PHYLOGENETIC DIVERSITY OF NITROGENASE (*nif*H) GENE WITHIN THE RHIZOSPHERE OF *Colocasia esculenta* OF EX-TIN MINING LAND

TAN KENG YEAN

BACHELOR OF SCIENCE (HONS)

BIOTECHNOLOGY

FACULTY OF SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

MAY 2011

PHYLOGENETIC DIVERSITY OF NITROGENASE (*nif*H) GENE WITHIN THE RHIZOSPHERE OF *Colocasia esculenta* OF EX-TIN MINING LAND

By

TAN KENG YEAN

A project report submitted to the Department of Biological Science

Faculty of Science

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements for the degree of

Bachelor of Science (Hons.) Biotechnology

May 2011

ABSTRACT

PHYLOGENETIC DIVERSITY OF NITROGENASE (*nif*H) GENE WITHIN THE RHIZOSPHERE OF *Colocasia esculenta* OF EX-TIN MINING SOIL

TAN KENG YEAN

Ex-tin mining land comprises nutrient-deprived soil containing high levels of toxic metals. This arid land condition is generally not suitable for agriculture and mostly abandoned. Nitrogen fixed microorganisms which are associated with plants in fixing nitrogen or improving nutrient uptake, have been noted to serve as possible commercial microbial inoculums for ex-tin mining rehabilitation. Therefore, soil microbial community in ex-tin mining land has piqued immense interest in studying the diversity of diazotrophic communities by culture-independent approach. Soil sample was collected in the ex-tin mining land within the rhizosphere of *Colocasia esculenta,* in the vicinity of Universiti Tunku Abdul Rahman, Kampar campus. Direct total microbial DNA was extracted from soil and was subjected to nested PCR amplification using two pairs of degenerate primers. The resulting PCR amplicons were gel-purified, ligated into pGEM®-T Easy Vector and consequently transformed into competent *Escherichia coli* JM109 cells. Recombinant clones with the correct insert size were selected based

on blue-white screening followed by colony PCR. Recombinant plasmids were purified and sequenced, yielding a total of 28 *nif*H homologous clone sequences. BLASTX analysis revealed high similarities to *nif*H gene fragments corresponding to unculturable nitrogenase genes. Molecular phylogenetic inference demonstrated that majority of *nif*H clones were clustered within the cluster III of Proteobacteria group, mainly in subdivision of δ-Proteobacteria. This study suggested predominance of *nif*H sequences from proteobacteria in soil sample rather than phylogenetically diverse *nif*H sequences from a broad range diazotrophs. From this study, several clones harboured *nif*H gene sequences from proteobacteria subdivision showed potential and having valuable properties for bioremediation.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to my research supervisor, Assistant Professor Dr Choo Quok Cheong for his patient guidance, valuable suggestions and relenting demeanour. Next, I would like to thank the postgraduate, Ms Ng Wei Ling and my research projects team members Cheah Hong Leong, Tan Wei Cheat, Toh Wai Keat and Ng Wei Kiang. Without their earnest advice and tireless support, this project would not be completed successfully within stipulated time. Last but not least, to my dearest family and friends, who has been supporting me every step to the successful culmination of this project - thank you all.

DECLARATION

I hereby declare that the project report is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

_____

TAN KENG YEAN

APPROVAL SHEET

This project report entitled **"PHYLOGENETIC DIVERSITY OF NITROGENASE (*nif*H) GENE WITHIN THE RHIZOSPHERE OF *Colocasia esculenta* OF EX-TIN MINING SOIL"** was prepared by TAN KENG YEAN and submitted as partial fulfillment of the requirements for the degree of Bachelor of Science (Hons) in Biotechnology at Universiti Tunku Abdul Rahman.

Approved by,

_____

(Assistant Professor Dr. CHOO QUOK CHEONG) Date:

Supervisor

Department of Biological Science

Faculty of Science

Universiti Tunku Abdul Rahman

FACULTY OF ENGINEERING AND SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: _____

PERMISSION SHEET

It is hereby certified that **TAN KENG YEAN** (ID No: **08ADB05886**) has completed this final year project entitled "**PHYLOGENETIC DIVERSITY OF NITROGENASE (*nif*H) GENE WITHIN THE RHIZOSPHERE OF *Colocasia esculenta* OF EX-TIN MINING SOIL**" under supervision of Dr. Choo Quok Cheong (Supervisor) _____ (Co-supervisor) from the Department of Biological Science, Faculty of Science.

I hereby give permission to my supervisors to write and prepare manuscript of these research findings for publishing in any form, if I did not prepare it within six (6) months time from this date provided that my name is included as one of the author for this article. Arrangement of the name depends on my supervisors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

A	Absorbance
ADP	Adenosine diphosphate
anf	Alternative nitrogen fixation gene
ATP	Adenosime triphosphate
BLAST	Basic Local Alignment Search Tool
BNF	Biological nitrogen fixation
bp	Base pair
CLPP	Community level physiological profiling
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
ExPaSy	Expert Protein Analysis System
FRIM	Forest Research Institute Malaysia
HGT	Horizontal gene transfer
IPTG	Isopropyl-β-D-thiogalactopyranoside
MEGA	Molecular Evolutionary Genetics Analysis
NCBI	National Center for Biotechnology Information
nif	Nitrogen fixation gene
NiFH	Translated *nif*H sequence
OD	Optical density
PCR	Polymerase chain reaction
rDNA	Ribosomal deoxyribonucleic acid

RISA	Ribosomal intergenic spacer length polymorphism
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
SSCP	Single strand conformation polymorphism
SSCU	Sole source carbon utilization
T-RF	Terminal restriction fragment
T-RFLP	Terminal restriction fragment length polymorphism
vnf	Vanadium nitrogenase
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1

INTRODUCTION

Nitrogen, being a constitute element in amino acid and nucleic acid, is a fundamental element that all living cells require. The nitrogen in the atmosphere, despite its abundance, is not readily accessible by the biosphere. Inert atmospheric dinitrogen has to be first converted into biologically available ammonia (NH_3) through the process of nitrogen fixation, which takes place by biological and non-biological means.

Biological nitrogen fixation (BNF) is an important source of fixed nitrogen for the biosphere, which is carried out exclusively by collective of prokaryotes termed diazotrophs. Many diazotrophs are part of the soil microbial communities, in a wide range of bacterial phyla, from *Archaea* to *Eubacteria* (Young, 1992). The soil microbial communities are crucial to the maintenance of soil function by involving in key processes including soil structure formation, decomposition of organic matter, toxin removal, the cycling of carbon, nitrogen, phosphorus, and sulphur (van Elsas & Trevors, 1997) as well as mycorrhizal associations (Barea, 1991).

Diazotrophs have long been recognized as important sources of reduced nitrogen ecosystem components in many environments (Ogunseitan, 2005). Diazotrophs in soil probably represent more undiscovered biological diversity than that found in any other environment on Earth (Tiedje, Asuming-Brempong, Nusslein, Marsh & Flynn, 1999). Thus, the diazotrophic diversity in soil is constantly being studied in order to discover new microbes and characterize their functional roles.

From a general perspective, the methods used to access the diversity of soil microorganism fall broadly into culture-dependent and culture-independent techniques. Due to the physiological diversity of diazotrophs and the documented unculturability of many prokaryotes, cultivation based strategies have severe limitations for underestimating the true diversity of diazotrophs (Korner & Laczkó, 1992). Therefore, several culture-independent approaches have been developed to bypass the need of cultivation in order to study the microbial diversity in soil (Kent & Triplett, 2002).

Using molecular approach to study the diversity of diazotroph organisms is primarily based on PCR amplification of various ecological markers. The presence of the gene *nif*H, which codes for the iron nitrogenase unit of the nitrogenase complex, has become a useful marker gene in culture-independent studies in providing evidence for potential nitrogen fixation. The highly conserved nature of the *nif*H gene makes it an ideal molecular tool to determine

the potential for biological nitrogen fixation in any environment (Zehr & Capone, 1996). By designing PCR primers that amplify the *nif*H gene and sequencing the amplified genes, the assemblage of microorganisms capable of nitrogen fixation can be elucidated.

To date, only little is known about the diazotrophic communities that inhabit the soil of ex-tin mining land or any other mining land. (Zehr, Jenkis, Short & Steward, 2003). By studying the diversity of diazotrophs of tailings of the ex-tin mining land, it is possible to elucidate the nitrogen fixing communities that are able to persist in such nutrient-deficient environment (Knabe, 1965; Fox, 1984). Hence, this study is aimed to describe the phylogenetic diversity of *nif*H genes of diazotrophs communities in a nitrogen-poor ex-tin mining land. This will help to add another entry to the expanding database of *nif*H sequences and revealed a wide diversity of uncultured diazotrophs.

The primary objectives of this study include:

1. To study the physicochemical properties of ex-tin mining soil
2. To identify the major phylogenetic classes constituting the nitrogen fixing community in ex-tin mining soil.
3. To assess *nif*H diversity in independent soil sites of the ex-tin mining land
4. To speculate on the possibility for diazotrophs of the ex-tin mining land to be used for land rehabilitation purpose.

CHAPTER 2

LITERATURE REVIEW

2.1 The Ecological Importance of Nitrogen Fixation

2.1.1 Biological Nitrogen Fixation

Nitrogen fixation is a life-sustaining process and a key link in the nitrogen cycle as it alleviates the limited biologically accessible nitrogen reserve of the ecosystem (Zehr et al., 2003). Biological nitrogen fixation (BNF) is the process of the enzymatic conversion of atmospheric di-nitrogen (N_2) to the non-gaseous N compound ammonium (NH_4^+). These compounds represent the starting molecules for the biosynthesis of amino-acids and other N containing biomolecules. Prior to the advent of the Haber-Bosch process, the input of fixed nitrogen into the biosphere was provided almost exclusively through biological nitrogen fixation (BNF) (Rees et al., 2005).

BNF has gained the limelight as the use nitrogenous fertilizers have seen severely polluted water bodies and the eutrophication of lakes and rivers (Sprent & Sprent, 1990). Nitrogen pool is continually drained by activities including denitrification, leaching, volatilization and removal of nitrogen-containing crop residues from the land (Millbank, 1974). This replacement of soil nitrogen is generally accomplished by the addition of chemically fixed nitrogen in the form of commercial inorganic fertilizers or by the activity of biological nitrogen fixation

system. BNF serves as a better means to replenish the loss nitrogen of the environment. This statement is supported by the facts that nitrogen fixing bacteria obtain the required energy for BNF from carbohydrates formed by the photosynthetic activity of plants, or from soil organic substrates – sources which are replenishable (Hubble & Kidder, 1992).

2.1.2 Diazotrophs

The ability of fixing nitrogen is found in a group of bacteria known as the diazotrophs, where paraphyletically distributed among the bacterial and archaea domains (Raymond, Siefert, Staples & Blankenship, 2003). There are about 38 genera of bacteria, 20 genera of cyanobacteria and 87 species of archaea which are classified as diazotrophs (Stolp, 2008). Diazotrophs are phylogenetically diverse and include organisms with vastly different physiological properties. The capacity to perform BNF has been detected in various phototrophic microorganisms (e.g. aerobic phototrophic Cyanobacteria (Young, 1992; Vaishampayan et al., 2001), anaerobic purple-sulfur phototrophs like Chromatium, and green-sulfur phototrophs, e.g. *Chlorobium* (Young, 1992), Chemolithotrophic microorganism e.g. *Alcaligenes, Thiobacillus, Methanosarcina,* (Madigan, Martinko & Parker, 2000; Young, 1992) or *Azospirillum lipoferum* (Malik & Schlegel, 1981) and in a great number of heterotrophic bacterial strains (e.g. anaerobes such as *Clostridium,* microaerophiles such as *Herbaspirillum,* and aerobes like *Azotobacter* (Hill, 1992; Paul & Clark, 1996).

2.1.3 Evolution of Biological Nitrogen-Fixation

The broad phylogenetic distribution of BNF raises the question of the evolutionary origin of this function. It has been assumed that the nitrogenase system is an ancient feature already present in an early 'common ancestor' of microbial life, which has subsequently evolved by vertical descent (Hennecke et al., 1985; Young, 1992; Zehr, Mellon, & Hiorns, 1997). However, other theories assume that BNF originated much more recently, where the genes for BNF were subsequently distributed mostly by horizontal gene transfer (Postgate, 1974; Ruvkun & Ausubel, 1980; Mancinelli & McKay, 1988; Normand & Bousquet, 1989). The time of origin and the mode of distribution are of considerable importance to the question of whether, and to what extent, the genes involved in BNF can be expected to contain phylogenetic information that allows identification of diazotrophic groups by use of probes or sequence analyses.

In addition to recent phylogenetic analysis of *nif*DK and *nif*EN, lend credence to the last common ancestor hypothesis, which postulates that nitrogenase, and BNF had already evolved in last common ancestor of the three domains in life (Fani, Gallo & Lio, 2000; Normand, Gouy, Cournoyer & Simonet, 1992). A discrepancy in this hypothesis is the sporadic distribution of nitrogenase within the archaeal and bacterial clades (Young, 1992) as well as its absence in eukaryotes. The second hypothesis has put forward by Raymond and other researchers (2003) explained the phenomenon predominantly revolving around gene loss. Besides, recent genomic evidence has signified the potential of horizontal gene transfer of

nitrogenase among microorganism. (Doolittle, 1999; Nesbo, L'Haridon, Stetter & Doolittle, 2001). The nitrogenase enzyme complex has clearly been mobile through non-vertical transmission, where instances of *nif* genes being selectively lost, duplicated and present on plasmids (Goodman & Weisz, 2002; Prakash, Schilperoort & Nuti, 1981; Thiel, 1993).

2.2 Biochemistry and Genetics of Biological Nitrogen-Fixation

2.2.1 Biochemistry of biological nitrogen-fixation

Biological nitrogen fixation can be summarized in the following equation:

$$N_2 + 10H^+ + 8e^- + nMgATP \rightarrow 2\ NH_4^+ + H_2 + nMgADP + nPi \qquad (n \geq 16)$$

N_2-fixation is a very costly process in terms of its energy requirements, consuming at least 8 mol ATP per mol NH_4^+ produced (Dean & Jacobson, 1992). This is due to the nature of nitrogen atoms are held by a triple bond which has high bond energy (945 kJ/mol). Molecules with higher bond order (eg. triple bond) have shorter bond length, thereby having higher bond energy (Kotz, Treichel & Weaver, 2006).

Biological nitrogen fixation is a complex process that involves a number of functional and regulatory gene products (Triplett, Roberts, Ludden, & Handelsman, 1989). The actual reduction of N_2 is performed by the nitrogenase protein complex, which consists of two metalloproteins: the nitrogenase, or nitrogenase molybdenum-iron protein (MoFe protein, *Nif*DK), and the nitrogenase reductase or nitrogenase iron protein (Fe protein, *Nif*H). The molybdenum-iron-

sulfur-homocitrate clusters of the MoFe protein are the actual sites of binding and reduction of the substrate N_2 and other alternative substrates, such as acetylene, protons and many others (Postgate, 1982). The Fe protein is responsible for shuttling electrons to the MoFe protein using at least two MgATP per electron (Halbleib, Zhang, & Ludden, 2000).

2.2.2 The Genes of Biological Nitrogen Fixation

The number and arrangement of genes involved in BNF varies between species. It takes about 20 gene products to assemble the nitrogenase protein complex and allow it to undergo maturation (Howard & Rees, 1996). The MoFe protein is a tetrameric protein ($\alpha_2\beta_2$) encoded by the *nif*DK genes, and the Fe protein is a homo-dimer (α_2) of the *nif*H gene product (Halbleib, Zhang & Ludden, 2000). These genes, along with regulatory genes and accessory genes coding for enzymes involved in electron transfer and metal cluster synthesis comprise the *nif* regulon (Dean & Jacobson, 1992). The nitrogen fixation genes are usually located at the chromosome, but in some cases, the presence of such genes on plasmids is documented (Stolp, 2008). Alternative nitrogenases have been shown to exist in which molybdenum in the MoFe-protein is substituted by vanadium (*vnf*DK, *vnf*H) or iron (*anf*DK, *anf*H). The alternative systems are transcriptionally regulated and in most studied organisms, *vnf* and *anf* are only expressed under conditions of Mo limitation (Bishop & Premakumar, 1992). However, little is known about the importance of these alternative nitrogenase systems in the soil environment and *anf*H-type sequences have rarely been detected in soil (Hamelin, Fromin,

Tarnawski, Teyssier-Cuvelle & Aragno, 2002; Poly, Ranjard, Nazaret, Gourbière & Monrozier, 2001; Shaffer, Widmer, Porteous & Seidler, 2000; Widmer, Shaffer, Porteous & Seidler, 1999).

2.3 Soil and Microbial Diversity

2.3.1 Microbial Diversity

The term biodiversity has been defined in various ways. In microbial term, it describes the number of different types and their relative abundance in a given community in a given habitat. In molecular-ecological term, it can be defined as the number and distribution of different sequence types present in the DNA extracted from the community in a given habitat. (Liesack, Janssen, Rainey, Ward-Rainey & Stackebrandt, 1997). These divergent terms are often used interchangeable to define soil microbial diversity. Microbial diversity has been defined as "the genetic variability within taxons (species) and the number (richness) and relative abundance of taxons and functional groups (guilds) within communities (Torsvik & Øvreås, 2002). The Convention on Biodiversity, agreed to in 1992 and now ratified by more than 140 countries (CIESIN, 1992) has stimulated the study of microbial diversity. Extensive efforts are currently being made to determine whether microbial species are uniquely placed and isolated geographically; the current evidence is that microbial communities are relatively similar among widely varying geographic regions and agricultural practices. (Supardiyono & Smith, 1997).

Microorganisms in soil probably represent more undiscovered biological diversity than that found in any other environment on earth (Tiedje et al., 1999). The study of soil bacterial diversity is important as the elucidation of their ecological roles will shed light on how the ecosystem operates. As important as they are to biosphere (Pace, 1997), soil bacterial could also benefit the field of biotechnology (Kellenberger, 2001; Whiteley, Donnell, Macnaughton, & Barer, 1996). As far as soil microbes are concerned, they are critical in maintaining the soil homeostasis as they are involved in soil structure formation, organic decomposition, biogeochemical cycles (including carbon, nitrogen phosphorus and sulphur) and soil fertility (O'Donnell, Seasman, Macrae, Waite, & Davies, 2001). Soil microbes also aid in curbing plant and promoting plant growth (van Elsas & Trevors, 1997; Doran, Sarrantonio, & Liebig, 1996).

Diazotroph diversity in soil environments has been less well-studied. Soil is a very heterogeneous medium, both physically and chemically (Ladd, Forster, Nannipieri & Oades, 1996) and diazotrophs can be distributed very differently from one cubic centimetre to the next (Vogel, Normand, Thioulouse, Nesme & Grundmann, 2003; Izquierdo, & Nüsslien, 2006). Distribution of diazotrophs in soil may be influenced by soil texture (Riffkin et al, 1999), nitrogen levels in the soil (Cejudo & Paneque, 1986; Limmer, & Drake, 1998) and the vegetation of the environment (Bardgett, Mawdsley, Edwards, Hobbs, Rodwell & Davis, 1999) as well as a host of other physical, chemical and biological factors.

2.3.2 The effect of Soil Structure and Environmental Conditions on Microbial Diversity

Soils are formed as a result of various factors. These 'soil forming' factors are parent material, climate, organisms, topography and time. Although generally considered as operating independently, they in fact interact to create a very diverse and complex habitat. The soil forming factors set the boundaries of the potential microbial diversity in soil. Within these boundaries, an almost endless variety of microniches exist that serve as unique environments where microbial diversity can develop and thrive. Thus, microbial diversity in soil results from the complex nature of soil (Wang & Dick, 2004).

Soil physicochemical properties (Kennedy & Smith, 1995) and soil particle size distribution (Ranjard & Richaume, 2001) are variables that define the soil microbial diversity. Various components of the soil, including the sand, silt, clay and organic matter composition, provide myriads of unique niches within the microhabitats that cater for a multitude of microbes of different taxa (van Elsas & Trevors, 1997). Most of the soil microbial community is shown to be particle specific (Sessitsch, Weilharter, Gerzabek, Kirchmann & Kandeler, 2001) specifically, samples with larger fractions of small soil particles corresponded to greater microbial diversity, thus emphasizing the impact of particle size on diversity. Diversity of vegetation may also influence microbial communities. Microbial biomass, activity levels (particularly regarding N mineralisation) and bacteria abundance have been positively correlated with plant species richness (Zak, Holmes, White, Peacock & Tilman, 2002). Bacteria usually have high

abundance in surface samples with dense plant covers. Moreover, soil is dynamic as its pH, carbon-nitrogen ratio, temperature and water content can vary over time. This indicates that the microbial communities fluctuate according to the soil environment at a specific period of time (Bölter, Blume, Schneider & Beyer, 1997).

