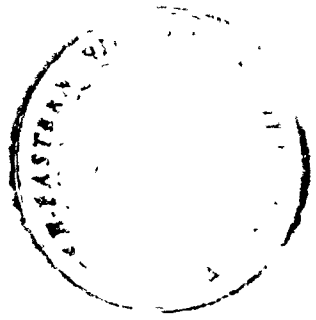
MOLECULAR TYPING OF *FRANKIA* IN THE FIELD COLLECTED ACTINORHIZAL ROOT NODULES FROM DIFFERENT ALTITUDES IN SIKKIM

By

ANINDITA KHAN



THESIS SUBMITTED
IN FULFILMENT OF THE DEGREE OF
DOCTOR OF PHILOSOPHY IN BOTANY

NORTH-EASTERN HILL UNIVERSITY
SHILLONG -793022, INDIA
2006

The North-Eastern Hill University

August, 2006

DECLARATION

I, Anindita Khan, hereby declare that the subject matter of this thesis entitled "Molecular typing of *Frankia* in the field collected Actinorhizal root nodules from different altitudes in Sikkim" is the record of work done by me, that the contents of this thesis did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other University/Institute.

This is being submitted to the North-Eastern Hill University for the award of the degree of Doctor of Philosophy in Botany.

(Anindita Khan)

(Head of the Department)

Head
Department of Botany
School of Life Sciences
N E H U, Shillong-22

(Supervisor)

ACKNOWLEDGEMENTS

The realization of a long cherished dream is no doubt wonderful, but it requires a lot of patience, hard work, honesty and help. When my doctoral work finally shapes into the form of a thesis, I would like to acknowledge those, who have made this difficult journey possible for me.

First and foremost, I express my sincere gratitude and indebtedness to my supervisor Prof. Arvind K. Misra, Department of Botany, North-Eastern Hill University, Shillong, for his enduring support, valuable advice, constant guidance, continuous inspiration and encouragement and most importantly the timely critical judgments to carry out this work.

My sincere thanks goes to Prof. N.K. Chrungoo (Head), Prof. A.K. Misra (former Head) and Prof. H.N. Pandey (former Head) for their kind permission to allow me to use all the laboratory facilities I required in the department during the course of my study.

I extend my heartiest thanks and gratefulness to all the faculty members of the Botany department, especially Prof. R.S. Tripathi (ex-faculty), Prof. H.N. Pandey, Prof. N.K. Chrungoo and Prof. S.K. Barik, for their timely help and useful suggestions.

I take this opportunity to express my deep sense of gratitude to Prof. D. Myrold, Department of Crop and Soil Science, Oregon State University, Corvallis, USA, for his much needed help and suggestions throughout the entire study period.

This acknowledgement would not be complete without thanking Dr. A.P. Krishna, scientist in-charge and Dr. K.K. Singh of GBPIHED, Sikkim Unit, Gangtok, for allowing me to avail their laboratory facilities to carry out ARA, Dr. Grace Dkhar, Research Officer, District and Local Research Station and Laboratories, Shillong for extending her kind help in carrying out soil analyses, Dr. Patiram, Head, Soil Science Division and Mr. Ritesh Saha, scientist, ICAR, Barapani, for allowing me to use their laboratory facilities.

I am also thankful to Dr. N.P. Goel, Centre for Adult Education, NEHU, for his kind help to carry out the statistical analyses.

I would like to express my thankfulness to the non-teaching staff of the Botany department for their help and cooperation.

I greatly appreciate the help, suggestions and encouragement I received from my seniors, Dr. Vineeta Chauhan, Dr. Sangeeta Bharali, Dr. Swapna Prabhu, Dr. Om Prakash Tripathi and Dr. Krishna Upadhyay.

I would be ever grateful to my labmates and juniors, Mr. Mhathung Yanthan, Mr. Syntiew Manik Siemlieh, Mr. David Lalsama Biate and Ms. Meena Elangbam for their all round help, encouragement and specially maintaining a friendly working atmosphere in the laboratory. I shall always cherish the sweet memories of being part of this laboratory in future.

I take this opportunity to express my special thanks to my friends, Ms. Anusuya Rout, Ms. Biswarupa Ghosh, Ms. Mahasweta Satpati, Dr. Aparesh Patra, Dr. Debipreeta

Dutta, Dr. Y. Premabati, Dr. T. Gomti Devi, Ms. Memma Singha, Ms. Cressida Jamir, Ms. Shiny Sangma, Ms. Nabanita Devdasan, Mr. Tia Moli and Mr. Laliatsanga, for their help and cooperation at various stages of my work.

I am forever indebt to my parents and to my brother for always being there for me. My special love, respect and thanks go to my father, Prof. S.K. Khan, for always being the ultimate source of inspiration for me to pursue a career in research.

Last, but not the least I shall always be thankful to my husband, Dr. B.L. Yadav for being the best friend in need with all his patience, care, understanding and constant encouragement throughout my entire study period.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

CHAPTER 1

INTRODUCTION

The growth of all organisms depends on the availability of mineral nutrients, and none is more important than nitrogen which is required in large amounts as an essential component of proteins, nucleic acids and other cellular constituents. Though there is an abundant supply of nitrogen in the atmosphere, molecular nitrogen is metabolically unavailable directly to higher plants and animals. The presence of a triple bond between the two nitrogen atoms of a molecule makes it almost inert. Nitrogen must be converted into ammonium (NH_4^+) or nitrate (NO_3^-) ions before it could be used by plants and animals. Conversion of molecular nitrogen to ammonium or nitrate forms is also called as nitrogen fixation. Only few microorganisms can 'fix' atmospheric nitrogen, making all other living organisms dependent on them for their requirements of 'fixed' nitrogen.

Micro-organisms that fix nitrogen are called diazotrophs. Biological nitrogen fixation is brought about by two types of micro-organisms – non-symbiotic or free living and symbiotic. The free living non-photosynthetic diazotrophs require a chemical energy source for bringing about nitrogen fixation, whereas the free living photosynthetic diazotrophs utilize light energy for the purpose. Free living micro-organisms like *Scytonema*, *Rivularia* or *Klebsiella* fix nitrogen utilizing fixed carbon as energy source. They therefore contribute very little fixed nitrogen to agricultural crops.

On the other hand diazotrophs that live in close proximity to plant roots and obtain fixed carbon from the plants are symbiotic or associative nitrogen-fixing microorganisms. Microorganisms like *Rhizobium* and *Frankia* form symbiotic association with host plant(s) and utilize fixed carbon supply of the host for fixing atmospheric di-nitrogen. These microorganisms make a substantial contribution of fixed

nitrogen to agriculture and forestry. It has been estimated that *Frankia* contributes about 2-362 kg N/ha/yr while the estimated contribution of Rhizobium-Legume symbiosis is about 24-584 kg N/ha/yr. (Shantharam and Mattoo, 1997).

Since the *Frankia* symbiosis results from an actinomycetic invasion of plant roots, it has been termed as "actinorhizal symbiosis" (Benson, 1988). Accordingly, the plants nodulated by *Frankia* are called "actinorhizal plants". Although these plants are taxonomically diverse, they have some common features. For example, all of them are dicotyledonous and perennial angiosperms (Baker and Schwintzer, 1990). Actinorhizal plants belong to four subclasses, eight families, twenty-five genera and more than 220 species (Wall, 2000). Well known genera are *Alnus* (Betulaceae), *Myrica* (Myricaceae), *Casuarina* (Casuarinaceae), *Elaeagnus, Hippophae* (Elaeagnaceae), etc. *Frankia* have attracted attention recently because they form root nodules on a broad range of non-leguminous plants and because such nodules fix N_2 as effectively as rhizobial nodules.

Among actinorhizal plants, the two genera *Alnus* and *Casuarina* exhibit highest nitrogen fixing potential (Dommergues, 1996) and *Alnus* is the most extensively studied among them. *Alnus* species are important among actinorhizal plants as pioneers in ecological succession of skeletal soils (Lawrence, 1951; Crocker and Major, 1955; Damière *et al.,* 1986). They are widely used for intense forest management (Tarrant and Trappe, 1971; Borman and Gordon, 1984), biomass production (Zavitkovski *et al.*, 1979) and regeneration of disturbed lands (Heilman, 1982). It is thought to have originated in Indo-China region (Furlow, 1979). About 47 species of *Alnus* are known (Swensen and Mullin, 1997). In India, only two species are found (*A. nepalensis* and *A. nitida*) and they are naturally distributed throughout the temperate Himalaya. *A. nepalensis* is confined to the higher elevations of Meghalaya and Arunachal Pradesh, and *A nitida* is found in

Himachal Pradesh. Trees of *A. nepalensis* are also found in some locations of Nagaland and Tamil Nadu (Varghese, 2000).

The genus *Hippophae*, commonly known as sea buckthorn is another actinorhizal plant which forms symbiotic association with *Frankia*. A fascinating plant species, it is a very attractive ornamental shrub. It has silvery deciduous leaves and colourful orange berries that persist through most of the winter. It is a native of Eurasia and used by humans for centuries (Akulinin, 1958; Ge *et al.*, 1985). Among all the species of this genus, *Hippophae rhamnoides* is the most widespread. It has been divided into approximately eight geographically separated subspecies, but some scientists think that some of these deserve the rank of species (Small *et al.*, 2000). The unusually hairy *H. gyantsensis* occurs only in a restricted part of Tibet adjacent to Sikkim (Small *et al.*, 2000).

Hippophae is believed to be a colonizer of open habitats. In India their distribution is strictly restricted to the higher altitudes of the Himalayas. The highly efficient relationship with *Frankia* results in the improved root growth which enhances the entire soil ecosystem.

Sea buckthorns have great economic value. The berries, leaves and seeds have tremendous nutritional and medicinal value. It has three to sixteen times as much vitamin C as Kiwi fruit. Its superoxide dismutase content of the fruit juice is four times higher than that found in Ginseng. It contains twenty four minerals, eighteen amino acids and eight vitamins (Small *et al.*, 2000). Sea buckthorn is also useful as soil enhancer, pollution reducer, source of firewood, and as a landscape management tool (Li and Schroeder, 1996).

The actinorhizal plants are found primarily in the temperate zone. Only some species of the Casurinaceae and the Myricaceae can be considered truly tropical. Some species of

Alnus and *Elaeagnus* are also found in the tropical zone but these species are restricted to the higher elevations of mountain regions, where the climate is essentially temperate. In high latitude countries such as Scandinavia, Canada and New Zealand conditions are not favourable for legumes to thrive. However, actinorhizal plants are abundant and capable of vigorous growth (Silvester, 1977).

These plants involved in symbiosis with *Frankia* are important pioneer species in nitrogen poor soils or disturbed environments (Benson *et al.*, 1984). They are extremely useful for rehabilitation of degraded and salt affected soils, where other plants may fail to grow. The high nitrogen content of the leaf litter of actinorhizal trees increases soil fertility and can pave the way for diversification of species within the ecosystem. Recent genetic studies show that possibly the eight different tree families of actinorhizal host plants developed the capacity for symbiosis at separate stages during evolution (Swensen, 1996; Myrold *et al.*, 1999).

Though the actinorhizal association is analogous to the much better studied *Rhizobium*-legume symbiosis, very little is known about *Frankia*, especially its ecology and genetics. Thus there is a need for obtaining basic information about the population sizes and diversity of *Frankia*. It may be noted that the first successful isolation of *Frankia* was reported only in 1978, when strain CpI1 was obtained from *Comptonia peregrina* by Callaham *et al.* (1978).

Over the past few decades, agriculture in the developed world has become increasingly dependant on industrially fixed chemical fertilizers and pesticides for achievement and maintenance of the high yields possible with modern crop types. In recent years, the adverse effects of these chemicals on the rural population, the environment and on food safety, has urged the need for changes in agricultural production methods, with the objectives of sustainability, economic production,

conservation of natural resources, and reduction in the use of synthetic chemicals. The challenge is to institute sustainable methods without compromising food-production levels, particularly in view of the need to increase agricultural yields worldwide to accommodate population growth.

There are several significant environmental reasons to seek alternative to chemically fixed nitrogen fertilizer. Chemical fixation of atmospheric nitrogen affects the balance of the global N_2 cycle, pollutes ground water, increases the risk of chemical spills and increases atmospheric nitrous oxide- a potent "green house" gas. On the other hand, biological nitrogen fixation offers an economically attractive and ecologically sound means of reducing external nitrogen input and improving the quality and quantity of internal resources.

Forest trees are essential for human society. They provide fuel, fiber, building materials, food and medicines among other things. Forests stabilize environments and trees are cornerstones of ecosystems. The importance of nitrogen-fixing actinorhizal plants in forestry has been recognized (Montpetit and Lalonde, 1988). These plants are widely used in such practices as reforestation, land reclamation and biomass production (Fessenden 1979; Gordon *et al.*, 1979; Ponder, 1983; Tarrant *et al.*, 1971).

The genetic resources of forest trees must be conserved to maintain and improve our forests. Various classification systems have been used to categorize modern biotechnologies used in forestry sector, mainly because of the potential genetic gains they could confer. This can be categorized as – biotechnology employing molecular markers, biotechnology aimed at enhancing plant propagation and biotechnology for modifying the genome of forest tree species. Application of biotechnologies in forests has been seen as a unique opportunity for obtaining new information on the extent, patterns and functioning of tree genetic diversity and for providing new tree varieties and

reproductive materials adapted to changing environmental, social and economic environments (Fenning and Gershenzon, 2002).

Biotechnology can provide tools for the identification of superior genotypes through the characterization of DNA markers. The same tools are used to study the genetic diversity of tree populations which is a component of biodiversity and is important in ensuring the sustainability of the forest resource.

Frankia-actinorhizal symbiosis has attracted a great attention in Biological Nitrogen Fixation research recently. Prior to this our understanding of the molecular genetics of *Frankia* was restricted because of the fact that frankiae have slow growth rates, and there was lack of availability of a suitable vector for introduction of alien DNA into *Frankia* cells. *Frankia* strains have tremendous diversity at molecular level as shown by DNA-DNA hybridization of total cellular DNAs. Moreover, plasmids have been found in some frankiae. Several of the *nif* genes have been identified in *Frankia* but progress in identifying *nod* genes has been limited.

Actinorhizal symbiosis is such an association where both the host and the microbe have very important roles to play. Therefore, a superior and efficient host microbe relationship could positively affect nitrogen-fixing capacity of the *Frankia* strain (Chauhan, 2000). And for establishing this superior and efficient relationship, most infective, effective and competitive *Frankia* strains have to be selected, which in turn require investigations on diversity existing within the species.

Study of genetic diversity of *Frankia* would become useful if this information is used in enhancing its natural abilities. There exists tremendous potential if we are able to enhance the nitrogen fixing ability of *Frankia* and are able to extend its host range to include some horticultural trees. However, we must understand the effect of ecological

parameters on molecular constitution and behaviour before we set out to achieve the above. It is in this context that the present study was taken up.

Objectives

The present study was taken up to study the impact of certain ecological parameters on distribution of different strains at different sites. To achieve this broad objective, the following specific tasks were laid out:

1. Molecular typing of *Frankia* strains present in the field collected nodules of some genera of actinorhizal trees from Sikkim.

2. To study relationship, if any, between ecological parameters (different soil characteristics, altitude etc.) and molecular diversity of *Frankia*.

Review of Literature

CHAPTER 2

REVIEW OF LITERATURE

Biological nitrogen fixation is an essential natural process that supports life on this earth. It is important on a global basis and is essential to the development of certain ecosystems (Akkermans and Roelofsen, 1980). The necessary information for biological nitrogen fixation is present only in the genotypes of some prokaryotes (Ruvkun and Ausbel, 1980), two of which have established endophytic symbioses with extensively represented groups of plants: *Rhizobium* with the legumes and *Frankia* with the actinorhizal plants. The symbiosis involving *Rhizobium* and legumes has been extensively studied due to its paramount importance in agriculture. However, the actinorhizal symbiosis is also regarded as a "dynamic biological system" where both the host plant and *Frankia* have significant roles to play (Verghese and Misra, 2002).

Actinorhizal associations rival *Rhizobium*-legume association with respect to the amount of nitrogen that they fix on a global basis (Schwintzer and Tjepkema, 1990). During the past two decades, several studies have been done to have a better understanding of the biology and genetics of the microsymbiont. Efforts dealing with the molecular ecology of the microsymbiont are still in the preliminary stages and hold a promising future in research.

2.1 The Actinomycete: *Frankia*

The microsymbionts within the root nodules of actinorhizal plants were first described in the second half of the nineteenth century. Later, the genus was named *Frankia* by Brunchorst in 1886 (Lechevalier and Lechevalier, 1989). *Frankia* has been

classified into the family Frankiaceae under order Actinomycetales (Hahn *et al.*, 1989; Normand *et al.*, 1996).

Actinomycetes are filamentous, branching, gram positive bacteria. The members of the genus *Frankia* taxonomically show some special distinguishing characters such as their host specificity, morphology of hyphae, vesicles and sporangia, biochemistry of cell walls, etc. (Lechevalier, 1984; 1994). 16S RNA cataloguing confirmed the classification to genus on the basis of cell chemistry, which reflects phylogenetic relations (Stackebrandt, 1986).

More than a century has passed between the time the presence of endophytes in root nodules of actinorhizal plants was suspected and their successful isolation. In general, the reason for this long delay was due more to the slow growth rate of frankiae, which may have a doubling time of up to 5 days. The nature of the endophyte remained controversial until the first successful isolation of strain CpI1 from *Comptonia peregrina* by Callaham *et al.* (1978). This opened a new era in research on actinorhizal symbiosis and especially on *Frankia*. Hundreds of isolates became available in pure cultures from varied range of actinorhizal plant species. Misra *et al.* (1991) characterized the microsymbiont without their isolation in pure culture on the basis of new techniques available for isolation and amplification of DNA directly from single nodule lobes. However, there are evidences that more than one *Frankia* strain can be present within a single nodule lobe (Diem *et al.*, 1983; Clawson *et al.*, 1998; Zhang *et al.*, 1984) and therefore it became necessary to obtain genetically pure cultures of *Frankia*. Techniques for generation of purified *Frankia* strains through single spore cultures were made available by Prin *et al.* (1991), and Sarma *et al.* (1998).

2.2 Actinorhizal plants: The Hosts

The plants that enter into a mutually beneficial symbiotic relationship with *Frankia* are in general termed as "actinorhizal plants".

As compared to the super specialized host range of *Rhizobium* confined only to the leguminous plants, *Frankia* is a more versatile and dynamic organism. It has a wider host range of plants belonging to eight families, twenty five genera and over two hundred species. All of them, except *Datisca,* are trees or woody shrubs. These trees can grow in marginal soils and are spread across every continent other than Antarctica. These plants play a major role in increasing the fertility of nitrogen-depleted soils generated due to flooding, fire, glacial activity, volcanic eruption, etc. They are also economically important as timber, fuel, medicine, windbreaks, sand dune stabilizers and in agro-forestry applications. Amongst the actinorhizal trees, alder together with *Frankia* is widely recognized as a good nitrogen fixer (Guofan and Tingxiu, 1987; Domenach *et al.*, 1989; Simonet *et al.*, 1991). It is thought to be responsible for a high level of soil nitrogen accretion worldwide (Tarrant and Trappe, 1971). Therefore, *Frankia*-alder associations have been extensively studied since the time the significance of actinorhizal symbiosis was fully comprehended.

2.3 Actinorhizal symbiosis

Nitrogen-fixation by *Frankia*-actinorhizal plants is often central to the dynamics of several ecosystems. The atmospheric nitrogen gets reduced by the bacterial enzyme nitrogenase and produces ammonia, which is later assimilated by the host plant. The ability of microsymbiont *Frankia* to reduce atmospheric nitrogen using nitrogenase enzyme is the driving force of this symbiosis. The ammonia thus produced by *Frankia* is partly made available to the host. In return, the host provides protection, carbohydrates

and other nutrients to the bacteria. This enables actinorhizal plants to thrive where soil nitrogen levels restrict plant growth.

2.4 Application of molecular techniques on actinorhizal symbiosis

For developing improved strains, the hidden characters of the natural heterogeneity of strains should be exploited (Normand and Lalonde, 1986). Initially, the diversity of *Frankia* strains were evaluated on the basis of serological analysis (Baker *et al.*, 1981), efficiency testing (Normand and Lalonde, 1982), analysis of total protein patterns (Benson and Hanna, 1983; Benson *et al.*, 1984), and SDS-PAGE (Benson and Hanna, 1983; Gardes and Lalonde, 1987). Later, when a wide range of strains became available, nodulation tests (Baker, 1987), isozyme analysis (Gardes *et al.*, 1987) carbon utilization studies (Shipton and Burggraf, 1982; Bloom *et al.*, 1989), etc. were utilized for the same purpose. Several molecular techniques have now become available for studying the natural heterogeneity in *Frankia* populations.

Polymerase Chain Reaction (Mullis *et al.*, 1986; Mullis and Faloona, 1987) based techniques have shown enormous potential for detection of naturally occurring DNA polymorphism. Restriction fragment length polymorphism (RFLP) analysis (An *et al.*, 1985b; Dobrista, 1985; Normand *et al.*, 1988), DNA hybridisation studies (An *et al.*, 1985a; Bloom *et al.*, 1989; Fernandez *et al.*, 1989) and 16S rDNA analysis (Hahn *et al.*, 1989; Harry *et al.*, 1991; Nazaret *et al.*, 1991; Mirza *et al.*, 1992; Ganesh *et al.*, 1994) were employed for assessing diversity of *Frankia* strains at molecular level. Restriction fragment length profiles obtained by the digestion of genomic DNA and PCR amplified products have helped characterize a number of organisms (Jamann *et al.*, 1993; Laguerre *et al.*, 1996; Verghese and Misra, 2000).

With the advancement in research on *Frankia*, Misra *et al.* (1991) showed the possibility of amplification and sequencing of DNA directly from nodule lobes without going for pure cultures. Further, the amplified PCR products could be subjected to digestion with appropriate restriction enzyme to get unique restriction patterns. This approach, variously called as PCR-RFLP (Nazaret *et al.*, 1991; Navarro *et al.*, 1992; Jamann *et al.*, 1993; Rouvier *et al.*, 1996) and CAPS (cleaved amplified polymorphic sequences) (Chen *et al.*, 1994; Ghareyazie *et al.*, 1995; Tsumura *et al.*, 1997; Perry *et al.*, 1999) has been used for discriminative studies of closely related *Frankia* strains (Navarro *et al.*, 1992; Jamann *et al.*, 1993; Rouvier *et al.*, 1996; Verghese and Misra, 2000). This approach was used by Sarma *et al.* (1998) in demonstrating the presence of multiple genomic strains in axenic cultures of *Frankia*.

2.4.1 Molecular Phylogenetic studies in *Frankia*

The use of ribosomal RNA as a taxonomic tool has been well demonstrated in bacteria. 16S rRNA sequence analyses have helped in characterization of phylogenetic relationships among various *Frankia* strains (Woese, 1987; Woese and Fox, 1977).

Because of the presence of more variable regions in addition to the highly conserved areas (used in study for distantly related taxa), 16S rDNA sequences are useful in determining relationships among organisms and the differentiation of genera and species (Jensen *et al.*, 1993). While 16S rRNA gene was used to distinguish between *Frankia* strains Ag45/Mut 15 (*nif*$^+$) and Ag B1.9 and Ag W1.1 (*nif*$^-$) by Hahn *et al.* (1989), 16S rDNA was used for studies including identification, characterization and differentiation of *Frankia* infecting different host species (Nick *et al.*, 1992; Mirza *et al.*, 1994; Murray *et al.*, 1997; Clawson *et al.*, 1998). The possibility of multiple origins of actinorhizal

symbioses has been demonstrated by molecular phylogenetic studies of both plants and *Frankia* (Jeong *et al.*, 1999; Swensen and Mullin, 1997).

Nazaret *et al.* (1991) used a rapid and reproducible protocol based on DNA amplification and estimated the phylogenetic relationship among the genomic species of *Frankia*, using double strand sequencing of the resulting partial 16S rDNA sequences. Earlier, many studies showed that *Frankia* had a single origin. Based on molecular phylogeny using full length 16S rDNA sequences, Normand *et al.* (1996) suggested that *Frankia* was monophyletic and was emended to a new family. The study also suggested that *Frankia* formed a well defined, coherent cluster and that the genus was further subdivided into four clusters. Molecular phylogenetic trees were reconstructed from nucleotide sequence of *nif* H and 16S rDNA for *Frankia*. Comparison of *Frankia* phylogenetic trees reconstructed using *nif* H and 16S rDNA sequences indicated that subgrouping of both trees corresponded with each other in terms of plant origins of *Frankia* strains. This result also suggested that 16S rDNAs can be utilized for co-evolution analyses of actinorhizal symbiosis (Jeong *et al.*, 1999). Varghese *et al.* (2003a) utilized the DNA sequences of the *rrn* region in working out the evolutionary trend in this symbiotic association.

2.4.2 PCR based studies on conserved regions

Invention of PCR technology has revolutionised research on molecular aspects of actinorhizal symbiosis. Since, PCR needs DNA primers, it is more useful to study the conserved genes. These genes have sequences, which are more or less conserved throughout the living kingdom. Any sequence change in these genes represents an evolutionary transition. The genomes of most of the organisms have regions of large sequence conservation and regions of substantial variability. The conserved regions can

be used for determining close relationships among various genera, whereas analysis of non-coding variable regions could extend the utility of the molecule at lower taxonomic level.

For *Frankia* strain characterization, different typing methods have demonstrated their usefulness (Jamann *et al.*, 1993). But new methods using defined PCR-amplified DNA fragments as substrate for restriction pattern analysis or sequencing permitted the identification of various microorganisms with less consumption of time and with little material.

2.4.3 Studies on *nif* genes of *Frankia*

The enzyme nitrogenase, involved in the fixation of molecular nitrogen, could be a useful tool for characterizing N_2-fixing microorganisms. The structural genes for nitrogenase (*nif*H, *nif*D, *nif*K) present both low and high sequence variability and have been useful for *Frankia* strain specific detection by PCR (Simonet *et al.*, 1990).

For developing a PCR-RFLP typing method, the *nif* D-K IGS seems to be more appropriate than the *nif* H-D IGS because it is much larger and probably more variable (Normand, 1992a). Jamann *et al.* (1993) investigated the feasibility of using a PCR-RFLP technique targeted on the *nif* D-K IGS for typing and eventually for estimating genetic divergence between relatively close *Frankia* strains.

2.4.4 16S-23S ITS region of *Frankia*

The polymorphic nature of internal *rrn* spacer regions can be used as a direct and rapid method to characterize organisms. In *Frankia* strain ORS020606 the organization of rRNA encoding *rrn* genes has the usual order of 16S-23S-5S as observed in other bacteria. This *Casuarina* infective strain showed the presence of two *rrn* clusters, which

couldn't be differentiated through RFLP studies, but showed a lot of variation in non-coding regions (Normand *et al.*, 1992b).

From studies carried out on this part of the gene, it is evident that the first intergenic spacer region between 16S-23S rRNA shows high degree of variation. Along with the ITS (internally transcribed spacer) the IGS (intergenic spacer) regions also accumulate more mutations compared to the flanking genes. Therefore, they can be used alone or in combination with flanking conserved regions to differentiate closely related organisms at both interspecific and intraspecific levels. PCR can be carried out for the spacer regions with the help of primers from flanked conserved sequences. ITS region between 16S-23S rRNA and IGS region between *nif* D-K were used by Rouvier *et al.* (1996) in *Casuarina* and *Allocasuarina* for diversity assessment at strain level. *Ceanothus*- infective *Frankia* strains were also studied for genetic diversity using PCR-RFLP of 16S-23S rRNA ITS region (Ritchie *et al.*, 1999). Varghese *et al.* (2003b) have advocated the use of the hypervariable ITS region for developing strain specific markers for selection of superior genotypes of both the microsymbiont as well as the hosts.

2.4.5 Other molecular approaches on the genetic diversity of *Frankia* and actinorhizal hosts

Molecular genetic markers had proved to be powerful tools to analyze genetic relationships and genetic diversity. In 1992, Hadrys and coworkers suggested applications of Random amplified Polymorphic DNA (RAPD) technique in molecular ecology to determine taxonomic identity, assess kinship relationships, analyse mixed genome samples and to create specific probes (Hadrys *et al.*, 1992). Using high resolution molecular fingerprinting techniques like RAPD, repetitive extragenic palindromic PCR and multilocus enzyme electrophoresis, a high bacterial diversity below the species and subspecies level (microdiversity) was revealed (Schloter

EASTERN Bal... 10 3802 Date

2000). It became apparent that bacteria of a certain species living in close association with different plants either as associated rhizosphere bacteria or as plant pathogens or symbiotic organisms, typically reflect this relationship in their genetic relatedness.

Recently, a study was carried out to analyse σ^{70} gene family in *Frankia*. σ^{70} family of transcriptional factors was stated to be essential for bacterial life. Members of this family have been duplicated through evolution and have evolved and specialized for all sorts of physiological responses (Blaha and Cournoyer, 2003).

PCR-RFLP study of IGS between *nifD* and *nifK* was applied to assess diversity of *Frankia* microsymbiont of nodule samples collected from five sites in the Hengduan Mountains. The study found presence of some nodule samples, which produced more than one PCR fragment, and compound RFLP patterns, indicating that *Frankia* strains with different PCR-RFLP patterns coexisted in the same host plant under natural conditions (Dai *et al.*, 2004).

Several segments of *Frankia* DNA have now been characterized. According to Lavire and Cournoyer (2003), at the end of July 2001, 293 kbp of *Frankia* DNA sequences were found in the databases. Thirty five percent of these sequences corresponded to full gene or gene cluster sequences. These genes could be divided according to their role into 6 key activities: gene translation (*rrnA* and *tRNA* ^{pro} gene), proteolysis (*pcr* genes), assimilation of ammonium (*glnA* and *glnII*), protection against superoxide ions (*sodF*), nitrogen fixation (*nif* cluster), and plasmid replication. These regions of DNA can be potentially used for characterizing various *Frankia* strains.

2.5 Molecular ecology of *Frankia*

Microbial ecology has been dependent on the use of conventional microbiological methods for a long time. Many microorganisms, like actinomycete *Frankia* are difficult

to cultivate. This problem has been approached by characterizing the microbes on the DNA level, without unreliable cultivation steps. This new approach in ecology is defined as "Molecular Ecology" and covers the range of applications of genetic and molecular techniques and theories in ecology. During the last few years, genetically modified microbes have been constructed by recombinant DNA techniques for putative use in the environment. The slow progress in this field is due to the lack of integration of microbial ecology and molecular biology.

Akkermans *et al.* (1994) showed the application of DNA probes and marker genes in the research on the ecology of Genetically Modified Micro-Organisms and recalcitrant or unculturable wild type microorganisms. Emphasis was given on the development and use of oligonucleotide probes to detect microbes in various engineered ecosystems, like *Frankia* in root nodules, etc. According to Hery *et al.* (2001), in New Caledonia, the unbalanced soils are abnormally rich in nickel and deficient in major elements and, thus constitute an original environment for studying microbial ecology. Some endemic plant species, such as *Gymnostoma* spp. (Casuarinaceae) are adapted to these extreme soils. A study of the genetic structure of *Frankia* populations present in root nodules of the eight *Gymnostoma* species has shown a correlation between *Frankia* genotypes and soil types. These findings suggest that soil conditions are important factors driving local bacterial adaptation: some *Frankia* populations are adapted to a nickel rich environment, while others are adapted to non-ultramafic soils. To determine whether the entire bacterial community follows the same trend, they spiked two soils with nickel on which *Gymnostoma* spp. grow. The result showed that there were few changes upon addition of nickel and that the main group of nickel resistant bacteria belonged to the actinomycete group.

Another recent study reported the effect of natural stress on genetic diversity of *Frankia* in *Alnus* nodules. By using rep-PCR technique, Tang *et al.* (2003) showed

abundant genetic diversity of *Frankia* in the mountains of Yunan province of China with different zones, altitudes and slopes. They concluded that diversity index was positively related to natural stresses.

Diversity of *Frankia* was also assessed in the Gaoligong Mountains of Yunan Province using PCR-RFLP technique by Dai *et al.* (2004). They found that more than one genotype of *Frankia* strains could form symbiosis with individual plants at the same time.

Materials and Methods

CHAPTER 3

MATERIALS AND METHODS

3.1 STUDY AREA:

The present study was conducted in Sikkim, which is a small mountainous state in the Eastern Himalayan region of India, extending approximately 114 km from North to South and 64 km from East to West, having a total geographical area of 7096 sq. km. The state is situated between 88°00'58" and 88°55'25" East longitudes and 27°04' and 28°07'48" North latitudes. It is surrounded by vast stretches of Tibetan plateau in North; Kingdom of Bhutan in the East; Darjeeling district of West Bengal in South and Kingdom of Nepal in West. The state has four districts namely East, West, North and South.

Sikkim is bestowed with abundant natural resources. Although Sikkim constitutes only 0.2% of the geographical area of the country, in terms of species richness it ranks very high. It has been identified as a Global Biodiversity Hot Spot. Hot Spots are areas that are extremely rich in species, have high endemism, and are under constant threat. Among the 25 hot spots of the world, two are found in India extending into neighbouring countries – the Western Ghats/Sri Lanka and the Indo-Burma region (covering the Eastern Himalayas).

