATTCAATATC
baca        AAGGACATTTAAACGTTCTTTTTCAAAATTACAGTGGAAAAAAGAATACCCCTTTCTATTCAATATC
ntab        AAGGGCATTTAAACGCTCATTTTCCAAATTATGGTGGGAAACAAAAAACCCCCTTCTATTCAATATC
```

ndhF data matrix (continued)

```
            11111111111111111111111111111111111111111111111111111111111111111111
            44444444444444444444444444444444444444444444444444444444444444444444
            00111111111122222222233333333334444444444555555555566666666666777777
            89012345678901234567890123456789012345678901234567890123456789012345
------------------------------------------------------------------------------
ccun        TCTATGGGGTAAAAAAGGGCCAAAGGTTATTAAAAAAAATTTCCGTTTATTAACTTTATTAAAAATG
cpu         TCTATGGGGTAAAAAAGGGCCAAAGGTTATTAAAAAAAATTTCCGTTTATTAACTTTATTAAAAATG
cpro        TCTATGGGGTAAAAAAGGGCCAAAGGTTATTAAAAAAAATTTCCGTTTATTAACTTTATTAAAAATG
cam         TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
csan        TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
cin         TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
cth         TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
cve         TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
cfen        TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
cco         TCTATGGGGTAAAAAAGGGCAAAAGGTTATTAAAACAAATTTTCGTTTATTAACTTTATTAAAAATG
rham        TCTATGGGGTAAAAAAGGACCAAAAGTTATTAAAAAAAAAAAATAGCTTGTGGGCTTTATTAACAATG
baca        TCTATGGGGTAAACGGGGGTTTAAAATAAGTAACAAAAACTTTCGTTTGGTAACTTTATTAAAAATG
ntab        TCTATGGGGTAAAAACGG..........AGTTAAGAAAAACTCTTG........TTTATTAACTATG
```

ndhF data matrix (continued)

```
            11111111111111111111111111111111111111111111111111111111111111111111
            44444444444444444444444445555555555555555555555555555555555555555555
            77777888888888899999999990000000000111111111122222222223333333333344
            56789012345678901234567890123456789012345678901234567890123456789012
------------------------------------------------------------------------------
ccun        AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cpu         AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cpro        AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cam         AATAATACTGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
csan        AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cin         AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cth         AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cve         AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cfen        AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
cco         AATAATAATGAAAGGACTTCTTTTTTTTCTAAAAAGGCATATCGA.....................
rham        AATAATATTCAAGGGATTTTTTTTTTTTGTAAAAAGACATATCGAATGGATAGTAATGATAGTAATG
baca        AATAATAATGAACGTGCTTCTTTTTTTTCAAATAAAGTATATAAA.....................
ntab        AATAATAATGAAAGTACTTATTTTTTTGCAAAAACTAAATATCCA.....................
```

ndhF data matrix (continued)

```
            11111111111111111111111111111111111111111111111111111111111111111111
            55555555555555555555555555555555555555555555555555555555555555555555
            44444444555555555566666666667777777777888888888999999999900000000
            23456789012345678901234567890123456789012345678901234567890123456789012345678
------------------------------------------------------------------------------
ccun        .................ATTGATAGTAATCTAAGAACCATGATAAGCCCTTTTATTACTATTAGTCA
cpu         .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cpro        .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cam         .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
csan        .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cin         .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cth         .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cve         .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cfen        .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
cco         .................ATTGATAGTAATCTAAGAAGCATGATAAGCCCTTTTATTACTATTAGTCA
rham        TAAGTCGAATTGATAGTATTCATAGTAATGTAAGAAGCATGGCGCAGCCTTTTAGTACTATTATTCG
baca        .................ATTGATGAGAATGTAAGAAATATGATCCAACCCTTTCTTTCTATTCCGCA
ntab        .................ATTGATAAAAATGGAAGAAAGATGACACGACCTTTTATGACTATTGCTCA
```

ndhF data matrix (continued)

```
            1111111111111111111111111111111111111111111111111111111111111111111111111
            6666666666666666666666666666666666666666666666666666666666666666666666666
            0111111111122222222223333333333444444444455555555556666666666777777
            9012345678901234567890123456789012345678901234567890123456789012345
-------------------------------------------------------------------------------
ccun        TTTTGGAACTAAGAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cpu         TTTTGGAACTAAGAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cpro        TTTTGGAACTAAGAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cam         TTTTGGAACTAAAAAAACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
csan        TTTTGGAACTAAAAAAACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cin         TTTTGGAACTAAAAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cth         TTTTGGAACTAAAAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cve         TTTTGGAACTAAAAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cfen        TTTTGGAACTAAAAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
cco         TTTTGGAACTAAAAACACTTTCTCATATCCGCAGGAATCGGACAATACTATGCTATTTCCTATGCTT
rham        TTTTGGTACTAAGAACACTTTCTCATATCCTCACGAGTCGGACAATACTATGCTATTTCCTATGCTT
baca        TTTTGGCAATACCAAGACTTATTCGTATCCTTATGAATCGGATAATACTATGTTATTCCCAATACTT
ntab        TTTTGAGCATAAAGCGGTTTATTCCTATCCTTATGAATCGGACAATACTATGCTATTCCCAATATTT
```

ndhF data matrix (continued)