2.3.3 General Limitation in Soil Microbial Studies

Soil microbial communities are often difficult to fully characterize, mainly because of their immense phenotypic and genotypic diversity, heterogeneity, and crypticity. With respect to the latter, bacterial populations in soil top layers can go up to more than 10^9 cells per g soil (Torsvik & Øvreås, 2002), and most of these cells are generally unculturable. The fraction of the cells making up the soil microbial biomass that have been cultured and studied in any detail are negligible, often less than 5% (Borneman & Triplet, 1997; Torsvik, Goksøyr, & Daae, 1990). As direct DNA-based methods offer the possibility to assess the total microbial diversity present, thus bypassing the limitations of cultivation-based studies, recent years have seen the rapid development of such cultivation-independent methods for analyzing the microbial communities in soil (Øvreås & Torsvik, 1998; Akkermans, van Elsas, & de Bruijn, 1995). However, difficulties still exist as the inherent diversity of soil microbes renders taxonomic categorization and identification of the microbes challenging.

The heterogeneity of soil type and spatial distribution of soil microbes complicate the study of the microbial communities (Trevors, 1998). Considering the fact that traditional soil sampling involves taking replicates of only 1 to 5 g of soil. Such limited amount of soil samples could result in high variability between replicates and low statistical power (Klironomos, Rillig & Allen, 1999). The small scale of soil at which microbial communities exist, can lead to biased results as dominant species may obscure the presence of minority populations. To overcome the problem of spatial heterogeneity and distribution, it has been suggested that soil sampling be done on a smaller scale, with more samples in order to assess microbial diversity in microhabitats (Grundmann & Gourbiere, 1999).

2.4 Methods of Assessing Soil Microbial Diversity

2.4.1 Cultivation-dependent Method

From a general perspective, the methods used to assess the diversity of soil microbes fall broadly into cultivation-dependent and cultivation-independent technique. Traditionally, methods to analyze soil microorganisms have been based on cultivation and isolation (van Elsas, Duarte, Rosado & Smalla, 1998); a wide variety of culture media has therefore been designed to maximize the recovery of diverse microbial groups. Culture-based method includes plate counting and community level physiological profiling (CLPP). Selective plating could be done along with direct viable counts in order to assess the diversity of soil bacteria. CLPP can be used to assess bacterial diversity based on sole source carbon utilization (SSCU) patterns via the use of specially designed 96-well

microlitre plates supplemented with different types of carbon source (Choi & Dobbs, 1999). Unfortunately, cultivation-based methods are time consuming and underestimate microbial soil diversity as only a small fraction of the microbial cells in soil are accessible to study (Janssen, Yates, Grinton, Taylor & Sait, 2002).

2.4.2 Cultivation-independent Method

Recent advances in molecular technology has aided in the development of cutting-edge studies of soil microbial communities. These methods involve direct analysis of metagenomic DNA extracted from microbial communities. Numerous protocols have been adopted to isolate soil-derived microbial DNA. DNA extraction methods are separated into two categories in general: the direct lysis method (Ogram, Sayler, & Barkay, 1987) and the cell extraction method (Faegri, Torsvik, & Goksøyr, 1977). Direct lysis achieved by using combination of enzymatic treatment, high temperatures and detergent treatment. Additionally, it includes physical disruption such as sonication, grinding-freezing-thawing, and bead beating. Meanwhile, the cell extraction method involves the separation of cells from soil particle through differential centrifugation prior to lysis. (de Lipthay, Enzinger, Johnsen, Aamand & Sørensen, 2004). Direct lysis method is reported by Gabor and other researchers (2003) to yield higher DNA recovery. A 10- to 100-fold increase in DNA yield using direct lysis compared to cell separation method may be explained in part by the capacity of the direct lysis approach to extract additional nonbacterial and extracellular DNA (Steffan, Goksøyr, Bej & Atlas, 1988). Despite the fact that cell separation method is less

efficient in terms of DNA recovery, DNA isolated using this method are less harsh and seem to be less contaminated with matrix compounds, including humic substances (Courtois et al., 2001). Nevertheless, the choice of DNA extraction methods relies largely on the soil type. (de Lipthay et al., 2004).

Analysis may use techniques such as PCR, denaturing gradient gel electrophoresis and restriction fragment polymorphism analysis to produce community fingerprints. Such techniques also permit phylogenetic relationships between organisms to be examined (Pace, 1996). Torsvik and other researchers (1990) reported the first extraction of DNA from soil that was suitable for downstream applications. This method created the opportunity to adapt or specifically develop an arsenal of molecular methods for soil microbial purposes. Since these pioneering works, a continuously increasing number of studies have sought to explore the diversity of soil microbial communities with the help of molecular methods, most prominently using the 16S rDNA genes or conserved sequences as genetic markers (Amann, Ludwig & Schleifer, 1995; Pace, 1996).

2.5 The Use of Molecular Methods in Soil Microbiology

2.5.1 PCR and Primer Design

The introduction of the polymerase chain reaction (PCR) (Mullis, Faloona, Scharf, Saiki, Horn & Erlich, 1986; Mullis & Faloona, 1987; Mullis, 1990) into molecular biology represented another milestone, allowing for the specific detection and investigation of even minor traces of genetic material. PCR permits the rapid,

exponential amplification of DNA sequences using three processes repeated cyclically: denaturation, annealing and extension. Selecting appropriate primers for PCR is crucial for valid and reliable results. Primers are designed to be complementary to the specific region of the DNA to be amplified and must be at least 70% homologous to the template in order for annealing and elongation to be successful (Steyn & Holland, 1993). "Universal" primers such as those used to amplify the conserved 16S rRNA genes are often used in phylogenetic studies, as DNA from many species is amplifiable (Baker, Ah Tow & Cowan, 2003). Since the ability to fix nitrogen is distributed widely and inconsistently among prokaryotes, it is difficult to assess the distribution and diversity of nitrogen-fixing microorganisms based solely on 16S rDNA phylogenetic diversity studies. The highly conserved nature of the *nif*H gene makes it an ideal molecular tool to determine the potential for biological nitrogen fixation in any environment (Zehr & Capone, 1996).

Designing PCR primers that amplify the *nif*H gene and sequencing the amplified genes can identify the assemblage of microorganisms capable of nitrogen fixation in any environment. This approach has been applied to environments that are nitrogen limited or are known to support nitrogen fixation, such as oligotrophic oceans (Zehr, Mellon & Zani, 1998), marine microbial mats (Olson, Litaker, & Paerl, 1999; Zehr et al., 1995), modern marine stromatolites (Steppe, Pinckney, Dyble & Paerl, 2001), tropical sea grass beds (Bagwell et al., 2002), rice roots (Engelhard, Hurek, & Reinhold-Hurek, 2000; Ueda, Suga, Yahiro & Matsuguchi,

1995), and termite hindguts (Ohkuma, Noda, & Kudo, 1999). PCR amplification of conserved genes using consensus bacterial primers and separation of the resultant PCR amplicons wither by cloning, by denaturing gradient gel electrophoresis (DGGE) constitute the most popular molecular techniques used to describe soil bacterial ecology to date (Muyzer & Smalla, 1998).

2.5.2 Clone Libraries

Clone libraries are useful to identify and characterize the dominant bacterial or fungal types in soil and thereby provide a picture of diversity. Construction of a clone library for genes encoding rRNA (rDNAs) and rapid screening of the library based on sequence differences have provide a useful means to access genomic information from as large a pool of soil microbes as possible, including those that are not readily culturable. The feasibility of cloning environmental DNA provides a route to study the phylogenetic, physical, and functional properties of the metagenome (Rondon et al., 2000).

However, to accurately describe the microbial diversity within a soil sample, clone libraries usually need to be quite large. There are as yet few studies in which the representativeness issue has been satisfactorily resolved, and hence microbial diversity has not been adequately covered in most studies to date. Rarefaction analysis, calculation of coverage values, or other statistical techniques are needed to evaluate whether the number of screened clones is sufficient to

realistically estimate the true diversity (Curtiss, Sloan & Scannell, 2002; Ravenschlag, Sahm, Pernthaler & Aman, 1999).

2.5.3 Microbial Community Fingerprinting Techniques

A range of techniques has been developed to fingerprint soil microbial communities. These include denaturing gradient gel electrophoresis (DGGE) (Heuer, Krsek, Baker, Smalla & Wellington, 1997; Muyzer, deWaal, & Uitterlinden, 1993; Muyzer & Smalla, 1998), amplified rDNA restriction analysis (ARDRA) (Massol-Deya, Odelson, Hickey & Tiedje, 1995), terminal restriction fragment length polymorphism (T-RFLP) (Liu, Marsh, Cheng & Forney, 1997), single-strand conformational polymorphism (SSCP) (Schmalenberger & Tebbe, 2002), and ribosomal intergenic spacer length polymorphism (RISA) (Ranjard et al., 2001).

PCR-DGGE is probably the most widely used among the methods to study microbial communities in environmental samples. The procedure is based on electrophoresis of PCR-amplified 16S rDNA fragments in polyacrylamide gels containing a linearly increasing gradient of denaturants; separation is dependent upon the electrophoretic mobility of partially-denatured DNA molecules held together by a 40-bp GC clamp in the 5' primer (Myers, Fischer, Lerman & Maniatis, 1985; Sheffield, Cox, Lerman & Myers, 1989). Mobility differs according to internal sequence variation instead of length. The resulting banding pattern would thus provide a community profile most likely representing

particular species and their relative abundances by virtue of band position and relative band intensity. (Lerman, Fischer, Hurley, Silverstein & Lumelsky, 1984). Besides, T-RFLP involves amplification of 16S rDNA fragment and subjected to restriction enzyme digestion, which generates terminal restriction fragments of different lengths. One of the primers used was labeled with a fluorescent dye so that when the preparation was analyzed with an automated DNA sequencer, the sizes of only the terminal restriction fragment (T-RF) could be determined and the amount could be quantified. Due to the terminal RFLPs (T-RFLPs) found in 16S rDNAs, the method can be used with DNA from complex microbial communities and can provide a sensitive and rapid means to assess community diversity and obtain a distinctive fingerprint of a microbial community. (Liu et al., 1997).

T-RFLP has been shown to have slightly higher resolution compared to DGGE, though the usage of restriction enzymes could lead to an overestimation of microbial diversity as a result of incomplete digestion (Osborn, Moore & Timmis, 2000). T-RFLP and DGGE offer advantages over conventional cloning and sequencing, as they are reproducible, less time-consuming, and allow for simultaneous analysis of multiple samples, which enables comparison between microbial communities.

2.5.4 Molecular detection of diazotrophs: Marker genes

Since diazotrophs are phylogenetically diverse, they cannot be targeted as a single group by rRNA gene-targeting probes or primers. Therefore, specialized

molecular approaches were developed to target the functional genes of BNF instead (Kirshtein, Paerl, & Zehr, 1991; Normand & Bousquet, 1989). The basis for this approach is the conserved nature of the genes involved, as well as their phylogenetic information content (Hennecke et al., 1985; Normand & Bousquet, 1989). Of the core genes involved in BNF, *nif*H, *nif*D and *nif*K have been shown to be homologous in all diazotrophs (Hennecke et al., 1985). The genes for nitrogenase (*nif*DK) and nitrogenase reductase (*nif*H) all contain highly conserved regions, reflecting the strict structural requirements of the nitrogenase enzyme complex for proper catalytic functioning (Hennecke et al., 1985).

While *nif*H, *nif*D and *nif*K genes have all been analyzed phylogenetically, it is *nif*H that provides the best phylogenetic resolution (Hennecke et al., 1985; Normand & Bousquet, 1989; Hirsch et al., 1995). The advantage of using the *nif*H gene as a molecular marker is twofold. First, the protein sequence of *nif*H is well conserved and clusters with many known prokaryotic families similarly to family wise clustering in 16S rRNA sequences (Postgate & Eady, 1988). Second, *nif*H remains the most thoroughly studied of these genes, with an extensive collection of sequences obtained from both cultured and uncultivated organisms from multiple environments (Zehr et al., 2003). To date, characterization of *nif*H genes from various soil types has unraveled sequence types that belonged to diverse unidentified diazotrophs (Ueda et al., 1995; Widmer et al., 1999; Piceno & Lovell, 2000; Poly et al., 2001). These studies have highlighted that most of the nitrogen

fixation of soil diazotrophs are carried out by the uncultivated phylotypes (Hamelin et al., 2002; Tan, Hurek & Reinhold-Hurek, 2003).

2.5.5 Limitation of Molecular Methods for the Study of Diazotrophs Diversity

Although molecular methods have been generally accepted to circumvent the major limitations posed by culturing, the presences of intrinsic biases possess new shortcomings to the study of soil microbial diversity. The initial step in molecular approach is total DNA extraction. The usage of particular cell breakage regimes tends to bias the total community DNA content toward specific groups of bacteria (Prosser, 2002). As Gram-negative cells tend to break more easily compared to Gram-positive cells, mild breakage regimes may not sufficiently break Gram-positive cells, enriching the extract with Gram-negative DNA. Conversely, harsher regimes contribute to the opposite due to the increased likelihood for shearing of gram-negative DNA (von Wintzingerode, Gobel & Stackebrandt, 1997).

Furthermore, the extraction efficiency can also be aggravated with the co-extraction of inhibitory substances which include organic matter, clay particles, metals and humic acids (Macrae, 2000). The latter, negatively interfere with PCR efficiency, restriction enzyme digestion, DNA transforming and other downstream processes (Romanowski, Lorenz, Sayler & Wackernagel, 1992; Steffan & Atlas, 1988; Trevors, Lee & Cook, 1992). Besides, the presences of

contaminants necessitate purification protocols which result in loss of nucleic acid yield introduces potential biases into microbial diversity analyses (Kirk et al., 2004).

PCR amplification biases raise particular interest due to its prevalent usage in molecular methods (Suzuki & Giovannoni, 1996; Frostegard et al., 1999; Becker, Boger, Oehlmann & Ernst, 2000; Qiu et al., 2001; Schmalenberger, Schwieger & Tebbe, 2001; Ishi & Fukui, 2001). Potential contributors to differential amplification include prefential amplification due to low overall GC content, higher binding efficiency of GC-rich permutations of degenerate primers and the correlation between amplification probabilities and gene copy numbers (Crosby & Criddle, 2003; Farrelly, Rainey & Stackebrandt, 1995). Additionally, the use of universal primers were designed using existing 16S rRNA database, which only represent a portion of total species diversity of microbial world (Amann, Ludwig & Schleifer, 1995). Thus, the data obtained do not accurately represent the total microbial diversity exist in environmental samples. Nevertheless, with careful interpretation of analyses, molecular-based techniques can be prominent in the study of soil microbial diversity.

2.5.6 The phylogeny of *nif*H

Molecular phylogenetic analysis is expressed as phylogenetic tree. The use of molecular phylogenetics has become widespread in various studies to demonstrate the relationships among species of a gene or among loci of multigene family (Bos

& Posada, 2005). Phylogenetic trees are being generated to derive information regarding the processes responsible for the observed pattern of evolutionary relationships (Normand, Gouy, Cournoyer & Simonent, 1992). Phylogenetic analysis of nitrogenase proteins and their homologs indicate that they segregate into distinct, topologically consistent clades.

All *nif*H genes fall into one of four clusters (Chien & Zinder, 1994): Cluster I includes standard molybdenum nitrogenases from cyanobacteria and proteobacteria (α, β, and γ), as well as γ-proteobacterial *vnf*H. Examples of diazotrophs grouped under this cluster are *Gluconacetobacter, Rhizobium, Burkholderia, Methlyocystis and Klebsiella*. Cluster II includes methanogen nitrogenases and bacterial *anf*H. Cluster III includes nitrogenases from diverse anaerobic bacteria such as *Clostridia* (low G+C, gram positive) and sulfate reducers (δ-proteobacteria), which is an example of the *nif*H phylogeny deviating from the 16S rRNA phylogeny. Cluster IV includes divergent group of *nif*-like sequences, which are similar to distantly related chlorophyllide reductase genes, of archaeal origin (Chien, Auerbuch, Brabban & Zinder, 2000).

To date, the heterogeneous communities of diazotrophs are described, via the phylogenetic study of *nif*H, in different environments including forest soil (Poly et al., 2001; Rösch, Mergel & Bothe, 2002), lakes (Zani, Mellon, Collier & Zehr, 2000), deep-sea vents (Mehta, Butterfield, & Baross, 2003), agricultural soil (Poly et al., 2001), wetland soil and rhizospheres (Lovell, Piceno, Quattro & Bagwell,

2000; Hamelin et al., 2002). Although *nif*H has been particularly useful in phylogenetic studies, discrepancies in the aspect of taxanomic placement are occasionally reported especially when topologies of *nif*H and 16S rRNA trees are compared (Young, 1992). Such discrepancies can arise, presumably due to lateral gene transfer, as postulated by Norman and Bousquet (1989).

CHAPTER 3

MATERIALS AND METHODS

3.1 Materials

The materials used and their sources were listed in Table 3.1. The bacterial strains and plasmids were listed in Table 3.2.

Table 3.1: Materials used and their supplier

Materials	Source/Reference
GeneRuler 100bp DNA ladder, Isopropyl-beta-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-Gal), GeneJET™ Gel Extraction Kit	Fermentas
2X Rapid Ligation, T4 DNA Ligase, 5X Colorless GoTaq® Flexi Buffer, Magnesium chloride ($MgCl_2$), Taq Polymerase, dNTPs mix, T7 promoter primers, SP6 promoter primers	Promega
Biotools DNA markers 1kb ladder	B&M labs
PowerSoil™ DNA Kit	MO BIO Laboratories Inc
DNA-spin™ Plasmid DNA Purification Kit	intRON Biotechnology

Table 3.2: Bacterial Strains and plasmids used in this study

Bacterial Strains/Plasmids	Description	Source/Reference
Bacterial strains:		
E.coli JM109	Plasmid host for *lacZ* expression. Genotype: *recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17, *supE*44, *relA*1, Δ(*lac-proAB*)/F' [*traD*36, *proAB*$^{+}$, *lacI*^{q}, *lacZ*ΔM15]	Bethasda Research Laboratories (1986)
Plasmids:		
pGEM®-T Easy	Commercial plasmid vector used in this study as cloning vector. A high-copy-number vector containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows identification of recombinants by blue/white screening on indicator plates.	Promega
pUC19	Plasmid vector with ampicillin resistant gene, multiple cloning site and *lac*Zα	Yanisch-Perron Vieira, & Messing (1985)

3.2 Soil Sampling

3.2.1 Study Area

The sampling site for this study was located on an ex-tin mining land, at the periphery of a tin-mine lake in Universiti Tunku Abdul Rahman, Kampar, Perak campus, Malaysia (approx N 04^{o} 20.459' E 101^{o} 08.497'). Vegetation consisted predominantly of Taro plant (*Colocasia esculenta*). The image of the location of sampling sites was shown in Appendix A.

3.2.2 Soil Sample Collection

The soil was sampled on October 26, 2010. There was sporadic rainfall the day before the sampling was conducted. About 100g soil sample, at a depth of about 10cm from the soil surface was collected at the site. The uppermost soil surface was avoided due to high amount of organic matter which would adversely affect subsequent analysis. The soil sample was inserted into sterile plastic bags, transported and preserved at 4^{o}C until further analysis.

3.3 Soil Physicochemical Properties Analysis

The soil sample was labeled accordingly and outsourced to Forest Research Institute Malaysia (FRIM) for physical and chemical properties of soil. Soil sample was analysed for wet pH; soil texture, in the percentages of silt, clay and sand; organic carbon content; and total nitrogen content.

3.4 Total DNA Extraction

An amount of approximately 1 g of soil sample was used for total metagenomic DNA extraction, using *PowerSoil DNA Isolation Kit* (MO BIO Laboratories, Inc). The protocol of DNA extraction was based on the proprietary method of the kit. The extracted DNA was eluted in 100µL of elution buffer provided by the kit. Co-extracted humic acids are the major contaminant when DNA is extracted from soil. These compounds absorb at 230 nm whereas DNA absorbs at 260 nm and protein at 280 nm. Concentration and purity of the extracted DNA was assessed using a spectrophotometer (Nanodrop) to determine the 280/260 nm and 230/260 nm absorbance ratios. The extracted DNA was electrophoresed in 0.7% (w/v) agarose gel to determine the size of genome via comparison to a corresponding 0.5µg/µL 1kb ladder DNA markers (Biotools).