3.2 COLLECTION SITES:

Ten sites located in the East and North Sikkim districts were selected for the present study. These sites were selected on the basis of their altitude and geographical location. Three different altitude zones were chosen, namely higher (>2000 metre from mean sea

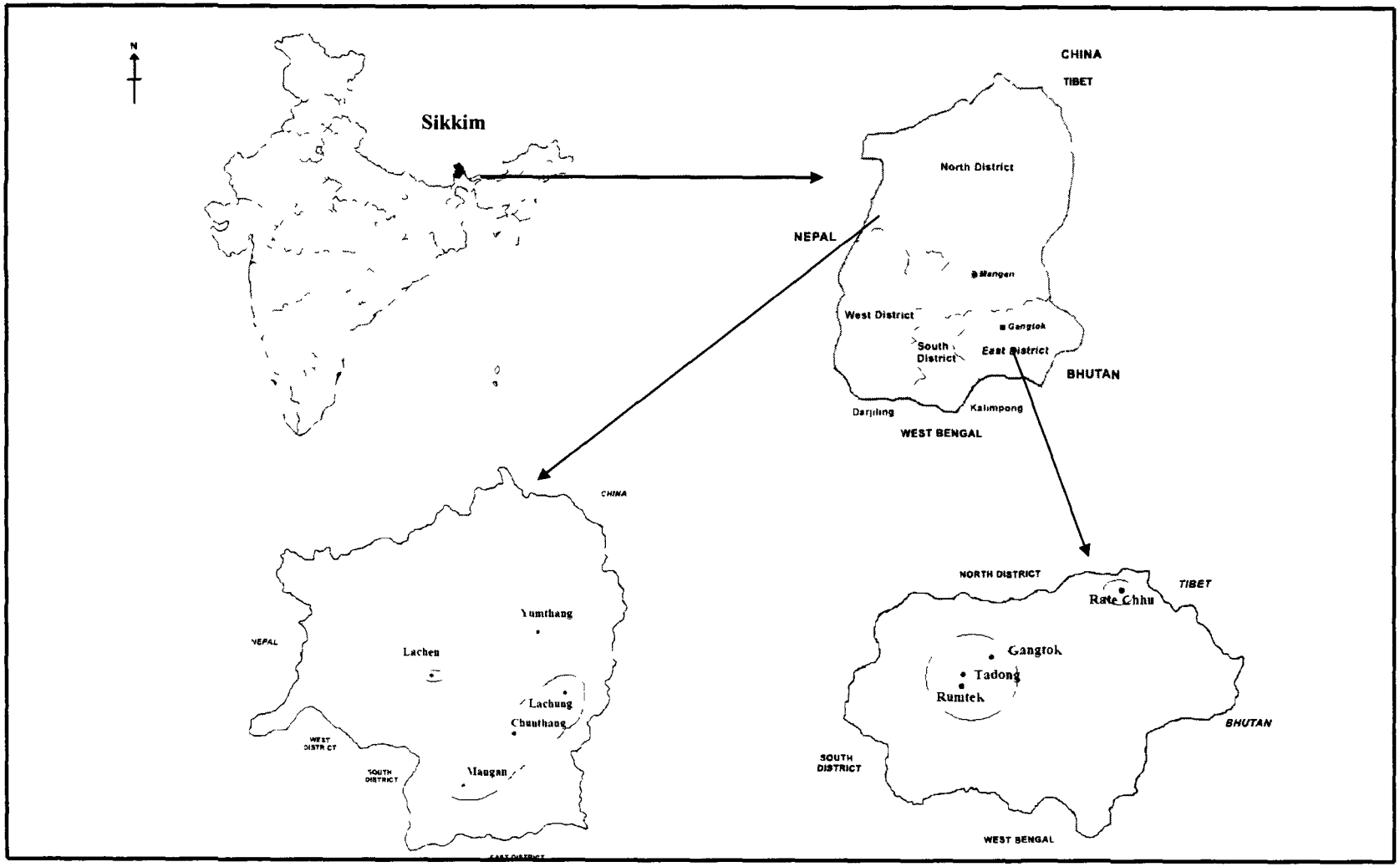

N
Sikkim
CHINA
TIBET
North District
NEPAL
Mangan
West District
Gangtok
South District
East District
BHUTAN
Darjiling
Kalimpong
WEST BENGAL
CHINA
Yumthang
NEPAL
Lachen
Lachung
Chunthang
WEST DISTRICT
SOUTH DISTRICT
Mangan
NORTH DISTRICT
TIBET
Rate Chhu
Gangtok
Tadong
Rumtek
BHUTAN
SOUTH DISTRICT
WEST BENGAL

Fig.1

level), middle (1500-1800 m from msl) and lower (<1000 m from msl) from each area. Samples from *Alnus* stands were collected from three different geographical areas, two from East Sikkim district and one from the North Sikkim district. Three sites from each geographical area were selected representing the three altitudes as above. *Hippophae* samples were colleted from ten trees from a single site in the high altitude range in North Sikkim (Fig.1).

3.2.1 COLLECTION OF *Alnus nepalensis* NODULES

3.2.1.1 SITES IN EAST SIKKIM DISTRICT:

- **High Altitude, Site-1 (On way to Hanumantok, Gangtok), 2100 m-**Samples were collected mainly from trees on the roadside slopes. The average age of the trees at this site was five to eight years. Solitary *Alnus* trees were found growing among other trees. Tree number 10 was located near one small streamlet. The soil was porous, loose, moist to dry and full of organic matter. In some places, the soil looked sandy and greyish brown.

- **High Altitude, Site-2 (Himalayan Zoological Park, Bulbulay, Gangtok), 2200 m-** Samples were collected from inside the zoological park. The average age of the trees was between seven to ten years. The soil was basically loose in texture, sandy in most places and greyish brown in colour.

- **Middle Altitude, Site-1 (Botanical Survey of India, Gangtok), 1709 m -** Samples were collected from trees in the BSI compound. Trees were about thirty to forty years old. Soil was blackish, more or less compact, and rich in organic matter.

Fig.2. *Alnus nepalensis* (near Hanumantok, Gangtok, East Sikkim) ⟶

Fig.3. Roots of *Alnus nepalensis* showing nodule clusters ⟶

Fig.2

Fig.3

Fig.4. Alder stands on the bank of river Teesta, North Sikkim ⟶

Fig.5. Alder stands on the mountain slopes, Gangtok, East Sikkim ⟶

Fig.4

Fig.5

- **Middle Altitude, Site-2 (Rateychhu, B1, Gangtok), 1700 m-** Samples were collected from trees near the Rateychhu bridge (B1) just before entering North Sikkim. It was a big stand of *Alnus* near one stream. Average age of the trees was fifteen to twenty years. Soil was mainly sandy, sometime rocky and dry. The colour of the soil was mostly greyish brown.
- **Low Altitude, Site-1 (Ranipool, Gangtok), 950 to 1000 m** - Samples were collected from trees of a big *Alnus* stand. Average age of the trees was around twenty-five years. Soil looked very fertile, rich in nutrients, wet and compact.
- **Low Altitude, Site-2 (Ranipool, Gangtok, On way to Rumtek monastery), 984 to 990 m** – Samples were collected from an *Alnus* stand located on a riverside. Very young (1 to 2 years) to old trees (20 to 25 years) were found. Soil was blackish, wet and compact. This site seemed to have a recent landslide.

3.2.1.2 SITES IN NORTH SIKKIM:

- **High Altitude, Site-3 (between Chungthang and Maltin), 2040 to 2067 m** – Samples were collected from young to middle aged trees from a site located on the roadside, approximately one kilometre before Maltin and eight kilometres from Chungthang. Age of the trees varied from 3 to 15 years. Soil was light brown in colour, sandy, mixed with rocks and loose.

- **Middle Altitude, Site-3 (Chungthang, near the bridge over Tista river), 1696 m**– Samples were collected from big trees of an *Alnus* stand. Profuse nodulation was found here. Average age of the trees was 20 years. Soil was wet, compact, blackish and full of organic matter.

Fig.6. *Hippophae* sp ⟶

Fig.7. *Hippophae* (with berries), near Lachen, North Sikkim ⟶

Fig.6

Fig.7

- **Low Altitude, Site-3 (between Mangan and Rang Rang), 1005 to 1024 m** – Samples were collected from young to middle aged trees from a site near Mangan. Age of the trees varied from 5 to 10 years. Soil was more or less compact, dark brown in colour and dry.

3.2.2 COLLECTION OF *Hippophae* NODULES:

- **High Altitude, Site 1 (Near Lachen), 2672 m** – Root nodules and soil samples were collected from a big *Hippophae* stand adjacent to a stream. The site was located approximately 1 kilometre before Lachen. Age of the trees was between 10 to 12 years. Soil was tight, compact, sandy and brownish in colour.

3.3 COLLECTION OF MATERIAL:

Root nodules and rhizosphere soil were collected for each individual tree. All together one hundred samples were collected, 90 for alder nodules and 10 for *Hippophae* nodules.

3.3.1 COLLECTION OF ROOT NODULES

Collections of root nodules were carried out in the month of October. This was based on the observation of Varghese (2000) that the nodule growth was maximum soon after onset of monsoon rains. Trees were randomly selected in each site. The lateral roots were traced out, and after digging the surrounding soil up to six inches, many nodule clusters were exposed.

The steps followed during collection were –

- The soil surface surrounding the selected tree was cleaned by removing weeds, etc.
- A sterilized hand spade was used for digging the soil.
- After removing the top soil lateral roots were found. Each time the rootlets were traced to their original tree before collection.
- The freshly collected nodules were kept in new plastic bags which were properly labelled giving name of the site, date of collection, sample number, altitude, etc.
- Nodule samples were brought to the laboratory and thoroughly washed under running tap water to clean all tightly adhering soil particles.
- They were further washed with double distilled water.
- The nodules were then surface sterilized for 2-3 minutes, using 30% Hydrogen Peroxide (H_2O_2).
- Finally this was followed by repeated rinsing using sterile double distilled water.
- After sterilization, the nodule samples were placed in freshly labelled new plastic bags and stored at -20°C for future use.

3.3.2 COLLECTION OF RHIZOSPHERE SOIL:

- From each tree rhizosphere soil surrounding the nodule clusters was also collected. About 250 gm of each soil sample was collected in properly labelled plastic bags.
- Two to three drops of toluene were added to it to stop further microbial activity.
- Once soil samples were brought to the laboratory, they were first air dried for two to three weeks.
- After soil samples were properly dried, they were sieved through a sieve to remove big pebbles and stones.

- Sieved soil samples were kept in new labelled plastic bags, and stored at room temperature for further analysis.

3.4 METHODS TO STUDY MOLECULAR DIVERSITY OF *Frankia*:

3.4.1 Isolation of genomic DNA from nodules:

Proper care was taken to avoid any contamination of DNA during the course of isolation. All the plastic wares and solutions were autoclaved prior to DNA extraction. Extraction of total genomic DNA from individual nodule lobes was carried out following the methodology given by Rouvier *et al.* (1996) with minor modifications.

- A single nodule lobe was first sterilized in 30% hydrogen peroxide (H_2O_2) for one to two minutes.
- It was repeatedly washed in autoclaved distilled water for three to four times.
- The outer epidermis was peeled off with a sterilized needle and the nodule lobe was transferred to a 1.5 ml micro centrifuge tube.
- The lobe was then crushed properly in 350µl of warm DNA extraction buffer and 10µl of 20% SLS (sodium lauryl sulphate) were added.
- The tube was then kept in a water bath for one hour at 65°C.
- The tube was then centrifuged at 8000 rpm for about 7 minutes at room temperature.
- The supernatant was transferred to another fresh tube and an equal volume of chloroform-isoamyl alcohol (24:1, v/v) was added to it. The tube was centrifuged again at 13,000 rpm for 30 minutes at room temperature to remove impurities like proteins, etc.

- The upper aqueous layer was transferred to another fresh tube and 2.5 volumes of ice cold ethanol were added.
- It was kept overnight at -20°C for precipitation.
- Next day the tube was centrifuged at 13,000 rpm for 30 minutes at 4°C. Ethanol was discarded and the DNA precipitate was washed twice with 350µl of 70% alcohol. Later the DNA precipitate was vacuum dried, dissolved in ultrapure water and stored at -20°C for future use.

3.4.2 Agarose gel electrophoresis:

The presence of isolated genomic DNA was detected through 0.8% (wt/vol.) agarose gel electrophoresis. Prior to electrophoresis the DNA samples were mixed with loading buffer (Sambrook *et al.*, 1989). Molecular size marker (λ DNA, Eco R1+Hind III double digest) was also run along with the samples. Electrophoresis was carried out in horizontal gel running tank using 1X TBE buffer, and run at 60-65 volts (5-10 v/cm) for an hour. The gel was then stained in ethidium bromide (0.5µg/ml) for about fifteen minutes and observed on a DNA transilluminator.

3.4.3 Quantification of DNA and Estimation of band size

Quantification of DNA was done either by UV fluorescence on an ethidium bromide stained agarose gel or by using BIO-RAD Gel Doc1000. Molecular weights of the bands were calculated by using Multi Analyst software (Version 1.1, build 34). Wherever it was not possible to estimate the molecular sizes using software, these were estimated using log graph paper and the following formula-

$$y = \log 1/x \quad \text{or} \quad [(9/500).\ x] = y$$

[x= molecular weight in base pair (bp) size and y= distance of band from the well which has to be plotted along the y axis].

3.4.3.1 Quantification by direct observation:

Direct quantification of DNA was carried out by visual observation on ethidium bromide stained agarose gel on a transilluminator. According to Sambrook *et al.*, (1989), the intensity of fluorescence is directly proportional to the total mass of DNA. Accordingly, the quantity of DNA sample was estimated visually by comparing the fluorescence yield by reference marker of known band size.

3.4.3.2 Quantification of DNA and estimation of band size by using Multi Analyst® software:

Using the BIO-RAD GelDoc1000, quantification of DNA samples was done by visually comparing the intensity of fluorescence of the bands and using Multi Analyst® software (Version 1.1, build 34) the molecular weight of each band of the marker was first calculated and then the molecular weight of the desired band was detected.

3.4.4 Amplification of genomic DNA by Polymerase Chain Reaction (PCR):

The PCR is an *in vitro* method for the enzymatic synthesis of specific DNA sequences, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA. It generates multiple copies of DNA of interest and thus facilitates a detailed and better analysis of the same. A repetitive series of cycles involving template denaturation, primer annealing, and the extension of the annealed primers by DNA polymerase (*Taq* DNA polymerase) results in the exponential accumulation of the target DNA (Mullis *et al.*, 1986; Mullis and Faloona, 1987).

PCR amplification was performed for the following regions –

i. Total 16S rRNA gene.

ii. 16S-23S rRNA ITS region.

iii. Total 23S rRNA gene.

Samples containing good amount of isolated DNA were subjected to PCR amplification. PCR was performed in a Thermal Cycler (GeneAmp PCR2400, Perkin Elmer, USA). Preparation of the PCR mix was done strictly under aseptic conditions using a UV hood. Only pre-autoclaved plastic wares and sterilized gloves were used during the preparation. The PCR mix was made in a total volume of 25µl in thin walled PCR tubes. Each reaction mixture contained the following amount of chemicals per tube-

- 2.5µl of 10µM primer,
- 2.5µl of 10X PCR buffer,
- 2.5µl of 25µM $MgCl_2$,
- 10µl of 1.25mM dNTP mix,
- 0.75µl of *Taq* polymerase (3u/µl),
- 1µl of template DNA
- Ultra pure water to make the total volume to 25µl of PCR mix per tube.

Each setup of amplification contained a negative control, where instead of adding the template DNA, 1µl of ultra pure water was added.

The annealing temperatures for the primers were roughly calculated using the following formula –

Annealing Temperature = [4 (G+C) + 2 (A+T)] – 5

However, annealing temperature was modified, i.e., decreased or increased, depending upon the results. Generally amplification was done at the lower annealing

temperature of the two primers used. An increase in the annealing temperature restricted the non specific amplifications and decrease in the temperature sometimes gave better results whenever it was difficult to get amplification.

Each amplification reaction was carried out for 35 cycles. Each cycle comprised of 1 min of denaturation at 94°C, 1 min of annealing at appropriate temperature and 1 min of elongation at 72°C. A hot start was given for five minutes at 94°C and at the end of the 35 cycles an additional 7 minute extension time at 72°C was added to allow complete extension of partial fragments (Fig 10).

3.4.4.1 Primer stock solutions:

The primers used for amplification were synthesized at Microsynth, Switzerland. They were desalted and PAGE purified. These lyophilized primers were first dissolved in ultra pure water to prepare a stock solution of 1M. A part of it was further diluted to make a 25mM stock solution. From 25mM stock, finally 10mM working solution was made from which the required amount was taken out to make the final PCR mix. All the stock solutions were stored at -20°C.

3.4.4.2 dNTP solutions:

The dNTPs (dATP, dGTP, dCTP, dTTP) in solution were acquired from Bangalore Genei, India. From the four individual deoxyribonucleotides, a final dNTP mix was prepared in a new tube by taking 50µl from each and adding 200µl of ultra pure water. This was working solution and it contained 1.25 mM of each dNTP. All dNTP solutions were stored at -20°C.

3.4.4.3 *Taq* DNA polymerase and Assay Buffer:

A 94kD thermostable enzyme, isolated from bacterium *Thermus aquaticus*, *Taq* DNA polymerase is used in PCR because it can withstand repeated exposure to the high temperatures required for strand separation. This enzyme lacks 3' to 5' exonuclease activity but has inherent 5' to 3' exonuclease activity. *Taq* DNA polymerase and assay buffer [10mM tris-HCl (pH 9), 1.5mM $MgCl_2$, 50mM KCl and 0.01% gelatin] were acquired from Bangalore Genei, India. The stock was stored at -20°C.

3.4.5 Agarose gel electrophoresis of the amplified products:

Amplified DNA samples were run in a 0.8% agarose gel at 70 volts for one and half hours. The gel was then stained in ethidium bromide for about fifteen minutes and scanned and photographed using Bio-Rad GelDoc1000 and the band sizes were calculated using the Multi-Analyst® software (version 1.1).

3.4.6 Amplicon Length Polymorphism (ALP):

Amplicons of different samples were subjected to gel electrophoresis together to detect the ALP. 1.2% agarose gel was used for this purpose. Known molecular weight marker (λ DNA, Eco R1+Hind III digested) was run with the samples to assess the variation in different amplicon length size. The ALP patterns were photographed using BioRad GelDoc1000.

3.4.7 Phenol purification of DNA:

Prior to restriction digestion, DNA was purified in few cases where presence of more than one band was found with the desired band (1200 bp).

The modified protocol of Byrnes *et al.* (1995) was used for this purpose.

- 20-30µl of amplified DNA was run on a 0.8% pre-stained agarose gel in 1X TBE buffer using horizontal electrophoresis.
- The movement of the bands was monitored frequently using a UV transluminator and when the band of interest was found separated from other bands, it was excised out using a clean sterile scalpel blade.
- All the excess agarose was removed and the slice containing the desired band was further cut into smaller pieces and transferred into a 0.5 ml micro-centrifuge tube.
- To this 40µl of saturated phenol (pH 8) was added and the tube was left overnight at 4°C.
- Next day the tube was kept at 60°C for half an hour till the agarose was completely melted.
- Then it was centrifuged at 13,000 xg for 10 min at room temperature.
- The aqueous phase was pulled and transferred to another fresh tube.
- To this equal volume of chloroform: isoamyl alcohol (24:1) was added.
- The tube was again centrifuged at 13,000 xg for five min.
- The aqueous phase was then transferred to a fresh tube and the DNA precipitated with two volumes of ice cold ethanol. The tube was kept overnight at 4°C for precipitation.
- Next day it was again centrifuged at 12,000 xg for fifteen min at 4°C.
- The DNA pellet was washed with 70% alcohol and vacuum dried.
- Then the DNA was dissolved in 4-5µl of ultra pure water.

To confirm the presence of eluted DNA the elutant was again electrophoresed.

3.4.8 Amplicon Restriction Pattern (ARP) or PCR/RFLP (Restriction fragment length profile) analysis of PCR products:

Restriction digestion was done for all the samples. The amplified 16S-23S ITS region of the gene was subjected to restriction digestion. Suitable restriction enzyme, *Rsa*1 (Rosche Pharmaceuticals) was selected for this purpose and the samples were digested overnight at 37°C as specified by the manufacturer. The restriction digestion mixture was prepared in a 0.5 ml tube in which 10µl of the amplicon was digested with 5 units of the restriction enzyme in the appropriate buffer (buffer B) solution (2 µl). Ultra pure distilled water was added to make the final volume to 20µl. After digestion was over, the digested samples were run in ethidium bromide stained 4% agarose gel using horizontal electrophoresis in 1X TBE buffer at 50 volt (2-3 v/cm) for 5-6 hrs. The gel was then scanned and photographed using BioRad GelDoc 1000 and the profiles were analyzed using the Multi Analyst® software.

3.4.9 Nucleotide Sequencing:

The distal part of 16S rRNA, ITS and proximal part of 23S rRNA region was considered for nucleotide sequencing. On amplification this region produced a single band of 1200bp. The samples were lyophilized in 0.2 ml thin walled PCR tube, which were then wrapped with parafilm^{R}, labelled properly and packed in small plastic bags.

These were sent to Oregon State University, Corvallis, USA, for sequencing. Sequencing was done based on Sanger's dideoxy chain termination method (Sanger *et al.*, 1977).

3.5 ACETYLENE REDUCTION ASSAY (ARA):

ARA was performed for nodule samples collected from *Alnus* trees. Only one site from each altitude in East Sikkim was considered for this experiment. This experiment was done at the G.B Pant Institute of Himalayan Environment and Development, Sikkim campus. Perkin Elmer Gas Chromatograph (Model 8700) was used for the purpose.

Nitrogenase activity was measured using the method described by Stewart *et al.* (1968). Apart from reducing nitrogen, the nitrogenase enzyme complex can also reduce acetylene (C_2H_2) gas to ethylene (C_2H_4). This property is utilized in Acetylene Reduction Assay where the quantity of ethylene produced or acetylene reduced is a reflection of the activity of the nitrogenase enzyme.

The nitrogenase activity was measured with very fresh nodules collected early in the morning and the experiment was completed on the same day.

1. Ten fresh, light brown nodule lobes were taken together for each tree.
2. Fresh weight of the nodules were recorded.
3. Two replicates were taken for each sample i.e. altogether twenty nodules were used for every tree.
4. The nodules were not surface sterilized, instead the soil particles were brushed off carefully from the surface of the nodule lobes.
5. Nodule lobes were taken in sterile, stoppered and airtight vials.
6. 1 ml of air in the vial was replaced with 1 ml of pure acetylene. One control was also kept with only acetylene gas without any nodules to check its purity.
7. The vials were incubated for 3 hours at 28 ± 2 °C.

After the incubation period was over the quantity of ethylene produced was measured with the help of the Gas Chromatograph.

1. From the air tight vial, 1 ml of the gas mixture was taken out with the help of an air tight syringe and was injected at the injection port of the gas chromatograph.
2. The injection port, oven and detector temperatures were 120°C, 90°C and 175°C.
3. The flow rates of H_2, air and N_2 were 50, 120 and 10 me/min respectively.
4. To confirm the retention time, 1ml of standard ethylene solution was injected into the gas chromatograph and the ethylene peak in the samples were compared.
5. Under the above said conditions, the retention time for ethylene was approximately 1.50 minutes and that of acetylene was 2.5 minutes.
6. After recording the area of ethylene gas of each of the replicates of each sample, the nodules from each vial were weighed.

Acetylene Reduction Assay was carried out for ten trees from a single site on a single day. The experiment was completed over a period of three days since altogether three sites (one from each altitude) were considered for this study.

The nitrogenase activity was estimated in terms of nmoles of ethylene produced per gram fresh weight per hour using the following formula-

$$N.A. = \frac{\text{nmole } C_2H_4\text{/unit area in the standard X area of } C_2H_4 \text{ in the sample X vol of vial}}{\text{Fresh weight x Incubation time}}$$

3.6 METHODS USED FOR SOIL ANALYSIS:

Following soil parameters were considered for analysis –

- Soil pH
- Total organic carbon
- Total nitrogen
- Available phosphorus

- Available potassium
- Exchangeable calcium
- Exchangeable magnesium
- Electrical conductivity

- **Soil pH:** One gm of soil was mixed with 25 ml of distilled water in 50 ml beaker. This was kept on an electrical shaker for half an hour to allow the soil and water to mix properly. The beaker was kept overnight to allow the soil to settle down. Next day the pH was measured with an electrical pH meter.
- **Total organic carbon:** Total organic carbon was measured by Walkley and Black (1934) method.
- **Total nitrogen:** Total nitrogen was measured by Kjeldahl method (Allen *et al.*, 1974).
- **Available phosphorus:** Available phosphorus was measured by Bray extractant method. (Allen *et al.*, 1974) using a Flame Photometer.
- **Available potassium:** Available potassium was measured by neutral ammonium acetate method (Allen *et al.*, 1974).
- **Exchangeable calcium:** Exchangeable calcium was measured by Ammonium acetate extractant method (Hesse, 1971).
- **Exchangeable magnesium:** Exchangeable magnesium was measured by direct titration method with EDTA (Hesse, 1971).
- **Electrical conductivity:** Electrical conductivity was measured in solution of soil and distilled water (1:2.5 w/v) with the help of salt (conductivity) bridge (Hesse, 1971).

3.7 COMPUTER ANALYSIS OF DATA:

- All the gel photographs and detailed information like banding patterns generated through restriction digestion, sequence information of each nodule microsymbiont, etc. were saved in separate computer files for further analysis.
- The band sizes of amplified and digested products were calculated using Multi Analyst® (version 1.1) software.
- Cluster Analysis was done using NTSYS software (version 2.1). To execute the programme, a table was prepared for selected banding patterns on the basis of presence or absence of particular bands with respect to each other. Using this table, a correlation matrix was prepared based on Jaccard's Coefficient, which has the following formula –

  Jaccard's coefficient = common bands / (total no. of bands – common bands)

  This matrix was entered into the particular executable format of the software and using the method SAHN (Sequencial Agglomerative Hierarchial Nested Cluster Analysis) the cluster dendogram was prepared.

- The nucleotide sequences obtained were individually fed into the GenBank for Blast Analysis using the website http://ncbi.nlm.nih.gov. The retrieved sequences along with the original sequences were then aligned using multiple sequence alignment program CLUSTAL W (1.75) (Thompson *et al.*, 1994). The online tool Gene Bee (http://www.genebee.msu.su) was used to construct phylogenetic trees.
- Data of Soil analysis were also saved in separate files for further analysis.
- Mean, Standard deviation and standard error of mean have been calculated site-wise for each altitude, entering raw data into Microsoft Excel Spreadsheet.
- ORIGIN 40 was used for graphical presentation of soil data.

- ANOVA and STUDENT NEWMAN KEULS TEST were performed on soil analysis using the SAS system. (Users guide. SAS Inst., Cary, NC., 1987).
- PRINCIPAL COMPONENT ANALYSIS (PCA) was performed for soil parameters, banding patterns, sites and altitudes to see if there was any relationship among them.
- PCA was performed using the software STATISTICA (version 5.0).
- Multiple regression analysis (MRA) was done to find out the highest significant variable which is playing the most important role in the distribution of genotypes.
- MRA was performed using the software STATISTICA (version 5.0).

Results and Discussion

CHAPTER 4

RESULTS AND DISCUSSION

Present study was designed to investigate whether the distribution of *Frankia* genotypes was dependent on altitude and soil characteristics or not. Actinorhizal root nodule samples were collected from three altitudes and three sites. Different *Frankia* genotypes were determined based on their molecular characteristics.

4.1 COLLECTION OF NODULES:

One hundred trees were selected in east and north Sikkim districts for collection of actinorhizal root nodules. Among these, ninety belonged to *Alnus nepalensis*, while ten belonged to *Hippophae sp.* Collection sites were divided according to altitudes (section 3.2). Thus for each altitude there were three locations and a total of nine locations were identified for the purpose of collection of root nodules and soil samples.

Alnus nepalensis nodules were found to be uniform in their morphology at all the sites. Very little variation was observed for the *Hippophae* nodules. Compared to alder nodules, *Hippophae* nodules were found to be lighter in colour with rounder lobes. In all the cases, fresh nodules were light brown in colour while the degenerating ones from the previous season were dark brown. Only the fresh nodules were collected for the experiments.

Differences were found in the amount of nodulation between the younger and older trees. Younger trees were found to produce more nodule clusters near the soil surface. It was also found that very young to middle aged trees on the slopes of the hills had more exposed lateral roots with profuse nodulation.

4.2 DNA EXTRACTION:

Total genomic DNA was extracted following the method of Rouvier *et al.* (1996) with minor modification as described in section 3.4.1. All samples consistently gave a single band of approximately 21,000 bp on agarose gel electrophoresis (Fig.8). The quantity of DNA extracted per nodule lobe varied from sample to sample. However, the amount of extracted DNA was sufficient for amplification. Appropriate dilutions were done for some samples prior to their use as templates for DNA amplification experiments.

4.3 PCR Amplification of 16S-23S rRNA ITS region:

The variable internal transcribed spacer (ITS) between 16S-23S rRNA (Fig. 9.a) was analyzed using two specific primers for the region (Fig. 9.b). Primer FGPS 989ac (Bosco *et al.*, 1992) in 16S rRNA gene was selected as the forward primer and a universal primer FGPL 2045′/FGPS 132′ (Simonet *et al.*, 1991., Ponsonet and Nesme, 1994,) in the 23S rRNA gene was used as the reverse primer. Primer FGPS 989ac is alder-casuarina group specific and amplifies *Frankia* DNA from these two host groups. Generally amplification was done at the annealing temperature of 49°C. In some cases, multiple bands were obtained at this annealing temperature due to priming at alternative sites.

Visual observation of the agarose gels showed varied intensity of the amplified bands. In cases where faint bands were obtained, the reaction was repeated to confirm the profiles.

Agarose gel electrophoresis generated almost uniform fragments, although there were minor size variations among different samples, which were revealed by amplicon length

Fig.8. Isolation of genomic DNA of *Frankia* from individual nodule lobes. M- Molecular weight marker in base pairs (Lambda DNA, Eco R1+Hind III double digest). Lanes are marked by sample numbers loaded. ⟶

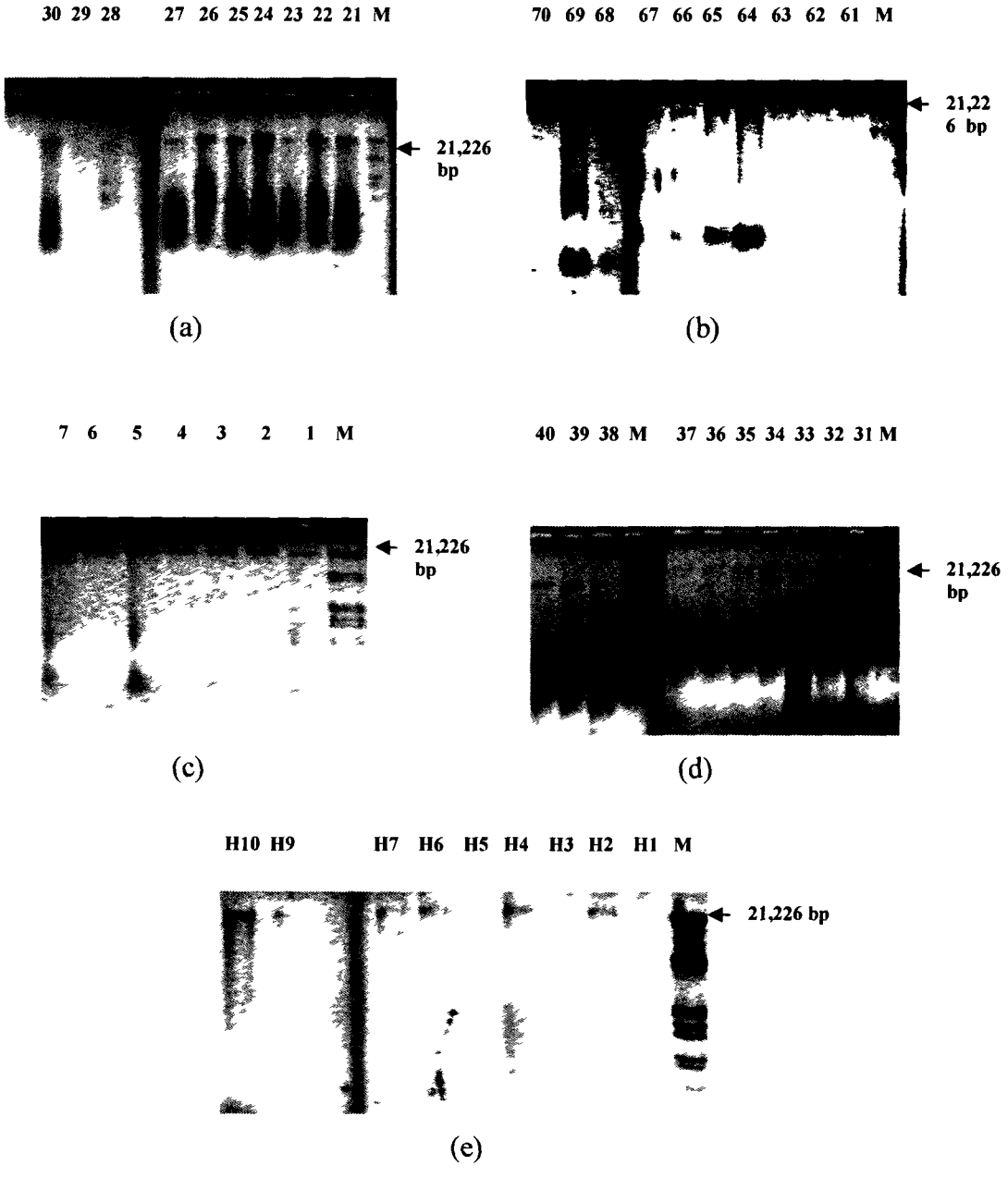

30 29 28 27 26 25 24 23 22 21 M
21,226 bp
(a)
70 69 68 67 66 65 64 63 62 61 M
21,22 6 bp
(b)
7 6 5 4 3 2 1 M
21,226 bp
(c)
40 39 38 M 37 36 35 34 33 32 31 M
21,226 bp
(d)
H10 H9 H7 H6 H5 H4 H3 H2 H1 M
21,226 bp
(e)

Fig.8.