```
            1111111111111111111111111111111111111111111111111111111111111111111111111
            6666666666666666666666667777777777777777777777777777777777777777777777777
            7777888888888899999999990000000000111111111122222222223333333333444
            6789012345678901234567890123456789012345678901234567890123456789012
-------------------------------------------------------------------------------
ccun        GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cpu         GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cpro        GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cam         GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
csan        GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cin         GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cth         GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cve         GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cfen        GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
cco         GTATTAGTACTATTTACTTTGTTCATTGGATTCATAGGAATTCCTTTCTTCAATCAATTCAATCAAG
rham        GTATTAGTTCTATTTACTTTGTTAATTGGAGTAATAGGAATTCCGCTC.........TTCAATCAAG
baca        ATATTGGTTCTATTTACTTTGTTCGTTGGATTCTTAGGAATTCCT.............TTGGATCAAG
ntab        GTATTAGGACTCTTTACTTTGTTTGTTGGATCTATAGGAATTCCT.............TTCAACCAAG
```

ndhF data matrix (continued)

```
            1111111111111111111111111111111111111111111111111111111111111111111111111
            7777777777777777777777777777777777777777777777777777777777777778888888888
            4444444555555555566666666667777777778888888888999999999990000000000
            3456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------
ccun        AAGCGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
cpu         AAGCGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
cpro        AAGCGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
cam         AAGCGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
csan        AAGCGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
cin         AAACGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCA
cth         AAACGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCA
cve         AAACGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCA
cfen        AAACGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
cco         AAACGATGGATTTGGATATATTATCAAAATTGTTAAACCCGGCTATAAACCTTTTACATCCAAATCC
rham        AAGGGGGGGATTCGGATATATTATCAAAATTGTTAACCCCGGCTATAAACCTTTTACATCAAAATCA
baca        ACGGGGTTGATTTGGATATATTATCCAAATGGTTAACCCCGTCTATAAATCTTTTACATAAAAATTC
ntab        AGGGAGGGAATTTGGATATATTATCGAAATGGTTAGCTCCATCTATAAATCTTTTGCATCAAAAGTC
```

ndhF data matrix (continued)

```
            1111111111111111111111111111111111111111111111111111111111111111111111111
            8888888888888888888888888888888888888888888888888888888888888888888888888
            1111111111222222222233333333334444444444555555555566666666667777777
            0123456789012345678901234567890123456789012345678901234567890123456
-------------------------------------------------------------------------------
ccun        AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTT
cpu         AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTT
cpro        AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTT
```

```
cam         AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
csan        AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
cin         AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
cth         AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
cve         AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
cfen        AAATACTTCTGTGGATTTGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
cco         AAATACTTCTGTGGATTGGGATGAATTTATTTCAAACGCAACTTTTTCAGTCAGTATAGCTTTTTTTT
rham        AGAAAATTCTGTGGATTGGTATGAATTTATGACAAACGCAACCTTTTCAGTCAGTATCGCTTTTTTC
baca        AAACTATTCAATAGATTGGTATGAATTTTGTAAAGATGCAGCTTTTTCAGTTAGTATAGCCTCTTTC
ntab        GAATAATTCGATGGATTGGAATGAATTTTTAAAGGATGCAGTTCTTTCAGTCAGTATAGCTTATTTC
```

ndhF data matrix (continued)

```
            1111111111111111111111111111111111111111111111111111111111111111111
            8888888888888888888888899999999999999999999999999999999999999999999
            7778888888888999999999900000000001111111111222222222233333333334444
            7890123456789012345678901234567890123456789012345678901234567890123
------------------------------------------------------------------------------
ccun        GGAATCCTTATAGCGGCCTTTTCATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cpu         GGAATCCTTATAGCGGCCTTTTCATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cpro        GGAATCCTTATAGCGGCCTTTTCATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cam         GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
csan        GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cin         GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cth         GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cve         GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cfen        GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
cco         GGAATCCTTATAGCGGCCTTTTTATATAAGCCGGTTTATTCATCTTTACAAAATTTGAATTTCTTTA
rham        GGAATCCTTATAGCGTCTTTTTTATATAAGCCGGCTTATTCATCTTTACAAAATTTGAATTTCTTTA
baca        GGAATATTTATAGCATTTTTTTTTATATAAACCTGTTTATTCATCTTTTCAAAATTTGGACTTAGTTA
ntab        GGAATATTTATAGCATCCTTTTTTATATAAACCCATTTATTCTTCTTTAAAAAATTTTGAGTTAATTA
```

ndhF data matrix (continued)