3.5 Nested PCR Amplification

3.5.1 Primary and Secondary PCR

Two pairs of degenerate primers were used for nested PCR amplification of a segment of the *nif*H gene. Both sets of primers were designed based on the conserved sequences in the *nif*H genes. Primary amplification was performed using outer primers, *nif*H3 (reverse) and *nif*H4 (forward) as detailed by Zani and co-researchers (2000) to amplify a region of approximately 470-bp fragment. Secondary amplification was performed using inner primers, *nif*H1(forward) and *nif*H2(reverse) as described by Zehr and McReynolds (1989) to amplify a region of approximate 360-bp fragment.

The annealing positions of the primers and the expected sizes of amplicon were illustrated in Figure 3.1. The primer sequences were shown in Table 3.3.

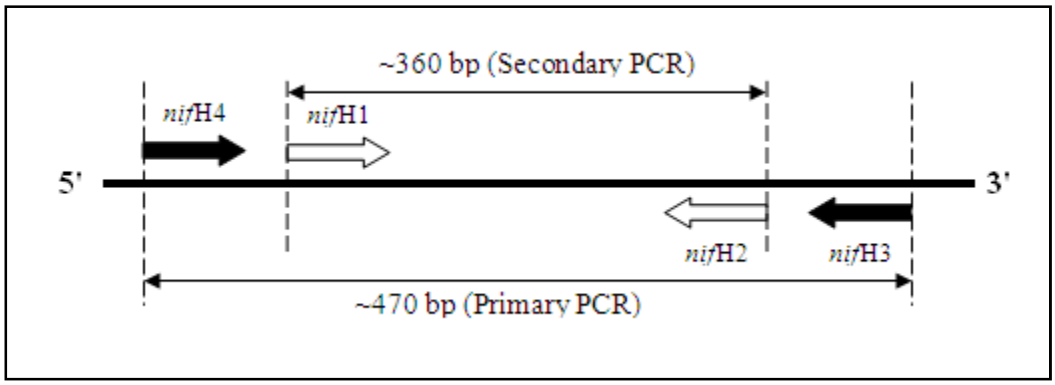

Figure 3.1: Schematic diagram showing the respective annealing positions of primers and the expected sizes of the PCR amplifications. Primary PCR reactions were expected to yield products of about 470 bp in size. Secondary PCR amplifications should yield amplicons with size corresponding to approximately 360 bp.

Table 3.3: Sequence of each degenerate primer used for *nif*H gene amplification.

Amplification	Primers	Sequences [1]
Primary PCR	*nif*H3	5' – ATR TTR TTN GCN GCR TA – 3'
	*nif*H4	5' – TTY TAY GGN AAR GGN GG – 3'
Secondary PCR	*nif*H1	5' – CTG YGA YCC NAA RGC NGA – 3'
	*nif*H2	5' – GDN GCC ATC ATY TCN CC – 3'

[1]Code of degeneracy: Y = T or C; R =A or G; D =A, G or T; and N = A, C, G or T

3.5.2 Nested PCR Parameter

The PCR reactions were carried out by using a non-gradient PCR thermal cycler (Biometra). All PCR reactions were conducted in 25µl aliquots containing 1X PCR Buffer, 2mM $MgCl_2$, 0.04U/µl Chromo *Taq* Polymerase, 0.2mM dNTP, 400 nM of each forward and reverse primer, 0.46ng/µl of DNA template and ddH_2O. The PCR parameters used for both primary and secondary PCR were tabulated in Table 3.4. An amount of 1µl of each primary PCR product served as template for secondary PCR amplification. The final nested PCR products were electrophoresed in 1.5% (w/v) agarose gel to determine for the presence of the approximate 360-bp fragment.

Table 3.4: Parameters for Primary and Secondary (Nested) PCR amplification.

Stage	Temperature (^oC)	Duration (minutes)	Cycle
Initial Denaturation	94	5	1
Denaturation	94	0.5	30
Annealing	48 (primary PCR)	0.5	
	59 (secondary PCR)	0.5	
Extension	72	0.5	
Final extension	72	10	1

3.5.3 Purification of PCR Products

The products from nested PCR amplification was pooled to a total of 50µL and electrophoresed in 1.5% (w/v) agarose gel**.** The band corresponding to the size of approximately 360bp was excised from gel and purified using *GeneJet*^{TM} *Gel Extraction Kit* (Fermentas) according to manufacturer protocol. The extracted DNA was eluted in 40µl elution buffer provided by the kit and stored in -20^{o}C. The purified DNA samples were electrophoresed on 1.5% (w/v) agarose gel to verify the presence of the approximate 360-bp fragment. Concentration and purity of the extracted DNA was then assessed spectrophotometrically.

3.6 Cloning and Colony PCR

3.6.1 Ligation of DNA fragments into pGEM-T Easy Vector

The purified products were ligated into pGEM-T Easy Vector (Promega) according to the manufacturer protocol. An amount of 3 µL of the purified PCR products were added to 1µL of pGEM-T Easy Vector (50ng), 1µL of T4 ligase (3U/µl) and 5µL of 2X Rapid Ligation Buffer. The ligation mixtures were incubated overnight at 4^{o}C for maximal yield of ligated products and subsequently stored in -20^{o}C.

3.6.2 Preparation of *Escherichia coli* strain JM109 Competent Cells

Competent cells were prepared using calcium chloride method as described by Sambrook, Fritsch & Maniatis (1989) with slight modification. *Escherichia coli* JM109 cells were streaked on LB agar and incubated overnight at 37^{o}C. A single

colony of *E. coli* strain was inoculated into 5mL LB broth and incubated overnight at 37^{o}C, with agitation at 200rpm. An amount of 500µL of the overnight inoculums was transferred into 25mL of the fresh LB broth and incubated at 37^{o}C, with agitation at 200rpm until it reached an OD_{600} of between 0.5-0.6. The cells were pelleted by centrifugation at 4000g for 5minutes at 4^{o}C. The supernatant was decanted and cells were resuspended in 3ml of ice-cold $CaCl_{2.}$ This was followed by incubation on ice for at least 2 hours before proceeding to transformation.

3.6.3 Transformation

An amount of three microlitre of ligation mixtures was gently mixed with 200µl of competent *E. coli* JM109 cells in a pre-chilled sterile 1.5mL microcentrifuge tube. Positive and negative controls were prepared using 100µL aliquots of competent cells each, with the addition of one microlitre of pUC19 to the former. The tubes were incubated on ice for one hour and subjected to heat-shock at 42^{o}C for exactly 90s. These were immediately incubated on ice for an additional 5 minutes. An amount of 900µL LB medium was added to each tube, followed by incubation at 37^{o}C for 45 minutes, with agitation at 80rpm. The cells were pelleted at 9000rpm for 10minutes. An amount of 900µl LB supernatant was decanted and the cells were resuspended in the remaining supernatant. Lastly, the cells suspensions were plated onto LB agar plates supplemented with 50µg/ml ampicillin, 20µL of 100mM IPTG, 20µL of 50mg/ml X-gal. The plates were then incubated for 16-18hours at 37^{o}C. The colonies developed were subsequently

screened for the colonies that possibly harbour the recombinant plasmids via blue-white screening. White colonies were randomly selected and patched onto a fresh LB agar plate supplemented with ampicillin (50 µg/mL), IPTG (100 mM) and X-Gal (50mg/mL) for further confirmation of the colour development of the colonies. Blue-white screening was made possible due to the α-complementation of *lacZ*ΔM15 gene in the *E. coli* JM109, by the *lacZα* gene present in the pGEM®-T Easy Vector.

3.6.4 Screening for clones with correct insert via Colony PCR

The white colonies were transferred into individual PCR mixtures using inoculation needle, with PCR mixtures consisting of 1X Go*Taq* Colorless Flexi buffer, 2mM $MgCl_2$, 0.04U/ml Chromo *Taq* Polymerase, 0.4µM of each T7 promoter primer (forward primer) and SP6 promoter primer (reverse primer), 200µM dNTP mix and ddH_2O to a final volume of 25µL. Both primers were universal primers and the sequences of the primers were shown in Table 3.5; The PCR parameters were shown in Table 3.6. The final PCR products were electrophoresed in 1.5% (w/v) agarose gel. A band corresponding to approximate 530-bp indicated the presence of insert.

Table 3.5: Sequences of T7 and SP6 universal primers

Primers	Sequences
T7 promoter (forward)	5'-TAA TAC GAC TCA CTA TAG GG-3'
SP6 promoter (reverse)	5'-ATT TAG GTG ACA CTA TAG-3'

Table 3.6: Colony PCR parameter

Stage	Temperature (^{o}C)	Duration (minutes)	Cycle
Initial Denaturation	96	3	1
Denaturation	96	0.5	
Annealing	50	0.25	30
Extension	60	2.5	
Final extension	60	10	1

3.6.5 Recombinant Plasmid Extraction

Recombinant colonies with plasmids of the correct insert size were inoculated into 5mL of LB medium supplemented with 50µg/mL ampicillin. The inoculum was incubated overnight at 37^{o}C, with agitation at 200rpm. The plasmids were harvested and purified using DNA-spin™ Plasmid DNA Purification Kit, according to the manufacturer's protocol. The purified plasmids were eluted in 50µL of elution buffer provided by the kit and stored at -20^{o}C. The plasmids were then electrophoresed on 1% (w/v) agarose gel to confirm presence of plasmid. Concentration and purity of the extracted DNA was then assessed spectrophotometrically to ensure the plasmids have met the requirements for sequencing.

3.7 DNA Sequencing and Phylogenetic Analysis

3.7.1 DNA Sequencing

Purified recombinant plasmids were sequenced with T7 and SP6 universal primers in both directions which were outsourced to First Base Laboratories Sdn. Bhd. The sequencing was carried out using *BigDye Terminator Cycle Sequencing* (Applied Biosystems).

3.7.2 Analysis of Nucleotide Sequences

The *nif*H clone sequences were examined and edited manually using Sequence Scanner v1.0 (Applied Biosystem). The resulting nucleotide sequences were aligned with BLASTX software hosted on the National Center for Biotechnology Information (NCBI) website. BLASTX is used to compare the newly determined DNA sequences against existing sequences in the NCBI non-redundant protein database. The software translated the nucleotide sequences and aligned the translated queries, based on local alignment algorithm, with known *nif*H polypeptide sequences. Hits with the highest score and lowest E-value were selected. The amino acid identities of the known bacteria with the nearest *nif*H polypeptide profile to the sequenced products were identified.

3.7.3 Multiple Sequence Alignment

All nucleotide sequences that were verified to be *nif*H homologous gene were translated into their respective polypeptide sequences via the ExPASy Proteomics tools. Compilation of the translated NifH polypeptide sequences were used to

perform the multiple sequence alignment using ClustalX 1.83 (Thompson, Gibson, Plewniak, Jeanmougin, & Higgins, 1997).

3.7.4 Construction of Phylogenetic Tree

The multiple sequence alignment of the NifH polypeptide sequences were used to construct the phylogenetic tree using MEGA 4.0 (Tamura, Dudley, Nei, & Kumar, 2007). A total of 52 known NifH polypeptide sequences were retrieved from the GeneBank database hosted on NCBI (refer to Table 3.7). These were then included in the construction of the dendrogram to aid in the categorization of diazotrophs. Evolutionary history was inferred using the neigbour-joining method (Saitou & Nei, 1987).

Table 3.7: The diazotrophs and their corresponding accession numbers of the nifH protein sequences.

Source description	Accession Number	Group
Acidithiobacillus ferroxidans	P00461	Beta-proteobacteria
Alcaligenes faecalis	Q44044	Beta-proteobacteria
Aminobacter sp. BA135	ACF19771	Alpha-proteobacteria
Anaeromyxobacter sp. Fw109-5	ABS27227	Delta-proteobacteria
Azospirillum brasilense strain SP7	P17303	Alpha-proteobacteria
Azotobacter vinelandii	P00459	Gamma-proteobacteria
Bradyrhizobium canariense	ACY38128	Alpha-proteobacteria
Bradyrhizobium elkanii	ABG74604	Alpha-proteobacteria
Bradyrhizobium japonicum	P06117	Alpha-proteobacteria
Bradyrhizobium sp. ANU289	P00463	Alpha-proteobacteria
Burkholderia mimosarum	AAT06092	Beta-proteobacteria
Clostridium cellulovorans 743B	ADL52532.1	Low G+C firmicute
Clostridium pasteurianum 1	P00456	Low G+C firmicute
Clostridium pasteurianum 3	P09553	Low G+C firmicute
Clostridium pasteurianum 4	P22548	Low G+C firmicute
Clostridium pasteurianum 6	P09555	Low G+C firmicute
Cyanothece sp Strain ATCC51142	AAB61408	Cyanobacteria

Table 3.7: (continued)

Source description	Accession Number	Group
***Desulfatibacillum alkenivorans* AK-01**	YP_002430688	Delta-proteobacteria
***Desulfovibrio aespoeensis* Aspo-2**	YP_004120783	Delta-proteobacteria
***Desulfovibrio magneticus* RS-1**	BAH75547	Delta-proteobacteria
Desulfovibrio vulgaris	AAS94434	Delta-proteobacteria
***Fischerella sp Strain* UTEX 1931**	AAC64642	Cyanobacteria
***Frankia alni strain* ArI3**	P08925	High G+C firmicute
***Frankia* sp. EUIK1**	AAC18640	High G+C firmicute
***Frankia* sp. Strain FaCl**	P46034	High G+C firmicute
Geobacter bemidjiensis Bem	ACH39087	Delta-proteobacteria
***Geobacter* sp. M21**	YP_003021955	Delta-proteobacteria
***Geobacter sulfurreducens* PCA**	AAR36215	Delta-proteobacteria
Gluconacetobacter diazotrophieus	AAD05046	Alpha-proteobacteria
Herbaspirillum seropedicae	CAA90932	Beta-proteobacteria
Klebsiella pneumonia	P00458	Gamma-proteobacteria
Methanococcus maripaludis	AAC45512	Archaea
Methanosarcina barkeri	P54799	Archaea

Table 3.7: (continued)

Source description	Accession Number	Group
Methanothermobacter marburgensis	Q50785	Archaea
***Methanothermobacter thermoautotrophicum* 1**	O27602	Archaea
***Methanothermobacter thermoautotrophicum* 2**	O26739	Archaea
***Methanothermococcus thermolithotrophicus* 2**	P08625	Archaea
Methylobacter luteus	CAD91849	Gamma-proteobacteria
***Methylocystis* sp. LW5**	AAK97419	Alpha-proteobacteria
Nostoc* sp. *PCC7120	P00457	Cyanobacteria
***Paenibacillus azotofixans* 1**	CAD56229	Low G+C firmicute
***Paenibacillus azotofixans* 2**	CAC27791	Low G+C firmicute
***Paenibacillus azotofixans* 3**	CAC27795	Archaea
***Pelobacter carbinolicus* DSM 2380**	ABA89338	Delta-proteobacteria
***Pelobacter propionicus* DSM 2379**	ABL01060	Delta-proteobacteria
Rhizobium leguminosarum	P00461	Alpha-proteobacteria
***Rhizobium* sp Strain NGR234**	P19068	Alpha-proteobacteria
Rhodobacter sphaeroides	AAB86864	Alpha-proteobacteria

Table 3.7: (continued)

Source description	Accession Number	Group
Sinorhizobium meliloti	P00460	Alpha-proteobacteria
Spirochaeta thermophile **DSM 6192**	ADN01326.1	Delta-proteobacteria
Syntrophobacter fumaroxidans **MPOB**	YP_845148	Delta-proteobacteria
Thermodesulfovibrio yellowstonii **DSM 11347**	ACI21914.1	Delta-proteobacteria
Trichodesmium **sp Strain IMS101**	AAS03796	Cyanobacteria

CHAPTER 4

RESULTS

4.1 Analysis of Soil Physicochemical Properties

A summary of the physicochemical property analysis of soil sample is shown in Table 4.1. Soil sample was slightly alkaline, where wet pH values at 7.94. The sampling site was found to have low nutrient availability, with the organic carbon and total nitrogen percentages of 1.16 and 0.10 respectively. Particle composition shows a high tendency toward large particle size. Sand predominates at 83%, followed by clay (12%) and silt (4%). The soil sample was categorized as loamy coarse sand by utilizing soil texture calculator available at http://soils.usda.gov/technical/aids/investigations/texture/. which incorporated the use of soil texture triangle shown in Figure 4.1.

Table 4.1: Physicochemical Properties of soil sample from study site

Physicochemical Properties	Soil sample
Wet pH	7.94
Total Nitrogen (%)	0.10
Total Organic Carbon (%)	1.16
Coarse sand (%)	69
Fine sand (%)	14
Silt (%)	4
Clay (%)	12
Texture	Loamy Coarse Sand

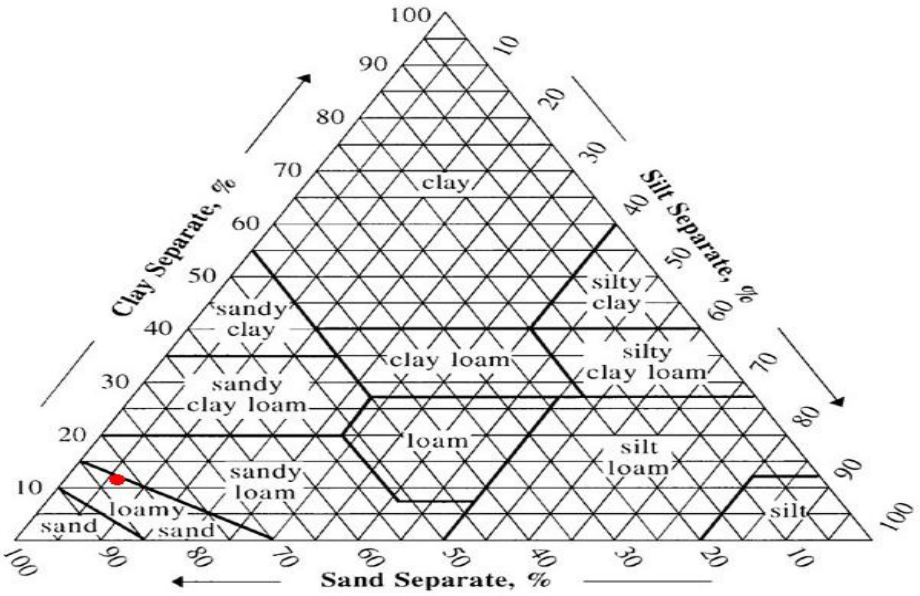

Figure 4.1: Soil texture triangle. The red dot positioned in the chart indicates the soil texture of sample according to its composition. (Source: http://soils.usda.gov/technical/aids/investigations/texture/triangle.jpg)

4.2 Total Soil DNA Extraction

Based on the gel image, as illustrated in Figure 4.2 (lane 2), the extracted total DNA molecules was more than 10-kb, indicating high molecular weight, with no shearing observed. The DNA purity ratio and concentration were tabulated in Table 4.2. The A_{260}/A_{280} purity indices of DNA extracted was below 1.8, which indicates presence of residual protein contaminants and other possible contaminants absorbing at particular 230nm wavelength. Meanwhile, the low A_{260}/A_{230} indices of DNA extracted was only 0.62, which was below 2.0, indicating residual humic acid contamination.

Table 4.2: Purity and concentration of total DNA extracted from soil measure spectrophotometrically.

Parameter	Soil sample
$\mathbf{A_{260/280}}$	1.48
$\mathbf{A_{260/230}}$	0.62
Concentration (ng/µL)	11.6

4.3 Nested PCR and Gel Purification of PCR products

The extracted total DNA from soil sample was amplified using nested PCR approach. The nested PCR amplicon include primary and secondary products. As illustrated in Figure 4.2, smearing was observed in lane 3 containing primary PCR product while distinct and resolved band at expected size of approximate 360-bp was observed in lane 5 containing secondary PCR product. Meanwhile, the

negative control lanes (lane 4 and 6) showed absence of contamination. The band corresponding to approximate 360-bp size was then gel-purified. Gel purified band in lane 7 showed much resolved band and devoid of smearing.