Fig.9(a). Diagrammatic map showing organization of *rrn* operon in prokaryotes. The underlined segment of the gene marks the analyzed region. ⟶

Fig.9(b). Amplification of *rrn* ITS was done using forward primer FGPS 989ac and reverse primer FGPL 2054'. ⟶

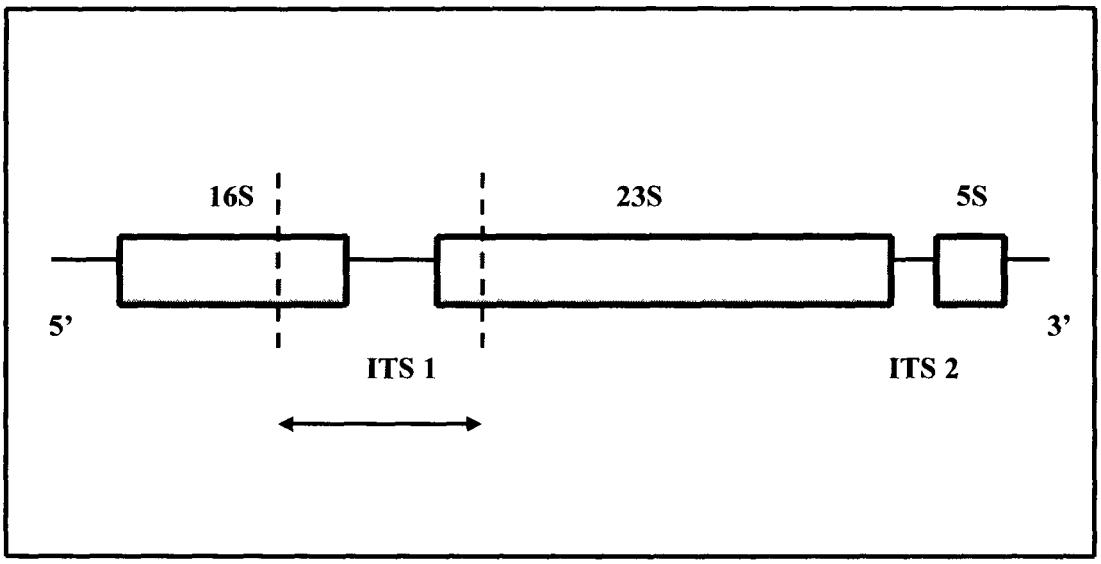


Fig. 9(a)

Primer	Sequence	Reference
FGPS 989ac	5'GGGGTCCGTAAGGGTC3	Bosco *et al.* 1992
FGPL 2054'	5'CCGGGTTTCCCCATTCGG3'	Simonet *et al.*, 1991

Fig. 9(b)

Fig.10. Diagrmmatic representation of the amplification reaction. ⟶

Diagram of the PCR cycle

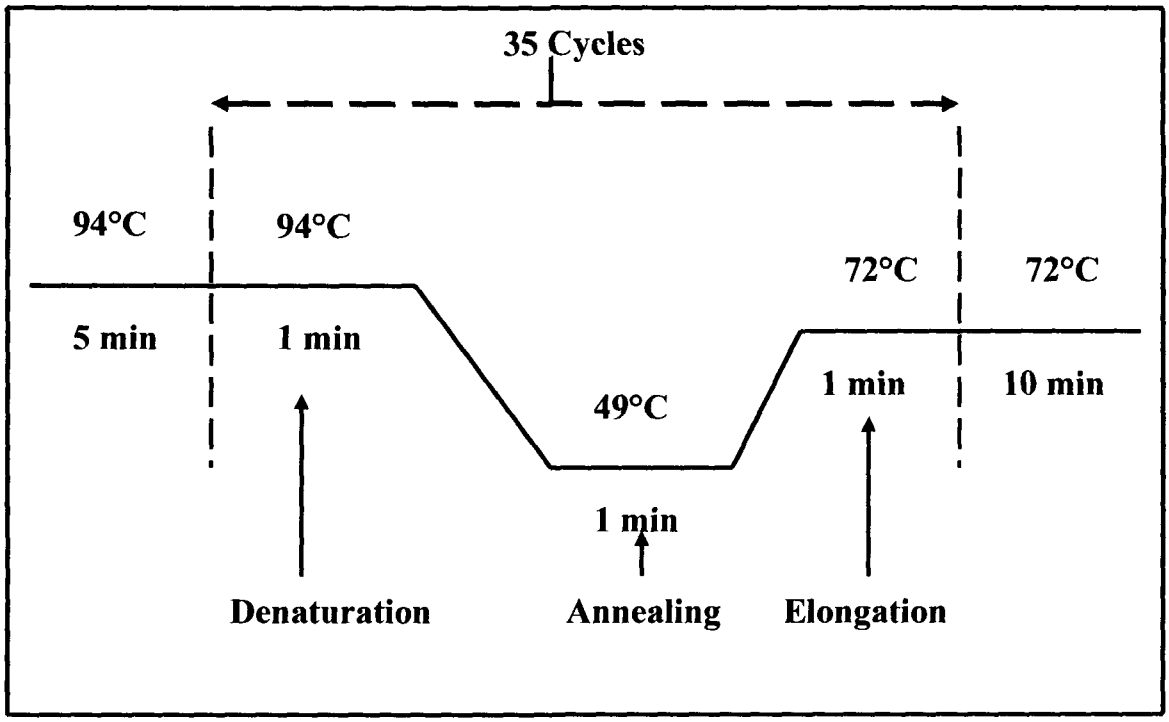
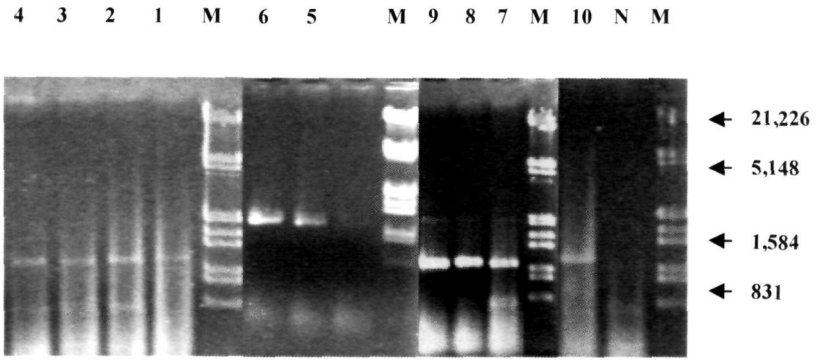
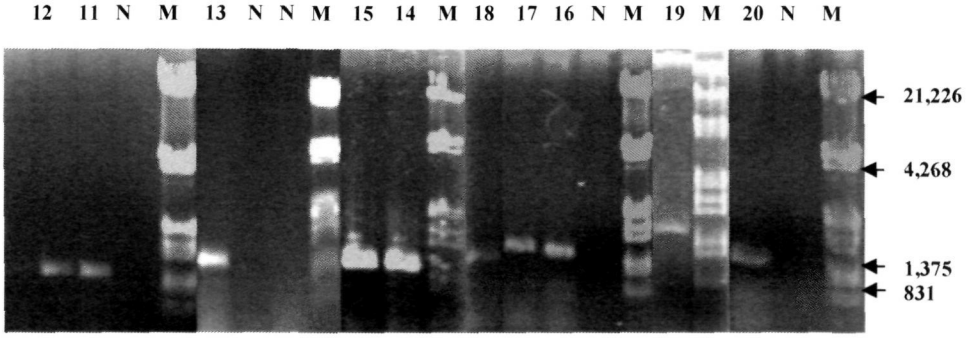


Fig.10

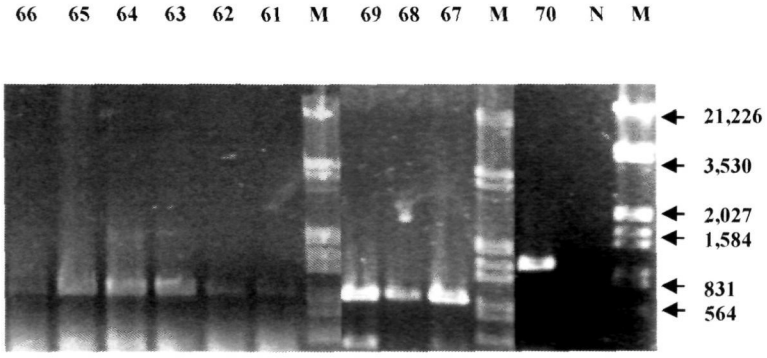
Fig.11. PCR amplification of 16S-ITS-23S rRNA region of *Frakia rrn* operon by primer pair-FGPS 989ac and FGPL 2045′. **(a) to (c)- Gel photographs showing amplified samples from high altitude sites. M- Molecular weight marker in base pairs (Lambda DNA, Eco R1+Hind III double digest). All lanes are marked with sample numbers loaded.** ⟶



(a) High Altitude, Site 1



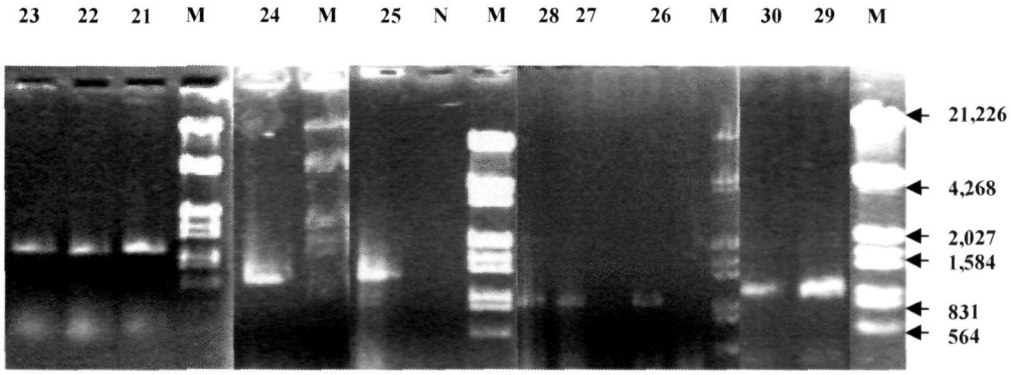
(b) High Altitude, Site 2



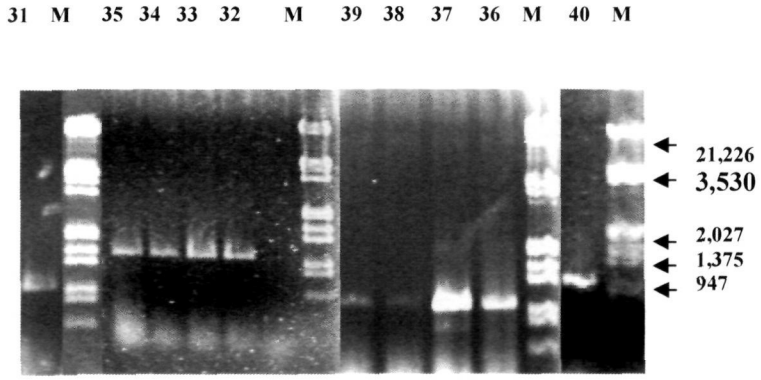
(c) High Altitude, Site 3

Fig.11

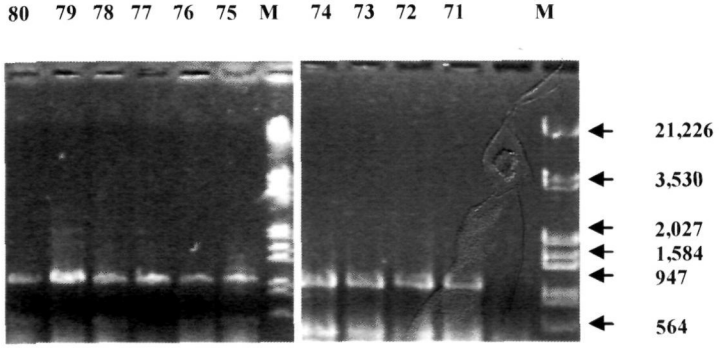
Fig.12. PCR amplification of 16S-ITS-23S rRNA region of *Frakia rrn* operon by primer pair-FGPS 989ac and FGPL 2045′. **(a) to (c)- Gel photographs showing amplified samples from middle altitude sites. M- Molecular weight marker in base pairs (Lambda DNA, Eco R1+Hind III double digest). All lanes are marked with sample numbers loaded.** ⟶



(a) Middle Altitude, Site 1



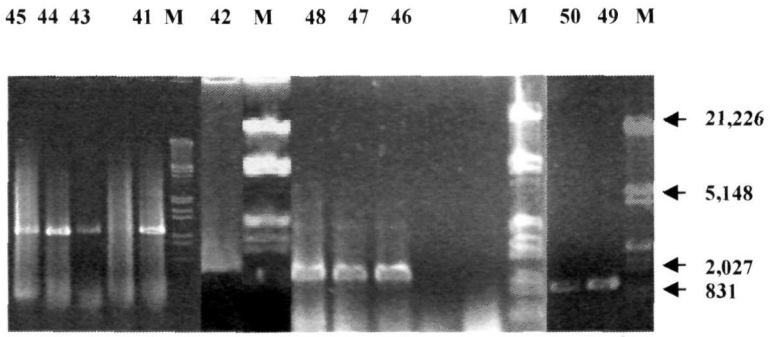
(b) Middle Altitude, Site 2



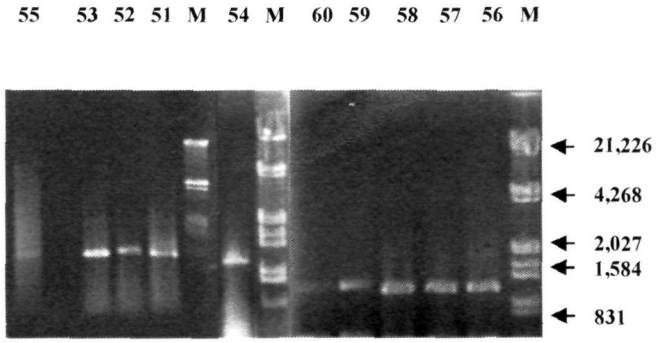
(c) Middle Altitude, Site 3

Fig.12

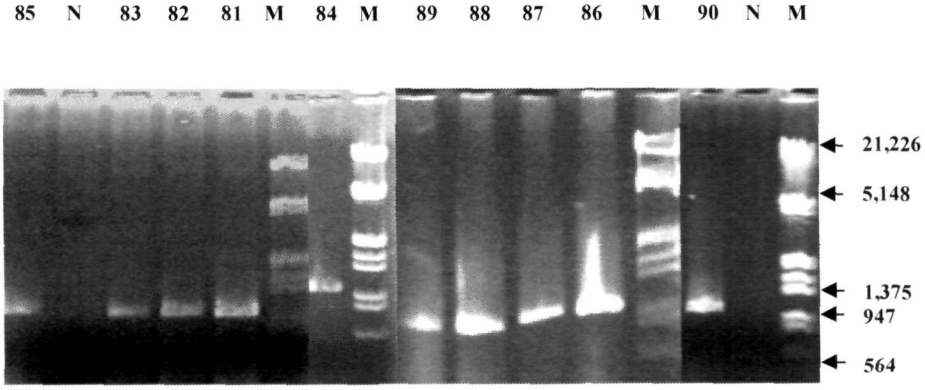
Fig.13. PCR amplification of 16S-ITS-23S rRNA region of *Frakia rrn* operon by primer pair-FGPS 989ac and FGPL 2045′. **(a) to (c)- Gel photographs showing amplified samples from low altitude sites. M- Molecular weight marker in base pairs (Lambda DNA, Eco R1+Hind III double digest). All lanes are marked with sample numbers loaded.** ⟶



(a) Low Altitude, Site 1



(b) Low Altitude, Site 2



(c) Low Altitude, Site 3

Fig.13

Fig.14. PCR amplification of 16S-ITS-23S rRNA region of *Frakia rrn* operon by primer pair-FGPS 989ac and FGPL 2045′. Gel photograph showing amplified samples from *Hippophae* sp. M- Molecular weight marker in base pairs (Lambda DNA, Eco R1+Hind III double digest). All lanes are marked with sample numbers loaded. ⟶

Fig.15. ALP gel photograph showing variations among amplicons of high altitude samples. M- Molecular weight marker in base pairs (Lambda DNA, Eco R1+Hind III double digest). All lanes are marked with sample numbers loaded. ⟶

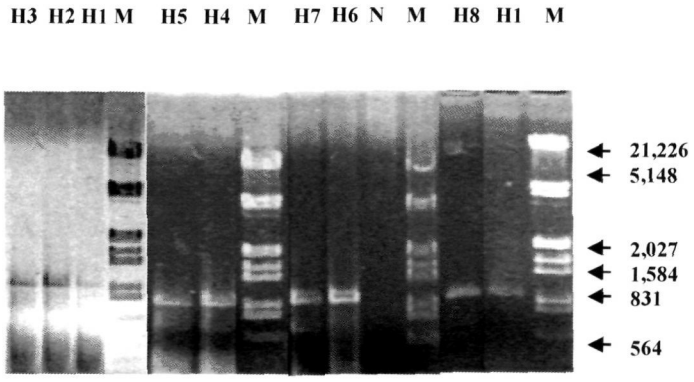

H3 H2 H1 M H5 H4 M H7 H6 N M H8 H1 M
21,226
5,148
2,027
1,584
831
564

Fig.14.

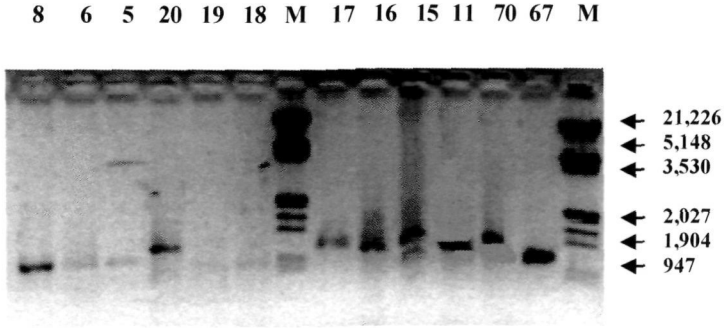

8 6 5 20 19 18 M 17 16 15 11 70 67 M
21,226
5,148
3,530
2,027
1,904
947

Fig.15.

profile (ALP)/amplicon fragment length profile (AFLP). This suggested for the presence of considerable variation among alder compatible Frankiae.

The molecular weight of the amplified band was estimated with the help of Multi-Analyst® software, and was found to be approximately 1200 bp (Fig.11 to 14). This included the distal part of the 16S rDNA, the ITS, which is about 420 bp, and the initial part of the 23S rDNA.

4.4 AMPLIFIED FRAGMENT LENGTH PROFILE (AFLP) OR AMPLICON LENGTH PROFILE (ALP).

Agarose gel electrophoresis of amplicons of different samples revealed differences in the number of amplified bands and their sizes. For examples, two bands were obtained in sample no. 15 of higher altitude, site 2. This indicated a size variation in the intervening region for the primer pair's annealing sites. This is not entirely unexpected as some deletions/ additions in the ITS region may not have any adverse impact on the organisms and may therefore get fixed. Further, *Frankia* is known to have two operons for *rrn* region (Normand *et al.*, 1992b). If an event of addition/ deletion in one of the operons occured, amplicons with different sizes would be generated. Gel photograph showing variations among amplicons of high altitude samples has been presented in Fig.15.

4.5 RESTRICTION DIGESTION OF AMPLIFIED PRODUCTS:

Greater genetic variability is found in the non-coding regions of a genome since they are subject to less selection pressure compared to the coding regions. This makes Amplicon Restriction Pattern (ARP) (also called as CAPS or PCR-RFLP) based fingerprinting method suitable for analyzing genetic diversity within any microbial

Sample numbers	P1	P2	P3	P4	P5	P6	P7	P8	P9	P 10	P 11	P 12	P 13	P 14	P 15
	1,2,3	4,5	6,7,8, 9,10	11 12 13 14 15 16 61 62 63 65 66 67 68 69	70	21,22 23,24 25,26 27,28 29,31 35,36 37,38 39,40	33, 34	32	71, 72, 74, 75	73, 76	77, 78	79, 80	41,42 43,45 47,50 52,53 54,56 57,58 60,82 83	44,46 48,49 51, 59 84 85 89	55
	P 16	**P 17**	**P 18**	**P 19**	**P 20**										
	81 86 90	88	H1 H3 H5 H7	H2 H4	H6										

Table 1.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
572 bp								+								+				
546 bp													+		+					
529 bp																				
518 bp												+								
467 bp																				
420 bp			+		+														+	
407 bp																		+		
396 bp						+														
382 bp															+					
366 bp																				+
342 bp																				
330 bp																			+	
297 bp																		+		
285 bp										+						+	+			+
254 bp			+																	
248 bp		+		+	+	+	+	+	+	+						+	+	+		
233 bp																				+
225 bp				+									+	+	+				+	
220 bp	+										+	+								
187 bp	+												+	+	+	+	+			
168 bp	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
161 bp				+																
150 bp			+	+					+	+	+	+								
144 bp	+						+	+										+	+	+
130 bp																+	+			
125 bp									+	+	+	+	+	+	+					
120 bp	+	+	+			+	+	+												
113 bp				+	+														+	

Table 2.

population. The ITS between 16S and 23S rRNA is one such region where the chances of detecting diversity are high.

As AFLP revealed variations in amplicon sizes, we decided to go in for ARP analysis to find out the variation at banding pattern level.

Restriction digestion of the amplicons was done site wise using the four base cutter restriction endonuclease *Rsa*I. This enzyme was chosen by mock digestion of the amplicon using MacVector® software. After restriction digestion, the samples were electrophoresed on a 4% agarose gel at 40V for 7-8 hrs. Different banding patterns emerged from the digested amplicons (Fig.16 to 19).

Restriction digestion of above amplicons with *Rsa*I revealed the presence of altogether **20** different patterns (Table 1).

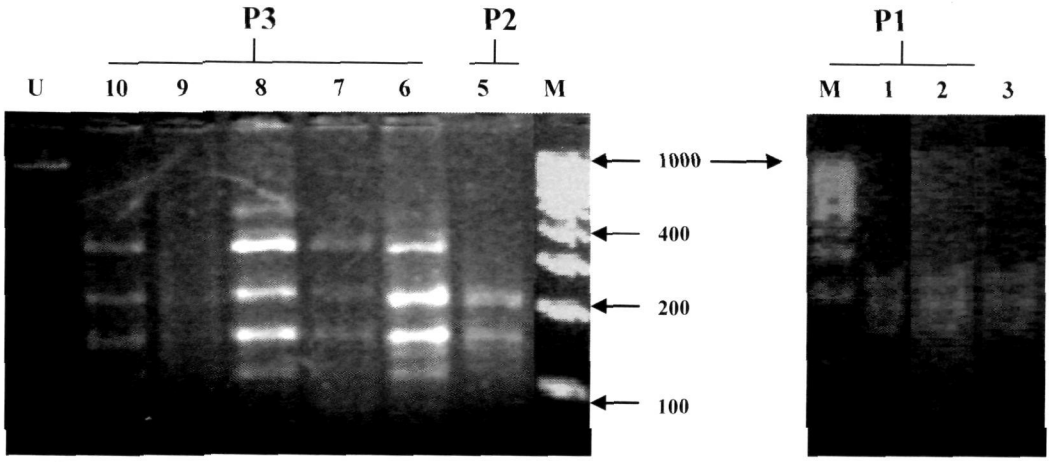
These patterns were classified on the basis of the presence or absence of different bands. The molecular sizes of the bands were determined by Multi Analyst® software or log graph paper (Section 3.4.3).

4.5.1 Patterns of high altitude, Site 1:

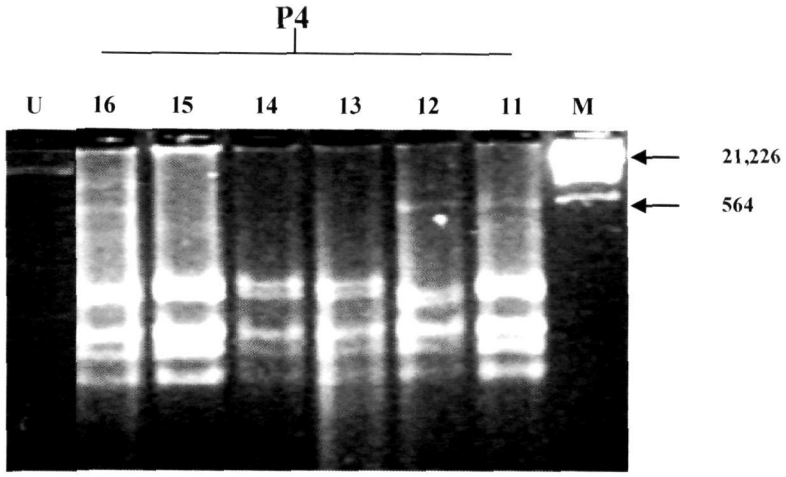
The patterns revealed by ARP after digesting the samples with *Rsa*I are presented in Fig.16a. From the visual observation of the gel and calculation of the band sizes using Multi Analyst® software, we were able to detect a total of three different patterns from this site.

Sample numbers 1, 2 and 3, represented a pattern (P1) with five bands. Sample numbers 4 and 5 represented a pattern (P2) with three visible bands in the gel, while sample numbers 6, 7, 8, 9 and 10 represented a pattern (P3) (see Tables 1 and 2).

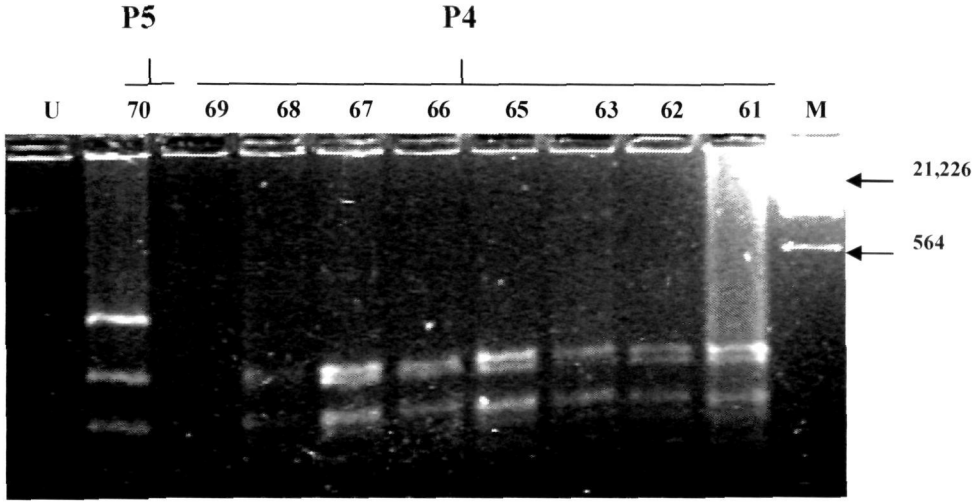
Fig.16. (a) to (c) showing RFLP gel photographs of Amplicon Restriction Patterns (ARPs) of amplified samples from high altitude sites. M – Molecular weight marker, Lambda DNA double digest for (b) and (c), 100 bp ladder for (a). All band sizes are in bp. U – Undigested sample run as control. All lanes are marked with sample numbers loaded. ⟶



(a) High Altitude, Site 1



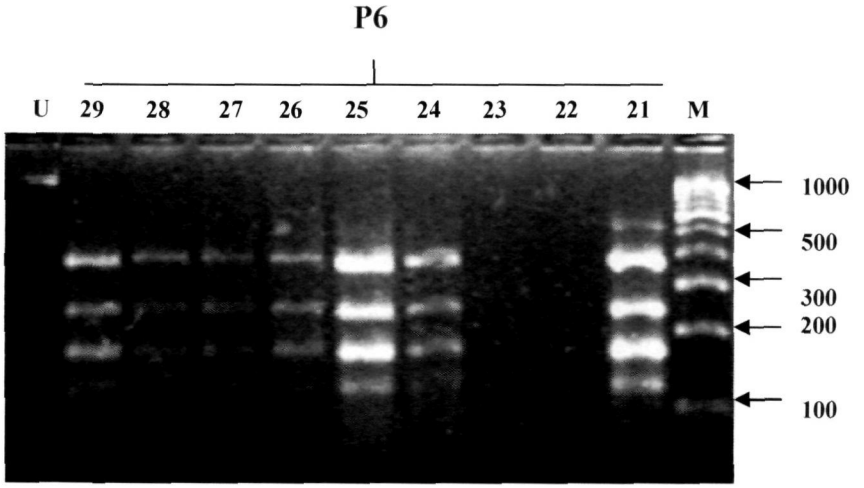
(b) High Altitude, Site 2



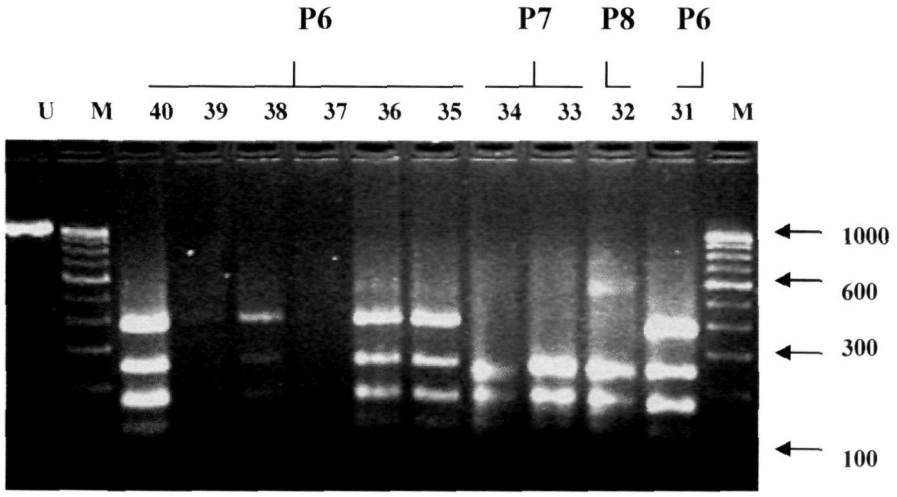
(c) High Altitude, Site 3

Fig 16

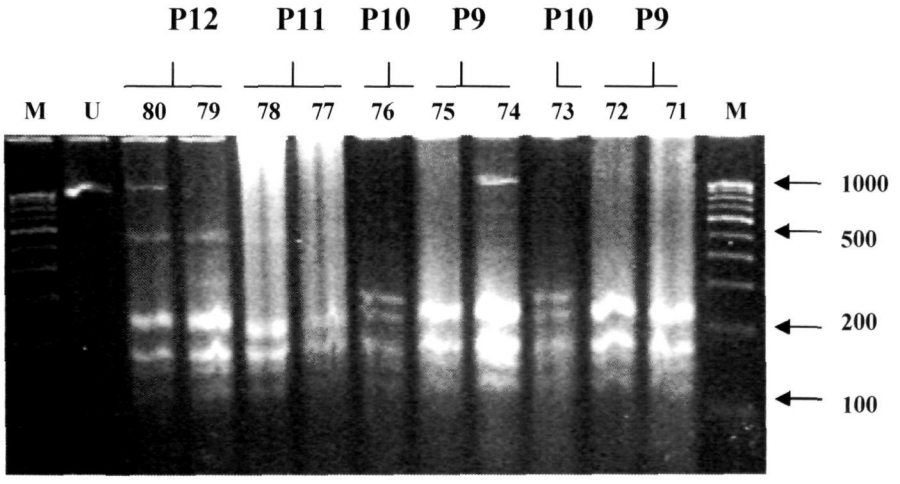
Fig.17. (a) to (c) showing RFLP gel photographs of Amplicon Restriction Patterns (ARPs) of amplified samples from middle altitude sites. M – Molecular weight marker, 100 bp ladder. All band sizes are in bp. U – Undigested sample run as control. All lanes are marked with sample numbers loaded. ⟶



(a) Middle Altitude, Site 1



(b) Middle Altitude, Site 2

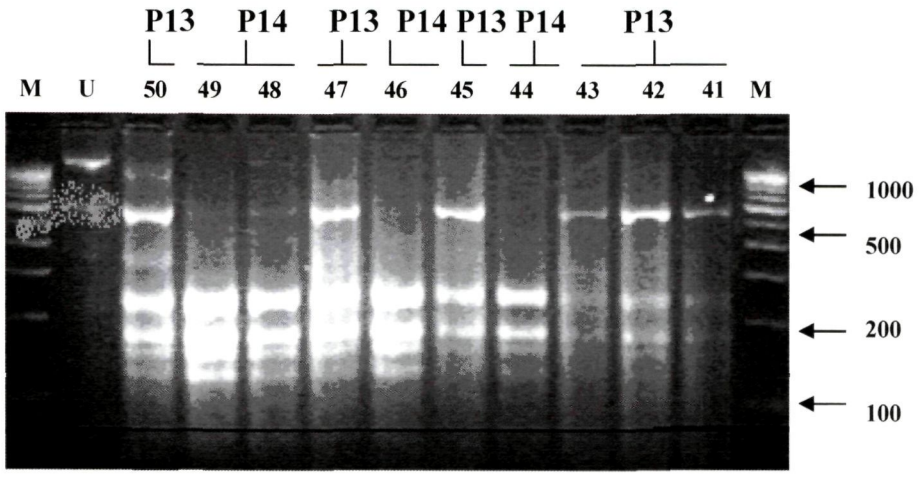


(c) Middle altitude, Site 3

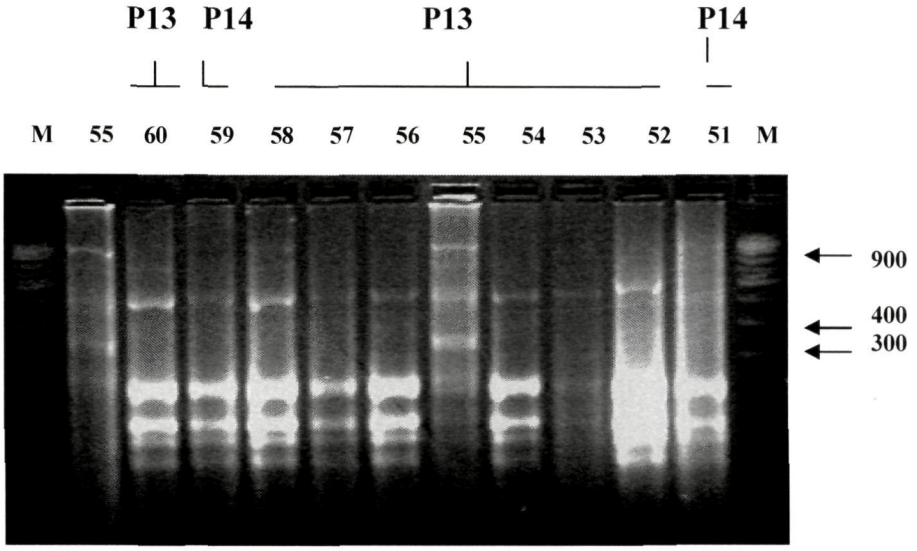
Fig.17

Fig. 18 ⟶

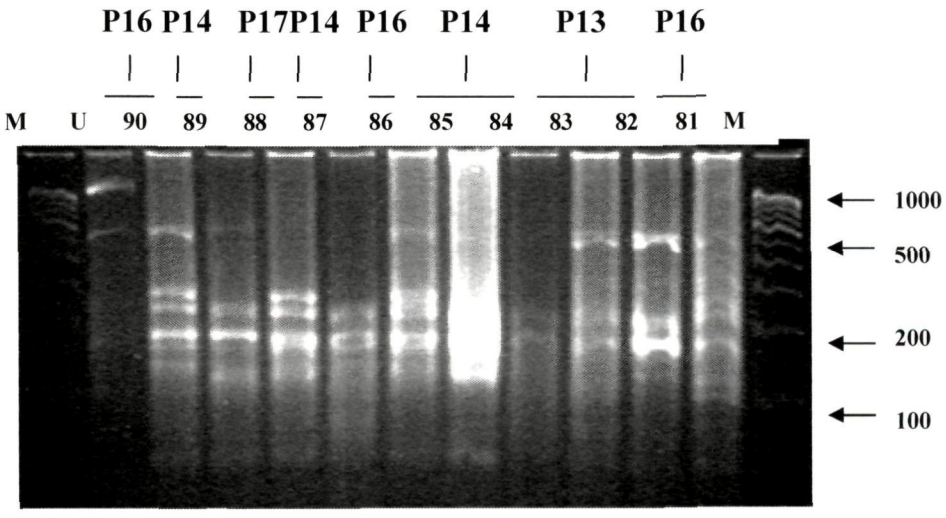
Fig.18. (a) to (c) showing RFLP gel photographs of Amplicon Restriction Patterns (ARPs) of amplified samples from low altitude sites. M – Molecular weight marker, 100 bp ladder. All band sizes are in bp. U – Undigested sample run as control. All lanes are marked with sample numbers loaded. ⟶



(a) Low Altitude, Site 1



(b) Low Altitude, Site 2



(c) Low Altitude, Site 3

Fig.18

Fig.19. RFLP gel photographs of Amplicon Restriction Patterns (ARPs) of amplified samples from *Hippophae* sp. M – molecular weight marker, 100 bp ladder. All band sizes are in bp. U – Undigested sample run as control. All lanes are marked with sample numbers loaded. ⟶

Fig. 19 ⟶

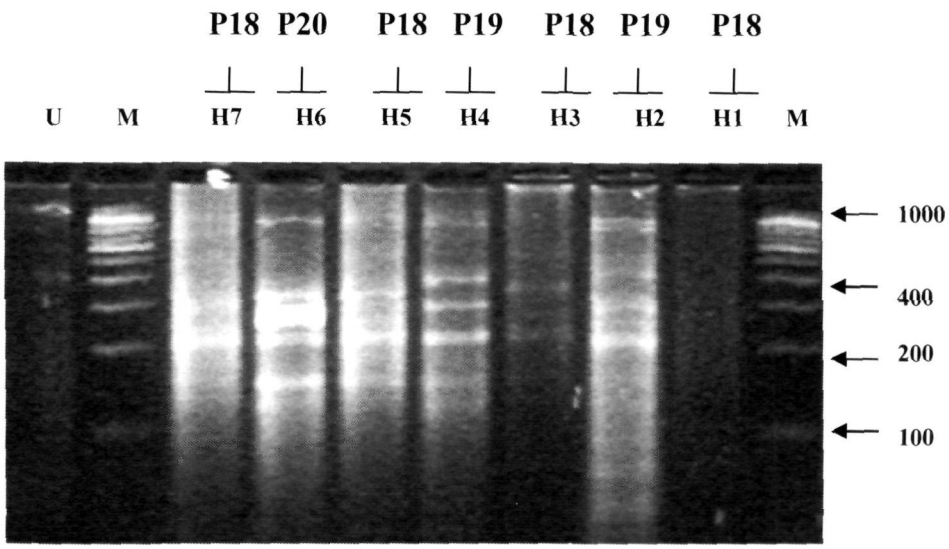

P18
P20
P18
P19
P18
P19
P18
U
M
H7
H6
H5
H4
H3
H2
H1
M
1000
400
200
100

Fig.19.