```
            1111111111111111111111111111111111111111111111111111111111122222222222
            9999999999999999999999999999999999999999999999999999999999900000000000
            4444445555555555666666666677777777778888888888999999999900000000001
            4567890123456789012345678901234567890123456789012345678901234567890
------------------------------------------------------------------------------
ccun        ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cpu         ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cpro        ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAAATACTAAATGGGATATA
cam         ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
csan        ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cin         ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cth         ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cve         ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cfen        ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
cco         ATTCATTTGTTAAAAGTATTCCTAAAAAAAGAAAAATTCTTTTGGAAAAACTACTAAATGGGATATA
rham        ATTCAGTTGTTAAAAGGATTCCGAAAAAAACAAAGATTTTGGGGGATAAACTACTAAATGTTATATA
baca        ATTCTGTGGTTAAAATAGGTCCTACGAGA......ATTTTTTCTGACAAAATAAAAAATGGTGTATA
ntab        ATTCTTTTGTTAAAAAGGGTCCTAAGAGA......ATTCTGTGGGACAAAATTATAAATGGCATATA
```

ndhF data matrix (continued)

```
            2222222222222222222222222222222222222222222222222222222222222222222
            0000000000000000000000000000000000000000000000000000000000000000000
            1111111112222222222333333333344444444445555555555666666666677777777
            1234567890123456789012345678901234567890123456789012345678901234567
------------------------------------------------------------------------------
ccun        TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cpu         TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cpro        TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cam         TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
csan        TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cin         TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cth         TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cve         TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cfen        TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
cco         TGATTGGTCATATAATCGTGGTTATATAGATGTTTTTTATGCAATATCCTTAACTAAAGGTATAAGA
rham        TGATTGGTCATATAATCGCGGTTACATAGATCTTTTTTATTCAATATCCTTAACTCGGGGTATAAGA
```

```
baca        TGATTGGTCATATAATCGTGGTTACATAGATGCCTTTTATGGAACATTCTTANNNNNNNNNNNNNNN
ntab        TGATTGGTCATATAATCGTGCTTACATAGATGCTTTTTATACAAGATTCTTAGTTGGTGGGATAAGA
```

ndhF data matrix (continued)

```
            22222222222222222222222222222222222222222222222222222222222222222222
            00000000000000000000011111111111111111111111111111111111111111111111
            77888888888899999999990000000000111111111122222222223333333333444444
            89012345678901234567890123456789012345678901234567890123456789012345
------------------------------------------------------------------------------
ccun        GGATTGGCTCAACTAACTCACTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cpu         GGATTGGCTCAACTAACTCACTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cpro        AGATTGGCTCAACTAACTCACTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cam         GGATTGGCTCAACTAACTCATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
csan        GGATTGGCTCAACTAACTCATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cin         GGATTGGCTCAACTAACTCATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cth         GGATTGGCTCAACTAACTCATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cve         GGATTGGCTCAACTAACTCATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cfen        GGATTGGCTCAACTAACTTATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
cco         GGATTGGCTCAACTAACTCATTTTTTTGATAAACGAATAATTGATGGAATTACGAATGGTGTCGGTA
rham        GGTTTGGCTCAACTAACTCATTTTTTTGATAAACGAGTAATTGATGGAATTACGAATGGGGTCGGTA
baca        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
ntab        GGATTAGCTGAATTTACTCATTTTTTTGATCGACGAGTAATTGATGGGATGACTAACGGGGTTGGGG
```

ndhF data matrix (continued)