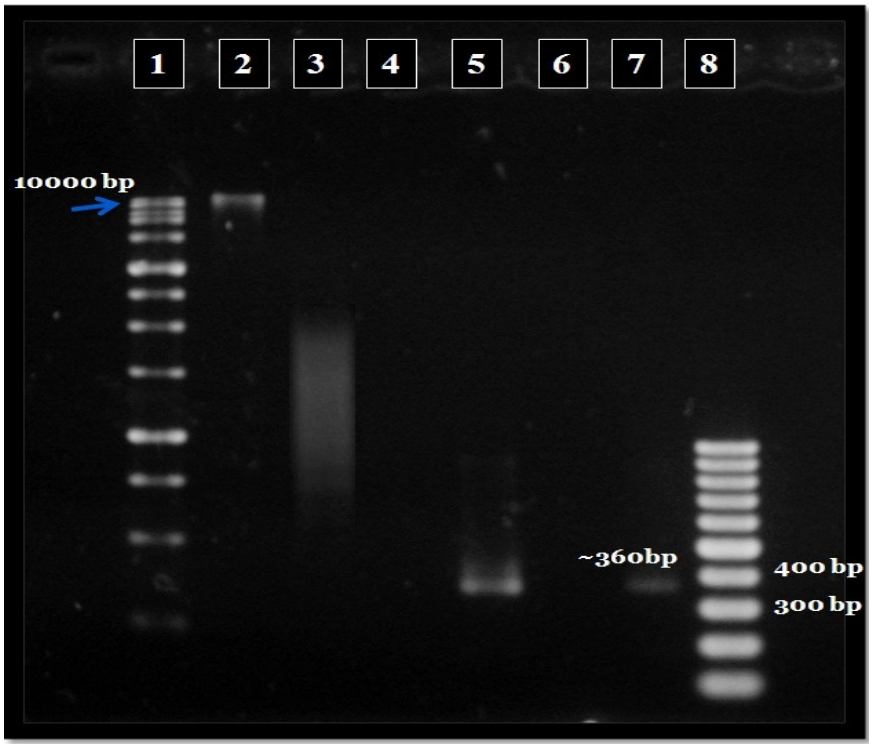

Figure 4.2: Photograph of ethidium bromide-stained 1.5% agarose gel image of total soil DNA, primary PCR product, secondary PCR product and gel purified secondary PCR product. Lane 1: 1kb DNA ladder; Lane 2: Total soil DNA; Lane 3: Primary PCR product **;** Lane 4: Negative control for primary amplification, using ddH_2O as template**;** Lane 5: Secondary PCR product**;** Lane 6: Negative control for secondary amplification, using ddH_2O as template; Lane 7: Gel purified secondary PCR product; Lane 8: 100bp DNA ladder.

4.4 Transformation of *E. coli* JM109

Blue-white screening was employed to screen for colonies that possibly harbor the recombinant plasmids. Blue bacterial lawn was observed on the positive control plate (Figure 4.2a) indicated high transformation efficiency. There was no growth of colonies observed on negative control plate (Figure 4.2b). This indicated contamination free during transformation. White colonies were observed amongst a background of blue colonies on plates of transformed cells which were incubated with recombinant plasmids harbouring gel purified product (Figure 4.2c). White colonies are cells containing inserted DNA ligated into pGEM®-T Easy Vector. This suggested that transformation of *E .coli* cells with recombinant plasmids was successful.

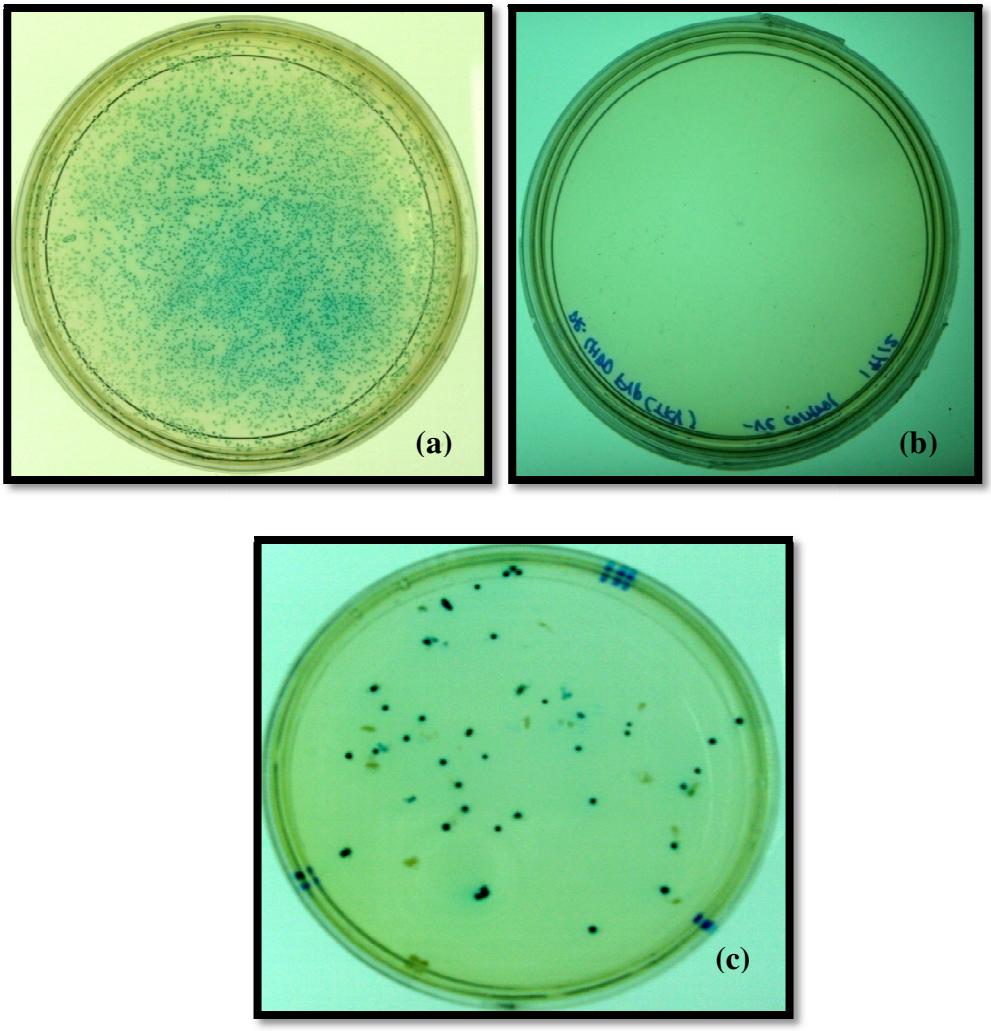

Figure 4.3: Positive and negative control plates and plates showing the growth of *E.coli* JM109 cells transformed with ligation mixture of gel purified secondary PCR products. (a) Blue bacterial lawn formed on positive control plate. (b) No colonies formed on negative control plate (c) Several white colonies and blue colonies formed on the plate which inoculated with cells that were transformed with ligated secondary PCR amplicons.

4.5 Colony PCR

All white colonies that underwent colony PCR showed distinct bands of slightly over 500-bp, corresponding to the expected fragment size (Figure 4.5). This reaffirmed that white colonies developed from transformation yielded insert of approximately 360-bp. Meanwhile, blue colonies that underwent colony PCR showed distinct band at approximate 170-bp. Transformed cells that produced blue colonies may contain self-ligated pGEM-T vectors.

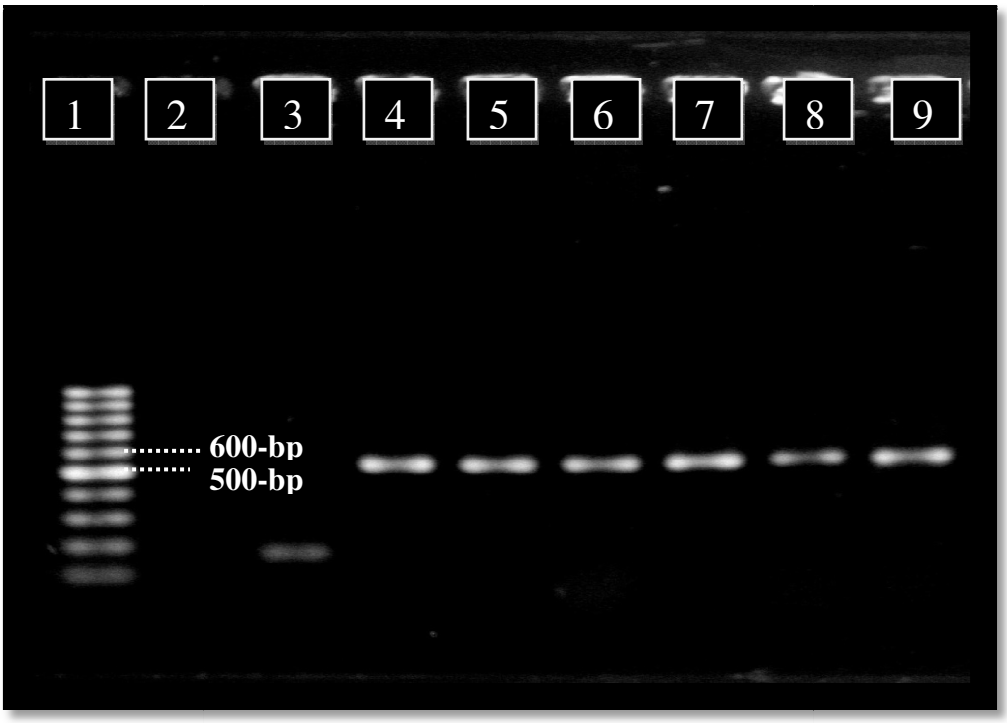

Figure 4.4: Photograph of ethidium bromide-stained 1.5% agarose gel image of colony PCR products of selected white colonies and blue colonies. Lane 1: 100bp DNA ladder; Lane 2: Negative Control (amplification without insertion of colony); Lane 3: pGEM-T (amplification with insertion of non-recombinant colony); Lane 4: pTKY2312-1; Lane 5: pTKY2312-2; Lane 6: pTKY2312-3; Lane 7: pTKY1612-14; Lane 8: pTKY1612-15; Lane 9: pTKY1612-16.

4.6 Recombinant Plasmid Extraction

Recombinant colonies were subjected to plasmid extraction using DNA-spin™ Plasmid DNA Purification Kit. The extracted plasmids were subjected to agarose gel electrophoresis and bands corresponded to approximately 2.5kb was observed. The band does not correspond to the expected fragment size of approximately 3.4kb since vector was 3kb in size while insert was 360bp in length. The possible explanation was bands of approximately 2.5kb consisted of plasmids of supercoiled conformation.

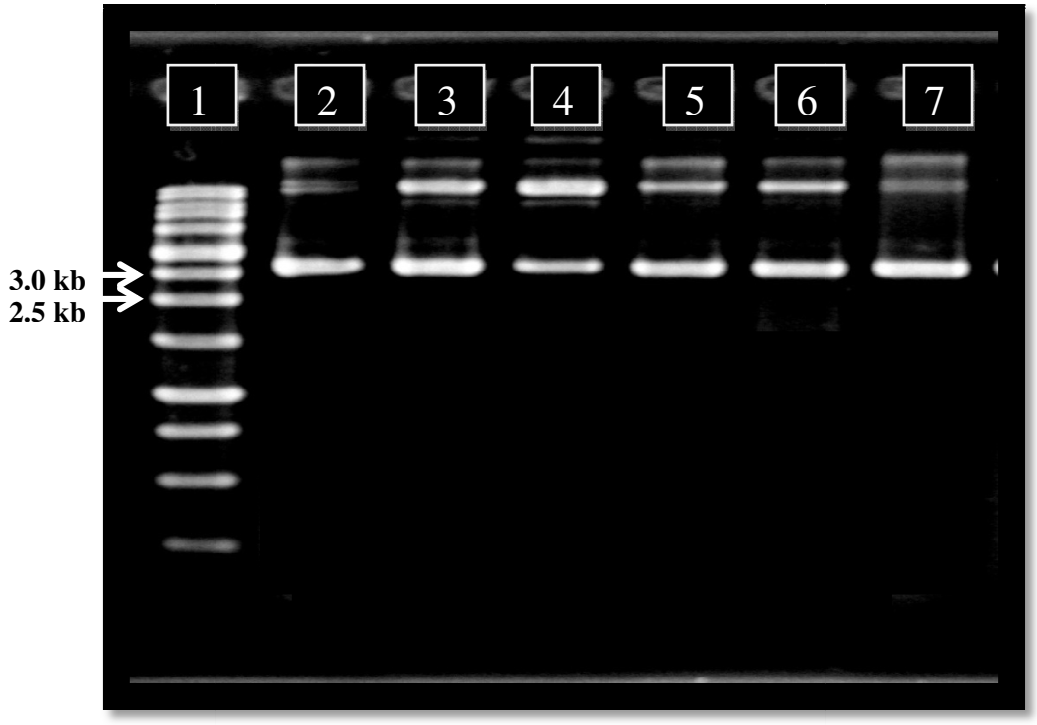

Figure 4.5: Photograph of ethidium bromide-stained 1% agarose gel image of extracted recombinant plasmids containing inserts of interest. Lane 1: 1kb DNA ladder; Lane 2: pTKY2312-1**;** Lane 3: pTKY2312-2**;** Lane 4: pTKY2312-3**;** Lane 5: pTKY1612-14**;** Lane 6: pTKY1612-15**;** Lane 7: pTKY1612-16

4.7 BLASTX Analysis

A total of 22 recombinant plasmid samples were obtained and sent for sequencing. Analysis of BLASTX showed that 19 DNA fragments out of 22 were confirmed to contain insert corresponding to *nif*H homology gene. BlastX alignment results were illustrated in Table 4.3, 4.4 and 4.5, displaying with the percentage of amino acid identities, E-values of the alignments and the hits on the known bacteria with the nearest NifH polypeptide sequences.

BLASTX homology search revealed that 15 clones which are pTKY2312-1, pTKY2312-2, pTKY2312-3, pTKY2312-7, pTKY2312-8, pTKY2312-9, pTKY2312-10, pTKY1612-16, pTKY1612-15, pTKY1412-13, pTKY1412-8 pTKY1412-2, pTKY0212, pTKY1612-3 and pTKY1612-9 had high amino acid identities (90%-99%) and low E-values (between $1e^{-54}$ and $9e^{-57}$) with *nif*H sequence of *Bradyrhizobium japonicum*. Of these, two clones which are pTKY2312-2 and pTKY2312-7 showed identical sequences.

Meanwhile, BLASTX homology search suggested that pTKY1612-17 clone shared 91% amino acid identity, E-values of $1e^{-57}$ with *nif*H sequence of *Desulfatibacillum alkenivorans* AK-01. Besides, pTKY1612-14 clone shared 96% amino acid identity with E-values of $9e^{-62}$ with *Geobacter* sp. M21. The alignment of pTKY1612-5 clone show highest similarity to *Desulfovibrio aespoeensis* Aspo2 with 88% amino acid identity and E-values of 7e-^{56}. The alignment of single clone, pTKY1612-2 has the highest hit with *Syntrophobacter fumaroxidans*

MPOB with 91% amino acid identity and E-values of $1e^{-56}$. All of these matches showed relatively low E-values (less than $3e^{-13}$) and high amino acid identities, suggesting that they were significant. This result indicates that the homologues obtained from BLASTX homology search were reliable.

Table 4.3: Recombinant plasmids and their corresponding amino acid identities of partial nifH sequences compared to that of the nearest known bacteria.

Plasmids	Identities	E-value	Phylogenetic	Nearest Known Bacteria
pTKY2312-2, pTKY2312-7	111/120 (93%)	$4e^{-59}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY2312-1	108/120 (90%)	$5e^{-57}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY2312-3	111/120 (93%)	$5e^{-58}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY2312-8	109/120 (91%)	$8e^{-58}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY2312-9	107/120 (90%)	$1e^{-56}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY2312-10	114/120 (95%)	$5e^{-60}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY1612-16	111/120 (93%)	$4e^{-58}$	alpha-proteobacteria	*Bradyrhizobium japonicum*

Table 4.3 (continued)

Plasmids	Identities	E-value	Phylogenetic	Nearest Known Bacteria
pTKY1612-15	119/120 (99%)	$9e^{-57}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY1612-9	109/120 (91%)	$4e^{-57}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY1612-3	116/120 (97%)	$2e^{-61}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY1412-13	110/120 (92%)	$2e^{-57}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY1412-8	110/120 (92%)	$3e^{-57}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY1412-2	118/120 (99%)	$2e^{-63}$	alpha-proteobacteria	*Bradyrhizobium japonicum*
pTKY0212	114/120 (95%)	$1e^{-54}$	alpha-proteobacteria	*Bradyrhizobium japonicum*

Table 4.3 (continued)

Plasmids	Identities	E-value	Phylogenetic	Nearest Known Bacteria
pTKY1612-17	109/120 (91%)	$1e^{-57}$	delta-proteobacteria	*Desulfatibacillum alkenivorans* AK-01
pTKY1612-14	115/120 (96%)	$9e^{-62}$	delta -proteobacteria	*Geobacter sp. M21*
pTKY1612-5	105/120 (88%)	$7e^{-56}$	delta -proteobacteria	*Desulfovibrio aespoeensis Aspo-2*
pTKY1612-2	109/120 (91%)	$1e^{-56}$	delta -proteobacteria	*Syntrophobacter fumaroxidans MPOB*

4.8 Multiple Sequence Alignment

The *nif*H nucleotide sequences of nineteen clones were translated into their corresponding polypeptide sequences. Multiple sequence alignment was performed on compilation of all resulting sequences using ClustalX 1.83 (Refer to Appendix D). The aligned sequences were used to construct NifH polypeptide phylogenetic tree.

4.9 Construction of Phylogenetic Tree

Along with 19 partial NifH polypeptide sequences obtained from this study, 52 NifH polypeptide sequences of known diazotrophs were retrieved from GenBank database to serve as reference sequences. All NifH sequences were used to construct dendrograms of the NifH polypeptide phylogenetic tree, as shown in Figure 4.6. Most of NifH clones were grouped in Cluster III while few clones were grouped in Cluster I.

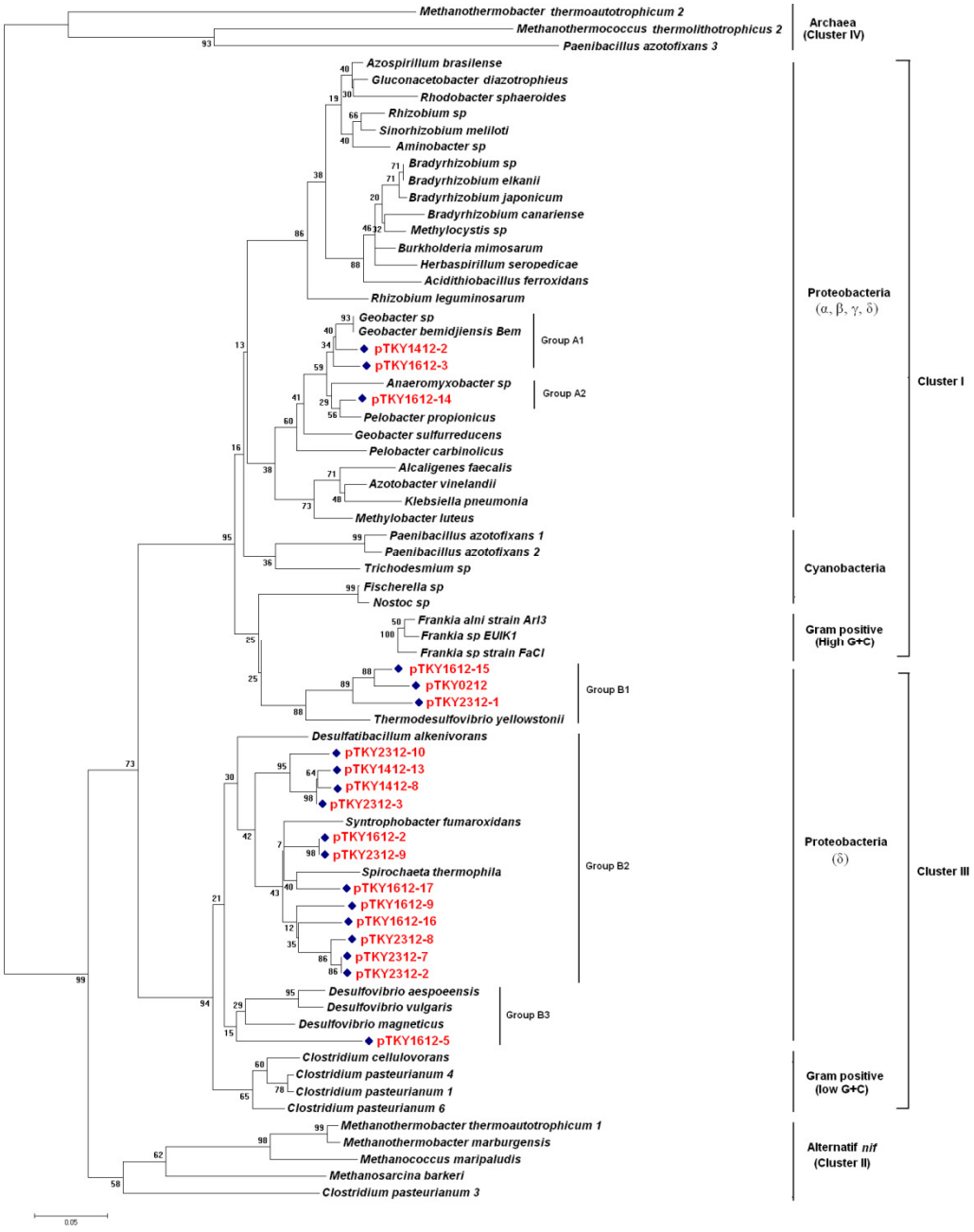

Figure 4.6: NifH phylogeny obtained using 52 known NifH amino acid sequences of representative diazotrophs and 19 translated amino acid sequences from this study (indicated by ♦). The evolutionary history was inferred using Neighbour-Joining method. The bootstrap consensus tree was inferred using 1000 replicates. The scale bar denoted 0.05 amino acid substitutions per site.