4.5.2 Patterns of high altitude, Site 2:

The patterns of this site are presented in Fig.16b. Only one banding pattern was observed at this site. Sample numbers 11, 12, 13, 14, 15 and 16 were found to have six bands and were grouped as pattern P4. Sample numbers 17, 18, 19 and 20 were rejected when nucleotide sequencing confirmed them to be contaminating DNAs (see Section 4.8 below).

4.5.3 Patterns of high altitude, Site 3:

Fig.16c shows the patterns obtained on digestion of the samples from this site. Here we could detect two different patterns of which the pattern represented by sample numbers 61 to 69 were found similar with pattern P4 of High altitude, site 2. The sample 70 showed four distinct bands and was grouped separately as P5.

4.5.4 Patterns of middle altitude, Site 1:

Fig.17a shows the patterns revealed for this site. Here we could detect only one pattern (P6), which was represented by sample numbers 21 to 29 with four clear visible bands. On nucleotide sequencing, sample number 30 was found to be contaminating DNA and was not considered for this study.

4.5.5 Patterns of middle altitude, Site 2:

Three different patterns were found at this site (Fig. 17b). Sample numbers 31, 35, 36, 37, 38, 39 and 40 were found to have profile P6, also found present at middle altitude, site 1. Sample numbers 33 and 34 showed four bands each, and were grouped as pattern P7. Sample number 32 was found to be unique and was grouped separately as P8.

4.5.6 Patterns of middle altitude, Site 3:

Here four different patterns could be detected on the basis of number and sizes of the bands (Fig. 17c).. Sample numbers 71, 72, 74, 75 were found to have four bands each

and were grouped as P9. Sample numbers 73 and 76 were similar with five bands and were grouped as P10. Sample numbers 77 and 78, showing four bands were grouped as P11 while sample numbers 79 and 80 with five bands each were grouped as P12.

4.5.7 Patterns of low altitude, Site 1:

Two different kinds of banding patterns were found at this site (Fig.18a). Sample numbers 41, 42, 43, 45, 47 and 50 showed similar profiles with five bands each and were considered as pattern P13. The sample numbers 44, 46, 48 and 49 were found to have similar profiles with four bands each and were grouped as P14.

4.5.8 Patterns of low altitude, Site 2:

Three patterns were detected for this site (Fig. 18b). Sample numbers 52, 53, 54, 56, 57, 58 and 60 belonged to pattern P13, also found at low altitude, site 1. Sample numbers 51, and 59 were found similar to P14 of low altitude, site 1. Only sample number 55 was different and unique and was designated as pattern P15.

4.5.9 Patterns of Low altitude, Site 3:

Here four different patterns could be detected (Fig.18c). Sample numbers 81, 86 and 90 with six visible bands were different from all others and were considered as P16. Sample numbers 82 and 83 were found to have profiles similar to P13. Sample numbers 84, 85, 87 and 89 were found to have four bands similar to P14. Sample number 88 alone represented pattern P17 with five bands.

4.5.10 Patterns of microsymbionts associated with *Hippophae* sp.

Interestingly, these samples showed (Fig.19) a considerable variation in banding patterns even within a very small population. Sample numbers H1, H3, H5 and H7 were found similar with five faint bands (P18) while, H2 and H4 were grouped as P19 with

five comparatively strong bands and H6 was found to be different from others and was considered as P20.

4.6 Analysis of the banding patterns on the basis of ARP / PCR-RFLP:

From the distribution of different patterns it was noticed that some sites had one or more banding patterns, which were exclusively specific to that particular site, i.e. some patterns appeared to be confined to a given site only. For example -

i. P1(1,2,3),P2 (4,5) and P3 (6,7,8,9,10) from high altitude, site 1

ii. P5 (70) from high altitude, site 3

iii. P7 (33,34) and P8 (32) of middle altitude, site 2

iv. P9 (71,72,74,75), P10 (73,76), P11 (77,78), P12 (79, 80) of middle altitude site 3

v. P15 (55) of low altitude, site 2

vi. P16 (81, 86, 90) and P17 (88) of low altitude, site 3

It was interesting to note that many of the genotypes found at site 3, located in north Sikkim, were not found in east Sikkim areas (sites 1 and 2). Clearly, geographic isolation was significant contributor to distribution of *Frankia* genotypes. In general, however, the distribution of genotypes was across the sites. It is likely that the genotypes that tended to be confined to a specific site, had evolved recently and had not yet moved to newer locations.

There were other patterns that seemed to have an altitude dependent distribution i.e. they were not confined to a particular site, but were found only in a specific altitude zone across the sites. For example-

i. High altitude - P4 (sites 2 and 3).

ii. Middle altitude - P6 (sites 1 and 2).

iii. Low altitude - P13 (sites 1 and 3) and P14 (sites 1, 2 and 3).

It seems that altitude plays a more significant role in determining the distribution of the genotypes. Altitude dependent climatic and soil factors were possibly responsible for distribution of more adapt genotypes at different altitudes.

We were constrained by the availability of *Hippophae* in high altitude areas only. Therefore, a comparative study of altitudinal distribution of *Frankia* found in *Hippophae* root nodules was not possible. However, the fact that we could identify at least three different Amplicon Restriction Patterns (ARPs) among the ten samples studied, indicated presence of substantial variability with respect to *Frankia* nodulating *Hippophae*.

4.7 Cluster Analysis:

Application of computer aided methods to image analysis of electrophoretic gels not only ensures accurate and reliable quantification of nucleic acids, but also enables statistical analyses of volumes of data (Spirovski *et al.*, 2005). One such application is the cluster analysis based on similarity measurements extensively used in population study. The term cluster analysis (first used by Tryon, 1939, to organize observed data into meaningful structures) encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. Clustering can thus be described as finding "natural groupings" in a set of data.

There are several ways to do this, but in general clustering methods can be divided into hierarchical (often called agglomerative or joining) and partitioning (also called divisive) methods.

In hierarchical clustering, first, individual items are joined to each other, and then the groups to each other, so that the result is a tree of cluster associations. In this tree, the different branches are the clusters.

In the present study, cluster analysis was performed on the banding patterns of all the samples which were confirmed to be *Frankia,* and that emerged from the ARP analysis to find out relationship among different strains. For this purpose, NTSYS (version 2.1) was used. The cluster dendrogram (Fig.20) was constructed using the method SAHN (sequential agglomerative hierarchical nested cluster analysis) and it was based on a similarity matrix which in turn was based on Jaccard's coefficients.

The relationship among different genotypes based on their banding patterns showed three broad clusters, I, II and III.

Cluster I grouped 33 samples representing eight different patterns into four sub clusters IA, IB, IC and ID, which were further divided into minor clusters. Cluster II included 50 samples representing nine patterns into two major sub clusters IIA and IIB, which were again divided into minor clusters. Cluster III represented *Frankia* nodulating *Hippophae* sp.

The cluster analysis revealed that RFLPs/ ARPs generally clustered along the lines of altitudes. Clusters 1A and 1C grouped high altitude samples. Cluster 1B had all the samples belonging to middle altitude, except two samples (4 & 5 representing pattern 2) from high altitude. Cluster 1D grouped samples from the lower altitude. Branch 'i' of cluster IIA grouped samples from high altitude, while branch 'ii' grouped samples belonging to middle altitude, except for sample 88 belonging to lower altitude and representing pattern P17. Cluster IIB grouped together samples distributed at middle and lower altitudes, but not at higher altitude. Therefore, it appears that altitude had a significant role in determining the distribution of *Frankia* genotypes.

Cluster III grouped together *Frankia* samples found associated with *Hippophae*. All of them belonged to high altitude. However, it was not possible to determine the role of

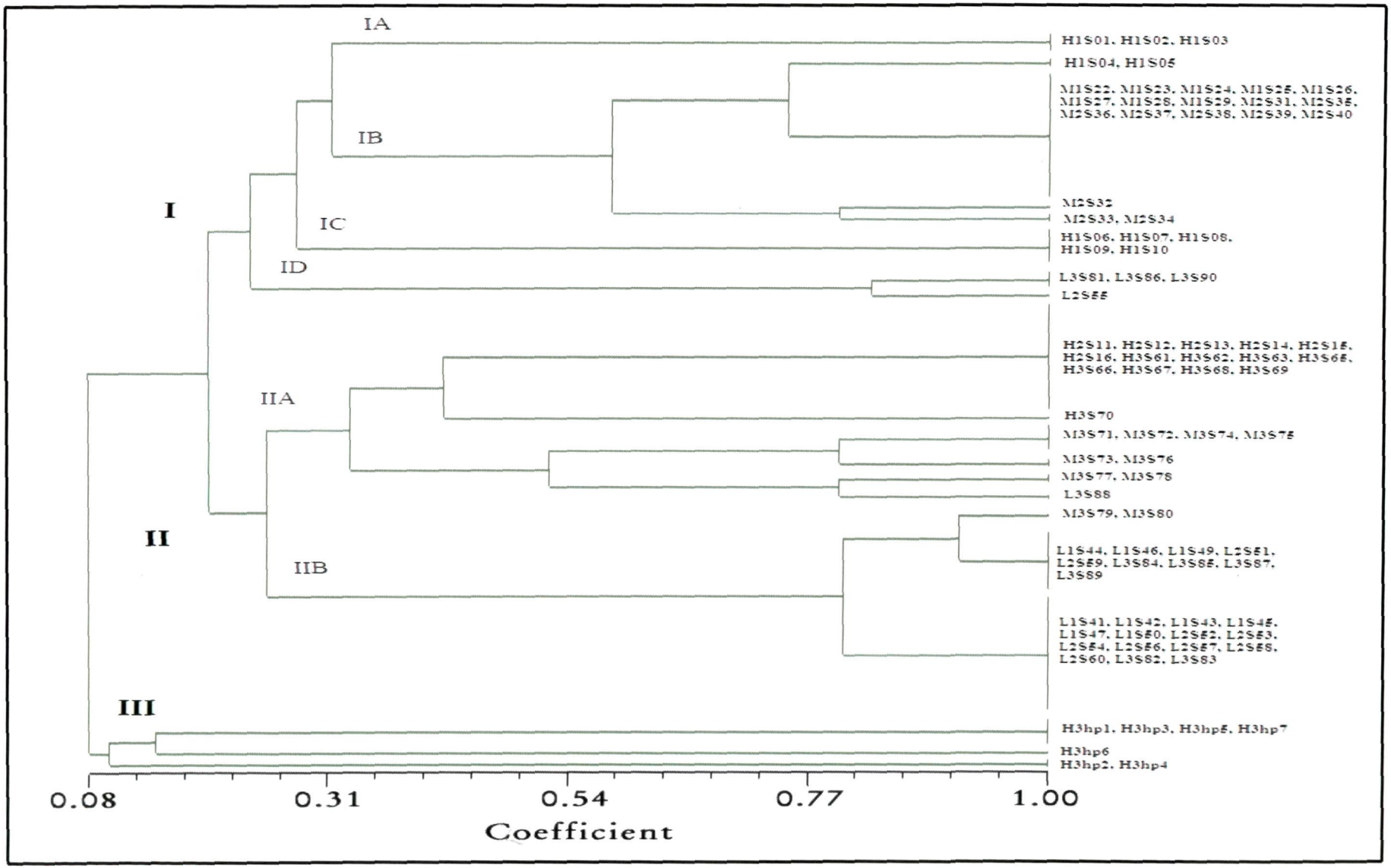
Fig.20. Cluster dendogram ⟶

I. For *Alnus nepalensis* samples-

H1S01 – High altitude, site 1, sample 1
M2S35 – Middle altitude, site 2, sample 35

II. For *Hippophae* sp. –

H3hp1 – High altitude, site 3, sample 1
H3hp6 – High altitude, site 3, sample 6


IA
IB
IC
ID
I
IIA
IIB
II
III
H1S01, H1S02, H1S03
H1S04, H1S05
M1S22, M1S23, M1S24, M1S25, M1S26, M1S27, M1S28, M1S29, M2S31, M2S35, M2S36, M2S37, M2S38, M2S39, M2S40
M2S32
M2S33, M2S34
H1S06, H1S07, H1S08, H1S09, H1S10
L3S81, L3S86, L3S90
L2S55
H2S11, H2S12, H2S13, H2S14, H2S15, H2S16, H3S61, H3S62, H3S63, H3S65, H3S66, H3S67, H3S68, H3S69
H3S70
M3S71, M3S72, M3S74, M3S75
M3S73, M3S76
M3S77, M3S78
L3S88
M3S79, M3S80
L1S44, L1S46, L1S49, L2S51, L2S59, L3S84, L3S85, L3S87, L3S89
L1S41, L1S42, L1S43, L1S45, L1S47, L1S50, L2S52, L2S53, L2S54, L2S56, L2S57, L2S58, L2S60, L3S82, L3S83
H3hp1, H3hp3, H3hp5, H3hp7
H3hp6
H3hp2, H3hp4
0.08
0.31
0.54
0.77
1.00
Coefficient

altitude, if any, in determining the distribution of *Frankia* associated with *Hippophae* because of non-availability of this plant at middle and lower altitudes.

4.8 Nucleotide sequence analysis:

Since it was not possible to sequence all the hundred samples, only selected samples from each site were sequenced. One or two samples representing each pattern were considered for this purpose. The number of samples, which could be finally sequenced, is given in Table 3. However, we could not sequence samples from low altitude site 1 and site 3 despite repeated attempts.

After receiving all the nucleotide sequences, they were individually fed into the GeneBank for BLAST (Basic Local Alignment Search Tool) search using the website http://ncbi.nlm.nih.gov. Except for sample numbers 17, 18, 19, 20 and 30, all other sequences blasted as *Frankia*.

These retrieved sequences along with the original sequences were aligned using multiple sequence alignment program CLUSTAL W (1.75) (Thompson *et al.*, 1994). Sequences were also aligned for each site, which gave a much clearer picture of particular strains and their banding patterns.

Meanwhile, individual sequences were also cut with *Rsa*I with the help of online tool Webcutter (2.0) [http://www.carolina.com/webcutter/carolina.asp]. This information was compared with the previous information we got through ARP analysis (Tables 4 to 8). Apparently, grouping of isolates on the basis of ARP was very effective since only few strains from different altitudes that showed almost similar ARP had differences at the sequence level.

Samples sequenced:

	Site 1	Site 2	Site 3
	Sample numbers		
High altitude	1, , 3, 4, 5, 6, 8, 8c	11, 13, 15, 16	65, 70
Middle altitude	21, 22, 24, 28, 29, 30	31, 32, 33, 36, 39	73, 75
Low altitude	N.S	58	N.S

Table 3.

4.8.1 High altitude, Site 1

From this site sample numbers 1, 4, 6 and 8 were sequenced. A comparison of these sequences revealed that their actual amplicon sizes differed slightly from each other. This is in line with our observations on AFLP/ALP.

Webcutter could identify six cutting sites on sequence no.1, producing seven bands (Table 4). However, only five bands were visible in the gel. Probably two bands of 169bp and 163bp merged together and appeared as a single 168bp band. The predicted 54 bp band could not be detected in the gel. Similarly for sample number 4, 169 bp and 164 bp bands got merged together and appeared as 164 bp band, and for sample number 6, 264 bp and 239 bp got merged together to appear as 254 bp band. All the five bands could be detected in the gel for sample number 8 (Fig.21). In high altitude site 1 samples, four cutting sites were found to be highly conserved, as they were present in all the sequences. Although very small differences were noticed between them (few base pairs), they always aligned together. Sites B, C, D and F were considered more conserved compared to sites A and E, which were not present in all the samples. The sequence alignment of this site revealed high sequence conservation, with minor substitutional variations and cases of insertion/deletions at a few positions.

- At positions 185 and 186 respectively, two bases 'G' and 'C' were inserted in sample 4.
- In the same sample another purine base 'A' was inserted at position 220.
- At position 231, pyrimidine base 'C' was found inserted in sample 1 and purine base 'A' in sample 4.
- At position 240 and 241 two pyrimidine bases 'C' and 'T' were detected in sample 4.

Fig.21. Diagrammatic representation of the sequences of High altitude, site 1, along with their restriction sites with *Rsa* 1. (Distances not in scale).⟶

Table.4. Comparison between RFLP and webcutter generated banding patterns of High altitude, site 1 ⟶

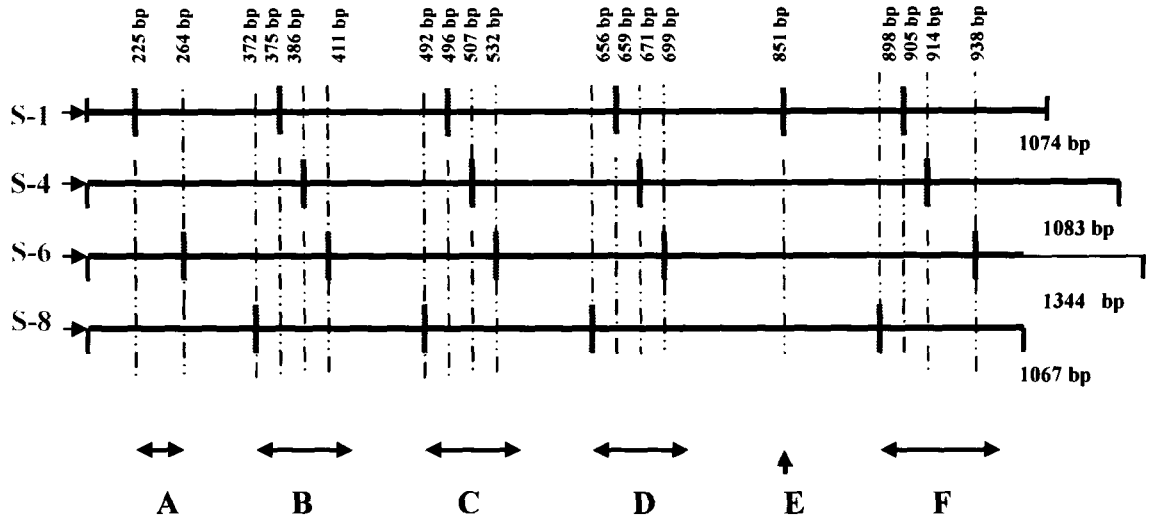


Fig.21

Sample number	Bands visible in the gel (bp)	Bands produced by Webcutter (bp)
1	218	225
	187	192
	168	169
		163
	144	150
	121	121
	N.V	54
4	N.V	386
	246	243
	164	169
		164
	N.V	121
6	420	406
	254	264
		239
	164	167
		147
	120	121
8	420	372
	254	242
	164	169
		164
	120	120

Table.4

- At position 274, purine base 'A' was found in both samples 1 and 4 but was absent in sample 8 and 6.
- Purine base 'A' was again found inserted in sample 4 at 292 and in sample 1 at 306.
- Three consecutive bases – 'C', 'G' and 'T' were found inserted in sample 4 at positions 316, 317 and 318, which were absent in all other samples.
- Purine base 'A' was found at positions 337, 356 and 369 in sample 4, and was absent in all other samples
- At position 399, one 'C' base was found inserted in sample 4.
- One purine base 'A' was found at position 430 in sample 1 but was absent in all other samples.
- In sample 6, one pyrimidine base 'C' was found at position 476, but was absent in other samples.
- At position 553 in sample 6 base 'C' was substituted for base 'T'.
- At 561, another base pair substitution was detected (transition between 'C' and 'T') but, since out of four samples, two were having 'C' and the other two 'T', it was not possible to conclude which base got substituted.
- Addition was found in sample 6 at positions 568 and 577, where 'T' and 'G' were inserted.
- Position 578 showed substitution (transition) of base 'T' by base 'C' in sample 8.

- Position 598 in sample 6 was also a case of substitution (transition) between two purine bases 'G' and 'A'. This was repeated in sample 8 at 601, where base 'G' present in other samples was substituted by base 'A'.
- Sample 6 showed case of insertion at position 638 where base 'A' was added, and in the same sample at 640, base 'G' was substituted by base 'A'.
- At position 651, in sample 8, base 'G' was substituted for base 'A'.
- At 657 and 660, in sample 6, base 'T' and base 'C' were substituted by base 'C' and 'T' respectively.
- There was deletion at position 662 in sample 1 and substitution (transversion) in sample 6, where base 'G' was replaced by base 'C'.
- Three consecutive nucleotides were substituted at positions 670, 671 and 672 in sample 6, where bases 'T', 'G' and 'C' were replaced by 'C', 'T' and 'T'. The 1st and 3rd replacements were transitions (between two pyrimidine bases) and the 2nd replacement was a case of transversion (between one purine and one pyrimidine).
- Base 'T' was substituted by base 'C' at 686 in sample 6 and base 'G' was substituted by base 'T' at 688 in sample 1, indicating cases of transition and transversion respectively.
- At position 691, an insertion took place as base 'T' was found in sample 6 and was absent in all others.
- At 709, transition occurred in sample 6, where base 'T' was found inserted in place of base 'C' in other samples. The same happened at position 715 and 717, but in the 1st case it was not clear as to which base got substituted (it was

transition between 'G' and 'A') but in the 2nd case, base 'G' present in other samples was substituted by base 'A' in sample 1.

- At position 719, again transition took place between 'C' and 'T' while at 720, base 'T' present in other samples was substituted by base 'G' (transversion) in sample 6.
- Position 734 and 763 showed cases of transition between purine bases, i.e. 'G' and 'A'.
- Insertion took place at position 766 where base 'T' was added in sample 1.
- Transitions occurred at 771 and 772 in sample 1, where base 'T' was substituted by 'C' and 'C' by 'T'. This could be a di-nucleotide inversion.
- Position 773 showed a case of transversion as base 'G' got substituted by base 'T' in sample 6. At position 774, transition was found between 'G' and 'A'. Position 775 showed interesting features, as deletion was noticed in sample 6 and transversion (from 'C' to 'G') was found in sample 1.
- Position 776 showed transition from 'G' to 'A' in sample 1 and position 777 saw a substitution of base 'C' by base 'T' (transition) in sample 6.
- At position 793 and 795, transition and transversion occurred in sample 6 as base 'T' was replaced by base 'C' in the first case and by base 'G' in the second case. Transversion also was found at position 796 between bases 'G' and 'C'.
- Deletions were detected at positions 814, 815 and 816 where three nucleotides ('C', 'T' and 'A') were deleted in sample 6. At position 816, base 'A' of sample 8 and 4 was substituted by base 'G' in sample 1, which was a case of transition.

- Positions 818 and 822 were also showing transition between purine bases 'G' and 'A', but at 818, it was not clear which base was substituted, while at 822, it was base 'A' in sample 6 that substituted base 'G' present in other samples.
- Positions 835 and 836 were cases of addition and deletion respectively. At 835 purine base 'G' was inserted in sample 4 while in sample 6 at 836, pyrimidine base 'T' was deleted.
- Transition occurred at position 837, where base 'C' present in other samples was substituted by base 'T' in sample 6.
- Position 840 was also a case of transition between pyrimidine bases 'C' and 'T' but it was unclear as to which base got substituted.
- Transversion occurred at position 854, where base 'G' (purine) present in other samples was substituted by base 'C' (pyrimidine) in sample 1. Immediately next to it at position 855, a case of insertion was found in the same sample (1) where an additional pyrimidine base 'T' was found.
- Transitions were found at positions 863, 864, 870 and 877. In the first, third and fourth cases, it was a substitution of base 'T' by base 'C' in sample 1, base 'A' by base 'G' in sample 1 and base 'C' by base 'T' in sample 6 respectively. At position 864 though it was a case of transition, it was not clear which base got substituted.
- At position 879, both transition and transvertion were noticed as base 'T' was replaced by base 'C' in sample 1 and by base 'G' in sample 6.
- Transitions were found again at positions 880, 881 and 883. In the first and second cases, it was unclear as to which base got substituted ('C'/ 'T'), but at 881, it was base 'T' that was substituted by base 'C' in sample 6. At position

884, either deletion or insertion occurred but again it could not be detected whether base 'T' of sample 1 and 6 were additions or some bases were deleted from sample 8 and 4. Position 886 and 888 were cases of transversions as base 'T' (pyrimidine) present in other bases was substituted by base 'G' (purine) in sample 6, while at 888, base 'A' present in all other samples was substituted by base 'C' (pyrimidine) in sample 6. Transition between two purines was noticed at position 887 in sample 6 as base 'G' found in other samples was replaced by base 'A'.

- Base 'G' was found inserted at position 892 in sample 1.
- At positions 898 and 919, though transition (between 'C' and 'T') occurred, it was not possible to find out which base was substituted. Position 899 was also transition as base 'C' was substituted by base 'T' in sample 6. Position 916 and 917, both were transversions as base 'T' (pyrimidine) got substituted with base 'G' (purine) in sample 6 and 1 respectively.
- Though position 1027 showed a case of transition, it could not be detected as to which base was substituted, as sample 6 and 1 showed bases 'G' while sample 8 and 4 showed base 'A'.
- Position 1032 was a case of transversion as base 'G (purine) present in other samples was replaced by base 'T' (pyrimidine) in sample 6.
- Position 1051 was showing transition (between 'C' and 'T') but again it was not clears as to which base got substituted.
- Transversions were found at positions 1054 and 1057 in samples 6 and 1, where both the purine bases 'A' and 'G' got substituted by pyrimidine base 'T' respectively.

To summarise the above observations, it was found that the distal part of 16S rRNA gene was highly conserved as no differences were observed in the initial part of the amplicon. There were several single nucleotide changes (SNP) especially in the middle part of the amplicon. Two specific regions with high density of SNPs were between 814-822 and 863-899.

4.8.2 High altitude, Site 2

From this site, sample numbers 11, 13, 15 and 16 were sequenced. The total amplicon length for sample number 11 and 13 was 1069 bp and for sample 16, it was 1034 bp. Sample number 15 could be sequenced only partially (521 bp), thus for the rest of its length, the actual information was not available. But all these four sequences blasted as *Frankia.*

Webcutter could identify altogether five cutting sites along all these sequences except sample number 16 (Fig.22), which, showed a similar pattern with the rest of the samples in the gel, but produced a very different kind of pattern when cut by Webcutter. For sample number 11, 13 and 15, these five cutting sites seemed conserved, as they were present in all the patterns. In sample number 15, although only three cutting sites (G, H and I) were recognized by the Webcutter because of its 521 bp length, we assumed it to be the same as sample numbers 11 and 13 for the rest two cutting sites, since difference was always found in the first cutting site i.e. G (either present or absent). For these three samples, all the five cutting sites aligned together, although there were minor differences (few base pairs) among them.

Webcutter could identify five cutting sites producing six bands, which were visible in the gel (Table 5).

Fig.22. Diagrammatic representation of the sequences of High altitude, site 2, with their restriction sites with *Rsa* 1. (Distances not in scale). ⟶

Table.5. Comparison between RFLP and webcutter generated banding patterns of High altitude, site 2 ⟶

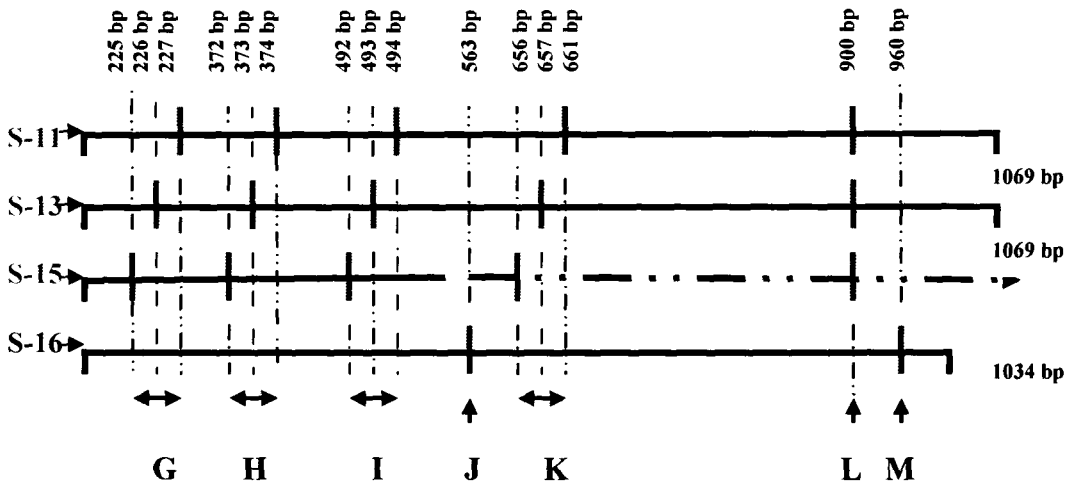


Fig.22

Sample number	Bands visible in the gel (bp)	Bands produced by Webcutter (bp)
11	246	239
		227
	164	169
		167
	150	147
	113	120
13	246	226
		243
	164	169
		164
	150	147
	113	120
15	246	
		225
	164	
	150	147
	113	120
16	246	563
	164	397
	150	74
	113	

Table.5

Sequence alignment of this site also revealed minor variations, which were restricted only to a few base pairs.