```
            22222222222222222
            11111111111111111
            44444555555555566
            56789012345678901
-----------------------------
ccun        TTGCGAGTTTTTTCTTG
cpu         TTGCGAGTTTTTTCTTG
cpro        TTGCGAGTTTTTTCTTG
cam         TTGCGAGTTTTTTCGTG
csan        TTGCGAGTTTTTTCGTG
cin         TTGCGAGTTTTTTCGTG
cth         TTGCGAGTTTTTTCGTG
cve         TTGCGAGTTTTTTCGTG
cfen        TTGCGAGTTTTTTCGTG
cco         TTGCGAGTTTTTTCGTG
rham        TTGTGAGTTTTTTCCTG
baca        NNNNNNNNNNNNNNNNN
ntab        TTATTAGTTTCATTGTA
```

Appendix 3. Binary data matrix of fingerprints generated by BOX and DR PCR of nodule DNA from *C. velutinus, C. sanguineus*, and *C. integerrimus* inoculated with soils from the *C. velutinus* and Douglas-fir stands. Ethidium bromide-stained agarose gel patterns were visually determined and converted into binary data; band presence (1), band absence (0), and ambiguous band (9). The fingerprinting patterns of nodules from *C. velutinus* inoculated with the soil from the *C. velutinus* stand correspond to nodule 1-13; those of *C. sanguineus*, nodule 27-37; those of *C. integerrimus*, nodule 49-63. The fingerprinting patterns of nodules from *C. velutinus* with the soil from the Douglas-fir stand correspond to nodule 14-26; those of *C. sanguineus*, 38-48; and those of *C. integerrimus*, 64-69.

Fingerprint data matrix

Band	Nodule																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2500b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2200b	9	9	9	1	1	1	1	1	1	1	1	1	1	9	9	9	1	1	1	1	1	1	1	1	9
1900b	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0
1800b	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1
1500b	1	1	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	1	0
1300b	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	1	1
1200b	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0
1100b	1	1	0	1	0	0	1	0	1	0	1	1	1	1	0	0	0	0	1	1	1	0	1	0	0
1050b	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	1
950b	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
900b	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
850b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
830b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
800b	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
770b	1	1	1	0	0	0	1	1	1	0	0	0	0	1	1	1	0	0	0	1	1	0	0	0	0
730b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
670b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
630b	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
600b	1	1	1	0	0	0	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	1	1	0	0
570b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
550b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
510b	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	0
490b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
470b	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1
390b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
370b	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1
350b	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	1	1	1	0	1	0
320b	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1
1400d	0	1	0	1	1	1	1	1	1	0	1	0	0	9	9	0	1	1	9	9	9	1	1	0	9
1200d	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	1
1100d	0	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
1000d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
950d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
850d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
770d	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
700d	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0
670d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
660d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
630d	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1
600d	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
570d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
550d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	1
510d	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
470d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
390d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
320d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
300d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
270d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Fingerprint data matrix (Continued)

Band	Nodule																									
	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
2500b	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
2200b	1	1	0	1	0	0	1	0	0	1	1	1	9	1	1	1	1	1	1	1	0	1	9	1	9
1900b	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1800b	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
1500b	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1300b	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0
1200b	1	1	0	0	0	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1	0	1	0	0	1
1100b	1	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1	0	0	0	1	1
1050b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
950b	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	1
900b	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
850b	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
830b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
800b	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
770b	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	0	1	1	1
730b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
670b	1	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0
630b	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	0	1	0	1
600b	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
570b	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
550b	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
510b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
490b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
470b	0	1	1	1	1	0	0	0	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1
390b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
370b	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
350b	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	0
320b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1400d	0	1	0	9	9	0	1	0	0	0	0	0	0	0	9	9	0	1	9	9	1	1	9	9	1
1200d	1	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	0
1100d	0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	0	0	1	1	0	0	1	1	1
1000d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0
950d	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
850d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
770d	1	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
700d	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0	0
670d	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1
660d	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
630d	1	1	0	0	0	1	0	1	1	0	1	1	1	1	0	1	0	0	1	1	0	0	0	0	1
600d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
570d	1	1	0	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
550d	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0
510d	0	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
470d	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
390d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
320d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
300d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
270d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Fingerprint data matrix (Continued)

Band	Nodule																		
	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
2500b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2200b	1	9	1	1	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1
1900b	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
1800b	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1500b	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1
1300b	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	1
1200b	1	0	1	1	1	1	1	0	0	1	0	1	1	1	1	1	0	1	1
1100b	0	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
1050b	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
950b	0	1	0	0	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0
900b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
850b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
830b	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
800b	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0
770b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
730b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
670b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
630b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
600b	1	1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	0
570b	0	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1
550b	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
510b	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
490b	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0
470b	1	1	0	0	0	1	0	1	1	0	0	0	0	1	1	1	1	0	0
390b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
370b	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0
350b	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1
320b	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1
1400d	1	1	1	1	1	1	1	1	1	1	1	1	1	9	1	9	9	1	0
1200d	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1	1
1100d	1	0	1	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0	0
1000d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
950d	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
850d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
770d	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	1	1	0
700d	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
670d	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
660d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
630d	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1	1
600d	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
570d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
550d	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1
510d	0	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	1
470d	1	1	9	1	1	9	1	1	1	0	0	0	0	1	1	1	0	0	0
390d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
320d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
300d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
270d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1