CHAPTER 5

DISCUSSION

5.1 Soil Sampling

5.1.1 Choice of study site

In this study, a single independent site of ex-tin mining land was selected for sampling. Since ex-tin mining land is commonly known to be deprived of nutrients (Awang, 1994; Majid, Hashim, & Abdol, 1994), the land may only harbored limited microorganism. Hence, to retrieve soil diazotrophs inhabiting the barren ex-tin mining land, it was justifiable that the soil sampling had to be done at location especially rhizospheres, which could support higher population density of microbes, including diazotrophs (Finan, O'Brian, Layzell, Vessey, & Newton, 2002).

The studies by Wang and Dick (2004) stated there is probably no portion of the soil richer in microbial abundance and diversity than the plant rhizosphere. Therefore, occurrence and distribution of nitrogen fixing bacterial community within rhizosphere of plants such as smooth cordgrass (*Spartina alterniflora*) (Lovell, Piceno, Quattro & Bagwell, 2000), oilseed rape (Kaiser, Püler & Selbitschka, 2001), oat (Avena sativa) (Soares, Roesch, Zanatta, de Oliveira Camargo & Passaglia, 2006), sorghum (Coelho et al., 2009), rice (Oryza sativa) (Wartiainen, Eriksson, Zheng & Rasmussen, 2008) were constantly being studied.

In addition, Shanta, Mubassara & Khan (2006) reported the nitrogen fixing potential of isolates of *Azospirillum* spp. were found to occur in rhizosphere soil of *Colocasia esculenta*. This had encouraged the sampling to be conducted around this plant rhizosphere to ensure higher probability of recovering diazotrophs from the soil.

5.1.2 Sampling Strategies

Soil sample was collected at a depth of 10cm from the soil surfaces, which labeled as A-horizon (5 – 25cm). (refer Appendix for schematic diagram of soil profile). Sampling at greater depth was precluded due to decreased diazotroph abundance in relation to increasing soil depth, which attributed to reduction of organic substrates that serve as microbial food sources (Mergel, Schmitz, Mallmann & Bothe, 2001; Rösch et al., 2002). Meanwhile, sampling on the topmost layer, O-horizon (0-5cm) was prevented as the layer of soil was dominated by organic material which may result in downstream complications if co-extracted (Food and Agriculture Organization of the United Nations, 1998).

A single gram of soil may contain 10^{10} bacterial cells (Faegri et al., 1977) and represent more than 13,000 species of bacteria as revealed by DNA analysis (Torsvik et al., 1990). Therefore, only 1g of the 100g of soil sample collected was used for analyses of diazotroph diversity in this study. However, Kirk and other researchers (2004) reported that the low amount of sampled might bias the study outcome since dominant species might eclipse the presence of the minority

factions. Since this study was meant to be an initial survey of the diazotrophic communities inhabiting nutrient inefficient ex-tin mining land, the low sampling amount was reasonable.

5.1.3 Soil Physicochemical Properties

The physicochemical property of soil is one of the key determinative factors in the study of soil microbial diversity (Kennedy & Smith, 1995). The analysis of the soil properties revealed that the total nitrogen and total organic carbon were 1.16% and 0.10%, respectively. The result obtained was consistent with the independent studies conducted by Awang (1994), which reported low total nitrogen (1.08%) and total organic carbon content (0.06%) in soil depth of 0cm-14cm. Organic carbon and nitrogen are far too low compared to normal soils under Malaysian conditions. This is somewhat related to the nature of the deposits which have just been exposed by the mining operation. Using data from Awang (1994) as references, it was reasonable to state that soil from ex-tin mining land in this study was deprived of carbon and nitrogen. The amount of organic carbon and total nitrogen in soil are important indicator of soil bacterial community. Low carbon and nitrogen content could cater low microbial populations and in turn affect the distribution of soil microbial which including diazotrophs.

From the soil texture triangle, the percentage of clay, silt and sand of soil sample was bordered on loamy sand. In terms of pH, soil from sampling site was slightly alkaline. According to a study by Liu and researchers (2010), it was found that

soil pH correlated strongly with both the total microbial and the functional microorganism community. Instead of low pH, alkalinity favours nitrogen fixation and this pH profile deemed beneficial for this study. The hypothesized reason was that low pH delays the development of nitrogenase activity (Roper & Smith, 1991; Limmer & Drake, 1998; Nelson & Mele, 2006). The alkalinity of soil from ex-tin mining land is possibly due to the influence of the limestone which underlies the deposits (Shamshuddin, Nik & Paramananthan, 1986).

5.2 Total DNA Extraction

As illustrated in Figure 4.1, gel image showed that the extracted total DNA molecules was more than 10-kb, indicating high molecular weight. Spectrophotomeric assessment of the total microbial DNA yielded $A_{260/280}$ ratios of 1.48, which was lower than the acceptable limits (between 1.8 and 2.0) of pure DNA. This indicated low residual protein contamination. Meanwhile, another purity ratio, $A_{260/230}$ indices gave 0.42, an indication of high contamination from humic substances, particularly humic acid, considering the acceptable ratio of pure DNA is of above 2.0 (Steffan, Goksøyr, Bej, & Atlas, 1988).

The frequent co-extraction of humic substances along with soil DNA is attributed to their similar size and charge characteristic. As addressed by Tsai & Olson (1992), the presence of humic acid would reduce efficiency of some molecular technique particulary with regards to PCR amplification due to their chelation of Mg^{2+} ions required by *Taq* polymerase. However, in this study, the high level of

humic acid contamination (high absorbance at A_{230}) does not tally with the success of the subsequent nested PCR amplification. Without further investigation, we can only hypothesize as to the possibility that presence of other substances that absorb A_{230}, which do not pose complications to the ensuing downstream processes.

5.3 Nested PCR

PCR amplification was performed in this study using a nested PCR approach, which has been adopted in several studies for *nif*H characterization (Choo, Samian, & Najimudin, 2003; Yeager et al., 2004). This approach was favoured as nested PCR is less affected by sample inhibition compared to single-stage PCR (Zani et al., 2000). In addition, Widmer et al. (1999) proven in their findings that nested PCR approach based on conserved primer target improved the specificity and sensitivity of *nif*H amplification.

Two sets of degenerate (universal) primers were used to amplify a relatively short (~360bp) but phylogenetically informative segment of *nif*H gene (Zehr & McReynolds, 1989). Degenerate primers were used to amplify most of the diazotrophs from representative taxa as it compromised specificity during amplification (Zehr & McReynolds, 1989). For the primary amplification, the outer primers (*nif*H 3 & 4) were used. As depicted in Figure 4.1, smearing was observed in primary PCR product, which attributed to the amplification of irrelevant sequences flanked within both priming sites (Roux, 1995). For

secondary PCR, inner primers (*nif*H 1 & 2) were used due to their complementarity to the gene fragement that was amplified by the outer primers (*nif*H 3 & 4). Secondary product was observed to yield specific and more resolved band, as only target DNA fragment of an appromixate 360bp was amplified.

The study by Widmer et al. (1999) showed that nested PCR approach would increase the specificity and sensitivity of *nif*H detection. The improvement resulted in highly specific detection of *nif*H genes in bulk environmental DNA samples. Meanwhile, increased sensitivity allowed the minority populations of diazotrophs to be detected as well, which is important in obtaining an accurate representation of the target community. Goto and researchers (2005) reported that great concern of primer contamination should be taken as this contamination resulted in false-positive results when searching for *nif*H genes in environmental samples. This problem is often caused by contamination with DNA from various sources including PCR reagents contained *nif*H gene and *nif*H-like DNA. For this reason, negative control was imperative. Gel image of Figure 4.1 showed the absence of any contamination.

5.4 Clone Libraries and Comparative Alignment of *nif*H Sequences

The PCR amplicons were cloned into pGEM®-T cloning vector prior to sequencing. This vector was used for PCR product ligation and cloning since it enables screening of PCR products by blue-white selection. The approach of subcloning the PCR product into a plasmid vector was used due to the advantage

that the amplified fragment can be sequenced with greater reliability (Alan & Tanner, 1997) On the other hand, direct sequencing was circumvented since noise in the earlier portion of sequencing would occur where the nucleotide sequences will most probably not be resolved. This could due to presence of primer dimers or possible contaminants (Applied Biosystems, 2010). It should be noted that only 22 clones were randomly selected among the myriads of others for DNA sequencing. This method did not ensure that all unique sequences would be discovered as rarer sequences might be missed. Therefore, the actual number of novel sequence could be even higher.

From BLASTX analysis, a total of 19 novel *nif*H sequences were obtained. BLASTX alignment showed that all of *nif*H clones showed high identities to unculturable *nif*H microorganism. Fifteen out of the nineteen *nif*H sequences aligned to the closest diazotroph, *Bradyrhizobium japonicum* with high amino acid identities ranging from 90-99%. *Bradyrhizobium japonicum* is a species of legume-root nodulating, microsymbiotic nitrogen-fixing bacterium species which well-known to form symbiotic relationship with soybean plant, *Glycine maz* (Stacey, Sanjuan, Luka, Dockendorff & Carlson, 1995). *Bradyrhizobium japonicum* belongs to the family, Rhizobiaceae which includes other nitrogen-fixing bacteria that develop symbiosis with legumes. These are bacteria commonly found in agriculture soils, habitats which are different from that of the ex-tin mining land. Nevertheless, the BLASTX analysis is not conclusive and the identity of the bacteria could only be speculated as the hits only suggested the

closest known identity of microorganism found. Thus, only preliminary assumption should be made that the exact identity of the bacteria might be species related to *Bradyrhizobium* genus.

BLASTX homology search suggested a single clone, pTKY1612-17 was likely to be associated with *Desulfatibacillum alkenivorans AK-01*. *Desulfatibacillum alkenivorans* is a mesophilic sulfate-reducer capable of utilizing C13 – C18 alkanes as growth substrates. This bacterium is useful as a tool of biotransformation by degrading alkanes in the process of bioremediation (So & Young, 1999). Another single clone, pTKY1612-14 was revealed to be closely similar to *Geobacter* sp. M21. *Geobacter* sp. (strain M21) is an anaerobic, chemolithotrophic Gram-negative bacterium. Bacteria from this genus are of interest because of their novel electron transfer capabilities, impact on the natural environment and their applications in the bioremediation of contaminated environments (Childers, Ciufo & Lovely, 2002). Similarly, these bacteria may possess potential use in the rehabilitation of ex-tin mining land.

According to BLASTX alignment, pTKY1612-2 showed closest similarity with *Syntrophobacter fumaroxidans* MPOB. *Syntropobacter fumaroxidans* is a syntrophic propionate-oxidizing bacterium, which was isolated from a culture enriched from anaerobic granular sludge. It oxidizes propionate syntrophically in co-culture with the hydrogen- and formate-utilizing *Methanospirillum hungateii*, and is able to oxidize propionate and other organic compounds in pure culture

with sulfate or fumarate as the electron acceptor (de Bok, Plugge, & Stams, 2004). An interesting case was noted where pTKY1612-5 sequence was revealed to be very similar to *Desulfovibrio aespoeensis* Aspo-2. *D. aespoeensis* strain Aspo-2 is recognized as a new species of the genus *Desulfovibrio, Desulfovibrio aespoeensis* is a sulfate-reducing, halotolerant bacterium that is able to adapt to conditions of high salinity and frequently found in the deep granitic rock (Motamedi, & Pedersen, 1998). The ability of this sulfate-reducing bacterium to reduce metals to a less toxic form, associated with its precipitation, is a potentially useful process for bioremediation. Thus, these bacteria may possess potential use in the rehabilitation of ex-tin mining land.

From the BLASTX alignment, fifteen out of the nineteen *nif*H sequences had shown to be closely similar to *Bradyrhizobium japonicum*. It could deduce that the sampling site in this study was predominated by diazotrophs that shared high similarity to *Bradyrhizobium* genus. This may imply that these microorganisms were largely dominating the sampling site and result prefentially amplified by PCR amplification. Although the nearest known bacteria were used as the hits for the BLASTX alignment results, the results frequently displayed the highest hit to uncultured nitrogen-fixing bacteria. Due to this limitation, the identities of most *nif*H sequences remained unknown. Hence, the evolutionary relationships and phylogenetic associations of the clones were inferred by using phylogenetic analysis.

5.5 Phylogenetic Analysis

The phylogenetic tree constructed in this study was based on NifH polypeptide sequences instead of nucleotide to eliminate distance artifacts due to different G+C ratios (Normand et al., 1992). As explained by Zehr and McReynolds (1989), vast phylogenetics differences were observed among diazotrophs due to the evolutionary timeline and the *nif*H gene sequences have diverged substantially. Although the sequences at the nucleotide level might be variable, the conserved regions were still observed at the amino acid levels. This discrepancy is due to the codon degeneracy of most amino acid levels. (Zehr & McReynolds, 1989). This approach is in agreement with studies such as Hamelin, Fromin, Tarnawski, Teyssier-Cuvelle, and Aragno (2002) and Burgmann, Widmer, Von Sigler & Zeyer (2003), which used amino acid sequences to construct the dendrograms.

Overall, nineteen *nif*H sequences obtained from this study were translated into amino acid sequence and aligned with the corresponding 52 known NifH amino acid sequences of representative diazotrophs extracted from NCBI database (adapted from Choo et al., 2003). The constructed NifH phylogenetic tree had a topology similar to that of the published phylogenetic tree constructed by Chien and Zinder (1996) and Choo et al., (2003), with respect to the NifH polypeptide sequences were grouped into four major clusters. Although several nodes were devoid of reliable bootstrap values (<50%), the consistency of the topology with established trees from other studies rendered the trees to be valid. The low bootstrap values were reported by Widmer et al. (1999) due to the relatively short

amino acid sequences (~120 residues) to construct the trees. Nevertheless, several studies by Young (1992) and Ueda and other authors (1995) has suggested that the phylogenetic trees constructed from the short NifH sequences were largely consistent with those built from the comparisons of large NifH sequences.

The constructed NifH phylogeny demonstrated that majority of the novel sequences clustered within the Proteobacterial diazotrophic groupings of Cluster I and III. The predominance of *nif*H sequences from Proteobacteria in soil samples was well documented in various studies on diverse soil samples (Ueda et al., 1995; Bürgmann et al., 2003; Toyota & Kuninaga, 2006; Chowdhurya, Schmidb, Hartmannb & Tripathia, 2009). The clustering of the *nif*H clones in the dendrogram depicted that 16 clones (84.2%) were clusted within Cluster III; a group containing phototrophic, anaerobic bacteria, and sulfate reducers such as chlorobi, δ-proteobacteria, and spirochaetes (Langlois, LaRoche & Raab, 2005). Only 15.8% of the recovered *nif*H sequences formed a monophyletic grouping within Cluster I.

The cluster representation yielded interesting observation, whereby the majority of novel sequence in this study cluster within δ-protobacteria groupings of Cluster III. Based on previous studies, the abundance of such group was unexpected. Widmer and other researchers (1999) showed that *nif*H sequences derived from forest soil commonly clustered with the α- and β-proteobacterial cluster and environmental *nif*H Cluster I. Additionally, Izquierdo and Nüsslein (2006)

reported that most of the sequences in GenBank database are from α-, β-, γ-Proteobacteria, while little information on *nif*H sequences from anaerobic organisms in soils (cluster III: Clostridia, δ-Proteobacteria, Spirochaeta). Discrepancies between our study and previous studies in cluster representation suggested that the abundance of diazotroph grouping depends greatly on the ecotype soil microenvironment explored. The detailed phylogenetic analysis and description are discussed based on individual groups.

5.5.1 Group A Diazotrophs

Diazotrophs in Group A showed some relatedness to NifH sequences of Cluster I. NifH sequences within this group were observed to be from α-, β-, γ- and δ-proteobacteria (excluding phototrophic, anaerobic bacteria, and sulfate reducers such as chlorobi and spirochaetes), which together with cyanobacteria and Gram positive (high G+C). The *nif*H gene product of pTKY1612-3 and pTKY1412-2, were labeled as Group A1, forming tight cluster with delta-proteobacteria (*Geobacter* sp. M21 and *Geobacter bemidjiensis Bem*). Group A2, on the other hand, consisted of the *nif*H gene products of pTKY1612-14 and they clustered with *Pelobacter propionicus* and *Anaeromyxobacter* sp.

5.5.2 Group B diazotrophs

Group B constitutes large number of clones which were related to NifH sequences of Cluster III. The Group B phylotype is further subdivided into B1-B3. Group B1 comprised of three NifH homologous sequences which were pTKY1612-15,

pTKY0212 and pTKY2312. The sequences formed a tight cluster with thermophilic sulfate-reducing bacterium**,** *Thermodesulfovibrio yellowstonii*. These bacteria are usually isolated from thermal vent water and grown optimally at high temperature (Henry, et al., 1994). Group B2 consisted of twelve NifH homologous sequences. The sequences may represent two possible novel diazotrophic groupings, as represented by the distinct monophyletic clades. Eight sequences formed a sub-cluster with delta-proteobacteria (*Syntrophobacter fumaroxidans* and *Spirochaeta thermophila*) while the remaining four sequences formed a second monophyletic sub-cluster with *Desulfatibacillum alkenivorans*. The *nif*H gene products of pTKY1612-5 was labeled as Group B3, and clustered within delta-proteobacteria which include *Desulfovibrio aespoeensis*, *Desulfovibrio vulgaris* and *Desulfovibrio magneticus*.

Based on the placement on the phylogenetic tree, pTKY1612-2 was closely related to *Syntrophobacter fumaroxidans*. The phylogeny placement is consistent with BLASTX results, revealing that NifH sequence of pTKY1612-2 showing highest homology with δ-proteobacteria, *Syntrophobacter fumaroxidans* MPOB. Nevertheless, most of the clones in Group B2 were observed to be oddly placed, contradicting BLASTX analysis result. The clones in Group B2 are phylogenetically associated with δ-proteobacteria, *Syntrophobacter fumaroxidans* and *Desulfatibacillum alkenivorans* AK-01. In contrast to the BLASTX analysis, the *nif*H sequences demonstrated relatedness to that of *Bradyrhizobium japonicum*. Odd placements and unanticipated clustering of diazotrophs on the tree may be

due to many factors such as gene duplication, gene fusion and horizontal gene transfer. This phenomenan are extremely difficult to verify especially with limited information on *nif*H sequences from anaerobic organisms in soil which grouping under Cluster III (Izquierdo & Nüsslein, 2006). Smith and other researcher (1992) described a phylogenetic congruency test based on the assumption that a *nif*H tree corresponds to conventional *nif*H phylogenies (Young, 1992; Zehr et al., 1995); if there was any odd placement, a horizontal-gene-transfer event may have occurred (Choo, Samian, & Najimudin, 2003). However, *nif*H phylogeny constructed in this study was not able to reveal any unanticipated features that could draw any assumption that the gene transfer phenomenon existed. The possible explanation of the odd placement of clones in dendrogram could be also due to the differences in comparative algorithm used by BLASTX homology search and the construction of phylogenetic tree. BLASTX analysis compares the clone sequences to its global database while the construction of phylogenetic tree take accounts on selected diazotroph NifH polypeptide sequences as basis of comparative analysis. The BLASTX alignments enabled the nearest known bacteria to be revealed. However, the results frequently displayed highest hit with unculturable instead. Thus, BLASTX alignment result was not conclusive and the identity of bacteria could only be speculated.