- At positions 23 and 50, sample 11 showed cases of insertion as pyrimidine base 'C' was found added and was absent in all other samples.
- Position 161 was a case of insertion again, but it was sample 13 where purine base 'A' was found while it was absent in other two samples.
- Positions 244 and 248 were cases of transitions as in the first case base 'G' (purine) of samples 13 and 11 were substituted by base 'A' (purine) in sample 15 and in the second case base 'T' (pyrimidine) was substituted by base 'C' (pyrimidine) in the same sample 15.
- At position 517, transversion occurred as base 'A' (purine) present in sample 13 and 11 was substituted by base 'C (pyrimidine) in sample 15.

As sample 15 could be sequenced only upto 521 bp, further comparisons were possible only between samples 13 and 11.

- Position 553 and 554 were cases of dinucleotide inversion (TG in sample 11 and GT in sample 13).
- Positions 556 and 557 can be cases of either addition or deletion for any of the two samples 13 and 11 as bases 'T' and 'G' of sample 11 were found missing in sample 13 . A case of possible transition was found at position 583 where base 'G' was found in sample 13 and base 'A' in sample 11.
- Position 603 was showing a case of transversion as base 'G' was noticed in sample 13 while base 'T' was present in sample 11. A case of either deletion or insertion occurred at 623 as base 'A' was present in sample 11 and it was absent in sample 13.

- Positions 624, 630, 639, 642 and 655 showed cases of transitions (purines at 625 and pyrimidines at 630, 639, 642 and 655).
- Transversion were detected at positions 647 and 656 where 'G' was found in sample 13 and bases 'C' and 'T' were found in sample 11.
- At position 675, it was again a case of transversion as pyrimidine 'C' was present in sample 13 and purine base 'G' was present in sample 11.
- At position 679 either deletion or addition occurred as base G was present in sample 13, while a gap was found in sample 11 at this point.
- Base pair substitutions were detected at positions 693, 695, 702 and 706. Out of all these only position 695 exhibited a transition as pyrimidine base C was found in sample 13 and purine base T was found in sample 11. In all others it was a case of transversion as pyrimidine base 'T' was found in sample 13, while purine base 'G' was found in sample 11.
- Positions 752, 753, 754 and 756 all exhibited cases of transversion as base 'T' was found in sample 13 at 752 and 754 and in sample 11 at 753 and 756, while base 'G' was found in sample 13 at 753 and 756 and in sample 11 at 752 and 754.
- Cases of deletion/ addition were detected at positions 761 and 762 where two purine bases 'G' and 'A' were found in sample 13 while they were absent in sample 11.
- Positions 763 and 779 were cases of transitions as pyrimidine bases 'C' and 'T' were found in sample 13 and 'T' and 'C' in sample 11.
- Position 781 showed transversion as pyrimidine base 'T' was found in sample 13 and purine base 'G' was found in sample 11.

- Trinucleotide deletion/addition was noticed at positions 800, 801 and 802, where pyrimidine bases 'C', 'T' and purine base 'G' were present in sample number 13, but absent in sample number 11.
- Position 808, showed a case of transition as purine base 'G' was found in sample 13 and 'A' in sample 11.
- Position 819 was showing transversion as base 'T' was present in sample 13 and base 'G' in sample 11.
- The next three positions 820, 821and 822 showed transitions. In the 1st and the 2nd cases base C was present in sample 13 and base 'T' in sample 11, while in the third case base 'T' was present in sample 13 and base 'C' in sample 11.
- Transitions were also found next at positions 846 and 850. In the first case it was between two pyrimidines , as 'C' was found in sample 13 and 'T' in sample 11 while in the second case purine base 'G' was present in sample 13 while 'A' in sample 11.
- Positions 860 and 863 were again cases of transition. At 860, base 'C' was present in sample 13 while base 'T' in sample 11, while at 863 it was the opposite.
- Positions 862, 864,865 and 866 were cases of transversions , as bases 'T', 'G', 'T' and 'G' were present in sample 13 while bases 'G', 'C', 'G' and 'T' were found in sample 11.
- Positions 870 and 871 were cases of addition or deletion as two nucleotides were missing in sample 13, while 'A' and 'C' were found in sample 11.
- Next positions for transitions were 880, 881, and 882 where bases 'T', 'C' and 'C' were found for sample 13 while 'C', 'T' and 'T' were present for sample 11.

- The last positions in these sequences where two cases of tranversions occurred were positions 848 and 899 were bases 'T' and 'G' were present in sample 13 while 'G' and 'T' in sample 11.

Therefore, it was found that the distal part of 16S rRNA gene was highly conserved as very few differences were observed in the initial part of the amplicon. There were several single nucleotide changes (SNP) especially in the middle part of the amplicon. Two specific regions with high density of SNPs were between 800-822 and 863-899.

4.8.3 High altitude, Site 3

Sequence alignment and cutting sites of the two different patterns found in this site are presented in Fig.23. Only sample numbers 65 and 70 were sequenced. When aligned together, their amplicon length was shown as 1070 bp. Both the sequences blasted as *Frankia*. In sample number 65, Webcutter could identify five cutting sites producing six bands, which were visible in the gel (Table 6). In sample number 70, the first cutting site N (226 bp) was missing, producing only five bands. 169 bp and 164 bp bands got merged together to appear as 164 bp band in the gel.

Sample number 65 (P4) showed similarity with sample number 13 from high altitude, site 2, which justified their grouping under P4 based on ARP data. For this site, cutting sites O, P, Q and R were considered as conserved, since site N was missing for sample number 70, it could not be considered as conserved. Sites O, P, Q and R aligned together for both the samples.

Sequence alignment of the two patterns revealed minor variations in this site. Since only two sequences were compared, it was not possible to conclude whether the cases were that of insertion or deletions. Even for base pair substitutions, it wasn't clear which two bases got substituted. Therefore, in this site, just the differences in their sequences have been highlighted.

Fig.23. Diagrammatic representation of the sequences of High altitude, site 3, with their restriction sites with *Rsa* 1. (Distances not in scale). ⟶

Table.6. Comparison between RFLP and webcutter generated banding patterns of High altitude, site 3 ⟶

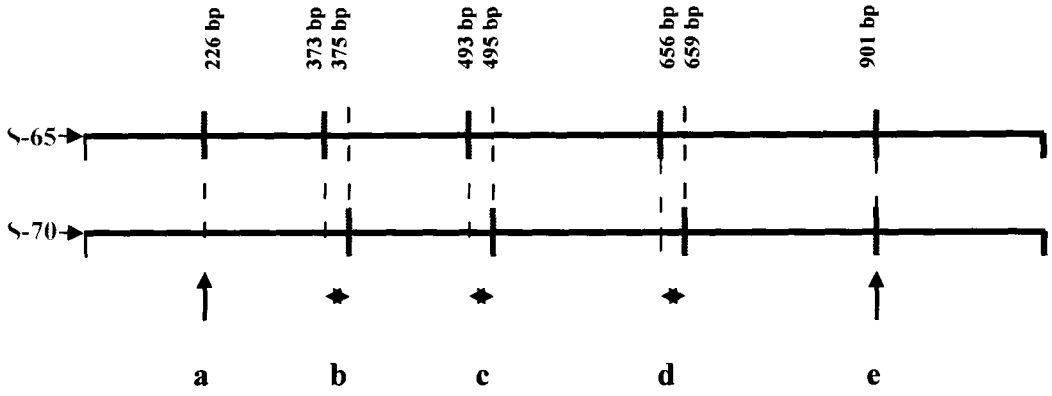


Fig.23

Sample number	Bands visible in the gel (bp)	Bands produced by Webcutter (bp)
65	246	245
	217	226
	161	169
		163
	150	147
	111	120
70	420	375
	246	242
	164	169
		164
	112	120

Table.6

- At position 1, base 'G' was present in case of sample 70, while it was absent for sample 65.
- At position 5, sample 65 had base 'G' and sample 70 had base 'T'. It appeared to be a case of transversion [from 'T' to 'G' as evident from all other samples from different sites].
- At positions 24, 25, 27 and 28, sample 65 had bases 'C', 'A', 'G' and 'T', while sample 70 had 'A', 'G', 'T' and 'G'. So, we can think of a possibility of transversion at positions 24, 27 and 28. At position 25, it was a case of transition between two purines (G and A).
- The next position where a possible insertion/deletion was found is 61. Here sample 70 had base 'A' which was found absent in sample 65.
- At position 67, another case of possible transversion was detected – base 'T' of sample 65 was found replaced by base 'G' in sample 70.
- Position 228 was most probably a case of transition as one pyrimidine (either 'T' or 'C') was replaced with another pyrimidine.
- Similar case was also noticed at positions 546, 627 and 639.
- At 640 transversion was found, as the substitution was between one pyrimidine ('T') and one purine ('G').
- Position 643 was again a case of insertion/deletion as an additional base 'T' was found in sample 70.
- Transition was noticed at position 668.
- At 672 and 673, insertion/deletion happened as bases 'C' and 'G' were found in sample 65 but were absent in sample 70.

- Position 698 was a case of transition while position 700 and 703 were transversions. Position 717 was again a case of transition.
- Insertion/deletion was detected at position 749.
- Positions 754 and 755 were showing cases of transition as the substitution were between pyrimidines (C and T).
- Transition was also noticed at 757 and 759 (between G and A). But at 758, transversion occurred between G and C. Transversion was again found at position 779 between G and C.
- Transition between bases 'G' and 'A' occurred at positions 799 and 801, and at 822 between pyrimidines 'C' and 'T'.
- At positions 844 and 845, transitions were noticed between pyrimidines 'C' and 'T' and at 848 between purines 'G' and 'A'.
- Again at 864 and 868 transition was detected between pyrimidines and purines respectively.
- Transition occurred at positions 879, 887 and 898 between pyrimidines ('C' and 'T') while at 896 it was a case of transversion between purine (G) and pyrimidine (T).
- At 1006 and 1030, transitions were found between purines ('G' and 'A') and pyrimidines ('C' and 'T') respectively, while at 1036, transversion occurred between pyrimidine (T) and purine (G).

In contrast to our above observations, in these two cases several SNPs were detected even in the distal part of the 16S rRNA gene. This is striking, as this portion of the gene

is generally considered highly conserved. The distribution of SNPs along the rest of the regions was more or less similar to the other cases.

4.8.4 Middle altitude, Site 1

Sample numbers 21, 22, 24, 28 and 29 were sequenced from this site. The diagrammatic representation of these aligned sequences along with their cutting sites has been shown in Fig. 11. The total length of the amplicons were 1072 bp, 1067 bp 1084 bp 1070 bp and 1071 bp respectively. All of these sequences blasted as *Frankia*.

Webcutter identified four restriction sites along all these sequences producing five bands. Restriction sites for all the patterns aligned together with minor differences (few base pairs) between them (Fig.24). As it was found in some earlier cases, the two bands around 165 bp to 170 bp always appeared to get merged together (Table 7). Restriction sites S, T, U and V appeared to be conserved at this middle altitude, site 1 and they aligned together.

- Cases of additions were found at position 23, in samples 28 and 29. Pyrimidine base 'C' was found in sample 28, while purine base 'A' was found in sample 29.
- In sample 24, 6 nucleotide bases were found inserted at positions 55, 56, 57, 58, 59 and 60. These were 'A', 'C', 'C', 'T', 'C' and 'G'.
- Positions 86, 102 and 109 exhibited base pair substitutions. The 1st case was a transition as purine base 'A' present in all other samples was replaced by purine base 'G' in sample 21. In 2nd and 3rd cases also transition mutations occurred, where pyrimidine base 'T' present in other samples was replaced by pyrimidine base 'C' in sample 22.
- In sample 28 at position 148 base 'C' was inserted.

Fig.24. Diagrammatic representation of the sequences of Middle altitude, site 1, with their restriction sites with *Rsa* 1. (Distances not in scale). ⟶

Table.7. Comparison between RFLP and webcutter generated banding patterns of Middle altitude, site 1 ⟶

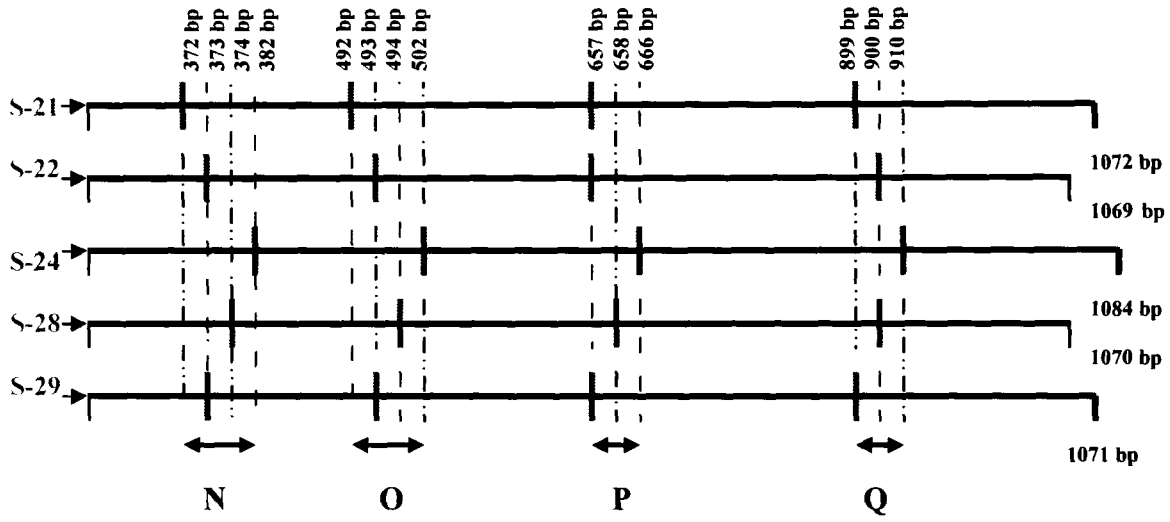


Fig.24

Sample number	Bands visible in the gel (bp)	Bands produced by Webcutter (bp)
21	396	372
	246	242
	162	173
		165
	116	120
22	390	373
	248	243
	166	169
		164
	123	120
24	396	382
	253	244
	166	174
		164
	123	120
28	392	374
	253	242
	162	170
		164
	116	120
29	396	373
	246	242
	157	172
		164
	116	120

Table.7

- Further downstream, at positions 271 and 272, two bases 'C' and 'A' were found inserted in sample 24 and were absent in all other samples.
- Transversion was detected at position 313, where purine 'A' was found in sample 24, whereas pyrimidine base 'C' was present in all other samples.
- Positions 317 and 328 were cases of insertions where base 'C' was found in sample 22 and 24 while it was absent in others. At position 328, base 'G' was present in sample 24 but was absent in others.
- Transition was found at 520 where base 'T' present in other samples was substituted by base 'C' in sample 29.
- Position 535 had an insertion for sample 21 where purine base 'A' was present which was not found in other samples.
- Two cases of insertions were found again at position 817 where base 'G' was present in sample 22 and base 'T' in sample 24.
- Transversion was detected in sample 24 at position 818 where 'T' present in all other samples got substituted by base 'G'.
- At position 941 base 'G' present in other samples was substituted for base 'A' in sample 22.
- Positions 972 and 973 had transition and transversion respectively for sample 21, as base 'C' present in other samples got substituted by base 'T' at 972 and base 'G' by base 'C' at 973.
- Insertions were found at 976 where base 'T' was found inserted in samples 28 and 21.

- Transition was found at position 991 as base 'C' of other samples got substituted by base 'T' in sample 28.
- In sample 21, transversion occurred at 1020 base 'C' was replaced by base 'A'. Transversion was found at 1021 as base 'C' was found in the place of base 'G' present in other samples. At position 1022 transition was found as base 'A' present in other samples was substituted by base 'G'. At position 1023 transversion was found as base 'T' was substituted by base 'A'.
- At position 1024, two bases ('T') were found inserted in samples 29 and 21. Bases 'G' and 'A' were found as a case of insertion in samples 29 and 21 respectively at position 1036. Position 1037 exhibited a case of substitution (transversion) as base 'A' was replaced by base 'C' in sample 21.
- At position 1051, base 'G' and base 'C' were found inserted in samples 21 and 24 respectively. Transition occurred at the next two positions, i.e. 1052 and 1053 as base 'T' and base 'C' were found in sample 24, while in other samples they were bases 'C' and 'T' respectively.
- A case of addition with four bases occurred in sample 24 at positions 1054, 1055, 1056 and 1057, where bases 'C', 'A', 'G' and 'A' were found.
- Last case of insertion could be detected at position 1070, where a base 'A' was inserted in sample 29 and was absent in all other samples.

It was therefore found that sample 21 had accumulated more SNPs than other samples. The distribution of SNPs along the compared region was seen through out.

4.8.5 Middle altitude, Site 2

The alignment and cutting sites identified by Webcutter for this site are presented in Fig.25. Sample numbers 31, 32, 33, 36 and 39 were sequenced for this site. All of them blasted as *Frankia*.

The total length of the amplicons revealed were 1070 bp for all, except sample 33, where it was 1067 bp. This indicated some minor changes along the sequence.

Webcutter could identify four cutting sites in sample number 31 and 36, producing five bands. In these two samples, site W was missing. In rest of the samples, this site was present producing five cutting sites with six bands. In sample number 31 only four bands out of five produced by Webcutter were clearly visible in the gel. Most probably 171 bp and 164 bp bands got merged together. In sample number 32, in addition to the bands produced by Webcutter an additional band of 572 bp could also be detected in the gel. It is well known that *Frankia* has two *rrn* operons and if during digestion one of the operons doesn't get digested, an additional band of high molecular weight could be seen.

Here also, the 239 bp and 227 bp bands and 170 bp and 167 bp bands merged together respectively. This was also found in sample number 33. In sample number 39, though the pattern in ARP gel looked similar to 31 and 36, the bands produced by Webcutter suggested a different pattern similar to 33.

Minor variations emerge from the five-sequence alignment from this site.

- Two purine bases (G) were found at positions 1 & 2 of sample 32, which were absent in all other samples. Substitution occured at position 3 in which base 'G' (purine) was substituted by base 'T' (pyrimidine) in sample 32.
- At position 80 in sample 31, there was an addition of one pyrmidine base 'C'.

Fig.25. Diagrammatic representation of the sequences of Middle altitude, site 2, with their restriction sites with *Rsa* 1. (Distances not in scale)⟶

Table.8. Comparison between RFLP and webcutter generated banding patterns of Middle altitude, site 2 ⟶

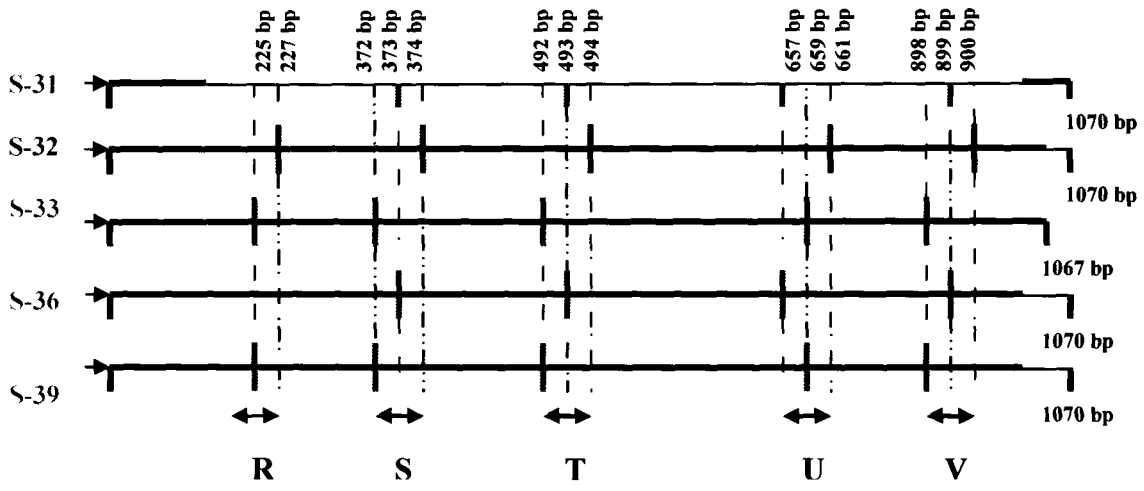


Fig.25

Sample number	Bands visible in the gel (bp)	Bands produced by Webcutter (bp)
31	371	373
	240	242
		171
	165	164
	125	120
32	572	
	240	239
		227
	171	170
		167
		147
	125	120
33	248	239
		225
	176	169
		167
		147
	129	120
36	394	373
	255	242
	181	171
		164
	137	120
39	382	
	233	239
		225
		172
	165	167
		147
	125	120

Table.8

- At position 144, another insertion was detected as base 'C' was found in sample 36.
- Transversion was noticed at position 178 where purine base 'A' was substituted by pyrimidine base 'T' in sample 36.
- At position 229, transition occurred. Base 'T' was replaced by base 'C' in both samples 31 & 36.
- At position 435 'C' substituted 'T' in sample 31.
- At 509, transition was detected as purine base 'G' was replaced by purine base 'A' in sample 36.
- Transitions were found at positions 547 & 584. In the first case base 'C' was replaced by baser 'T' in sample 31 & 36, while at second place base 'A' was replaced by base 'G' in the same samples.
- Deletions of 'G' were found at positions 554 and 563 and deletion of 'A' at position 624 in samples 31 & 36.
- Positions 626, 631, 643 & 646, were found to have transitions. AT 626, base 'A' was replaced by base 'G' in sample 31 & 36. At 631, base 'T' was replaced by base 'C' in sample 31 & 36. At position 643, base 'C' was replaced by base 'T' in the same samples. In the last position base 'T' was replaced by base 'C'.
- Transversion was found at position 648 where base 'C' was substituted by base 'G' in samples 31 & 36.

- Transition occurred at position 656 as base 'C' was substituted by base 'T'. Transversion was detected at 657 as base 'T' was replaced by base 'G' and base 'T' was substituted by base 'C' at 658 in samples 31 and 36.
- Base 'C' to 'T' transition was found at position 672 in samples 31 & 36.
- Deletion of base 'T' was found at position 677 in samples 31 & 36.
- In samples 31 & 36, T to C transition was found at position 695, G to A transition at position 701 and at positions 705 and 706, bases 'C' and 'G' were both substituted by base 'T'.
- G to A transitions were found at positions 720 and 749 in samples 31 & 36.
- T to G transversion was found at position 758 and G to A transition was found at position 759 in samples 31 & 36.
- Two pyrimidine bases 'C' were inserted at positions 760 and 761 in samples 31 & 36.
- T to C transition was found at position 762 in samples 31,33 and 36
- At position 778, base 'C' was substituted by base 'T' in samples 31 and 36.
- G to C transversion occurred at position 781 in samples 31 and 36.
- Three bases were inserted at positions 799, 800 and 801 in samples 31 and 36.
- G to A transition was found at position 803 and A to G at 807 in samples 31 and 36.
- Insertion of base 'C' was noticed in samples 31 and 36 at position 822.

- At positions 824 (C to T), 847 (C to T) and 860 (T to C) transitions occurred in samples 31 and 36.
- Three bases were simultaneously substituted again at positions 862, 863 and 864 in samples 31 and 36 where bases 'G', 'T' and 'C' were replaced by bases 'T', 'C' and 'T'.
- Base substitutions were found at positions 866 (T to C) and 867 (T to G) in samples 31 and 36. In same samples deletion of base 'C' occurred at position 871, dinucleotide inversion was found at 880 and 881 (TC to CT) and transversion at 898 (G to T).
- In samples 31 and 36, C to T transition was found at position 901.
- At positions 954 (A to T), 955 (T to A) and 957(A to T), transversions were detected in sample 36.
- Insertion of 'A' was detected at position 959 in samples 31 and 36.
- G to A transition was found at 1010 in samples 31 and 36.
- Position 1029 was an interesting site, where deletion of base 'A' was noticed in samples 33 and 39 and A to G base substitution was found in sample 32.
- C to T transition was found at 1035 in samples 31 and 36.
- At sites 1046, 1047 and 1048 bases 'C', 'G and 'A' were found inserted only in sample 39.

In general majority of changes were found in samples 31 and 36. Other samples had fewer changes. Apparently, samples 31 and 36 represented genotypes that had evolved later than others.

4.8.6 Middle altitude, Site 3

For this site, the diagram for sequence alignment and restriction sites could not be constructed because only two samples (73 and 75) from this site could be sequenced. Out of these two sequences only a length of 633 bp could be sequenced for sample 73 and 636 bp for sample number 75. Though both the sequences blasted as *Frankia*, surprisingly, sample number 75 produced four cutting sites with Webcutter, while no cutting sites could be detected in sample number 73 along its 633 bp length. Sample number 75 also could not give the actual banding pattern with Webcutter because of its partial sequence length. So for this site we could only depend on the ARP data according to which we found four different banding patterns. Sample numbers 71, 72, 74, 75 represented similar pattern and were grouped as P10. Each of these samples showed four bands – 247 bp, 168 bp, 148 bp and 123 bp, all of which were visible in the gel. Sample 73 and 76 showed similar pattern (P11) with five visible bands in the gel. These bands were 285 bp, 247 bp, 168 bp, 148 bp and 123 bp. Sample numbers 77 and 78 also showed similar pattern and were grouped as P12. This pattern showed four distinct bands- 220 bp, 168 bp, 148 bp, 123 bp. Pattern 13, comprised of sample number 79 and 80 that showed five different bands of 518 bp, 220 bp, 168 bp, 148 bp, 123 bp.

4.8.7 Low altitude, Site 1 and Site 2

We were not successful in sequencing any samples from lower altitude site 1. Sample numbers 41, 42, 43, 45, 47 and 50 from this location and sample numbers 52, 53, 54, 56, 57, 58 and 60 of lower altitude, site 2, showed five bands each and this profile was considered as pattern P14. Only sample number 58 could be sequenced partially (479 bp) and this sequence blasted as *Frankia*. When it was cut with Webcutter, it could detect only one restriction site (at 312 bp) along the sequence, since the length of the sequence was too short for detecting all other cutting sites. This 312 bp band was visible in the gel.

It appeared that had we succeeded in sequencing the whole amplicon, other restriction sites could have been detected and it would have shown the same pattern as visible in the gel. For P15, represented by sample numbers 44, 46, 48 and 49 of lower altitude, site 1 and sample numbers 51, 57 and 59 of lower altitude, site 2 sequencing was not successful even after repeated efforts. For these samples, we considered only the ARPs which showed four bands for all these samples.

4.8.8 Lower altitude, Site 3

No samples could be sequenced for this site. Here four different patterns could be detected (section 4.5.9).

4.8.9 Samples of *Hippophae* sp.

Three different banding patterns (P18, P19, and P20) were observed in samples from *Hippophae* sp. One sample representing each banding pattern was selected for sequencing. H1, H2 and H6 were sequenced and all of them blasted as *Frankia*. Sample H2 could be sequenced up to a length of 1040 bp, while in H1 and H6 nucleotide sequence lengths were relatively short, being 418 bp and 566 bp respectively.

When these three sequences were aligned together, interestingly they showed a lot of differences among them. Similarity in their nucleotide bases was found to be comparatively less than among the samples from *Alnus nepalensis*.

Webcutter could identify four cutting sites along the sequence of sample H2, producing five bands. Four of these bands were visible in the gel. Another band (420 bp) which was visible in the gel was not produced by webcutter, instead it showed a band of 211 bp which was not visible in the gel.

In sample H6, webcutter could identify three cutting sites, but since it was partially sequenced, only two bands (147 bp and 120 bp) were found, but only 147 bp band was visible in the gel.

Sample H1 showed no cutting sites along its 418 bp length.

While comparing the alignment of nucleotide sequences of the samples, it was noticed that the position of the restriction sites along the sequences varied in different samples (Fig.21 to 25), even for the samples showing the same banding pattern from one particular site.

These variations were sometimes caused due to insertions or deletions of one, two or three base pairs, which were not so significant. But sometimes, these variations were long enough and were caused by additions or deletions of several base pairs. High altitude, site 1 was considered as an example as to explain the differences in various restriction sites in detail.

Four samples were aligned from this site, out of which sample number 1 and 6 showed an additional restriction site 'A' which was absent in other two samples. This cutting site was present at position 225 in sample no. 1 while in sample no. 6, it was found at position 264. This long gap of 39 bases was mainly caused by a repetition of the starting sequence of 39 bases in sample 6. Because of this, the other restriction sites also changed their positions in this sample. Therefore in sample no. 6, second (B), third (C), fourth (D) and sixth (F) restriction sites were also found to be distantly placed from the same restriction sites of the other samples.

The second cutting site B was present in all the four samples but at different positions. In sample 8, it was found at 372 but in sample 1 it was found at position 375.

This shifting was caused by the presence of base 'A' at two places (position 274 and 306) and base 'C' at 231 in sample no. 1.

When compared between sample 8 and sample 4, this cutting site was found at a distance of 14 bases upstream in sample 4 (at position 386). Exactly 14 insertions were found in sample 4, which easily explains this change in position. Bases 'G' and 'C' at positions 185 and 186, base 'A' at 220 and 231, bases 'C' and 'T' at 240 and 241, base 'A' at both 274 and 292. A short sequence of three bases ('C', 'G' and 'T') was found at positions 316, 317 and 318. At positions 337, 356 and 369, there were additions of base 'A'.

Difference between sample 1 and 6 was reduced to 36 bases which was caused by three additions in sample 1 at positions 231 (base 'C'), and 274 and 306 (both 'A's).

Restriction site C was present at 492, 496, 507 and 532 in sample no. 8, 1, 4 and 6 respectively. When sample 8 and 1 were compared, it was found that the difference between their cutting sites had increased by one base than the difference of their previous cutting site B. This was because sample 1 had an insertion at position 430. Sample 8 and 4 had a difference of 14 bases for site B which had increased by one base in case of site C. It was found that an insertion of base 'C' occurred at position 399. Difference between 1 and 6 remained the same. Though there was an addition of base 'A' at position 430 in sample 1, it was adjusted by another addition of base 'C' in sample 6.

Restriction site D was present at positions 656, 659, 671 and 699 in samples 8, 1, 4 and 6 respectively. Difference between sample no. 8 and 1 was again reduced to 3 bases because of one deletion at position 662 in sample no. 1. Between sample 8 and 4, the difference remained the same as site C (i5 bases). Difference between 1 and 6 was increased by 4 bases (altogether a difference of 40 bases), which were found inserted at positions 568 (base 'T'), at 577 ('G') at 638 ('A') and at 662 ('C') in sample 6.

Site E was only present in sample 1.

Site F was found at positions 898, 905, 914 and 938 in samples 8, 1, 4 and 6 respectively. Difference between sample 8 and 1 was increased to 7 bases due to addition of four bases in sample 1. These were at 766 (base 'T'), 855 (base 'T'), 884 (base 'T') and 892 (base 'G'). Difference between sample 8 and 4 was reduced by one base which was caused by the insertion of base 'G' at position 835 in sample 4.

Difference between sample 1 and 6 was again reduced to 33 bases. There was an addition of base 'T' at position 691 in sample 6 which was adjusted by another addition of base 'T' at position 766 in sample 1. But the difference of 7 bases was caused by 7 deletions at position 775, 3 consecutive bases ('C', 'T' and 'A') at 814, 815 and 816, base 'T' at 836 in sample 6 and additions of bases 'T' and 'G' at 855 and 892 in sample 1.

This same trend was also noticed in the position difference of different restriction sites among samples of other sites. Not many variations were noticed for high altitude site 2 samples, where the amplicon lengths were found to be same for sample 11 and 13. Sample 15 could be sequenced partially, so full information was not available for it. Strikingly, though the amplicon length of sample 16 was much smaller (1034 bp) only two cutting sites were found along it at positions 563 and 960 which were absent in other samples. For samples 11 and 13, differences in cutting sites such as G, H, I, J and L were not so significant being mostly of one to two bases. Only site K in sample 11 showed a difference of 4 bases with sample 13, which otherwise was always one base pair.

For sample 65 and 70 of high altitude, site 3, the amplicon lengths were exactly the same (1070 bp). Restriction site 'a' was present only in sample 65. For sites 'b' and 'c' difference in their sites were only for 2 bases whereas for site 'd' it was a difference of 3 bases. Site 'e' was found at same position 901.

Middle altitude site 1 was an interesting site, where sample 24 had a longer amplicon because of several insertions, specially 6 bases at position 55 and four insertions at position 1054. This would also affect the distribution of restriction sites resulting in changes in ARP. Other samples mostly showed similar changes as discussed earlier for samples of previous sites.

Middle altitude site 2 also did not show any striking difference. All the differences were found to be mostly of one or two bases and thus the amplicon lengths also did not vary much.

4.9 Phylogenetic Analysis

The variable internally transcribed spacer (ITS) region between 16S-23S rRNA was analysed. Very limited information is available for this specific region from *Frankia* or any related organism. The data obtained after nucleotide sequencing were aligned together (site wise) to find out the variation among them at nucleotide level. Phylogenetic tree (rectangular) was constructed using cluster algorithm method (Fig.26). Nucleotide sequence of some strains of *Frankia* submitted in the GenBank by Varghese and Misra (2000) were used as out species. Strains AntKR (Accession no. - AJ404870), AnpOR (Acc. no. - AJ404868), AnpHR (Acc. no. – AJ404866), AnpNHR (Acc. No. – AJ404867), AnpUSR (Acc. no.-AJ404869), M88466, AgTR (Acc. no. –AJ404871) and M55343 were used for comparison. All these strains were found clustering in the same pattern as found in earlier studies (Varghese *et al.*, 2003).

Website www.genebee.msu.su was used for constructing the phylogenetic tree. The tree formed (Fig.26), showed close relationship between several *Frankia* strains. The numbers at the forks correspond to the bootstrap values out of 100.