5.6 Current Limitation and Future studies

As a pioneer study, this research has provided a novel insight with regards to the phylogenetic affiliation and abundance of the diazotrophic community within ex-tin mining soil. Due to time constrain, only 19 *nif*H sequences were determined and analysed in this study. The small gene pool of 19 sequences probably represents a tiny fraction of the community, renders the diazotrophic diversity of ex-tin mining land underrepresented. Future studies should be done at large scale, by soil sampling at various intervals, over a greater spatial distance to obtain a more comprehensive picture of diazotrophic communities residing in the ex-tin mining land.

Currently, there are encouraging results from present methodology complemented with molecular methods such as DGGE (which separates PCR products of same size but different sequences by chemical denaturation) and T-RFLP (which separates PCR products by recognizing only the terminal fragment of restriction digestion) provide a means for rapid and reliable determination and comparison of diazotrophic community composition. The use of nitrogenase gene sequences to describe diazotroph diversity is somewhat limited by the use of *nif*H gene in this study, as it is very highly conserved and the use of one gene for phylogenetic analysis can be misleading (Zehr, Jenkins, Short, & Steward, 2003). Therefore, future studies by using other marker genes, such as 16S rDNA, *nif*DK genes could be used concurrently to provide additional supporting or contrasting information for evaluating *nif*H phylotypes on ex-tin mining land soil.

Metagenomics has changed the way microbiologists approach many problems, redefined the concept of a genome and accelerated the rate of gene discovery (Helen & Wolfgang, 2005). A future study on diazotrophic communities by polyphasic approach is strongly suggested. Both culture-dependent and non-culture dependent techniques are used to complement each other in order to obtain the accurate description of nitrogen-fixing communities barren on ex-tin mining land. If the culturing of the "unculturables" is successful, inoculants of diazotrophs could be mass-produced to use for restoration of polluted ex-tin mining land. Furthermore, this study can be extended to isolation and investigation of diazotroph adaptive mechanisms on harsh environment, i.e. ex-tin mining land to unlock valuable information of nitrogen-fixing bacteria functionality in nitrogen limiting soil.

CHAPTER 6

CONCLUSION

A total of nineteen *nif*H clone sequences were obtained from the soil samples within the rhizosphere of *Colocasia esculenta* within the periphery of ex-tin mining land vicinity to Universiti Tunku Abdul Rahman, Kampar campus. Phylogenetic tree constructed revealed that most of the *nif*H clones obtained in this study were clustered within the Cluster I and III of Proteobacteria group, mainly in subdivision δ- proteobacteria. These result indicated there was a predominance of *nif*H sequences from proteobacteria in soil sample rather than phylogenetically diverse *nif*H sequences from a broad range diazotrophs. Several clones obtained in this study harboured *nif*H gene sequences that affiliate to proteobacteria which have valuable properties in bioremediation, such as metal reducers, sulphate reducer, etc. However, the potential microbial candidates might not be ascertained till further studies conducted.

The discovery of nineteen different possible species of diazotroph in this study has added another entry to the expanding database of *nif*H sequences to provide further references and serve to complete the missing information on genome context, in overcoming the challenges to ascertain the accurate diversity of diazotrophic communities on ex-tin mining land.

REFERENCES

Akkermans, A. D. L., van Elsas, J. D., & de Bruijn, F. J. (1995). *Molecular Microbial Ecology Manual*. Dordrecht: Kluwer.

Alan, R. S., & Tanner, K. (1997). Rapid (Ligase-Free) Subcloning of PCR Products. *Methods in Molecular Biology*, *67*, 69-78.

Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995). Phylogenetic Identification and *In Situ* Detection of Individual Microbial Cells without Cultivation. *Microbiological Reviews*, *59*, 143-169.

Applied Biosystems. (2010). *DNA sequencing: Frequently Asked Questions and Troubleshooting*, Retrieved January 31, 2011, from http://www.3.appliedbiosystems.com/cms/groups/mcb_support/documents/generalddocuments/cms_041258.pdf.

Awang, K. (1994). Growth of three multipurpose tree species on tin tailings in Malaysia. *Journal of Tropical Forest Science*, *7*(1), 106-112.

Bagwell, C. E., La Rocque, J. R., Smith, G. W., Polson, S. W., Friez, M. J., Longshore, J. W., et al. (2002). Molecular Diversity of Diazotrophs in Oligotrophic Tropical Seagrass Bed Communities. *The Federation of European Materials Societies Microbiology Ecology*, *39*, 113–119.

Baker, G.C., Ah Tow, L., & Cowan, D. A. (2003). PCR-based Detection of Non-indigenous Microorganisms in "Pristine" Environments. *Journal of Microbiological Methods*, *53*, 157-164.

Bardgett, R. D., Mawdsley, J. L., Edwards, S., Hobbs, P. J., Rodwell, J. S., & Davis, W. J. (1999). Plant Species and Nitrogen Effects on Soil Biological Properties of Temperate Upland Grasslands, *Functional Ecology*, *13*, 650-660.

Barea, J. M. (1991). Vesicular-arbuscular Mycorrhizae as Modifiers of Soil Fertility. *Advances in Soil Sciences*, *15*(1), 2-40.

Becker, S., Boger, P., Oehlmann, R., & Ernst, A. (2000). PCR Bias in Ecological Analysis: A Case Study for Quantitative *Taq* Nuclease Assays in Analyses of Microbial Communities. *Applied and Environmental Microbiology*, *66*, 4945-4953.

Bethesda Research Laboratories. (1986). BRL pUC host: *E. coli* DH5α competent cells. *Focus*, *8*(2), 9.

Bishop, P. E., & Premakumar, R. (1992). Alternative Nitrogen Fixation Systems. In *Biological Nitrogen Fixation*. (p. 736-762). New York: Chapman & Hall.

Bölter, M., Blume, H. P., Schneider, D., & Beyer, L. (1997). Soil Properties and Distributions of Invertebrates and Bacteria from King George Island (Arctowski Station), Maritime Antarctic, *Polar Biology, 18*, 295-304.

Borneman, J., & Triplett, E. W. (1997). Molecular Microbial Diversity in Soil from Eastern Amazonia: Evidence for Unusual Microorganisms and Microbial Population Shifts Associated with Deforestation. *Applied and Environmental Microbiology, 63*, 2647–53.

Bos, D. H. & Posada, D. (2005). Using Models of Nucleotide Evolution to Build Phylogenetic Trees. *Developmental and Comparative Immunology*, *29*, 221-227.

Bürgmann, H., Widmer, F., Von Sigler, W., & Zeyer, J. (2003). The New Molecular Screening Tools for Analysis for Free-Living Diazotrophs in Soil. *Applied and Environmental Microbiology*, *70*(1), 240-247.

Cejudo, F. J., & Paneque, A. (1986). Short-term Nitrate (Nitrite) Inhibition of Nitrogen Fixation in *Azotobacter chroococcum*, *Journal of Bacteriology, 165*, 240-243.

Chien, Y. T., Auerbuch, V., Brabban, A. D., & Zinder, S. H. (2000). Analysis of Genes Encoding an Alternative Nitrogenase in the Archaeon *Methanosarcina barkeri* 227. *Journal of Bacteriology*, *182*(11), 3247-3253.

Chien, Y. T., & Zinder, S. H. (1994). Cloning, DNA Sequencing, and Characterization of a *nifD*-homologous Gene from the Archaeon *Methanosarcina barkeri* 227 which Resembles *nifD1* from the Eubacterium *Clostridium pasteurianum*. *Journal of Bacteriology*, *176*, 6590–6598.

Chien, Y. T., & Zinder, S. H. (1996). Cloning, Functional Organization, Transcript Studies, and Phylogenetic Analysis of the Complete Nitrogenase Structural Genes (*nif*HDK2) and Associated Genes in the Archeon *Methanosarcina barkeri* 227. *Journal of Bacteriology, 178*(1), 143-148.

Choi, K. H., & Dobbs, F. C. (1999). Comparison of Two Kinds of Biolog Microplates (GN and ECO) in their Ability to Distinguish among Aquatic Microbial Communities. *Journal of Microbiological Methods*, *36*(3), 202-213.

CIESIN, (1992). Planning Future Sources of Environmental Information; Meeting Information Needs of the U.S. Environmental Protection Agency. *Preliminary Report, Consortium for International Earth Science Information Network (CIESIN)*

Childers, S. E., Ciufo, S., & Lovley, D. R. (2002). Geobacter metallireducens accesses insoluble Fe(III) oxide by chemotaxis. *Nature*, *41*(6), 767-769.

Choo, Q. C., Samian, M. R., & Najimudin, N. (2003). Phylogeny and characterization of three *nif*H homologous genes from *Paenibacillus azotofixans*, *Applied and Environmental Microbiology*, *69*(6), 3658-3662.

Chowdhurya, S. P., Schmidb, M., Hartmannb, A., & Tripathia, A. K. (2009). Diversity of 16S-rRNA and *nif*H genes derived from rhizosphere soil and roots of an endemic drought tolerant grass, *Lasiurus sindicus*. *European Journal of Soil Biology*, *45*(1), 114-122.

Coelho, M. R. R., Marriel, I. E., Jenkins, S. N., Lanyon, C. V., Seldin, L. & O'Donnell, A. G. (2009). Molecular detection and quantification of *nif*H gene sequences in the rhizosphere of sorghum (Sorghum bicolor) sown with two levels of nitrogen fertilizer. *Applied Soil Ecology*, *42*(1), 48–53.

Courtois, S., Frostegård, A., Göransson, P., Depret, G., Jeannin, P., & Simonet, P. (2001). Quantification of Bacterial Subgroups in Soil: Comparison of DNA Extracted Directly from Soil or from Cells Previously Released by Density Gradient Centrifugation. *Environmental Microbiology*, *3*, 431-439.

Crosby, L. D., & Criddle, C. S. (2003). Understanding Bias in Microbial Community Analysis Techniques due to *rrn* Operon Copy Number Heterogeneity. *BioTechniques*, *34*(4), 2-9.

Curtiss T. P., Sloan, W. T., & Scannell, J. W. (2002). Estimating Prokaryotic Diversity and its Limits. *Proceedings of the National Academy of Sciences USA, 99*(16), 10234–10236.

de Bok, F. A. M., Plugge, C. M., & Stams, A. J. M. (2004). Interspecies Electron transfer in methanogenic propionate degrading consortia. *Water Research*, *38*, 1368-1375.

de Lipthay, J., Enzinger, C., Johnsen, K., Aamand, J., & Sørensen, S. (2004). Impact of DNA Extraction Method on Bacterial Community Composition Measured by Denaturing Gradient Gel Electrophoressis. *Soil Biology and Biochemistry*, *36*(10), 1607-1614.

Dean, D., R., & Jacobson, M. R. (1992). Biochemical Genetics of Nitrogenase. In *Biological Nitrogen Fixation*. (p.763-787). New York: Chapman and Hall.

Doolittle, W. F. (1999). Lateral genomics. *Trends in Biochemical Sciences*, *24*, M5-M8.

Doran, J. W., Sarrantonio, M., & Liebig, M. A. (1996). Soil Health and Sustainability. In *Advances in Agronomy* (p. 1-54). California: Academic Press.

Engelhard, M., Hurek, T., & Reinhold-Hurek, B. (2000). Preferential Occurrence of Diazotrophic Endophytes, *Azoarcus* spp., in Wild Rice Species and Land Races of *Oryza sativa* in Comparison with Modern Races. *Applied and Environmental Microbiology*, *2*, 131–141.

Faegri, A., Torsvik, V. L., & Goksøyr, J. (1977). Bacterial and Fungal Activities in Soil: Separation of Bacteria by a Rapid Fractionated Centrifugation Technique. *Soil Biology and Biochemistry*, *9*(2), 105-112.

Fani, R., Gallo, R., & Lio, P. (2000). Molecular Evolution of Nitrogen Fixation: The Evolution History of the *nif*D, *nif*K, *nif*E and *nif*N Genes. *Journal of Molecular Evolution*, *51*, 1-11.

Farrelly, V., Rainey, F. A., & Stackebrandt, E. (1995). Effect of Genome Size and *rrn* Gene Copy Number on PCR Amplification of 16S rRNA Genes from a Mixture of Bacterial Species. *Applied and Environmental Microbiology*, *61*, 2798-2801.

Finan, T. M., O'Brian, M. R., Layzell, D. B., Vessey, J. K., & Newton, W. (2002). *Nitrogen Fixation: Global Perspectives*. UK: CABI Publishing.

Food and Agriculture Organization of the United Nations. (1998). *World soil resources reports*. Retrived January 23, 2011, from http://www.fao.org/docrep/W8594E/W8594E00.htm

Fox, J. E. D. (1984). Rehabilitations of Mined Lands. *Forestry Abstracts*, *45*(9), 555-560.

Frostegard, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Gall, F. L., *et al.* (1999). Quantification of Bias Related to the Extraction of DNA Directly from Soils. *Applied and Environmental Microbiology, 65*, 5409-5420.

Gabor, E. M., de Vries, E. J., & Janssen, D. B. (2003). Efficient Recovery of Environmental DNA for Expression Cloning by Indirect Methods. *The Federation of European Materials Societies Microbiology Ecology*, *44*, 153-163.

Goodman, R. M., & Weisz, J. B. (2002). Plant-microbe Symbioses: An Evolutionary Survey. In *Biodiversity of Microbial Life: Foundation of Earth's Biosphere*. New York: Wiley.

Goto, M., Ando, S., Hachisuka, Y., & Yoneyama, T. (2005). Contamination of Diverse *nif*H and *nif*H-like DNA into Commercial PCR primers. *Federation of European Microbiological Society*, *246*, 33-38.

Grundmann, L. G., & Gourbiere, F. (1999). A Micro-sampling Approach to Improve the Inventory of Bacterial Diversity in Soil. *Applied Soil Ecology*, *13*, 123-126.

Halbleib, C. M., Zhang, Y. P., & Ludden, P. W. (2000). Regulation of Dinitrogenase Reductase ADP-ribosyltransferase and Dinitrogenase Reductase-activating Glycohydrolase by a Redox Dependent Conformational Change of Nitrogenase Fe Protein. *Journal of Biological Chemistry*, *275*, 3493-3500.

Hamelin, J., Fromin, N., Tarnawski, S., Teyssier-Cuvelle, S., & Aragno, M. (2002). *nif*H Gene Diversity in the Bacterial Community Associated with the Rhizosphere of *molinia coerulea*, an Oliginitrophilic Perennial Grass. *Environmental Microbiology*, *4*(8), 477-481.

Helen L. S., & Wolfgang, R. S. (2005). Metagenomics: Advances in ecology and biotechnology. *The Federation of European Materials Societies Microbiology Letter*, 247, 105-111.

Hennecke, H., Kaluza, K., Thony, B., Fuhrmann, M., Ludwig, W., & Stackebrandt, E. (1985). Concurrent Evolution of Nitrogenase Genes and 16S Ribosomal RNA in *Rhizobium* Species and Other Nitrogen-fixing Bacteria. *Archives of Microbiology*, *14*(23), 342-348.

Henry, E. A., Devereux, R., Maki, J. S., Gilmour, C. C., Woese, C. R., Mandelco, L., Schauder, R., Remsen, C. C., & Mitchell, R. (1994). Characterization of A New Thermophilic Sulfate-Reducing Bacterium *Thermodesulfovibrio Yellowstonii*, Gen. Nov. and Sp. Nov.: Its Phylogenetic Relationship to Thermodesulfobacterium Commune and Their Origins Deep Within The Bacterial Domain. *Archives of Microbiology*, *161*(1), 62-69.

Heuer, H., Krsek, M., Baker, P., Smalla, K., & Wellington, E. M. H. (1997). Analysis of Actinomycete Communities by Specific Amplification of Gene Encoding 16S rDNA and Gel-electrophoretic Separation in Denaturing Gradient. *Applied and Environmental Microbiology*, *63*, 3233–3241.

Hill, S. (1992). Physiology of Nitrogen Fixation in Free-living Heterotrophs. In *Biological Nitrogen Fixation.* (p. 87-134). New York: Chapman & Hall.

Hirsch, A. M., McKhann, H. I., Reddy, A., Liao, J. Y., Fang, Y. W. & Marshall, C. R. (1995). Assessing Horizontal Transfer of *nif*HDK Genes in Eubacteria: Nucleotide Sequence of *nifK* from *Frankia* Strain HFPCcI3. *Molecular Biology and Evolution*, *12*, 16-27.

Howard, J. B., & Rees, D. C. (1996). Structural Basis of Biological Nitrogen Fixation. *Chemistry Review*, *96* (7), 2965-2982.

Hubbell, D. H., & Kidder, G. (1992). *Biological nitrogen Fixation*. Retrieved January 10, 2011, from http://edies.ifas.ufl.edu/SS180.

Ishi, K., & Fukui, M. (2001). Optimization of Annealing Temperature to Reduce Bias Caused by a Primer Mismatch in Multitemplate PCR. *Applied and Environmental Microbiology, 67*(8), 3753-3755.

Izquierdo, J. A., & Nüsslein, K. (2006). Distribution of Extensive *nif*H Gene Diversity Across Physical Soil Environments. *Microbial Ecology*, *51*, 441-452.

Janssen, P. H., Yates, P. S., Grinton, B. E., Taylor, P. M., & Sait, M. (2002). Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Applied Environment Microbiology*, *68*, 2391–2396.

Kaiser, O., Pühler, A. & Selbitschka, W. (2001). Phylogenetic Analysis of Microbial Diversity in the Rhizoplane of Oilseed Rape (Brassica napus cv. Westar) Employing Cultivation-Dependent and Cultivation-Independent Approaches. *Microbial Ecology*, *42*(2), 136-149.

Kellenberger, E. (2001). Exploring the Unknown: The Silent Revolution of Microbiology. *European Molecular Biology Organization Reports*, *2*(1), 5-7.

Kennedy, A. C., & Smith, K. L. (1995). Soil Microbial Diversity and the Sustainability of Agricultural Soil. *Plant and Soil*, *170*(1), 75-86.

Kent, A. D., & Triplett, E. W. (2002). Microbial Communities and their Interactions in Soil and Rhizosphere Ecosystem. *Annual Reviews in Microbiology, 56*(1), 211-236.

Kirk, J. L., Beaudette, L. A., Hart, M., Moutoglis, P., Klironomos, J. N., Lee, H., et al. (2004). Methods of Studying Soil Microbial Diversity. *Journal of Microbiological Methods*, *58*(2), 169-188.

Kirshtein, J. D., Paerl, H. W., & Zehr, J. (1991). Amplification, Cloning, and Sequencing of a *nif*H Segment from Aquatic Microorganisms and Natural Communities. *Applied and Environmental Microbiology*, *57*, 2645-2650.

Klironomos, J. N., Rillig, M. C., & Allen, M. F. (1999). Designing Belowground Field Experiments with the Help of Semi-variance and Power Analyses. *Applied Soil Ecology*, *12*, 227-238.

Knabe, W. (1965). Observation on World-wide Efforts to Reclaim Industrial Waste Land. In *Ecology and the Industrial Society* (p. 263-296). Oxford: Blackwell.

Koner, J., & Laczkó, E. (1992). A New Method for Assessing Soil Microorganism Diversity and Evidence of Vitamin Deficiency in Low Diversity Communities. *Biology and Fertility of Soils, 13*(1), 58-60.

Kotz, J. C., Treichel, P., & Weaver, G. C. (2006). Bonding and Molecular Structure: Fundamental Concepts. In *Chemistry and Chemical Reactivity* (p. 421). Canada: Thompson Brooks/Cole.

Langlois, R. J., LaRoche, J., & Raab, P. A. (2005). Diazotrophic Diversity and Distribution in the Tropical and Subtropical Atlantic Oceon. *Applied and Environmental Microbiology*, *71*(12), 7910-7919.

Ladd, J. N., Forster, R. C., Nannipieri, P., & Oades, J. M. (1996). Soil Structure and Biological Activity. *Soil Biochemistry, 9*, 23-78.

Lerman, L. S., Fischer, S. G., Hurley, I., Silverstein, K., & Lumelsky, N. (1984). Sequence-determined DNA Seperations. *Annual Review of Biophysics and Bioengineering*, *13*, 399-423.

Liesack, W., Janssen, P., Rainey, F. A., Ward-Rainey, N. L., & Stackebrandt, E. (1997). Microbial Diversity in Soil: The Need for a Combined Approach Using Molecular and Cultivation Techniques. In *Modern Soil Microbiology* (p. 375-439). New York: Marcel Dekker Inc.

Limmer, C., & Drake, H. L. (1998). Effect of Carbon, Nitrogen, and Electron Acceptor Availiablity on Anerobic N_2 Fixation in Beech Forest Soil, *Soil Biology and Biochemistry, 30*, 153-158.

Liu, L., Duan, Z. H., Xu, M. K., Hu, J. C., Wang, S. L., Hu, Z. G., et al. (2010). Effect of Monospecific and Mixed Cunninghamia Lanceolata Plantations on Microbial Community and Two Functional Genes Involved in Nitrogen Cycling. *Plant and Soil*, *327*(1), 413–428.

Liu, W. T., Marsh, T. L., Cheng, H., & Forney, L. J. (1997). Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. *Applied and Environmental Microbiology*, *63*(11), 4516-4522.

Lovell, C. R., Piceno, Y. M., Quattro, J. M., & Bagwell, C. E. (2000). Molecular Analysis of Diazotroph Diversity in the Rhizosphere of the Smooth Cordgrass, Spartina Alterniflora. *Applied and Environmental Microbiology*, *66*(9), 3814-3822.

Macrae, A. (2000). The Use of 16S rDNA Methods in Soil Microbial Ecology. *Brazilian Journal of Microbiology*, *31*(1), 77-82.

Madigan, M. T., Martinko, J. M., & Parker, J. (2000). *Brock Biology of Microorganisms*. (9th ed). Upper Saddle River, NJ: Prentice Hall.

Majid, N. M., Hashim, A., & Abdol, I. (1994). Rehabilitation of Ex-tin Mining Land by Agroforestry Practice. *Journal of tropical Forest Science*, *7*(1), 113-127.

Malik, K. A., & Schlegel, H. G. (1981). Chemolithoautotrophic Growth of Bacteria Able to Grow Under N_2-fixing Conditions. *The Federation of European Materials Societies Microbiology Letters*, *11*, 63-67.

Mancinelli, R. L., & McKay, C. P. (1988). The Evolution of Nitrogen Cycling. *Origins of Life and Evolution of the Biosphere*, *18*, 311-325.

Massol-Deya, A. A., Odelson, D. A., Hickey, R. F., & Tiedje, J. M. (1995). Bacterial Community Fingerprinting of Amplified 16S and 16-23S Ribosomal RNA Gene Sequences and Restriction Endonuclease Analysis (ARDRA). In *Molecular Microbial Ecology Manual* (p. 1-8). The Netherlands: Kluwer Academic Publishers.

Mehta, M. P., Butterfield, D. A., & Baross, J. A. (2003). Phylogenetic Diversity of Nitrogenase (*nif*H) Genes in Deep-sea and Hydrothermal Vent Environment of the Juan de Fuca Ridge. *Applied and Environmental Microbiology*, *69*(2), 960-970.

Mergel, A., Schmitz, O., Mallmann, T., & Bothe, H. (2001). Relative Abundance of Denitrifying and Dinitrogen-Fixing Bacteria in Layers of A Forest Soil. *The Federation of European Materials Societies Microbiology Ecology*, 36, 33-42.

Millbank, J. W. (1974). Associations with Blue-green Algae. In *The Biology of Nitrogen Fixation* (p. 238-264). North Holland: Amsterdam.

Motamedi, M. & Pedersen, K. (1998). *Desulfovibrio aespoeensis* sp. nov., A Mesophilic Sulfate Bacterium from Deep Groundwater at Aspo Hard Rock Laboratory, Swedan. *International Journal of Systematic Bacteriology*, *48*(1), 311-315.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific Enzymatic Amplification of DNA *In Vitro*: The Polymerase Chain Reaction, *Cold Spring Harbour Symposium on Quantitative Biology*, 51, 263-273.

Mullis, K. B. (1990). The Unusual Origin of the Polymerase Chain Reaction. *Scientific American*, *262*, 56-65.

Mullis, K. B., & Faloona, F. A. (1987). Specific Synthesis of DNA In Vitro via a Polymerase-catalyzed Chain Reaction. *Methods in Enzymology*, *155*, 335-351.

Muyzer, G., & Smalla, K. (1998). Application of Denaturing Gradient Gel Electrophoresis (DGGE) and Temperature Gradient Gel Electrophoresis in Microbial Ecology. *Antonie van Leeuwenhoek. Journal* of *Microbiology*, *73*, 127-141.

Muyzer, G., deWaal, E. C., & Uitterlinden, A. G. (1993). Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-amplified Genes Coding for 16S rRNA. *Applied Environmental Microbiology, 59*, 695–700.

Myers, R. M., Fischer, S. G., Lerman, L. S., & Maniatis, T. (1985). Nearly All Single Base Substitutions in DNA Fragments Joined to a GC-clamp can be Detected by Denaturing Gradient Gel Electrophoresis. *Nucleic Acids Research*, *31*, 3131-3145.

Nelson, D. R., & Mele, P. M. (2006). The Impact of Crop Residue Amendments and Lime on Microbial Community Structure and Nitrogen-Fixing Bacteria in the Wheat Rhizosphere. *Australian Journal of Soil Research*, *44*(4), 319-329.

Nesbo, C. L., L'Haridon, S., Stetter, K. O., & Doolittle, W. F. (2001). Phylogenetic Analyses of Two 'Archaeal' Genes in *Thermotoga maritima* Reveal Multiple Transfers between Archaea and Bacteria. *Molecular Biology and Evolution*, *18*, 362-375.

Normand, P., & Bousquet, J. (1989). Phylogeny of Nitrogenase Sequences in *Frankia* and other Nitrogen-fixing Microorganisms. *Journal of Molecular Evolution*, *29*, 436-447.

Normand, P., Gouy, M., Cournoyer, B., & Simonet, P. (1992). Nucleotide Sequence of *nif*D from *frankia alni* Strain ArI3: Phylogenetic Inferences. *Molecular Biology and Evolution*, *9*, 495-506.

O'Donnell, A. G., Seasman, M., Macrae, A., Waite, I., & Davies, J. T. (2001). Plants and Fertilizers as Drivers of Change in Microbial Community Structure and Function in Soils. *Plant and Soil*, *232*(1), 135-145.

Ogram, A., Sayler, G. S., & Barkay, T. (1987). The Extraction and Purification of Microbial DNA from Sediments. *Journal of Microbiological Methods*, *7*(2), 57-66.

Ogunseitan, O. (2005). *Microbial Diversity: Form and Function in Prokaryotes*. Oxford: Wiley-Blackwell.

Ohkuma, M., Noda, S., & Kudo, T. (1999). Phylogenetic Diversity of Nitrogen Fixation Genes in the Symbiotic Microbial Community in the Gut of Diverse Termites. *Applied and Environmental Microbiology*, *65*, 4926–4934.

Olson, J. B., Litaker, R. W., & Paerl, H. W. (1999). Ubiquity of Heterotrophic Diazotrophs in Marine Microbial Mats. *Aquatic Microbial Ecology*, *19*, 29–36.

Osborn, A. M., Moore, E. R. B., & Timmis, K. M. (2000). An Evaluation of Terminal-restriction Fragment Length Polymorphism (T-RFLP) Analysis for the Study of Microbial Community Structure and Dynamics. *Environmental Microbiology*, *2*, 39-50.

Øvreås, L., & Torsvik, V. (1998). Microbial Diversity and Community Structure in Two Different Agricultural Soil Communities. *Microbial Ecology*, *36*, 303–315.

Pace, N. R. (1996). New Perspective on the Natural Microbial World: Molecular Microbial Ecology. *American Society for Microbiology News*, *62*, 463-470.

Pace, N. R. (1997). A Molecular View of Microbial Diversity and the Biosphere. *Science*, *276*(5313), 734-740.

Paul, E. A., & Clark, F. E. (1996). *Soil Microbiology and Biochemistry* (2nd ed.). San Diego, CA: Academic Press.

Piceno, Y. M., & Lovell, C. R. (2000). Stability in Natural Bacterial Communities: Nutrient Addition Effects on Rhizosphere Diazotroph Assemblage Composition. *Microbial Ecology*, *39*(1), 32-40.

Poly, F., Ranjard, L., Nazaret, S., Gourbière, F., & Monrozier, L. J. (2001). Comparison of *nif*H Gene Pools in Soils and Soil Environments with Contrasting Properties. *Applied and Environmental Microbiology*, *67*(5), 2255-2262.

Postgate, J. R. (1974). Evolution Within Nitrogen-fixing Systems. In *Evolution in the Microbial World*. (p. 263-292). Cambridge: Cambridge University Press.

Postgate, J. R. (1982). *The Fundamentals of Nitrogen Fixation*. Cambridge: Cambridge University Press.

Postgate, J. R. (1998). *Nitrogen Fixation* (3rd ed.). Baltimore: Edward Arnold Publishers.

Postgate, J. R., & Eady, R. R. (1988). The Evolution of Biological Nitrogen Fixation. In *Nitrogen Fixation: Hundred Years After* (p. 31-40). Stuttgart: Gustav Fischer.

Prakash, R. K., Schilperoort, R. A., & Nuti, M. P. (1981). Large Plasmids of Fast-growing Rhizobia: Homology Studies and Location of Structural Nitrogen Fixation (*nif*) Genes. *Journal of Bacteriology*, *145*, 1129-1136.

Prosser, J. I. (2002). Molecular and Functional Diversity in Soil Microorganisms. *Plant and Soil*, *244*, 9-17.

Qiu, X., Wu, L., Huang, H., McDonel, P. E., Palumbo, A. V., Tiedje, J. M., et al. (2001). Evaluation of PCR-generated Chimeras, Mutations, and Heteroduplexes with 16S rRNA Gene-based Cloning. *Applied and Environmental Microbiology, 62*, 880-887.

Ranjard, L., & Richaume, A. S. (2001). Quantitative and Qualitative Microscale Distribution of Bacteria in Soil. *Research in Microbiology*, *152*(8), 707-716.

Ranjard, L., Poly, F., Lata, J. C., Mougel, C., Thioulouse, J., & Nazaret, S. (2001). Characterization of Bacterial and Fungal Soil Communities by Automated Ribosomal Intergenic Spacer Analysis Fingerprints: Biological and Methodological Variability. *Applied Environmental Microbiology, 67*, 4479-4487.

Ravenschlag, K., Sahm, K., Pernthaler, J., & Aman, R. (1999). High Bacterial Diversity in Permanently Cold Marine Sediments. *Applied and Environmental Microbiology, 65*, 3982–3989.

Raymond, J., Siefert, J. L., Staples, C. R., & Blankenship, R. E. (2003). The Natural History of Nitrogen Fixation. *Molecular Biology and Evolution*, *21*(3), 1-8.

Rees, D. C., Tezcan, F. A., Haynes, C. A., Walton, M. Y., Andrade, S., Einsle, O., et al. (2005). Structural Basis of Biological Nitrogen Fixation. *Philosophical Transactions of the Royal Society*, *363*, 971-984.

Riffkin, P. A., Quigley, P. E., Kearney, G. A., Cameron, F. J., Gault, R. R., Peoples, M. B., et al. (1999). Factors Associated with Biological Nitrogen Fixation in Dairy Pastures in South-western Victoria. *Australian Journal of Agricultural Research, 50*, 261-272.

Romanowski, G., Lorenz, M. G., Sayler, G., & Wackernagel, W. (1992). Persistence of Free Plasmid DNA in Soil Monitored by Various Methods, Including a Transformation Assay. *Applied and Environmental Microbiology*, *58*, 3012-3019.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T.H, Liles, M. R., et al. (2000). Cloning the Soil Metagenome: A Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Applied and Environmental Microbiology*, *66*(6), 2541.

Roper, M. M., & Smith, N. A. (1991). Straw Decomposition and Nitrogenase Activity by Free-living Microorganism from Soil Effects of pH and Clay Content. *Soil Biology and Biochemistry*, *23*(3), 275-283.

Rösch, C., Mergel, A., & Bothe, H. (2002). Biodiversity of Denitrifying and Dinitrogen-fixing Bacteria in an Acid Forest Soil. *Applied and Environmental Microbiology*, *68*(8), 3818-3829.

Roux, K. H. (1995). Optimisation and troubleshooting in PCR. *Genome Research*, *4*(5), 185-194.

Ruvkun, G. B., & Ausubel, F. M. (1980). Interspecies Homology of Nitrogenase Genes. *Proceedings of the National Academy of Sciences of the United States of America*, *77*, 191-195.

Saitou, N., & Nei, M. (1987). The neighbour-joining method: A new method for reconstruction phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406-425.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning – A laboratory Manual* (2nd edition). New York: Cold Spring Habour Labortory Press.

Schmalenberger, A., & Tebbe, C. C. (2002). Bacterial Community Composition in the Rhizosphere of Transgenic, Herbicideresistant maize (*Zea mays*) and Comparison to its Non Transgenic Cultivar Bosphore. *The Federation of European Materials Societies Microbiology Ecology, 40*, 29–37.

Schmalenberger, A., Schwieger, F. & Tebbe, C. C. (2001). Effect of Primers Hybridizing to Different Evolutionarily Conserved Regions of the Small-subunit rRNA Gene in PCR-based Microbial Community Analyses and Genetic Profiling. *Applied and Environmental Microbiology, 67*(8), 3557-3563.

Sessitsch, A., Weilharter, A., Gerzabek, M. H., Kirchmann, H., & Kandeler, E. (2001). Microbial Population Structures in Soil Particle Size Fractions of a Long Term Fertilizer Field Experiment. *Applied and Environmental Microbiology*, *67*(9), 4215-4224.

Shaffer, B. T., Widmer, F., Porteous, L. A., & Seidler, R. J. (2000). Temporal and Spatial Distribution of the *nif*H Gene of N_2 Fixing Bacteria in Forests and Clearcuts in Western Oregon. *Microbial Ecology*, *39*, 12-21.

Shamshuddin, J., Nik, M., & Paramananthan, S. (1986). Morphology, Mineralogy and Chemistry of Ex-mining Land in Ipoh, Perak. *Pertanika*, *9*(1), 89-97.

Shanta, N. A., Mubassara, S., & Khan, Z. U. M. (2006). Occurrence and distribution of Heterotrophic Bacteria and Azospirillum in Hydrophytes. *Journal of Life Science*., *18*, 59-62.

Sheffield, V. C., Cox, D. R., Lerman, L. S., & Myers, R. M. (1989). Attachment of a 40-base Pair G+C Rich Sequences (GC-clamp) to Genomic DNA Fragments by the Polymerase Chain Reaction Results in Improved Detection of Single-base Changes. *Proceedings of the National Academy of Sciences*, *86*, 232-236.

Smith, M. W., Feng, D. F., & Doolittle, R. F. (1992). Evolutionary by Acquisition: The case for Horizontal Gene Transfer. *Trends in Biochemical Sciences, 17*, 489-493.

Soares, R. A., Roesch, L. F. W., Zanatta, G., de Oliveira Camargo, F. A., & Passaglia, L. M. P. (2006). Occurrence and Distribution of Nitrogen Fixing Bacterial Community Associated with Oat (*avena sativa*) Assessed by Molecular and Microbiological Techniques. *Applied Soil Ecology*. *33*(1), 221–234

So, C. M., & Young, L. Y. (1999). Isolation and Characterization of A Sulfate Reducing Bacterium that Anaerobically Degrades Alkanes. *Applied and Environmental Microbiology*, *65*(7), 2969-2976.

Sprent, J. I., & Sprent, P. (1990). *Nitrogen Fixing Organisms: Pure and Applied Aspects*. United Kingdom: Chapman & Hall.

Stacey, G., Sanjuana, J., Luke, S., Dockendorffa, T., & Carlsonb, R. W. (1995). Signal Exchange in the *Bradyrhizobium*-Soybean Symbiosis. *Soil Biology and Biochemistry*, *27*(4), 473-483.

Steffan, R. J., & Atlas, R. M. (1988). DNA Amplification to Enhance Detection of Genetically Engineered Bacteria in Environmental Samples. *Applied and Environmental Microbiology*, *54*, 2185-2191.

Steffan, R. J., Goksøyr, J., Bej, A. K., & Atlas, R. M. (1988). Recovery of DNA from Soils and Sediments. *Applied and Environmental Microbiology*, *54* (12), 2908-2915.

Steppe, T., Pinckney, J. L., Dyble, J., & Paerl, H. W. (2001). Diazotrophy in Modern Marine Bahamian Stromatolites. *Microbial Ecolology*, *41*, 36–44.

Steyn, C. D., & Holland, P. W. (1993) *Essential Developmental Biology: A Practical Approach*. Oxford: Oxford University Press.

Stolp, H. (2008). *Microbial Ecology: Organisms, Habitats, Activities.* New York: Cambridge University Press.

Supardiyono, E. K., & Smith, D. (1997) Technical Report–Microbial Diversity: *Ex Situ* Conservation of Indonesian Microorganisms. *World Journal of Microbiological Biotechnology*, *13*, 359-361.

Suzuki, M. T., & Giovannoni, S. J. (1996). Bias Caused by Template Annealing in the Amplification of Mixtures of 16S rRNA Genes by PCR. *Applied and Environmental Microbiology*, *62*, 625-630.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolution Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*, 1596-1599.

Tan, X. Y., Hurek, T., & Reinhold-Hurek, B. (2003). Effect of N-fertilization, Plant Genotype and Environmental Conditions on *nif*H Gene Pools in Roots of Rice. *Environmental Microbiology*, *5*(10), 1009-1015.

Thiel, T. (1993). Characterization of Genes for an Alternative Nitrogenase in the Cyanobacterium *Anabaena variabilis*. *Journal of Bacteriology*, *175*, 6276-6286.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Research, 25*(24), 4876-4882.

Tiedje, J. M., Asuming-Brempong, S., Nusslein, K., Marsh, T. L., & Flynn, S. J. (1999). Opening the Black Box of Soil Microbial Diversity. *Applied Soil Ecology, 13*, 109-122.

Torsvik, V., & Øvreås, L. (2002). Microbial Diversity and Function in Soil: From Genes to Ecosystem. *Current Opinion in Microbiology, 5*(3), 240-245.

Torsvik, V., Goksøyr, J., & Daae, F. L. (1990). High Diversity in DNA of Soil Bacteria. *Applied and Environmental Microbiology, 56*, 782-787.

Toyota, K., & Kuninaga, S. (2006). Comparison of Soil Microbial Community between Soil Amended With or Without Farmyard Manure. *Applied Soil Ecology, 33* (1), 39-48.

Trevors, J. T. (1998). Bacterial Biodiversity in Soil with an Emphasis on Chemically-contaminated Soils. *Water Air Soil Pollution, 101*(1), 45-67.

Trevors, J. T., Lee, H., & Cook, S. (1992). Direct Extraction of DNA from Soil. *Microbiology Releases, 1*, 111-115.

Triplett, E. W., Roberts, G. P., Ludden, P. W., & Handelsman, J. (1989). What's New in Nitrogen Fixation. *American Society for Microbiology News, 55*, 15-21.

Tsai, Y. L., & Olson, B. H. (1992). Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology*, *58*, 2292-2295.

Ueda, T., Suga, Y., Yahiro, N., & Matsuguchi, T. (1995). Remarkable N_2 Fixing Bacterial Diversity Detected in Rice Roots by Molecular Evolutionary Analysis of *nif*H Gene Sequences. *Journal of Bacteriology*, *177*(5), 1414-1417.

Vaishampayan, A., Sinha, R. P., Hader, D. P., Dey, T., Gupta, A. K., Bhan, U., & Rao, A. L. (2001). Cyanobacterial Biofertilizers in Rice Agriculture. *Botanical Review*, *67*, 453-516.

van Elsas, J. D., & Trevors, J. T. (1997). *Modern Soil Microbiology*. New York: Marcel Dekker.

van Elsas, J. D., Duarte, G. E., Rosado, A. S., & Smalla, K. (1998). Microbiological and Molecular Methods for Monitoring Microbial Inoculants and their Effects in the Environment. *Journal of Microbiological Methods*, *32*, 133–154.

Vogel, J., Normand, P., Thioulouse, J., Nesme, X., & Grundmann, G. L. (2003). Relationship between Spatial and Genetic Distance in *Agrobacterium* spp. in 1 cubic centimeter of Soil, *Applied and Environmental Microbiology, 69*, 1482-1487.

von Wintzingerode, F., Gobel, U. B., & Stackebrandt, E. (1997). Determination of Microbial Diversity in Environmental Samples: Pitfalls of PCR-based rRNA Analysis. *The Federation of European Materials Societies Microbiology Ecology*, *21*, 213-229.

Wang, P., & Dick, W. (2004). Microbial and Genetic Diversity in Soil Environments. *Journal of Crop Improvement*, *12*(1), 249 - 287.