Sample 13 of high altitude, site 2, showed very close relationship with sample number 65 from high altitude, site 3 (bootstrap value of 80). In RFLP analysis, both

these patterns were grouped under P4. The sequence alignments also showed a high degree of similarity except some minor differences at some places, which were mainly due to mutations (base pair substitutions, additions or deletions, Section 4.7.2). These two samples were in turn related to sample 1 from high altitude, site 1 (P1) with a bootstrap value of 79. All these three samples, representing three different patterns formed a separate cluster, exclusively of high altitude samples. Although a major branch with bootstrap value of 100 was clustered with these three, the internal bootstrap values were not significant. Therefore, the associations within the cluster appear to be more due to chance. A closer look found that samples mostly from the middle altitude were present in this cluster. The notable exceptions being samples 4 (high altitude, site 1), 8c (high altitude site 1) and 70 (high altitude site 3). However, since the bootstrap values for the clustering of these high altitude samples were very low, the presence of these in the clusters may be ignored. One significant observation is that AntKR, sequence of a sample earlier collected by Dr. Rajani Varghese from the Kulu area (approximately 1200 mt, which also falls under the middle altitude range in this study) of Himachal Pradesh (Varghese, 2000), clustered with the group of middle altitude samples with a bootstrap value of 90.

The second major branch comprised of two major clusters. One of them represented sequences retrieved from the GenBank and the other had three samples from high altitude and two from the middle altitude. Once again the low bootstrap values indicate chance associations.

Special mention must be made of the clustering of sample 15 (high altitude, site 2) and sample 58 (low altitude, site 2) with a very high bootstrap value. These samples did not cluster together on the basis of their ARP. This clustering must have occurred because only partial sequence was available for these two samples. Therefore, it could be ignored.

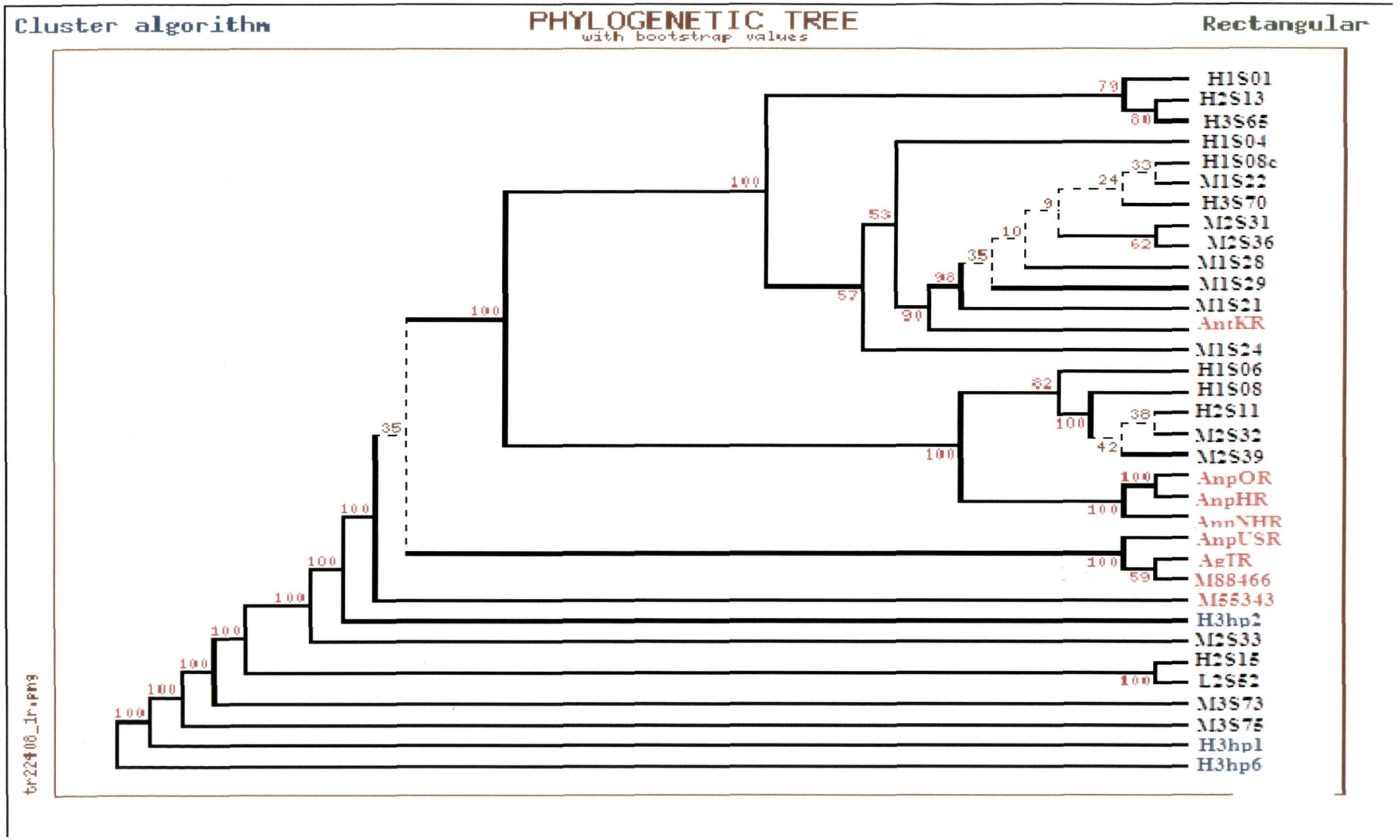
Fig.26. Phylogenetic tree obtained from GeneBee (http://www.genebee.msu.su) , constructed on the basis of nucleotide sequences. The numbers at the forks correspond to the bootstrap values out of 100.

i. For *Alnus nepalensis* samples-

H1S01 – High altitude, site 1, sample 1
M2S36 – Middle altitude, site 2, sample 36

ii. For *Hippophae* sp. –

H3hp1 – High altitude, site 3, sample 1
H3hp6 – High altitude, site 3, sample 6


Cluster algorithm
PHYLOGENETIC TREE
with bootstrap values
Rectangular
H1S01
H2S13
H3S65
H1S04
H1S08c
M1S22
H3S70
M2S31
M2S36
M1S28
M1S29
M1S21
AntKR
M1S24
H1S06
H1S08
H2S11
M2S32
M2S39
AnpOR
AnpHR
AnnNHR
AnpUSR
AgTR
M88466
M55343
H3hp2
M2S33
H2S15
L2S52
M3S73
M3S75
H3hp1
H3hp6
79
80
100
53
24
33
9
10
62
35
98
57
90
100
62
100
38
42
100
100
100
100
100
59
35
100
100
100
100
100
100
100
100
100
tr224t8_1r.phy

From the foregoing, it appeared that the distribution of *Frankia* strains in the soil was significantly affected by altitude, with some exceptions. This further reinforces the inference made on the basis of ARP (Section 4.7).

4.10 Acetylene Reduction Assay (ARA), (Stewart *et al.*, 1968)

ARA was performed for nodule samples collected as above from *Alnus* trees (Section 3.5).

The nitrogenase activities of nodule samples from different sites estimated in terms of ethylene gas produced /gm nodule fresh weight / hour are presented in Table 9.

Low altitude nodules were found to show highest nitrogenase activity compared to middle and high altitude nodules. Nitrogenase activity in the nodules from high altitude was so less that it was considered as negligible.

4.11 Soil Analysis

Eight soil parameters were analysed in the present study (Section 3.6), The values of mean and standard deviation for all the soil parameters are given in Table 10. Based on these mean data, column graphs were plotted with standard error of mean, using ORIGIN 40 [Fig 27 (a to h)].

i. **pH** –The soils were found to be acidic in nature. The pH ranged between 4.73 to 5.6. The only exception was the low altitude site 2 where the pH was found to be near neutral (6.78) (Fig.27.a.).

ii. **Total organic carbon** –It was found to be in the higher range in most of the sites. When total organic carbon is > 0.75 (%), it is considered as high. Except in low altitude site 1, where the value was 0.198, and hence considered as low, all other sites

Table.9. Nitrogenase activities of nodule samples from different sites estimated in terms of ethylene gas produced /gm fresh weight of nodule / per hour. ⟶

Acetylene Reduction Assay:

Altitude	Lowest (N.A)	Highest (N.A)
High(>2000mt)	Negligible	Negligible
Middle(1700mt)	1.277	6.96
Low(1000mt)	1.061	43.32

Table.9

had values >1.00. In two cases (middle altitude, site 1 and site 3) it was even >2.00 (Fig.27.b.).

iii. **Total nitrogen** – The amount of total nitrogen, expressed in percentage was in the lower range, indicating the rhizosphere soils were poor in nitrogen content. The amounts of nitrogen found were >1.00 (%) (Fig.27.c.). Abundant nodulation occurs in nitrogen poor soils.

iv. **Available phosphorus** - The available phosphorous ranged between 11.778 to 19.576 µg/g which can be considered as medium (10 to 24.6 µg/g = medium range). Only middle altitude site 1 was exceptional, showing a value of 9.668 (µg/g) which is considered as low (<10 = low range) (Fig.27d.).

v. **Available potassium** – The range of available potassium varied considerably for different altitudes and sites. In high altitude sites 2 & 3, middle altitude sites 2 & 3 and low altitude sites 1 and site 2 available potassium values were within the medium range (118-280 kg/ha). High altitude site 1, middle altitude site 1 and low altitude site 3 had very high values, the highest being middle altitude site 1 (642.9) (Fig.27.e.).

vi. **Exchangeable calcium-** The exchangeable calcium was found to be within normal range i.e. 1.3 (meq/100g) to 2.7. Only exception was middle altitude site 3 where it was very high (7.0) Fig. 27.f.

vii. **Exchangeable magnesium** – Magnesium was also found to be within the normal range i.e. 0.51(meq/100g) to 1.88. Fig. 27.g.

viii. **Electrical conductivity** – Values of <0.8 are considered as normal and >0.8 are considered as critical. All the values were under the normal range (Fig. [illegible]).

UNIVERSITY LIBRARY 103802

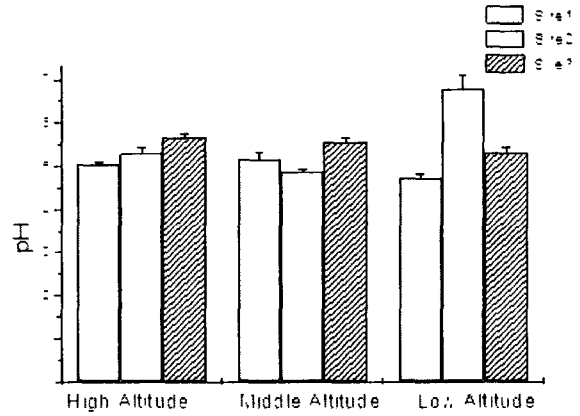
Table.10. Mean and standard deviation values of different soil parameters ⟶

	Mean									Standard deviation								
	High alttitude			Mıddle altitude			Low altitude			High alttıtude			Middle altitude			Low altitude		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
pH	5.03	5.29	5.67	5.16	4.87	5.55	4.73	6.78	5.31	0.27	0.42	0.29	0.55	0.16	0.27	0.28	0.97	0.33
C (%)	1.54	1.11	1.71	2.26	1.60	2.28	0.19	1.77	1.68	0.59	0.33	0.51	0.37	0.46	0.22	0.55	0.38	0.39
N (%)	0.10	0.007	0.04	0.05	0.04	0.08	0.05	0.03	0.06	0.09	0.00	0.05	0.03	0.04	0.02	0.02	0.02	0.03
P (µg/g)	19.57	13.3	11.77	9.66	17.6	13.79	12.36	14.87	13.79	7.99	5.21	2.08	5.10	2.90	1.07	1.17	6.08	1.07
K (kg/ha)	437.2	188.9	206.8	642.9	174.9	281.4	277.3	193.36	370.76	273.7	109.2	163.9	337.6	117.3	143.1	175.9	234.6	302.7
Ca (meq/100gm)	1.6	1.3	1.6	2.7	1.6	7.0	2.4	2.6	1.9	1.27	2.15	1.61	2.79	1.19	1.06	1.33	1.77	1.27
Mg (meq/100gm)	1.04	0.51	0.75	1.88	1.1	1.56	1.62	1.05	1.17	0.52	0.37	0.33	2.61	0.81	0.30	0.75	0.93	0.51
El-Con (meq/100gm)	0.06	0.02	0.06	0.07	0.04	0.27	0.05	0.49	0.08	0.02	0.01	0.03	0.06	0.02	0.02	0.02	0.65	0.03

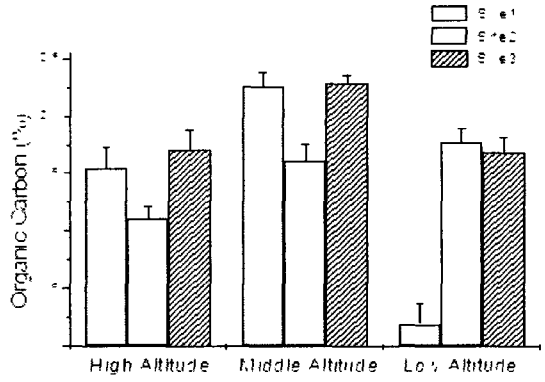
Table.10.

Fig.27. Graphical representation of soil analysis (Graphs plotted with mean and standard deviation of mean values).⟶

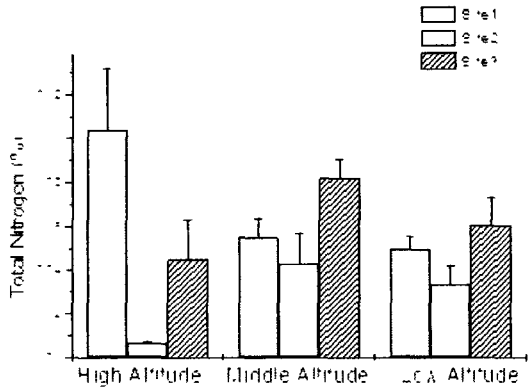
(a) – pH
(b) - Organic carbon
(c) – Total nitrogen


pH
High Altitude
Middle Altitude

(a)


High Altitude
Middle Altitude

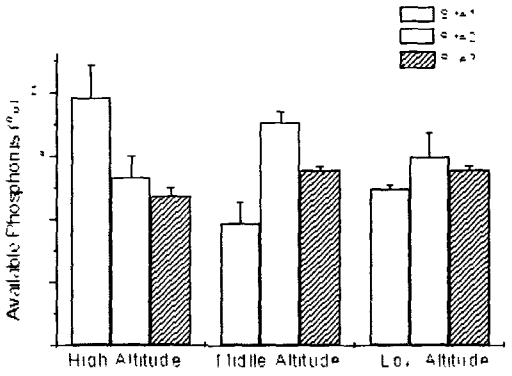
(b)


High Altitude
Middle Altitude

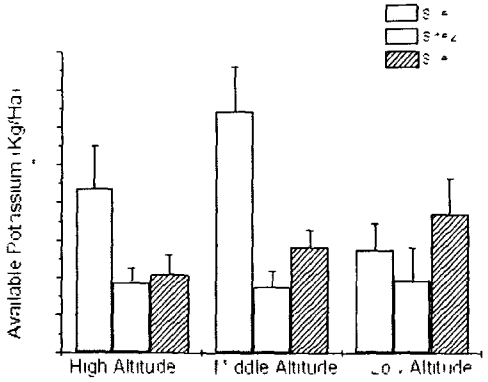
(c)

Fig.27. Graphical representation of soil analysis (Graphs plotted with mean and standard deviation of mean values).⟶

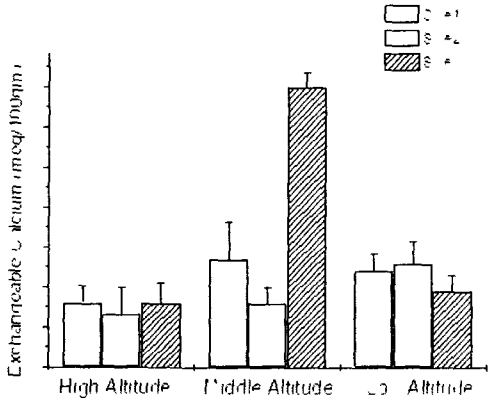
(d) - Available phosphorus
(e) – Available potassium
(f) – Exchangeable calcium


High Altitude
Middle Altitude
Low Altitude

(d)


High Altitude
Middle Altitude

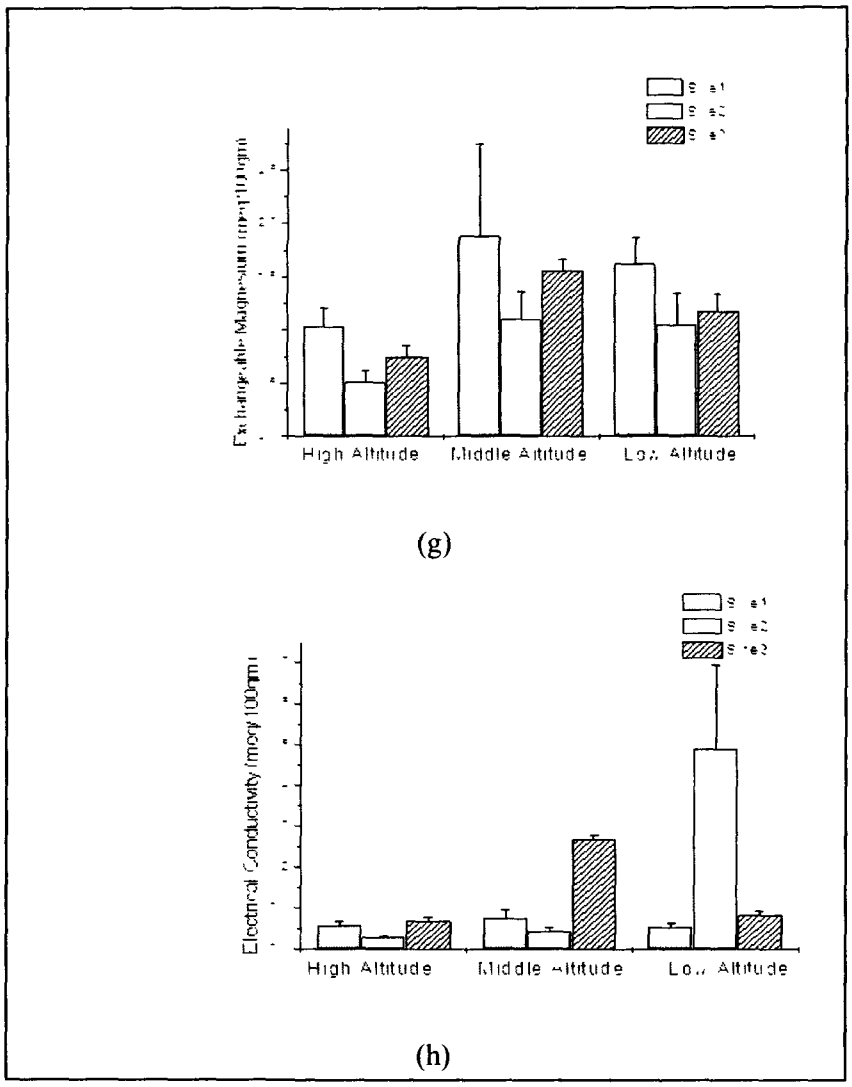
(e)


High Altitude
Middle Altitude

(f)

Fig.27. Graphical representation of soil analysis (Graphs plotted with mean and standard deviation of mean values). ⟶

(g) – Exchangeable magnesium
(h) - Electrical conductivity


High Altitude
Middle Altitude
High Altitude
Middle Altitude

(g)

(h)

Fig.27

In general, the soil types were mainly "mineral" having an acidic pH range. The soils were generally rich in organic carbon content and poor in nitrogen content. Phosphorus level was found to be normal. Available potassium also was found to be within the normal range for most of the sites. Calcium, magnesium and electrical conductivity, were also found to be within the normal range.

4.12 Analysis of Variance: ANOVA

One- way ANOVA was performed on raw soil data using the SAS system.

In an effort to determine the relationship of altitude and sites with composition of soils, ANOVA was done taking sites as blocks. At 0.05% significance level, pH, Carbon, Calcium and Magnesium were found to be significantly affected by altitude. It was found that the altitude, site and their interaction had significant influence on soil pH. The altitude and site variations had significant effect on soil organic carbon content, but their interaction had no such effect. In contrast to this, levels of phosphorus and potassium and electrical conductivity seemed to be affected only by the interaction of site and altitude. At 1% significance level, following soil parameters were found to be significant for sites

i. pH
ii. Organic carbon
iii. Total nitrogen
iv. Combined effect of Organic carbon and nitrogen
v. Available Potassium
vi. Exchangeable Calcium

Following soil parameters were found to be significant for altitudes at 1% significance level.

i. Organic carbon
ii. Exchangeable Calcium

Table.11. Results of ANOVA at 1% (*) and 5% (*****) significant level. ⟶

ANOVA for altitude and site dependent variations in soil characteristics

	Model			Block			Altitude		
	Degree of freedom	F Value	Significance level	Degree of freedom	F Value	Significance level	Degree of freedom	F Value	Significance level
pH	4	5.61	*	2	7.69	*	2	3.52	*****
C	4	9.72	*	2	7.66	*	2	11.79	*
N	4	3.27		2	6.07	*	2	0.46	
P	4	0.94		2	1.33		2	0.55	
EC	4	2.29		2	1.77		2	2.82	
CN	4	5.00	*	2	8.73	*	2	1.27	
K	4	4.29	*	2	7.42	*	2	1.16	
Ca	4	5.83	*	2	4.27	*****	2	7.40	*
Mg	4	3.12	*****	2	2.54		2	3.70	*****
CaMg	4	1.31		2	2.55		2	0.07	

Table 11.

Following soil parameters were found to be significant for altitudes at 5% significant level –

i. pH

ii. Exchangeable Magnesium

Thus, it appears that the specificity for altitudes found in the distribution pattern of *Frankia* genotypes were mainly influenced by the organic carbon and exchangeable calcium content of the soils, while soil pH and exchangeable magnesium also had some effect on the distribution pattern of the genotypes.

When sites and altitudes were considered together, following soil parameters were found to be significant at 1% and 5% significance level –

i. pH

ii. Organic carbon

iii. Combined effect of organic carbon and nitrogen

iv. Available Potassium

v. Exchangeable Calcium

Soil parameters those were significant for both sites and altitudes only at 5% significance levels were–

i. Total nitrogen

ii. Exchangeable Magnesium

Thus, when sites and altitudes were considered together, distribution of *Frankia* genotypes were found to be influenced mainly by pH, organic carbon, CN, available potassium and exchangeable calcium. Though total nitrogen and exchangeable magnesium were also found to have some influence, their effects were less.

Values of different soil parameters at 1% and 5% significance level are given in Table 11.

4.13 Student-Newman-Keuls test (SNK):

The SNK test is a method for separations tests that can be used with ANOVA. Since ANOVA showed a significant effect of treatment for altitude we tried this test to determine which altitudes were different from others. It was found that for pH low altitudes were showing higher values than middle altitudes i.e. low altitudes were more significant when we consider variation of pH with altitudes (Table 12). Similarly, for organic carbon middle and low altitudes did not differ among each other but both of them differed with high altitudes. For exchangeable calcium and magnesium middle altitudes were found to be more significant than low and high altitudes.

4.14 Principal Components Analysis (PCA):

Principal components analysis is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data. The other main advantage of PCA is that once the patterns have been found in the data, the data is compressed, i.e. by reducing the number of dimensions, without much loss of information.

The principal components analysis was performed following the method described by Sneath and Sokal (1973) [SAS Institute, Inc. (1985)]. This analysis helps in explaining the proportion of relative contribution of various characters to the total variance of a population (Abadie *et al.*, 1998).

This multivariate statistical computation for PCA was performed to find out if different genotypes have a relationship with soil characteristics by using the statistical software package STATISTICA (version 5.0).

Soil Parameter	SNK Grouping	Mean	Altitude	Data Interpretation
pH	A	5.33	H	The low and middle altitudes differ.
	BA	5.19	M	
	B	5.64	L	
C	A	1.45	H	Middle and high and low and high altitudes do differ.
	A	2.05	M	
	B	1.81	L	
N	A	0.05	H	Low altitude differ with high and middle altitudes ?
	A	0.06	M	
	A	0.48	L	
P	A	14.88	H	Altitudes do not differ among themselves.
	A	13.69	M	
	A	13.50	L	
EC	A	0.05	H	Altitudes do not differ among themselves.
	A	0.13	M	
	A	0.20	L	
K	A	257.73	H	Altitudes do not differ among themselves.
	A	359.89	M	
	A	291.78	L	
Ca	A	1.56	H	Middle and low and middle and high differ.
	B	3.70	M	
	B	2.34	L	
Mg	A	0.75	H	Middle and high altitudes do differ.
	BA	1.50	M	
	B	1 26	L	

Table 12.

The principal components for all the parameters are presented in Table 13. The eigenvalues and total variance in percentage and cumulative percentage are given in the Table 14.

Principal components analysis was performed to find out if different genotypes had any relationship with soil characteristics, altitudes and sites. STATISTICA (version 5.0) was used for this purpose.

The analysis gave eleven principal components (PCs) or factors.

The largest contribution to the variation (22.96%) was made by the first component. The first four components accounted for 66.54% of the total variability and they were considered for discussion.

The highest proportion of total variability exposed by the first component (22.96%) was contributed mainly by parameters like organic carbon, potassium, electrical conductivity, nitrogen, sites and pattern. High negative values were observed for parameters like magnesium and altitude followed by calcium and phosphorus.

18.42% of the total variability was contributed by the second PC. High positive values were observed for parameters like pH, electrical conductivity and magnesium. High negative values were observed for pattern, carbon, nitrogen followed by altitude.

The third component accounted for 14.00% of the total variability. The traits showing greater influence in this component were sites and magnesium followed by pH. High negative values were found in available potassium, nitrogen, altitude, phosphorus followed by calcium, electrical conductivity and carbon.

The fourth PC contributed 11.14% to the total variability. The parameters contributing the highest positive values were phosphorus, pattern, calcium followed by electrical conductivity, while the high negative contributions were from parameters like organic carbon, potassium, altitude, nitrogen, magnesium followed by sites and pH.

Table.13. Principal components for all the parameters along with the eigenvectors.

Table.14. Eigenvalues, total variance in percentage and cumulative percentage.

Factor Loadings (Unrotated) (new.sta)
Extraction: Principal components
(Marked loadings are > .700000)

	Factor 1	Factor 2	Factor 3	Factor 4
pH	.217536	.877903	.041223	-.000606
CARB	.626816	-.281315	-.215135	-.336331
PHOS	-.054528	.228601	-.375217	.732732
POT	.426200	.271850	-.540177	-.273252
ELEC_CON	.474216	.750217	-.223512	.086208
CAL	-.287537	.125502	-.293154	.195837
MAG	-.696464	.353989	.103808	-.207867
NITRO	.405077	-.263288	-.538449	-.226432
ALTITUDE	-.638879	-.069768	-.524363	-.264193
SITES	.554034	.188173	.589864	-.156049
PATTERN	.478975	-.487192	-.029335	.516547
Expl.Var	2.526087	2 026850	1.540918	1.225915
Prp.Totl	.229644	.184259	.140083	.111447

Table.13.

Eigenvalues (new.sta)
Extraction: Principal components

	Eigenval	% total Variance	Cumul. Eigenval	Cumul. %
1	2.526087	22.96443	2.526087	22.96443
2	2.026850	18.42591	4.552937	41.39034
3	1.540918	14.00835	6.093855	55.39868
4	1.225915	11.14468	7.319770	66.54337

Table.14.

For studying the relationship of pattern with various other variables, different factors were considered. But since this study was aimed to find out the effect of altitude on the distribution of patterns in particular, the factors which accounted most for these two traits were considered first. Factor 2 accounted for the highest value for altitude and factor 4 accounted for the highest value for pattern. Therefore, the graph which was plotted with factor 2 and factor 4, gave a clear indication of the relatedness of these two traits.

When factor 1 was plotted against factor 2 (fig.28.a), it was found that total nitrogen and organic carbon were closely related to pattern and sites. Altitude and magnesium were also found to be grouping together. Calcium was found to be distantly related to them. Fig.28.b shows factor 1 against factor 3. Here pattern was found to be related with carbon, electrical conductivity, nitrogen and potassium and also distantly with sites. Altitude was related to magnesium.

When factor 1 was plotted against factor 4 (fig.28.c.), total nitrogen and organic carbon were found to be more closely related to sites than to pattern, but we can consider them within the same group. Magnesium and altitude were more closely related in this case, while calcium was distantly grouping as observed in earlier cases.

Factor 4 accounted for largest variability for patterns and factor 2 accounted for largest variability for altitude. Therefore, plotting these two factors revealed the relationship between patterns and altitudes, which were found to be related (Fig. 28d). Factor 3 accounted for largest variation for sites. Therefore, plotting factor 3 and factor 2 revealed the relationship between altitude and sites (Fig. 28e). Comparing these two plots, it became clear that altitude and patterns were related. However, patterns were more often related to carbon and nitrogen status of soil. Therefore, possibly, the altitude exerted its effect through these two soil parameters.

Fig.28. Graphical representation of Principal components analysis (PCA). ⟶

(a) – Factor 1 vs Factor 2
(b) - Factor 1 vs Factor 3
(c) - Factor 1 vs Factor 4

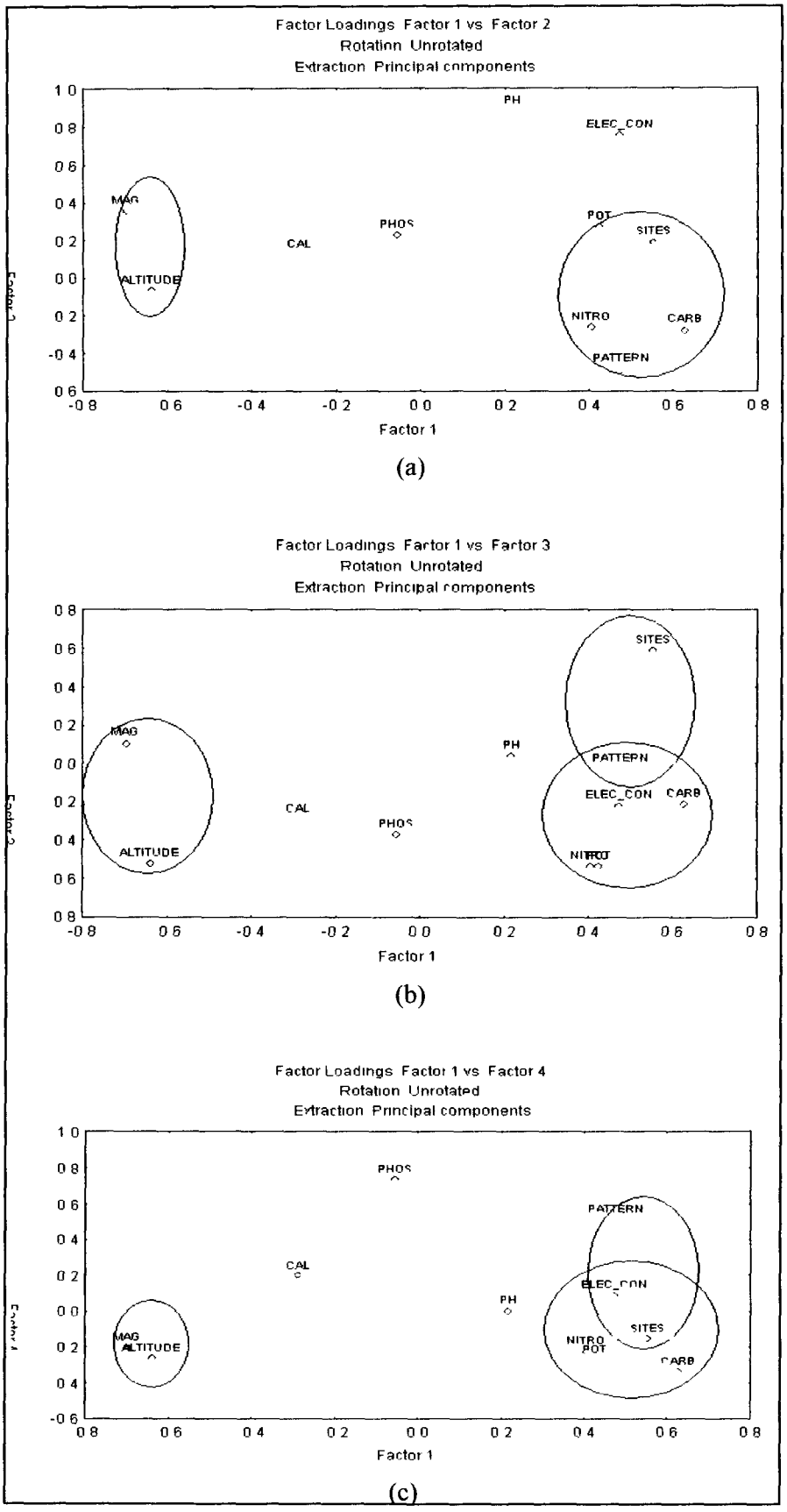

Factor Loadings Factor 1 vs Factor 2
Rotation Unrotated
Extraction Principal components
PH
ELEC_CON
MAG
PHOS
POT
SITES
CAL
ALTITUDE
NITRO
CARB
PATTERN
Factor 1
(a)
Factor Loadings Factor 1 vs Factor 3
Rotation Unrotated
Extraction Principal components
SITES
MAG
PH
PATTERN
ELEC_CON
CARB
CAL
PHOS
ALTITUDE
NITRO
Factor 1
(b)
Factor Loadings Factor 1 vs Factor 4
Rotation Unrotated
Extraction Principal components
PHOS
PATTERN
CAL
ELEC_CON
PH
MAG
ALTITUDE
SITES
NITRO
POT
CARB
Factor 1
(c)

Fig.28. Graphical representation of Principal components analysis (PCA). ⟶

(d) – Factor 2 vs Factor 3
(e) - Factor 2 vs Factor 4

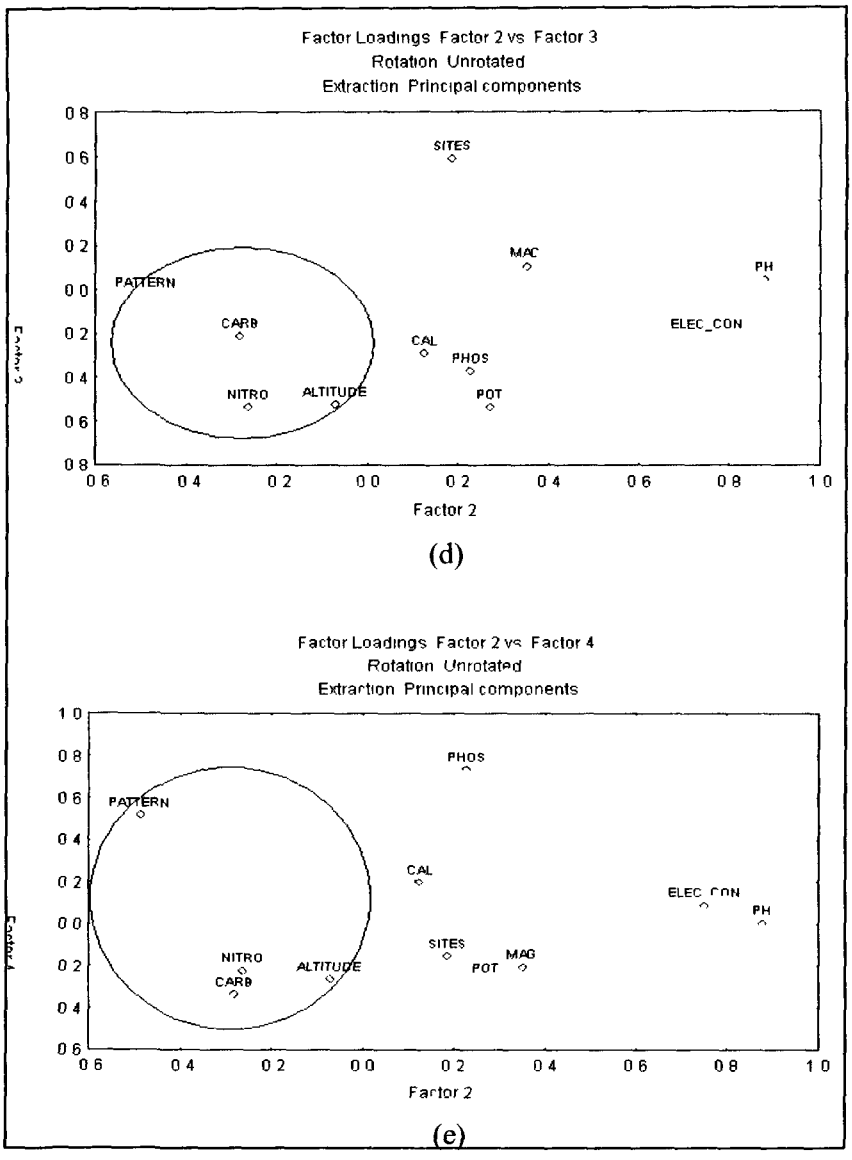

Factor Loadings Factor 2 vs Factor 3
Rotation Unrotated
Extraction Principal components
SITES
MAC
PH
PATTERN
CARB
ELEC_CON
CAL
PHOS
NITRO
ALTITUDE
POT
Factor 2
(d)
Factor Loadings Factor 2 vs Factor 4
Rotation Unrotated
Extraction Principal components
PHOS
PATTERN
CAL
ELEC_CON
PH
SITES
MAG
POT
NITRO
ALTITUDE
CARB
Factor 2
(e)

Fig.28

Table 15. Multiple Regression Analysis between Banding Patterns and Environmental Parameters ⟶

Environmental parameters	Beta	Std. Error of Beta	P - level
pH	- 0.089443	0.133389	0.504467
Organic Carbon	0.175712	0.105691	0.100373
Available Phosphorus	0.239101	0.097588	0.016492
Available Potassium	0.117575	0.106178	0.271511
Electrical Conductivity	- 0.267485	0.138934	0.057795
Exchangeable Calcium	- 0.022803	0.091795	0.804458
Exchangeable Magnesium	- 0.328012	0.104506	0.002386
Total Nitrogen	- 0.066537	0.099423	0.505298
Altitude	-.400762	0.107576	0.000364
Sites	- 0.172907	0.112945	0.129789

Table 15.

Therefore, from the above observations it became clear that the distribution of patterns varied with altitudes but their relationship appeared to be determined by the status of organic carbon and total nitrogen in the soil.

4.15 Multiple Regression Analysis (MRA):

To find out the highest significant variable which played the most important role in the distribution of genotypes, Multiple regression analysis was carried out considering patterns as dependent variable and all others parameters as independent variables. This analysis revealed altitude as the most significant parameter with the highest beta value of -.400762 at 0.000364% significance level (Table 15). The second most significant parameter was magnesium with a beta value of - 0.328012 at 0.002386% significance level and the third most important parameter was phosphorus with a beta value of 0.239101 at 0.016492% significance level. Organic carbon though was not shown as a significant parameter, a beta value of 0.175712 at 0.100373% significance level showed it as an indicative parameter.

From above observations it became clear that:

- *Frankia* genotypes tended to be distributed in an altitude specific manner.
- Although site dependent distribution was not very strong, some genotypes were found to be distributed in a site-specific manner.

Conclusion

CONCLUSION

1. Certain *Frankia* genotypes were found to be site specific, having narrow geographical distribution.
2. Some *Frankia* genotypes were found to cross boundaries of sites within the same altitude, showing wide distribution.
3. Most often *Frankia* strains were altitude specific. Some strains were not found to cross the boundaries of altitude. Therefore, altitude played the most important role in determining the distribution of *Frankia* genotypes.
4. Among soil parameters, carbon and nitrogen were the key determinant factors which were playing important roles in determining the distribution pattern of *Frankia* genotypes.
5. Status of magnesium and phosphorus content of soils had coincidental associations with altitude.
6. Efficient *Frankia* genotypes can be selected from low altitude sites since better nitrogen fixers were found at lower altitudes.

References

REFERENCES

Abadie T., J.R. Magalhaea, C. Cordeiro, S.N. Parentoni. and R.V. de Andrade. (1998). A classification for Brazilian maize landraces. Plant Genetic Resources Newsletter. **114:**43-44.

Akkermans, A. D. L., M. S. Mirza, H. J. M. Harmsen, H. J. Blok, P. R. Herron, A. Sessitsch and W. M. Akkermans. (1994). Molecular ecology of microbes: A review of promises, pitfalls and true progress. FEMS Microbiology Reviews **15**:185-194.

Akkermans, A. D. L. and W. Roelofsen. (1980) Symbiotic nitrogen fixation by actinomycetes in *Alnus*-type root nodules. In: Nitrogen fixation. Proc. Symp. Phytochemical Society of Europe, W. D. P. Stewart and J. R. Galon (eds.), Academic Press London. pp. 279-299.

Akulinin, I.A. (1958). Using sea buckthorn oil in treating burns. *Sov. Med.* **11:**137-138.

Allen, S. E., H. M. Grimshaw, J. A. Parkinson and C. Quarmby. (1974). Chemical Analysis of Ecological Materials. Blackwell Scientific Publications, Oxford, England.

An, C. S., W. S. Riggsby and B. C. Mullin. (1985a). Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. Int. J. Syst. Bacteriol. **35:** 140-146.

An, C. S., W. S. Riggsby and B. C. Mullin. (1985b). Restriction pattern analyses of genomic DNA of *Frankia* isolates. Plant and Soil. **87:** 43-48.

Baker D., L. P. Willium and J. G. Torrey. (1981). Immunochemical analysis of relationship among isolated frankiae (Actinomycetales). Int. J. Syst. Bacteriol. **31**: 148-151.

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

Baker, D. D. and C. R. Schwintzer. (1990). Introduction. In: The Biology of *Frankia* and Actinorhizal plants. C.R.Schwintzer and J. Tjepkema (eds.), Academic Press, California. pp. 3-11.

Benson, D. R. (1988). The genus *Frankia*: actinomycete symbionts of plants. Microbial Sci. **5(1):** 9-12.

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

Benson, D. R., S. E. Buchholz and D. G. Hanna. (1984). Identification of *Frankia* strains by 2-D-polyacrylamide gel electrophoresis. Appl. Envron. Microbiol, **47:** 489-494.

Blaha D. and B. Cournoyer. (2003). Cloning, characterization and phylogenetic analysis of the *sigA* σ^{70} factor gene sequence from the actinomycete *Frankia*. Plant and Soil, **254:** 97-106.

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

Bormann, B. T. and J. C. Gordon. (1984). Study of density effects in young red alder plantations: productivity photosynthate partitioning and nitrogen fixation. Ecology **65:**394-402.

Bosco M., M. P. Fernandez, P. Simonet, R. Materassi and P. Normand. (1992). Evidence that some *Frankia* sp. strains are able to cross boundaries between

Alnus and *Elaeagnus* host specificity groups. Appl. Environ. Microbiol. **58 (5):** 1569-1576.

Byrnes, L., B. Enenkel, F. Gannon and T. Smith. (1995). Preparation of DNA for use as probes. In: Gene Probes 1: A practical approach. B. D. Hames and S. J. Higgins (eds.), IRL Press, Oxford. pp. 17-55.

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

Chauhan, V. (2000). Generation of Molecular Signatures for *Alnus nepalensis* genotypes with high nitrogenase activity in symbiotic association with *Frankia.* Ph.D thesis, NEHU, Shillong, India.

Chen, H. B., J. M. Martin, M. Lanvin and L. E. Talbert. (1994). Genetic diversity in hard red spring wheat based on Sequence-Tagged-Site PCR markers. Crop Sci. **34:** 1628-1632.

Clawson M. L., M. Caru and D. R. Benson. (1998). Diversity of *Frankia* strains in root nodules of plants from the families Elaeagnaceae and Rhamnaceae. Appl. Environ. Microbiol. **64 (9):** 3539-3543.

Crocker, R. L., and J. Major. (1955). Soil development in relation to vegetation and surface age at Glacier Bay, Alaska. J. Ecol.**43:**427-448.

Dai Y., J.Cao, X. Tang, and C. Zhang. (2004). Diversity of *Frankia* in nodules of *Alnus nepalensis* at Gaoliong Mountains revealed by IGS, PCR-RFLP analysis. Ying Yong Sheng Tai Xue Bao, **15(2):** 186-90.

Damière, C., A. Capellano, and A. Moiroud. (1986). Dynamique de l'azote dans un peuplement natural d' *Alnus incana* (L.) Moench. Acta Oecol. Plant **7(2):**165-175.

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

Dobrista, S. V. (1985). Restriction analysis of the *Frankia* spp. Genome. FEMS Microbiol. Lett. **29:**123-128.

Domenach A. M., F. Kurdali, and R. Bardin. (1989). Estimation of symbiotic dinitrogen fixation in alder forest by the method based on Natural ^{15}N abundance. Plant Soil **118:**51-59.

Dommergues Y. R. (1996). Nitrogen fixation in actinorhizal plants: Application. Acta Botanica Gallica. **143:** 663-679.

Fenning T. M. and J Gershenzon. (2002): Where will the wood come from? Plantation forests and the role of biotechnology. Trends in Biotechnology **20**: 291-296

Fernandez, M. P., H. Meugnier, P. A. D. Grimont, and R. Bardin. (1989). Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int. J. Syst. Bacteriol. **39**: 424-429.

Fessenden, R. J. (1979). Use of actinorhizal plants for land reclamation and amenity planting in the USA and Canada. In: Symbiotic Nitrogen Fixation in the management of Temperate Forests J. C. Gordon, C. T. Wheeler, D. A. Perry (eds.). Oregon State University, Corvallis,OR, pp. 403-419.

Furlow J. J. (1979). The systematics of the American species of *Alnus* (Betulaceae). Rhodora **81:** 1-121, 151-248.

Ganesh, G., A. K. Misra, C. Chapelon and P. Normand. (1994). Morphological and molecular characterization of *Frankia* sp. isolate from nodules of *Alnus nepalensis*. Don. Arch. Microbiol. **161:** 152-155.

Gardes, M., and M. Lalonde. (1987). Identification and subgrouping of *Frankia* strains using sodium dodecyl sulphate-polyacrylamide gel electrophoresis. Physiol. Plant. **70**: 237-244.

Gardes, M., J. Bousquet, and M. Lalonde. (1987). Isozyme Variation among 40 *Frankia* Strains. Appl. Environ. Microbiol. **53(7):**1596–1603.

Ge, X., G. Shi, and Y. Zhang. (1985). Application of sea-buckthorn in medicine. *Shanxi Med. Res.* **2:**9-14.

Ghareyazie, B., N. Huang, G. Second, and G. S. Khush. (1995). Classification of rice germplasm. I. Analysis using AFLP and PCR based RFLP. Theor. Appl. Genet. **91:** 218-227.

Gordon, J. C., J. O Dawson. (1979). Potential uses of nitrogen fixing trees and shrubs in commercial forestry. Bot. Gaz. **140** (Suppl):s88-s90.

Guofan L., and D. Tingxiu. (1987). A study of nodulation and nitrogen fixation of alder on the purplish soils in China. Plant Soil, **99:** 285-290.

Hadrys, H., M. Balic and B. Schierwater. (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecology, **1:** 55-63.

Hahn, D., M. Dorsch, E. Stackbrandt and A. D. L. Akkermans. (1989). Synthetic oligonucleotide probes for identification of *Frankia* strains. Plant Soil. **118:** 211-219.

Harry, D. E., D. C. Yang and J. O. Dawson. (1991). Nucleotide sequence and diversity in the 16S ribosomal RNA from *Frankia.* Plant Soil. **131:** 143-146.

Heilman, P. (1982). Rates of nitrogen and organic matter accumulation in coal mine spoils supporting red alder stands. Can. J. For. Res. **12:**809-813.

Hery, M., S. Nazaret, J. Tanguy, P. Normand, and E. Navarro. (2001). Adaptation of bacterial communities from Caledonian soils under *Gymnostoma* sp. to nickel

spiking. Presented at the 12th International Meeting on *Frankia* and Actinorhizal Plants, at Carry-le-Rouet, France.

Hesse, P. R. (1971). A Textbook of Soil Chemistry Analysis, Chem. Publishing Co., N.Y. pp 88-104.

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.

Jensen, M. A., J. A. Webster and N. Straus. (1993). Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Appl. Environ. Microbiol. **59:** 945-952.

Jeong S. C., N. J. Ritchie and D. D. Myrold. (1999). Molecular phylogenies of plants and *Frankia* support multiple origins of actinorhizal symbioses. Mol. Phylogenet. Evol. **13(3):** 493-503.

Laguerre, G., P. Mavingui, M. Allard, M. Charnay, P. Louvrier, S. Mazurier, L. Rigottier-Gois and N. Amarger. (1996). Typing of Rhizobia by PCR DNA fingerprinting and PCR Restriction Fragment Length Polymorphism analysis of chromosomal and symbiotic gene regions: Application to *Rhizobium leguminosarum* and its different biovars. Appl. Environ. Microbiol. **62 (6):** 2029-2036.

Lavire, C., and B. Cournoyer. (2003). Progress on the genetics of the N_2-fixing actinorhizal symbiont *Frankia*. Plant and Soil, **254:** 125-137.

Lawrence, D. B. (1951). Recent glacier history of Glacier Bay, Alaska, and development of vegetation on deglaciated terrain with special reference to the importance of alder in the succession. Yearb. Am. Phil. pp:175-178.

Lechevalier M. P. (1984). The taxonomy of genus *Frankia*. Plant Soil. **78:**1-6.

Lechevalier M. P. (1994). Taxonomy of the genus *Frankia* (Actinomycetales). Int. J. Syst. Bacteriol. **44:** 1-8.

Lechevalier M. P. and H. A. Lechevalier. (1989). Genus *Frankia* Brunchrost (1886), 174 AL. In: Bergey's manual of systematic bacteriology. S.T. Williams, M.E Sharp and J.G. Holt (eds.). The Williams and Wilkins Co., Baltimore, USA. pp 2410-2417.

Li, T. S. C. and W. R. Schroeder. (1996). Sea buckthorn (*Hippophae rhamnoides* L.): a multipurpose plant. Hort. Technology, **6:**370-380.

Mirza, M. S., D. Hahn and A. D. L. Akkermans. (1992). Isolation and characterization of *Frankia* strains from *Coriaria nepalensis*. Syst. Appl. Microbiol. **15:**289-295.

Mirza, M. S., S. Hameed and A. D. L. Akkermans. (1994). Genetic diversity of *Datisca cannabina*-compatible *Frankia* strains as determined by sequence analysis of the PCR-amplified 16S rRNA gene. Appl. Environ. Microbiol. **60:** 2371-2376.

Misra A. K., E. Paget, P. Simonet and P. Normand. (1991). Isolation and amplification of *Frankia* DNA from nodules of *Alnus nepalensis* and studies on overcoming foreign DNA contamination. Indian J. Expt. Biol. **29:** 946-948.

Montpetit, D. and M. Lalonde. (1988). *In vitro* propagation and subsequent nodulation of the actinorhizal *Hippophae rhamnoides* L. Plant Cell, Tissue and Organ Culture. **15:**189-199.

Mullis, K. B. and F. A. Faloona. (1987). Specific synthesis of DNA *in vitro* via polymerase catalysed chain reaction. Methods Enzymol. **155:** 335-350.

Mullis, K. B., F. A. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. Cold spring Harbor Symp. Quant. Biol. **51:** 263-273.

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 endophytes of the actinorhizal shrub *Ceanothus*. Physiol. Plant. **99:**714-721.

Myrold, D. D. and N. J. Ritchie. (1999). Geographic distribution and genetic diversity of *Ceanothus*-infective *Frankia* strains. Appl. Environ. Microbiol; **65(4)**: 1378-1383

Navarro, E., P. Simonet, P. Normand and R. Bardin. (1992). Characterization of natural populations of *Nitrobacter* spp. Using PCR-RFLP analysis of the ribosomal intergenic spacer. Arch. Microbiol. **157:** 107-115.

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(13)**: 4072-4078.

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:** 175-181.

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 *Frankia*: a review. Plant Soil. **90:** 429-453.

Normand, P., M. Gouy, B. Cournoyer and P. Simonet. (1992a). Nucleotide sequence of *nif* D from *Frankia alni* strain ArI3: Phylogenetic inferences. Molecular Biology and Evolution. **9**: 495-506.

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

Normand, P., P. Simonet and R. Bardin. (1988). Conservation of *nif* sequences in *Frankia.* Mol. Gen. Genet. **213:** 238-246.

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. Syst Bacteriol. **46:**1-9.

Perry, D. J., Isabel, N. and J. Bousquet. (1999). Sequence-Tagged-Site (STS) markers of arbitrary genes: the amount and nature of variation revealed in Norway spruce. Heredity. **83:** 239-248.

Ponder, F. (1983). Effect of autmn-olive on the mineral composition of black walnut leaves. Commun. Soil. Sci. Plant. Anal. 14: 1253-1263.

Ponsonet, C. and X. Nesme. (1994). Identification of *Agrobacterium* strains by PCR-RFLP analysis of pTi and chromosomal regions. Arch. Microbiol. **161(4):**300-309.

Prin Y., L. Maggia, B. Picard, H. G. Diem and P. Goullet. (1991). Electrophoretic comparison of enzymes from 22 single spore cultures obtained from *Frankia* strain ORS140102. FEMS Microbiol. Lett. **77:** 223-228.

Ritchie, N. J. and A. M. David. (1999). Geographic distribution and genetic diversity of *Ceanothus*-infective *Frankia* strains. Appl. Environ. Microbiol. **65:** 1378-1383.

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

Ruvkun, G. B. and F. M Ausubel. (1980). Interspecies homology of nitrogenase genes. Proc. Natl. Acad. Sci. U. S. A. **77(1):**191-5.

Sambrook, J., E. F. Fritsch and T. Maniatis. (1989). Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen and A. R. Coulson. (1977). DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci., USA, **74:** 5463-5467.

Sarma G., A. Sen, R. Varghese and A. K. Misra. (1998). A novel technique for isolation of *Frankia* and generation of single spore cultures. Can. J. Microbiol. **44:** 490-492.

SAS Institute Inc. (1996). SAS/SAT User's guide, Cary, NC.

Schloter M., M. Lebuhn, T. Heulin, and A. Hartmann. (2000). Ecology and Evolution of bacterial microdiversity. FEMS Microbiol. Reviews. **24(5):** 647-660.

Schwintzer, C. R. and J. D. Tjepkema. (eds.). (1990). The Biology of *Frankia* and Actinorhizal Plants. Academic Press, New York, 389pp.

Shantharam, S. and A. K. Mattoo. (1997). Enhancing biological nitrogen fixation: An appraisal of current and alternative technologies for N input into plants. Plant Soil, **194:** 205-2167.

Shipton, W. A. and A. J. P. Burggraf. (1982). A comparison of the requirement for various carbon and nitrogen sources and vitamins in some *Frankia* isolates. Plant Soil. **69:** 149-161.

Silvester,W. B. (1977). Dinitrogen fixation by plant associations excluding legumes. In: R. W. F. Hardy and A. H. Gibson (eds.). A treatise on nitrogen fixation. Agronomy and Ecology, New York, John Wiley & Sons. **4:** 141-190.

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(11):** 3278-3286.

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

Small E., P. M. Catling, and T. S. C. Li. (2000). Blossoming Treasures of Biodiversity: 5. Sea Buckthorn (*Hippophae rhamnoides*)-an ancient crop with modern virtues. Biodiversity **3(2)**: 25-27.

Sneath, P. H. A. and R. R. Sokal. (1973). Numerical Taxonomy: The Principles and Practice of Numerical Classification, W.H. Freeman and Company, San Francisco.

Spirovski F., K. Stojanoski and A. Mitrevski. (2005). Cluster Analysis Method in Electrophoretic Pattern Classification. Journal of the Iranian Chemical Society. **2 (1):** 26-31.

Stackebrandt, E. (1986). The significance of "wall types" in phylogenetically based taxonomic studies on actinomycetes. In: Biological, Biochemical and Biomedical Aspects of Actinomycetes. G. Szabo, S Biro and M Goodfellow (eds.). Akademiai Kiado, Budapest. pp. 497-506.

Stewart W. P. D., G. P. Fitzgerald and R. H. Burris. (1968). Acetylene reduction in nitrogen fixing blue-green algae. Arch. Microbiol. **62:** 336-348.

Swensen, S. M. and B. C. Mullin. (1997). The impact of molecular systematic on hypotheses for the evolution of root nodule symbioses and implications for expanding symbioses to new host plant genera. Plant Soil, **194:** 185-192.

Swensen, S.M. (1996). The evolution of actinorhizal symbiosis: Evidence for multiple origins of the symbiotic association. Am. J. Bot. **83 (11):** 1503-1512.

Tang X., Y. Dai, Z. Xiong, Z. Zhang and C. Zhang. (2003). Effect of natural stress on genetic diversity of *Frankia* in *Alnus* nodules. Ying Yong Sheng Tai Xue Bao, **14(10):** 1743-6.

Tarrant R. F. and J. M. Trappe. (1971). The role of *Alnus* in improving the forest environment. Plant and Soil **(special vol.):** 335-348.

Thompson J. D., D. G. Higgins and T. J. Gibson. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighing position specific gap penalties and weigh matrix choice. Nucleic Acid Res. **22:** 4673-4680.

Tyron R. C. *Cluster analysis*. Ann. Arbor, MI: Edwards Brothers, 1939.

Tsumura, Y., Y.Suyama, K.Yoshimura, N. Shirato and Y. Mukai. (1997). Sequence-Tagged-Sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. Theor. Appl. Genet. **94:** 764-772.

Varghese R. 2000. Investigations on molecular diversity in alder compatible Frankiae. Ph.D. thesis, NEHU,Shillong, India.

Varghese, R., V. S. Chauhan and A. K. Misra. (2003a). Evolutionary implications of nucleotide sequence relatedness between *Alnus nepalensis* and *Alnus glutinosa* and between corresponding *Frankia* micro-symbionts. Plant Soil **254:** 219-227.

Varghese, R., V. S. Chauhan and A. K. Misra. (2003b). Hypervariable spacer regions are good sites for developing specific PCR-RFLP markers and PCR primers for screening actinorhizal symbionts. J. Bioscience **48(4):** 437-442.

Verghese S. K. and A. K. Misra. (2002). *Frankia*-Actinorhizal symbiosis with special reference to host-microsymbiont relationship. Current Science. **83(4)**:404-408.

Verghese, S. K. and A. K. Misra. (2000). PCR-RFLP based screening of *Frankia* in alder nodules having different levels of nitrogenase activity. Symbiosis. **28**: 337-350.

Walkley, A. and I. A. Black, (1934). An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Science, **37:** 28-29.

Wall, L. G. (2000). The Actinorhizal Symbiosis. J. Plant Growth Regulation. **19(2)**: 167-182.

Woese C. R. (1987). Bacterial evolution. Microbiol. Rev. **51:** 221-271.

Woese, C. R. and G. E. Fox. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA. **74:** 5088-5090.

Zavitkovski, J., E. A. Hansen and M. A. McNiel. (1979). Nitrogen-fixing species in short rotation systems for fiber and energy production. In: Symbiotic nitrogen fixation in the management of temperate forests. J. C. Gordon, C. T. Wheeler, and D. A. Perry (eds.). Oregon State University, Corvallis. pp. 388-402.

Zhang, Z., M. F. Lopez and J. G. Torrey. (1984). A comparison of cultural characteristics and infectivity of *Frankia* isolates from root nodules of *Casuarina* species. Plant and Soil. **78:** 79-90.

Appendices

APPENDIX 1

BUFFERS USED FOR EXTRACTION OF DNA AND AGAROSE GEL ELECTROPHORESIS

1. EXTRACTION BUFFER FOR ISOLATION OF DNA (pH 8.0)

COMPONENT	AMOUNT/100ML
Tris base (0.1M/100mM)	50 ml
EDTA (0.1M/100mM)	4 ml
NaCl (1.4M)	8.18gm
CTAB	2% (wt./vol.)
PVP	1% (wt./vol.)

CTAB: Cetyl Trimethyl Ammonium Bromide
PVP : Polyvinylpyrrolidone

2. TBE BUFFER [5X] (pH 8.0)

COMPONENT	AMOUNT/LITRE
Tris base	54.0g
Boric acid	27.5g
EDTA (0.5M)	20 ml

3. TYPE III LOADING BUFFER (6X)

Bromophenol blue	0.25% (wt./v.)
Xylene Cyanol FF	0.25% (wt./v.)
Glycerol	30% in water (wt./v)

APPENDIX 2

BUFFERS USED FOR PCR AND RESTRICTION ANALYSIS

1. **10X PCR BUFFER** (pH 8.3 at 25°C)

Tris-HCl	100 mM
KCl	500 mM
$MgCl_2$	15 mM
Gelatin	0.01% (wt./v.)

2. **Buffer L [Roche]** (pH 7.5 at 37°C)

Tris-HCl	10 mM
$MgCl_2$	10 mM
Dithioerythritol	1 mM

APPENDIX 3

High Altitude, Site 1

(http://align.genome.jp/)

```
CLUSTAL W (1.83) multiple sequence alignment


1               ----------------------------------------GGGGTCCGTAAGGGTCCTGCA
8               ----------------------------------------GGGGTCCGTAAGGGTCCTGCA
4               ----------------------------------------GGGGTCCGTAAGGGTCCTGCA
6               GGGGTCCGTAAGGGTCCCCGAATGGGGAAACCCGGAGGTGGGGTCCGTAAGGGTCCTGCA
                                                        ********************

1               CAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
8               CAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
4               CAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
6               CAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
                ************************************************************

1               AGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGTCGGGGACTCATAGGAGACTGCCG
8               AGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGTCGGGGACTCATAGGAGACTGCCG
4               AGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGTCGGGGACTCATAGGAGACTGCCG
6               AGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGTCGGGGACTCATAGGAGACTGCCG
                ************************************************************

1               GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATG--CCCCTTACGTCCTGG
8               GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATG--CCCCTTACGTCCTGG
4               GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCCCTTACGTCCTGG
6               GGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATG--CCCCTTACGTCCTGG
                ****************************************  ***************

1               GCTGCACACATGCTACAA-TGGCCGGTACCAAAGGGCT--GCGATGCCGTGAGGTGGAGC
8               GCTGCACACATGCTACAA-TGGCCGGCAC-AAAGGGCT--GCGATGCCGTGAGGTGGAGC
4               GCTGCACACATGCTACAAATGGCCGGCACAAAGGGGCTCTGCGATGCCGTGAGGTGGAGC
6               GCTGCACACATGCTACAA-TGGCCGGTAC-AAAGGGCT--GCGATGCCGTGAGGTGGAGC
                ***************** ******* ** ** *****  ********************

1               GAATCCCAAAAAAGCCGGTCTCAGTTCGGA-TCGGGGTCTGCAAACTCGACCCC---GTG
8               GAATCCCAAAAA-GCCGGTCTCAGTTCGGA-TCGGGGTCTGCAA-CTCGACCCC---GTG
4               GAATCCCAAAAAAGCCGGTCTCAGTTCGGAATCGGGGTCTGCAA-CTCGACCCCCGTGTG
6               GAATCCCAAAAA-GCCGGTCTCAGTTCGGA-TCGGGGTCTGCAA-CTCGACCCC---GTG
                ************ ***************** ************* *********   ***

1               AAGTCGGAGTCGCTA-GTAATCGCAGATCAGCAA-TGCTGCGGTGAA-TACGTTCCCGGG
8               AAGTCGGAGTCGCTA-GTAATCGCAGATCAGCAA-TGCTGCGGTGAA-TACGTTCCCGGG
4               AAGTCGGAGTCGCTAAGTAATCGCAGATCAGCAAATGCTGCGGTGAAATACGTTCCCGGG
6               AAGTCGGAGTCGCTA-GTAATCGCAGATCAGCAA-TGCTGCGGTGAA-TACGTTCCCGGG
                *************** ****************** ************ ************

1               CCTTGTACACACCGCCC-GTCACGTCACGAAAGTCGGTAACACCCGAAAGCCGGTGGCCT
8               CCTTGTACACACCGCCC-GTCACGTCACGAAAGTCGGTAACACCCGAA-GCCGGTGGCCT
4               CCTTGTACACACCGCCCCGTCACGTCACGAAAGTCGGTAACACCCGAA-GCCGGTGGCCT
6               CCTTGTACACACCGCCC-GTCACGTCACGAAAGTCGGTAACACCCGAA-GCCGGTGGCCT
                ***************** ***************************** ***********

1               AACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACC-GGCGATTGGGACGAAGTCGTAACAA
8               AACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACC-GGCGATTGGGACGAAGTCGTAACAA
4               AACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACC-GGCGATTGGGACGAAGTCGTAACAA
6               AACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCCGGCGATTGGGACGAAGTCGTAACAA
                ********************************** ************************
```

```
1               GGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGGCTGGTC
8               GGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGGCTGGTT
4               GGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGGCTGGTT
6               GGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCCGGCTGGTC
                ************************************************** *******

1               TGTCCC-GTTGGGGG-TGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCGGTCTGGT
8               TGTCCC-GTTGGGGG-CGGGCTGGTCCAGGGCCAGGGCCAGATGTGTAGGCCGGTCTGGT
4               TGTCCC-GTTGGGGG-TGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCGGTCTGGT
6               TGTCCCTGTTGGGGGGTGGGCTGGTCCAGGGCCAGGACCGGATGTGTAGGCCGGTCTGGT
                ****** ********  ******************** ** ********************

1               TGCTCGAGGGTGTGGA-CGCTGATGATGTGGCTGCTGGCT-TGTCTGGTGCCTAGTACTC
8               TGCTCGAGGGTGTGGA-CGCTGACGATGTAGCTGCTGGCTGTGTCTGGTGCCTAGTACTC
4               TGCTCGAGGGTGTGGA-CGCTGACGATGTGGCTGCTGGCTGTGTCTGGTGCCTAGTACTC
6               TGCTCGAGGGTGTGGAACACTGATGATGTGGCTGCCGGTTCTGTCTGGCTTCTAGTACTC
                **************** * **** ***** ***** ** * *******   *********

1               CTGCTTTTG-GGTGGGGTGGAACGGGGCTGGGTGGAGCTGGTGGTCCGTGGCGCGCTGTT
8               CTGCTTGTG-GGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCGTGGCACGCTGTT
4               CTGCTTGTG-GGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCGTGGCACGCTGTT
6               CTGCCTGTGTGGTGGGGTGGAACGGGGTTGGGTGGGGCGGGTGGTCCGTGGCGCGCTGTT
                **** * ** ***************** ***** * *  ************* *******

1               GGGTCCTGAGGGAGTCAGGCCGGTTGTGGCTGGGACTTTCTTGATCCGCTTTTTGTGGGT
8               GGGTCCTGAGGGAGTCAGGCCAGT-GTGGTCGACGCTTTCTTGATCCGCTTTTTCTGGGT
4               GGGTCCTGAGGGAGTCAGGCCAGT-GTGGTCGACGCTTTCTTGATCCGCTTTTTCTGGGT
6               GGGTCCTGAGGGAGTCAGGCCGGT-GTGGTCTG-GTTTTCTTGATCCGCTTCTGGTGGGT
                ********************** ** ****       *************** *  *****

1               GCCCCTGTCCTCCTGTGTGGGGGGTGGGGGTGT-TCCGCGAGGGTGGTGCGGCTCTGGTC
8               GCCCCTGTCCTCCTATATGGGGGGTGGGGGTGT-TCCGTGAGGGTGGTGCGGG-CTGGTC
4               GCCCCTGTCCTCCTATATGGGGGGTGGGGGTGTGTCCGTGAGGGTGGTGCGGG-CTGGTC
6               GCCCCTGTCCTC---TGTGGAGGGTGGGGGTGT--TCGCGAGGGTGGTGCGGG-CTGGTC
                ************   * *** ************   ** *************  *******

1               GCCGCAGGGTACTGCCTCTTGTTGTGAGGGGTGAGTCCTGCTGGTGCGGCTGGTTGCCGT
8               GTTGCAGGATACTGCCTTCTGC-GTGAGGG-TGAGTTCTGCTGGTGCGGCTGGTTTCTGT
4               GTTGCAGGATACTGCCTTCTGC-GTGAGGG-TGAGTTCTGCTGGTGCGGCTGGTTTCTGT
6               GTCGCAGGATACTGCTTGTCGTTGGACGGG-TGAGTCTTGCTGGTGCGGCTGGTGTCCGT
                *  ***** ****** *   *  *   *** *****  ***************  * **

1               CCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGGGCG
8               CCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGGGCG
4               CCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGGGCG
6               CCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGGGCG
                ************************************************************

1               CACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATGCCTC
8               CACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGCGATATGCCTC
4               CACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGCGATATGCCTC
6               CACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCTATATGCCTC
                ********************************************** **** *********

1               GGGGAGCTGCCAACCTAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG-------
8               GGGGAGCTGTCAACCGAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG-------
4               GGGGAGCTGTCAACCGAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG-------
6               GGGGAGCTGCCATCCGAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGGAGTCCGT
                ********* ** ** ************************************

1               ----------------------------------------------------------
8               ----------------------------------------------------------
4               ----------------------------------------------------------
6               AAGGGGTCCGAATGGGGTCCGTAAGGGTCCGAATGGGGAAACCCGGAGGTGGGGTCCGTA

1               ----------------------------------------------------------
8               ----------------------------------------------------------
4               ----------------------------------------------------------
6               AGGGTCCGAATGGGGAAACCCGGCCGAATGGGGAAACCCGGAGGTGTGGTCCGTAAGGGT
```

```
1               ----------------------------------------------------------
8               ----------------------------------------------------------
4               ----------------------------------------------------------
6               CCGAATGGGGAAACCCGGAGGTGGGGTCCGTAAGGGTCCGAATGGGGAAACCCGGAGGTG


1               ---------------------------------------------------
8               ---------------------------------------------------
4               ---------------------------------------------------
6               GGGTCCGCAAGGGTCCGAAAGGGGAAACCCAGCCGAATGGGGAAACCCGG
```

High Altitude, Site 2

```
CLUSTAL W (1.83) multiple sequence alignment


13              GGGGTCCGTAAGGGTCCTGCAC-AGGTGGTGCATGGCTGTCGTCAGCTC-GTGTCGTGAG
15              GGGGTCCGTAAGGGTCCTGCAC-AGGTGGTGCATGGCTGTCGTCAGCTC-GTGTCGTGAG
11              GGGGTCCGTAAGGGTCCTGCACCAGGTGGTGCATGGCTGTCGTCAGCTCCGTGTCGTGAG
                ********************* ************************* **********

13              ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGT
15              ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGT
11              ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGT
                ************************************************************

13              CGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAAGGTGGGGATGACGTCAAGT
15              CGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAA-GGTGGGGATGACGTCAAGT
11              CGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAA-GGTGGGGATGACGTCAAGT
                ************************************** ******************

13              CATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTGC
15              CATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTGC
11              CATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTGC
                ************************************************************

13              GATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA
15              GATACCGCGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA
11              GATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA
                *** *** **************************************************

13              CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATAC
15              CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATAC
11              CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATAC
                ************************************************************

13              GTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCC
15              GTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCG
11              GTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCG
                ************************************************************

13              GTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTCG
15              GTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTCG
11              GTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTCG
                ************************************************************

13              TAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGG
15              TAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCCCCTCCTT----------------
11              TAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGG
                ********************************** *******

13              CTGGTCTGTCCCGTT--GGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCGG
15              ------------------------------------------------------------
11              CTGGTCTGTCCCTGTTGGGGGGTGGGCTGGTCCAGGGCCAGGACCGGATGTGTAGGCCGG
```

```
13              TCGGGTTGCTCGAGGGTGTGGA-CGCTGACGATGTGGCCGCTGGTTGTGTCTGGTGTCTA
15              ------------------------------------------------------------
11              TCTGGTTGCTCGAGGGTGTGGAACACTGATGATGTGGCTGCCGGTTCTGTCTGGCTTCTA


13              GTACTCCTGCCTGTCTGGGTGGGGTGGAACGGTGCTGGGTGTGGCTGGTGGTCCGTGGCG
15              ------------------------------------------------------------
11              GTACTCCTGCCTGTGTGG-TGGGGTGGAACGGGGTTGGGTGGGGCGGGTGGTCCGTGGCG


13              CGCTGTTGGGTCCTGAGGGAGTCAGGCCGGTTGTGGCTGGGACTTTCTTGATCCGCTTTT
15              ------------------------------------------------------------
11              CGCTGTTGGGTCCTGAGGGAGTCAGGCCGGTGTGGTCTGG--TTTTCTTGATCCGCTTCT


13              TGTGGGTGCCCCTGTCCTCCTGTGTGGGGGGTGGGGGTTCCTGCGAGGGTGGTGCGGGCT
15              ------------------------------------------------------------
11              GGTGGGTGCCCCTGTCCTC---TGTGGAGGGTGGGGGTGTTCGCGAGGGTGGTGCGGGCT


13              GGTCGCCGCGGGATACTGCCTTCGTGTGG--GGGTGAGTTCCGCTGGTGCGGCTGGTTGC
15              ------------------------------------------------------------
11              GGTCGTCGCAGGATACTGCTTGTCGTTGGACGGGTGAGTCTTGCTGGTGCGGCTGGTGTC


13              CGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGG
15              ------------------------------------------------------------
11              CGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGG


13              GCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATGC
15              ------------------------------------------------------------
11              GCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATGC


13              CTCGGGGAGCTGCCAACCGAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG
15              ------------------------------------------------------
11              CTCGGGGAGCTGCCAACCGAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG
```

High Altitude, Site 3

```
CLUSTAL W (1.83) multiple sequence alignment


65              -GGGGTCCGTAAGGGTCCTGCACCAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
70              GGGGTTCCGTAAGGGTCCTGCACAGGTGGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
                 *** ***************** *  ********************************


65              ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGT
70              ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATGT
                ************************************************************


65              CGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAA-GGTGGTGATGACGTCAAGT
70              CGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAAGGTGGGGATGACGTCAAGT
                ************************************** ***** ************


65              CATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTGC
70              CATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGGGCTGC
                ********************************************** ************


65              GATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA
70              GATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAA
                ************************************************************
```

```
65              CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATAC
70              CTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATAC
                ************************************************************

65              GTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCG
70              GTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCG
                ************************************************************

65              GTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTCG
70              GTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTCG
                ************************************************************

65              TAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGG
70              TAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGG
                ************************************************************

65              CTGGTCTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCGGTC
70              CTGGTTTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCGGTC
                ***** ********************************************************

65              TGGTTGCTCGAGGGTGTGGACGCTGATGATGTGGCTGCCTGC-GTGTCTGGTGCCTAGTA
70              TGGTTGCTCGAGGGTGTGGACGCTGACGATGTGGCTGCTGGCTGTGTCTGGTGCCTAGTA
                *************************** ***********  ** ******************

65              CTCCTGCCTTCCGGGGTGGGGTGGAACGGGGCTGGGTGGTGTGGGTGGTCCGTGGCGCGC
70              CTCCTGCTTGT--GGGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCGTGGCACGC
                ******* *    ********************** * ** ************* ***

65              TGTTGGGTCCTGAGGGAGTCAGGCCGGTTGTGGCTGGGACTTTCTTGATCCGCTTTTTGT
70              TGTTGGGTCCTGAGGGAGTCAGGCCAGT-GTGGTCGACGCTTTCTTGATCCGCTTTTTCT
                ************************* ** ****  *   ******************* *

65              GGGTGCCCCTGTCCTCCTGTGTGGGGGGTGGGGGTGTTCCGCGAGGGTGGTGCGGGCTGG
70              GGGTGCCCCTGTCCTCCTATATGGGGGGTGGGGGTGTTCCGTGAGGGTGGTGCGGGCTGG
                ****************** * ******************** ******************

65              TCGCCGCGGGATACTGCCTTCTGTGTGGGGGTGAGTTCCGCTGGTGTGGCTGGTTGCCGT
70              TCGTTGCAGGATACTGCCTTCTGCGTGAGGGTGAGTTCTGCTGGTGCGGCTGGTTTCTGT
                ***  ** ***************** *** ********** ******* ******** * **

65              CCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGGGCG
70              CCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAGGGCG
                ************************************************************

65              CACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATGCCTC
70              CACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGCGATATGCCTC
                ******************************************** ***************

65              GGGGAGCTGCCAACCTAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG
70              GGGGAGCTGTCAACCGAGCTGTGATCCGAGGATTTCCGAATGGGGAAACCCGG
                ********* ***** ************************************
```

Middle Altitude, Site 1

```
CLUSTAL W (1.83) multiple sequence alignment

28              GGGGTCCGTAAGGGTCCTGCACCAGGTGGTGCATGGCTGTCGTCAGCTCGTGTC------
29              GGGGTCCGTAAGGGTCCTGCACAGGGTGGTGCATGGCTGTCGTCAGCTCGTGTC------
21              GGGGTCCGTAAGGGTCCTGCAC-AGGTGGTGCATGGCTGTCGTCAGCTCGTGTC------
22              GGGGTCCGTAAGGGTCCTGCAC-AGGTGGTGCATGGCTGTCGTCAGCTCGTGTC------
24              GGGGTCCGTAAGGGTCCTGCAC-AGGTGGTGCATGGCTGTCGTCAGCTCGTGTCACCTCG
                ********************  ******************************
```

```
28          GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGT
29          GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGT
21          GTGAGATGTTGGGTTAAGTCCCGCAGCGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGT
22          GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCTATGCTGCCAGCGAGT
24          GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGT
            ************************* ************** ****** **********

28          AATGTCGGGGACTCATAGGAGACTGCCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGT
29          AATGTCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGT
21          AATGTCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGT
22          AATGTCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGT
24          AATGTCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGT
            ************************** ******************************

28          CAAGTCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGG
29          CAAGTCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGG
21          CAAGTCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGG
22          CAAGTCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGG
24          CAAGTCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGG
            ************************************************************

28          GCTGCGATGCCGTGAGGTGGAGCGAATCCC--AAAAAGCCGGTCTCAGTTCGGATCGGGG
29          GCTGCGATGCCGTGAGGTGGAGCGAATCCC--AAAAAGCCGGTCTCAGTTCGGATCGGGG
21          GCTGCGATGCCGTGAGGTGGAGCGAATCCC--AAAAAGCCGGTCTCAGTTCGGATCGGGG
22          GCTGCGATGCCGTGAGGTGGAGCGAATCCC--AAAAAGCCGGTCTCAGTTCGGATCGGGG
24          GCTGCGATGCCGTGAGGTGGAGCGAATCCCCAAAAAAGCCGGTCTCAGTTCGGATCGGGG
            ******************************  **************************

28          TCTGCAACTCGACCCC-GTGAAGTCGG-AGTCGCTAGTAATCGCAGATCAGCAATGCTGC
29          TCTGCAACTCGACCCC-GTGAAGTCGG-AGTCGCTAGTAATCGCAGATCAGCAATGCTGC
21          TCTGCAACTCGACCCC-GTGAAGTCGG-AGTCGCTAGTAATCGCAGATCAGCAATGCTGC
22          TCTGCAACTCGACCCCCGTGAAGTCGG-AGTCGCTAGTAATCGCAGATCAGCAATGCTGC
24          TCTGCAACTCGAACCCCGTGAAGTCGGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGC
            ************ *** ********** ********************************

28          GGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACAC
29          GGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACAC
21          GGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACAC
22          GGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACAC
24          GGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACAC
            ************************************************************

28          CCGAAGCCGGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGG
29          CCGAAGCCGGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGG
21          CCGAAGCCGGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGG
22          CCGAAGCCGGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGG
24          CCGAAGCCGGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGG
            ************************************************************

28          ACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTT-CTAAG
29          ACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCCGGATCACCTCCTTT-CTAAG
21          ACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTACTAAG
22          ACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTT-CTAAG
24          ACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTT-CTAAG
            ************************************* ************** *****

28          GAGCGTCTGGCTGGTTTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTG
29          GAGCGTCTGGCTGGTTTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTG
21          GAGCGTCTGGCTGGTTTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTG
22          GAGCGTCTGGCTGGTTTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTG
24          GAGCGTCTGGCTGGTTTGTCCCGTTGGGGGTGGGCTGGTCCAGGGCCAGGGCCGGATGTG
            ************************************************************

28          TAGGCCGGTCTGGTTGCTCGAGGGTGTGGACGCTGACGATGTGGCTGCTGGCTGTGTCTG
29          TAGGCCGGTCTGGTTGCTCGAGGGTGTGGACGCTGACGATGTGGCTGCTGGCTGTGTCTG
21          TAGGCCGGTCTGGTTGCTCGAGGGTGTGGACGCTGACGATGTGGCTGCTGGCTGTGTCTG
22          TAGGCCGGTCTGGTTGCTCGAGGGTGTGGACGCTGACGATGTGGCTGCTGGCTGTGTCTG
24          TAGGCCGGTCTGGTTGCTCGAGGGTGTGGACGCTGACGATGTGGCTGCTGGCTGTGTCTG
            ************************************************************
```

```
28              GTGCCTAGTACTCCTGCTTGTGGGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCG
29              GTGCCTAGTACTCCTGCTTGTGGGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCG
21              GTGCCTAGTACTCCTGCTTGTGGGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCG
22              GTGCCTAGTACTCCTGCTTGTGGGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCG
24              GTGCCTAGTACTCCTGCTTGTGGGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCG
                ************************************************************

28              TGGCACGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGC
29              TGGCACGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGC
21              TGGCACGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGC
22              TGGCACGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGC
24              TGGCACGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGC
                ************************************************************

28              TTTTT-CTGGGTGCCCCTGTCCTCCTATATGGGGGG-TGGGGGTGTTCCGTGAGGGTGGT
29              TTTTT-CTGGGTGCCCCTGTCCTCCTATATGGGGGG-TGGGGGTGTTCCGTGAGGGTGGT
21              TTTTT-CTGGGTGCCCCTGTCCTCCTATATGGGGGG-TGGGGGTGTTCCGTGAGGGTGGT
22              TTTTT-CTGGGTGCCCCTGTCCTCCTATATGGGGGGGTGGGGGTGTTCCGTGAGGGTGGT
24              TTTTTTCTGGGTGCCCCTGTCCTCCTATATGGGGGGTGGGGGTTGTTCCGTGAGGGTGGT
                ***** ****************************  **** ******************

28              GCGGGCTGGTCGTTGCAGGATACTGCCTTCTGCGTGAGGGTGAGTTCTGCTGGTGCGGCT
29              GCGGGCTGGTCGTTGCAGGATACTGCCTTCTGCGTGAGGGTGAGTTCTGCTGGTGCGGCT
21              GCGGGCTGGTCGTTGCAGGATACTGCCTTCTGCGTGAGGGTGAGTTCTGCTGGTGCGGCT
22              GCGGGCTGGTCGTTGCAGGATACTGCCTTCTGCGTGAGGGTGAGTTCTGCTGGTGCGGCT
24              GCGGGCTGGTCGTTGCAGGATACTGCCTTCTGCGTGAGGGTGAGTTCTGCTGGTGCGGCT
                ************************************************************

28              GGTTTCTGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTA
29              GGTTTCTGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTA
21              GGTTTCTGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTA
22              GGTTTCTGTCCGTACGTTGAGAACTGCACAGTGGACGCGAACATCTTTGTGGCCAAGTTA
24              GGTTTCTGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTA
                ************************************** ******************

28              TTAAGGGCGCACGGTTGGATGCCTTGGCACTAGGAGCCGATGAAGGACGTGGGAGACTGC
29              TTAAGGGCGCACGGT-GGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGC
21              TTAAGGGCGCATCGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGA
22              TTAAGGGCGCACGGT-GGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGC
24              TTAAGGGCGCACGGT-GGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGC
                ***********  *  ************** ****************************

28              GAT-ATGCCTCGGGG-AGCTGTCAACCGAG-CT----GTGATCCGAGGA-TTTCCGAATG
29              GATTATGCCTCGGGGGAGCTGTCAACCGAG-CT----GTGATCCGAGGAATTTCCGAATG
21              CGATATGCCTCGGGGACGCTGTCAACCGAGGCT----GTGATCCGAGGA-TTTCCGAATG
22              GAT-ATGCCTCGGGG-AGCTGTCAACCGAG-CT----GTGATCCGAGGA-TTTCCGAATG
24              GAT-ATGCCTCGGGG-AGCTGTCAACCGAGCTCCAGAGTGATCCGAGGA-TTTCCGAATG
                    ***********  *************       ************ **********

28              GGGAAACCCGG
29              GGGAAACCCGG
21              GGGAAACCCGG
22              GGGAAACCCGG
24              GGGAAACCCGG
                ***********
```

Middle Altitude, Site 2

```
CLUSTAL W (1.83) multiple sequence alignment


33              --GGGGTCCGTAAGGGTCCTGCACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
32              GGTGGGTCCGTAAGGGTCCTGCACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
39              --GGGGTCCGTAAGGGTCCTGCACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
31              --GGGGTCCGTAAGGGTCCTGCACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
36              --GGGGTCCGTAAGGGTCCTGCACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAG
                  ************************************************************

33              ATGTTGGGTTAAGTCCCGC-AACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATG
32              ATGTTGGGTTAAGTCCCGC-AACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATG
39              ATGTTGGGTTAAGTCCCGC-AACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATG
31              ATGTTGGGTTAAGTCCCGCCAACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATG
36              ATGTTGGGTTAAGTCCCGC-AACGAGCGCAACCCTCGTCCTATGTTGCCAGCGAGTAATG
                ******************* ************************************

33              TCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAG
32              TCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAG
39              TCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAG
31              TCGGGGACTCATAGGAGACTGCC-GGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAG
36              TCGGGGACTCATAGGAGACTGCCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCTAG
                ********************** ********************************** **

33              TCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTG
32              TCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTG
39              TCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTG
31              TCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGGGCTG
36              TCATCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGCACAAAGGGCTG
                ************************************************ ***********

33              CGATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCA
32              CGATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCA
39              CGATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCA
31              CGATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCA
36              CGATGCCGTGAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCA
                ************************************************************

33              ACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATA
32              ACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATA
39              ACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATA
31              ACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATA
36              ACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATA
                ************************************************************

33              CGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCC
32              CGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCC
39              CGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCC
31              CGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCC
36              CGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCC
                ************************************************************

33              GGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTC
32              GGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTC
39              GGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTC
31              GGTGGCCTAACCCTCGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTC
36              GGTGGCCTAACCCTTGTGGGAGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTC
                ************** ********************************************

33              GTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTG
32              GTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTG
39              GTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTG
31              GTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAGCGTCTG
36              GTAACAAGGTAGCCGTACCGGAAGGTGCAGCTGGATCACCTCCTTTCTAAGGAGCGTCTG
                ***************************** ********************************
```

```
33          GCTGGTCTGTCCCTGTTGGGGGGTGGGCTGGTCCAGGGCCAGGACCGGATGTGTAGGCCG
32          GCTGGTCTGTCCCTGTTGGGGGGTGGGCTGGTCCAGGGCCAGGACCGGATGTGTAGGCCG
39          GCTGGTCTGTCCCTGTTGGGGGGTGGGCTGGTCCAGGGCCAGGACCGGATGTGTAGGCCG
31          GCTGGTTTGTCCC-GTTGGGGG-TGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCG
36          GCTGGTTTGTCCC-GTTGGGGG-TGGGCTGGTCCAGGGCCAGGGCCGGATGTGTAGGCCG
            ****** ****** ******** ******************** ****************

33          GTCTGGTTGCTCGAGGGTGTGGAACACTGATGATGTGGCTGCCGGTTCTGTCTGGCTTCT
32          GTCTGGTTGCTCGAGGGTGTGGAACACTGATGATGTGGCTGCCGGTTCTGTCTGGCTTCT
39          GTCTGGTTGCTCGAGGGTGTGGAACACTGATGATGTGGCTGCCGGTTCTGTCTGGCTTCT
31          GTCTGGTTGCTCGAGGGTGTGGA-CGCTGACGATGTGGCTGCTGGCTGTGTCTGGTGCCT
36          GTCTGGTTGCTCGAGGGTGTGGA-CGCTGACGATGTGGCTGCTGGCTGTGTCTGGTGCCT
            ********************** * **** ********** ** * *******   **

33          AGTACTCCTGCCTGTGTGGTGGGGTGGAACGGGGTTGGGTGGGGCGGGTGGTCCGTGGCG
32          AGTACTCCTGCCTGTGTGGTGGGGTGGAACGGGGTTGGGTGGGGCGGGTGGTCCGTGGCG
39          AGTACTCCTGCCTGTGTGGTGGGGTGGAACGGGGTTGGGTGGGGCGGGTGGTCCGTGGCG
31          AGTACTCCTGCTTGTG-GGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCGTGGCA
36          AGTACTCCTGCTTGTG-GGTGGGGTGGAACGGGGCTGGGTAGGGTTGGTGGTCCGTGGCA
            ********** **** ***************** ***** ***  ************

33          CGCTGTTGGGTCCTGAGGGAGTCAGGCCGGTGTGGTCTG-GTTCTCTTGATCCGCTTCTG
32          CGCTGTTGGGTCCTGAGGGAGTCAGGCCGGTGTGGTCTG-GTTTTCTTGATCCGCTTCTG
39          CGCTGTTGGGTCCTGAGGGAGTCAGGCCGGTGTGGTCTG-GTTTTCTTGATCCGCTTCTG
31          CGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGCTTTTT
36          CGCTGTTGGGTCCTGAGGGAGTCAGGCCAGTGTGGTCGACGCTTTCTTGATCCGCTTTTT
            ***************************** ********   * * ************ *

33          GTGGGTGCCCCTGTCCTC---TGTGGAGGGTGGGGGTGTTC-GCGAGGGTGGTGCGGGCT
32          GTGGGTGCCCCTGTCCTC---TGTGGAGGGTGGGGGTGTTC-GCGAGGGTGGTGCGGGCT
39          GTGGGTGCCCCTGTCCTC---TGTGGAGGGTGGGGGTGTTC-GCGAGGGTGGTGCGGGCT
31          CTGGGTGCCCCTGTCCTCCTATATGGGGGGTGGGGGTGTTCCGTGAGGGTGGTGCGGGCT
36          CTGGGTGCCCCTGTCCTCCTATATGGGGGGTGGGGGTGTTCCGTGAGGGTGGTGCGGGCT
             *****************   * *** ************** * ****************

33          GGTCGTCGCAGGATACTGCTTGTCGTTGGACGGGTGAGTCTTGCTGGTGCGGCTGGTGTC
32          GGTCGTCGCAGGATACTGCTTGTCGTTGGACGGGTGAGTCTTGCTGGTGCGGCTGGTGTC
39          GGTCGTCGCAGGATACTGCTTGTCGTTGGACGGGTGAGTCTTGCTGGTGCGGCTGGTGTC
31          GGTCGTTGCAGGATACTGCCTTCTGCGTGA-GGGTGAGTTCTGCTGGTGCGGCTGGTTTC
36          GGTCGTTGCAGGATACTGCCTTCTGCGTGA-GGGTGAGTTCTGCTGGTGCGGCTGGTTTC
            ****** ************ *   *   ** ********  **************** **

33          CGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAA~G
32          CGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAA~G
39          CGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAA~G
31          TGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTATTAAAG
36          TGTCCGTACGTTGAGAACTGCACAGTGGACGCGAGCATCTTTGTGGCCAAGTTTATTAAG
             ************************************************  * * *

33          GGCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATG
32          GGCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATG
39          GGCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGGCTGCGATATG
31          GGCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGCGATATG
36          GGCGCACGGTGGATGCCTTGGCACCAGGAGCCGATGAAGGACGTGGGAGACTGCGATATG
            ************************************************** **********

33          CCTCGGGG-AGCTGCCAACCGAGCT---GTGATCCGAGGATTTCCGAATGGGGAAACCCG
32          CCTCGGGGGAGCTGCCAACCGAGCT---GTGATCCGAGGATTTCCGAATGGGGAAACCCG
39          CCTCGGGG-AGCTGCCAACCGAGCTCGAGTGATCCGAGGATTTCCGAATGGGGAAACCCG
31          CCTCGGGGAAGCTGTCAACCGAGCT---GTGATCCGAGGATTTCCGAATGGGGAAACCCG
36          CCTCGGGGAAGCTGTCAACCGAGCT---GTGATCCGAGGATTTCCGAATGGGGAAACCCG
            ******** ***** **********   ********************************

33          G
32          G
39          G
31          G
36          G
            *
```

Middle Altitude, Site 3

```
CLUSTAL W (1.83) multiple sequence alignment


73              ------------------------------------------------------------
75              GGGGTCCGTAAGGGTTAATCGCAGATCAGCAATGCTGCGGTGAATACGTTCCCGGGCCTT


73              --------------GGGGTCCGTAAGGGTCCTGCACAGGTGGTGCATGGCTGTCGTCAGC
75              GTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGC-CGGTGGCCCAACCC
                           *   *  ***    ***   *  **  **  * *     *


73              TCGTGTCGTGAGATGTTGG-GTTAAGTCCCGCAAC--GAGCGCAACCCTCGTCCTATGTT
75              TTGTGGGGGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGAAGTCGTAACAAGGTGGC
                * ***  * ***  ** *  * *  * ** ** *   *  ** *  * *       **


73              GCCAGCGAGTAATGTCGGGGACATCCCCACCTTCCTCCGAGTTGACC--CCGG--CAGTC
75              GGTACCGGAAGGTG-CGGCTGGATCACCTCCTTTCTAAGGAGCGTCTGGCTGGTTCGTCC
                *  * **     ** ***    *** ** **** **  *    * *   * **  *   *


73              TCCTATGAGTGGGCTGGTCCAGGGCCAGGACCGGATGTGTAGGCCGGTCTGGTTGCTCGA
75              TGTTGGGGATGGGCTGGTTCAGGGCCAGGGCCGGATGTGTAGGCCGGTCTGGTTGCTCAA
                *  *  *  ********* ********** ****************************** *


73              GGGTGTGGAACACTGATGATTTGGCTGCCGGTTCTGTCTGGCTTCTAT-GTGTTCTTGAG
75              GGGTGTGGA-CGCTGACGATG-GGCTGCTGGTTGGTGCTGGGTGTTCTAGTACTCCTG--
                ********* * **** ***  ****** ****    **** *  * * **  ** **


73              ACCGTGTTTAATGTTTTTCAATGATTA-AGGATTGAATGTTAGGGATGATTTTAATTTGA
75              --TGTTTCTGGTGTTCTGTTCTCTTCGGAGGGTGGGGTGCTGGGGGTG--TGGGGTGGAG
                   ** * *  **** *    *  *   *** * *  ** * *** **  *    *


73              GAGATATTTCATGTTGACAAGTTTTTAATTCATTATCATTTTTGTTTATATCAAATGCTT
75              GGGATGTCTGGTGGTGGCTGGTGGTTCGTGGCACGCTGTTGG-GTCCTGAGGGAGTTAGG
                * *** * *  ** ** *  **  **  *          **   **    *   * *


73              ACTTTTTGTTTGATTTGTTATGAGGGTCAGCTGTCAACCAAGCTGTAATCCGAGGATGGG
75              CCTGTGGGTCTGGTTCCTTCT-----TTTTCTATGTGCT--GCTGTCTGTCCCTGTGGGG
                 ** *  ** ** **  ** *     *   ** *   *   *****    *   *  ***


73              GTCCGTAAGGGTCTCTCCTTGCATCCAAACGGGTGCGACTCAGACGGCCGGACGGTAAGG
75              GTGGGTGGTGGCGCTGGTTCGTCTGTGGTGGGGTACCCCACGTCCTTACGGACCCCAA--
                **  **   **       * *  *      **** *  * *   *   *****   **


73              TTACATGTCCGGACTCGAGTGGCTCAGACAGCCGGACGGTAAGGTTACATGTCCGGACGA
75              TCACTAGT--GAATTCGCG-GCCGCCTGCAGGTCGACCATATGGG-AGAGCTCCCAACGC
                * **  **  * * *** * * * *   ***   ***  ** **  * *  ***  ***


73              GCGCGGTCTGTTGCATGCTTATGTGGCAACGTGCGTCCGAATGGGGAAACCCGGAA
75              ---------------------------------------------------------
```

Hippophae sp.

```
CLUSTAL W (1.83) multiple sequence alignment


H2              GGGGTCCGTAAGGGTCTCTGCACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGA
H6              ------------------------------------------------------------
H1              ------------------------------------------- ------------
```

```
H2              TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTATGTTGCCAGCGCGTCATGGC
H6              ------------------------------------------------------------
H1              ------------------------------------------------------------

H2              GGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCA
H6              ----------AGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCA
H1              ------------------------------------------------------------

H2              TCATGCCCCTTACGCCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTGCGA
H6              TCATGCCCCTTACGTCCTGGGCTGCACACATGCTACAATGGCCGGTACAAAGGGCTGCGA
H1              ------------------------------------------GTCCGTAAGGGTCCGA
                                                           ** *  * ** * ***

H2              TACCGCAAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACT
H6              TACCACAAGGTGGAGCGAATCCCAAAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACT
H1              ATGGGGAAACCCGGACGAATGGGGAAACCCGGAGGTGGGGTCCG-TAAGGGTCTGCGGCT
                      **    *  *****    ***  * *   *  * ** * *  ******** **

H2              CGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATACGT
H6              CGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAATGCTGCGGTGAATACGT
H1              GGATCCCCTCCTTACTAAGGAGCGTCTG---GCTGGTCTGTCCTGTTGGGGTCCGTAAGG
                 ** *** *     *  **  **   *    ** * ** *   ** ** ***   ** *

H2              TCCCGGGC-------CTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGA
H6              TCCCGGGC-------CTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTGACACCCGA
H1              GTCCGAATGGGGAAACCCGGAGGTGTGGTCCTTAAGGGTCCGAATG--GGGAAGCCCGGA
                  ***          *  * *      * ** * * *   **** *  **  *  ** **

H2              AGCCGGTGGCCCAACCCTTGTGGGGGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGA
H6              AGCCGGTGGCCCAACCCTTGTGGGGGGAGCCGTCGAAGGTGGGACCGGCGATTGGGACGA
H1              GGTGGGTCCGTAAGGGTCCGAATGGGAAACC--CGGAGGTGGGGTCCATAAG-GGTCCGA
                 *  ***     *      *   *** * **  ** *******  *    *  **  ***

H2              AGTCGTAACAAGGTAGCCGTACCGG--AAGGTGCGGCTGGATCACCTCCTTTCTAAGGAG
H6              AGTCGTAACAAGGTAGCCGTACCGG--AAGGTGCGGCTGGATCACCTCCTTTCTAAGGAG
H1              ATGGGGAACCCGGAGGTGGGGTCCGTAAGGGTCCGAATGGGGAAGCCCG-----GAGGTG
                *   * ***  **  *  *   * *  * *** **  ***   * * *       *** *

H2              CTTTCTGACTGGGGCCTTCTGTAGTGGGGGGTTTCTGGTGCAGAGCGAAGACCGGCCGGA
H6              CTTTCTGACTGGGGCCTTCTGTAGTGGGGGGTTTCTGGTGCAGAGCGAAGACCGGCCGGA
H1              GGGTCCGTAAGGGTCC-----GAATGGGGAAACCCGGGGGGTGGGGTCGTAAGGGTCCGAA
                   ** *   *** **      * *****     * ** *  * *     *  * *** *

H2              TGGTTGG-TCTGGGGTGCTCATGGGTGTGGAACGCTGACTATGTTCTTTCTGGTCTGGTT
H6              TGGTTGG-TCTGGGGTGCTCATGGGTGTGGAACGCTGACTATGTTCTTTCTGGTCTGGTT
H1              TGGGGAAACCCGGAGTCCGTAAGGGGTCCGAATGGGGTCCGTAAGGGTCGGGGTCCGAAT
                ***      * ** ** *  * ***    *** *  * *  *     *   **** *  *

H2              GTTGTCCGGGGTTCCTCTGTGTTCTTCGGGTGGTGGCTGGTCTGGGTGGGCGGTACGCTG
H6              GTTGTCCGGGGTTCCTCTGTGTTCTTCGGGTGGTGGCTGGTCTGGG--------------
H1              G-----------------------------------------------------------
                *

H2              TTGGGTTTTGAGGGAGTGAGATGCCCTCGTGTATGACCCTGCCTTCTGCTGGCCCGTTGG
H6              ------------------------------------------------------------
H1              ------------------------------------------------------------

H2              GGTTGGTGGGAGGCAGGTTTGCGGGCTGGTTCGTCGTGGGATACCGCTGTCTTTTATGGT
H6              ------------------------------------------------------------
H1              ------------------------------------------------------------

H2              GGTGGGTCCTGCAGGTGCGGCTGGTTTCCGCTCGTAGTTTGAGAACTGCACAGTGGACGC
H6              ------------------------------------------------------------
H1              ------------------------------------------------------------
```

```
H2              GAGCATCTTTATCTTTGTGGCCAAGTTATTAAGGGCGCACGGTGGATGTCTTGGCACCAG
H6              ------------------------------------------------------------
H1              ------------------------------------------------------------

H2              GAGCCGATGAAGGACGTGGGAGGCTGCGATATGCCTCGGGGAGCTGCCAACCTAGCTGTG
H6              ------------------------------------------------------------
H1              ------------------------------------------------------------

H2              ATCCGAGGATTTCCGAATGGGGAAACCCGG
H6              -----------------------------
H1              -----------------------------
```

NEHU LIBRARY 103802
Acc
Ac
D 7-9-07
Cl
S
Ex
P