Wartiainen, I., Eriksson, T., Zheng, W. W. & Rasmussen, U. (2008). Variation in the Active Diazotrophic Community in Rice Paddy—*nif*H PCR-DGGE Analysis of Rhizosphere and Bulk Soil. *Applied Soil Ecology*, *39*(1), 65-75.

Whiteley, A. S., O'Donnell, A. G., Macnaughton, S. J., & Barer, M. R. (1996). Cytochemical Colonization and Quantitation of Phenotypic and Genotypic Characteristics in Individual Bacterial Cells. *Applied and Environmental Microbiology*, *62*(6), 1873-1879.

Widmer, F., Shaffer, B. T., Porteous, L. A., & Seidler, R. J. (1999). Analysis of *nifH* Gene Pool Complexity in Soil and Litter at a Douglas Fir Forest Site in the Oregon Cascade Mountain Range. *Applied and Environmental Microbiology*, *65*, 374-380.

Yanisch-Perron, C., Vieira, J., & Messing, J. (1985). Improved M13 Phage Cloning Vectors and Host Strain: Nucleotide Sequences of The M13 mp18 and pUC19 Vectors. *Gene*, *33*(1), 103-119.

Yeager, C. M., Kornosky, J. L., Housman, D. C., Grote, E. E., Belnap, J., & Kuske, C. R. (2004). Diazotrophic Community Structure and Function in Two Successional Stages of Biological Soil Crusts from the Colorado Plateau and Chihuahuan Desert. *Applied and Environmental Microbiology*, *70*(2), 973-983.

Young, J. P. W. (1992). Phylogenetic Classification of Nitrogen-fixing Organisms. In *Biological Nitrogen Fixation* (p. 43-86). New York: Chapman & Hall.

Zak, D. R., Holmes, W. E., White, D. C., Peacock, A. D., & Tilman, D. (2002) Plant Diversity, Soil Microbial Communities and Ecosystem Function: Are There any Links? *Ecology, 84*, 2042–2050.

Zani, S., Mellon, M. T., Collier, J. L., & Zehr, J. P. (2000). Expression of *nif*H Genes in Natural Microbial Assemblages in Lake George, New York, Detected by Reverse Transcriptase PCR. *Applied and Environmental Microbiology*, *66*(7), 3119-3124.

Zehr, J. P., & Capone, D. G. (1996). Problems and Promises of Assaying the Genetic Potential for Nitrogen Fixation in the Marine Environment. *Microbial Ecolology*, *32*, 263–281.

Zehr, J. P., Jenkins, B.D., Short, S. M., & Steward, G. F. (2003). Nitrogenase Gene Diversity and Microbial Community Structure: A Cross-system Comparison. *Environmental Microbiology*, *5*(7), 539-554.

Zehr, J. P., & McReynolds, L. A. (1989). Use of Degenerate Oligonucleotides for Amplification of the *nif*H Gene from the Marine Cyanobacterium *Trichodesmium spp*. *Applied and Environmental Microbiology*, *55*(10), 2522-2526.

Zehr, J. P., Mellon, M. T., & Hiorns, W. D. (1997). Phylogeny of Cyanobacterial *nif*H genes: Evolutionary Implications and Potential Applications to Natural Assemblages. *Microbiology*, *143*, 1443-1450.

Zehr, J. P., Mellon, M. T., & Zani, S. (1998). New Nitrogen-fixing Microorganisms Detected in Oligotrophic Oceans by Amplification of Nitrogenase (*nif*H) Genes. *Applied and Environmental Microbiology*, *64*, 3444–3450.

Zehr, J. P., Mellon, M. T., Braun, S., Litaker, W., Steppe, T., & Paerl, H. W. (1995). Diversity of Heterotrophic Nitrogen Fixation Genes in a Marine Myanobacterial Mat. *Applied and Environmental Microbiology*, *61*, 2527–2532.

Zehr, J. P., Jenkins, B. D., Short, S. M., & Steward, G. F. (2003). Nitrogenase Gene Diversity and Microbial Community Structure: A Cross-system Comparison. *Environmental Microbiology*, *5*, 539-554.

Appendix A

A Location of soil sampling site

Figure 1: (a) Sampling site within rhizophere of *Colocasia esculenta*. (b) Close-up view of taro plant, *Colocasia esculenta*

Appendix B

B.1 pGEM-T® Easy vector map and sequence reference points

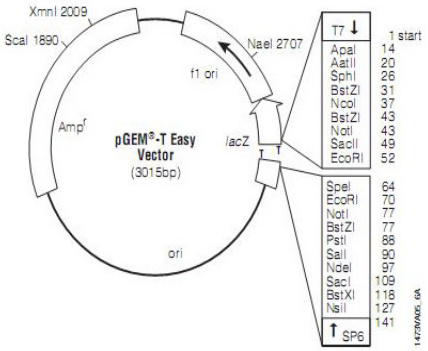

Figure 2: pGEM-T® Easy Vector Map

B.2 pGEM-T® Easy Vector promoter and multiple cloning site sequence

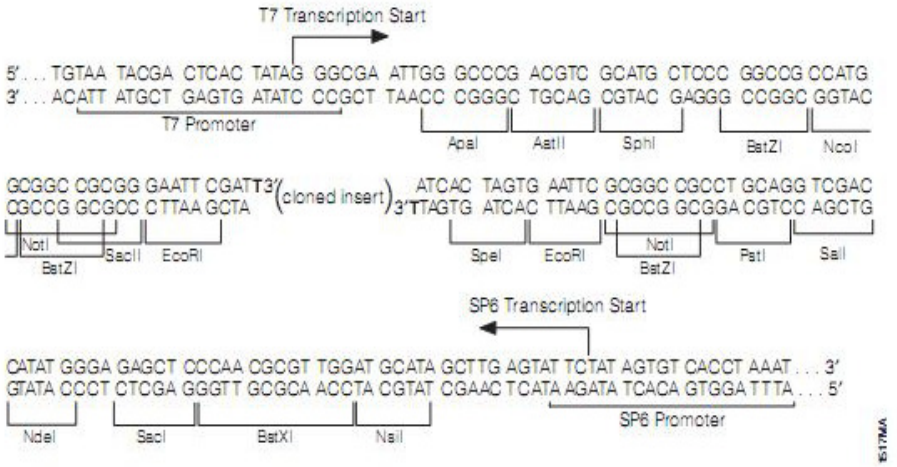

Figure 3: The promoter and multiple cloning sequence of the pGEM-T®-T Easy Vector.

Appendix C

C Compilation of the *nif*H sequences obtained in this study (22 in total). Coloured bases represent in the forward and reverse primers respectively.

>1st_BASE_540839_pTKY2312_1_SP6

GTCGCCATCATTTCGCCGGAGGTGACGATGTATATCTCCTTGGCTTTGC
CCTCGCGTATGGGCATCGCGAAGCCGCCGCATACGACATCGCCCAAG
ACGTCATAGAACACGAAGTCGAGGTCCTCTGAGTAGGCTCCGTTTTCC
TCAAGGAAGTTGATGGCCGTGATGACGCCGCGTCCTGCACAGCCCACG
CCGGGCTCAGGGCCACCGCTTTCTGCGTACTTTATATCCTTAAACCCG
GTAAGCAGAACCTCGTCGATCTCAAGGTCTTCGACAGTGCCCTTTTCG
CGCGCCATGTCCATGACCGTGTTTTGCGCCTTCTTATGCAGTATCAGCC
TCGTCGCGTCTGCCTTGGGGTCGCAG

>1st_BASE_540841_pTKY2312_2_SP6

GGGGCCATCATTTCCCCGGACACGACGATGTAGATTTCCTGGGCTTTG
CCTTCACGAATGGGCATTGCGAAGCCTCCACAGACGACGTCGCCCAAA
ACATCGTAGAAGGCGTAATCGAGCTTCTCGCTCTCCGAGTACGCCCCC
AATTGTTCCAGCAAGTTGATTGAAGTTATTATTCCTCGTCCCGCGCAAC
CCACTCCAGGTTCGGGGCCGCCTGATTCGACGCAGATGGTGCCGTTAA
AGCCAGTTTTGCGGATGTCTTCCAGATCCACGTCCTCCCCTTCCGAGCG
TAGTGTATCCAGCACTGTCTTTTGCGACAGGCCGCCCAGCAGGAGCCG
CGTGGAATCCGCCTTCGGGTCACAG

>1st_BASE_540843_pTKY2312_3_SP6

GGGGCCATCATTTCCCCGGATACGACTATGTATATCTCCTGAGCTTTGC
CGTCGCGTATCGGCATCGCGAAACCGCCGCACACGACATCTCCGAGG
ACGTCATAGAATACATAATCAAGGCCGATCTTTTCAGAGTAAGCTCCA
AGCTGCTCAAGCAGGTTGATTGAAGTGATTATGCCTCTGCCGGCGCAG
CCCACACCCGGTTCCGGCCCTCCGGATTCAACGCAACGGGTGTTCTTG
AATCCCGGTTTAACCACGTCCTCAAGGTCAAGGTCTTCGCCTTCCATG
CGAAGCGTATCGAGCACTGTTTTCTGGGACAGGCCTCCAAGCAGAAGT
CTGGTAGAGTCCGCTTTCGGGTCACAG

>1st_BASE_540845_pTKY2312_7_SP6

GGGGCCATCATTTCACCGGACACGACGATGTAGATTTCCTGGGCTTTG
CCTTCACGAATGGGCATTGCGAAGCCTCCACAGACGACGTCGCCCAAA
ACATCGTAGAAGGCGTAATCGAGCTTCTCGCTCTCCGAGTACGCCCCC
AATTGTTCCAGCAAGTTGATTGAAGTTATTATTCCTCGTCCCGCGCAAC
CCACTCCAGGTTCGGGGCCGCCTGATTCGACGCAGATGGTGCCGTTAA
AGCCAGTTTTGCGGATGTCTTCCAGATCCACGTCCTCCCCTTCCGAGCG
TAGTGTATCCAGCACTGTCTTTTGCGACAGGCCGCCCAGCAGGAGCCG
CGTGGAATCCGCTTTCGGATCACAG

>1st_BASE_540847_pTKY2312_8_SP6

GGGGCCATCATTTCACCGGACACGACGATGTAGATTTCCTGGGCTTTG
CCTTCACGAATGGGCATTGCGAAGCCTCCACAGACGACGTCGCCCAAA
ACATCGTAGAAGGCGTAATCGAGCTTCTCGCTCTCCGAGTACGCCCCC
AATTGTTCCAGCAAGTTGATTGAAGTTATTATTCCTCGTCCCGCGCAAC
CCACTCCAGGTTCGGGGCCGCCTGATTCGACGCAGATGGTGCCGTTAT
AGCCAGTTTTGCGGATGTCTTCCAGATCCACGTCCTCCCCTTCCGAGCG
TAGTGTATCCAGCACTGTCTTTTGCGACAGGCCGCGCAGCAGGAGCCG
CGTGGAATCCGCTTTCGGATCACAG

>1st_BASE_540849_pTKY2312_9_SP6

GGGGCCATCATTTCACCCGATACCACGATGTAGATCTCCTGGGCCTTG
CCTTCCCGGATGGGCATGGCAAATCCACCGCAGACAACGTCGCCCAG
GACGTCGTAAAACACGTAGTCCAGGCTCTCGCTCTCCGCATAGGCCCC
GAGCTGTTCCAACAGGTTGATCGAGGTGATGATGCCGCGGCCGGCACA
CCCCACCCCCGGCTCCGGGCCTCCCGACTCAACACAGATCGATCCCTT
GAAGCCTCGCTTCCGGATGTCCTCCAGCTCCACGTCCTCCCCTTCCGAT
CTCAGGGTGTCGAGTACCGTGTTCTGAGAAAGCCCGCCGAGAAGAAG
CCTGGTCGAATCCGCCTTGGGATCACAG

>1st_BASE_540851_pTKY2312_10_SP6

GAGGCCATCATTTCACCGGACACAACGATGTAGATCTCTTTTGCTTTGC
CGTCGCGTATAGGCATGGCGAATCCTCCGCACACAACGTCCCCCAAAA
CATCATAGAAAACGTAGTTCAGCCCTATCTTCTCGGAGTAAGCGCCAA
GCTGCTCGAGCAGATTGATAGACGTTATTATGCCTCTTCCGGCACATC
CTACTCCCGGCTCCGGCCCGCCGGATTCAACGCACCTTGTACCCTTGA
AACCGATTTTCACTACATCCTCAAGATCAAGGTCCTCGCCCTCGGCGC
GAAGAGTATCAAGTACGGTCTTTTGAGAGAGCCCTCCCAGGAGCAGCC
TCGTGGAATCAGCCTTAGGATCGCAG

>1st_BASE_539065_pTKY2112_10_SP6

GGGGCCATCATCTCACCAGAAAGTTCACCAACATGCGCATTCTGCTGG
TGGAAGATGAGCCGAATGCGGCCCGGATGTTGGCAAAGGGCCTGCGT
GAGCAATCATATGCTGTCGATATTGCCGCGGACGGCGAAGCCGCCCTT
TACCAATTAAACGTCTACGATTATGACCTGACGATCCTTGGCGTCATG
CACCCTCACAAGGACGGGCTTCAGGTCTGCCGTGAACTGCGCGGCCTG
GGCCTGTCCACTCCGGTGCTCATGCTCACCGCAAGGGACGCGGTTGTC
GATCGCATTGCCGGGCTCGACGCCGGGGCTGACGACTACCTGATCAAG
CCATTCGATTTCGGCGAAATGATGGCCTC

>1st_BASE_539063_pTKY1612_17_SP6

GGGGCCATCATTTCTCCGGACACCACGATGTAGATTTCCTGGGCTTTG
CCTTCCCGGATCGGCATGGCGAACCCACCGCAGACCACGTCGCCCAGG
ACATCATAGAAAGCGTAGTCGAGCTTTTCGCTTTCCGCGTAGGCGCCC
AGTTGTTCCAGGAGATTGATGGAGGTGATGATCCCACGCCCGGCGCAG
CCCACGCCCGGTTCCGGTCCGCCCGATTCCTCGCAGATGGTGCCTTTG
AAACCCACCTTGCGGATGTCCTCCAGCTCCACGTCCTCCCCCTCGGCA
CGGAGAGTGTCGAGTACCGTGCTCTGGGACAGGCCGCCCAGTAGCAG
CCGGGTGGAATCCGCTTTGGGGTCGCAG

>1st_BASE_539061_pTKY1612_16_SP6

GTTGCCATCATTTCGCCGGAGACCACGATGTAGATTTCCTGGGCCTTG
CCTTCACGTATGGGCATTGCGAAGCCGCCGCAGACAACGTCGCCGAG
GACGTCGTAGAATGCATAATCGAGCTTTTCGCTCTCTGCATAGGCTCC
GAGTTGCTCCAGCAGGTTGATGGAAGTGATGATGCCGCGACCTGCGCA
GCCCACTCCCGGTTCCGGCCCGCCCGATTCCACGCAGACAACGCCTTT
GAAGCCGGTCTTGCGGATATCTTCCAGATCAACGTCTCCGCCTTCTTCC
CGGAGGGTGTCGAGCACCGTTTGCTGGGAGAGCCCGCCGAGCAGGAG
CCGGGTGGAGTCCGCTTTTGGGTCGCAG

>1st_BASE_539059_pTKY1612_15_SP6

GAAGCCATCATTTCGCCGGAGGTCACGATATAGATCTCCTTGGCCTTT
CCTTCCCGGATCGGCATGGCGAACCCTCCGCAGACCACGTCGCCGAGC
ACGTCGTAGAACACGAAATCGGTATCGTCGCCGTAGGCGCCGTTTTCT
TCCAGAAAATTAATGGCTGTGATCACGCCGCGGCCGGCGCATCCGACC
CCGGGCTCCGGTCCGCCCGACTCGGCGCATTTGATGTTCTTGAATCCG
TGCAGCAGCACTTCATCGAGTTCCAGGTCCTCCACCGTGCCCTTTTCAC
GCGCCATGTCCATGACCGTGTTCTGCGCTTTTCTGTTCAGAATCAGCCG
CGTCGAATCCGCTTTTGGATCACAG

>1st_BASE_539057_pTKY1612_14_SP6

GAGGCCATCATTTCGCCGGAACAGACGATATAGATCTCTTCCGCCTTC
CCTTCGCGGATCGGCATGGCGAACCCGCCGCAGACAACGTCGCCGAG
AACGTCATAGAAGACGAAGTCAAGGTCGGGGGTGTAGGCGCCGTTCT
CTTCCAGGAAGTTGATGGCGGTGATGACGCCGCGGCCGGCGCAGCCG
ACGCCCGGCTCGGGACCACCGGATTCAACGCACTTCACATCCCCGTAA
CCCACCTTGAGGACGTCTTCCAGTTCCAGATCCTCAACCGTACCCCGC
TCCCGGACCAGATCCATGACGGTCTCCTGGGCCTTGGTGTGGAGCATC
AGGCGTGTCGAGTCGGCTTTCGGGTCGCAG

>1st_BASE_539055_pTKY1612_12_SP6

GAGGCCATCATCTCCCCCCACGCGACCATCAACATCATCCGGAACTTC
AAGGTCTTCGAGAAAAAAGGCGTCGAGACCCCCACCCTGATCGAGGG
AGTTGTCCGCTGCCCGAACCCCGGCTGCATCTCCAATACCCACGAACC
CATCCGGAGCAAGTTCGCGGTTAAGCCCAACGGGCTCCACTGTATCTA
CTGCGACTGGGTTATCTCGGGAGATATTCCGAGCCACATCATCTGATC
CTTCCCTGCCGGATCCGCCCGCAGTTCTCTGACACTCAGGCTTGGACC
CCCGGGGAAATGATGGCCCC

>1st_BASE_539053_pTKY1612_9_SP6

GCCATCATTTCCCCGGAGACGACTATGTAGATCTCCTGGGCCTTGCCTT
CTCTGATAGGCATAGCGAACCCCCCGCAAACAACGTCCCCGAGAACG
TCGTAGAACGCATAATCGAGCTCTTCTTTTTCCGCGTAAGCGCCTAGCT
GTTCCAGCAGGTTGATAGAGGTGGTTATTCCCCGGCCCGCGCATCCTA
CGCCTGGCTCGGGTCCTCCTGATTCAACGCAGACAGTGCCGCCGAACC
CTTTTTTTCTTATATCTTCCAGCTCAACGTCCTCGCCTTCGGACCGCAG
AGTGTCCAGCACGGTATTTTGCGAAAGCCCTCCTAGAAGAAGCCTTGT
GGAGTCCGCCTTAGGGTCGCAG

>1st_BASE_539051_pTKY1612_5_SP6

GAGGCCATCATTTCGCCCGAGCAGACGATGTAGATTTCCTGGGCCTTG
CCTTGGCGAATCGGCATGGCGAAATCACCACACACCACATCGCCAAG
AACATCGTAGAAGGCATAGTCGAGGCCCTCTGACTCCTCGTAGGCTCC
CAGTTGCTCCAGGAGGTTGATGCTGGTGGTGATACCGCGGCCCGCACA
GCCCACCCCGGGTTCTGGACCGCCCGATTCCACGCAGTAGCAATTGCC
ATAACCGTCGCGACGAATGTCCTCGAGTTCGACGTCCTCGCCCTCTTC
GCGTATCGTGTCGAGGACTGACTTCTGCGCAAGGCCATGCAACAGCAA
ACGGGTAGAGTCAGCCTTCGGATCACAG

>1st_BASE_537427_pTKY_1612_2_SP6

GACGCCATCATTTCGCCCGATACCACGATGTAGATCTCCTGGGCCTTG
CCTTCCCGGATGGGCATGGCAAATCCACCGCAGACAACGTCGCCCAG
GACGTCGTAAAACACGTAGTCCAGGCTCTCGCTCTGTGCATAGGCCCC
GAGCTGTTCCAACAGGTTGATCGAGGTGATGATGCCGCGGCCCGCACA
TCCCACCCCCGGCTCCGGGCCTCCCGACTCCACGCAGATCGATCCCTT
GAAGCCTCGCTTCCGGATGTCCTCCAGCTCCACGTCCTCCCCTTCCGAT
CTCAGGGTGTCGAGTACCGTGTTCTGAGAAAGCCCGCCGAGAAGAAG
CCTGGTCGAATCAGCCTTAGGGTCGCAG

>1st_BASE_537426_pTKY_1412_13_SP6

GTGGCCATCATTTCCCCGGATACGACTATGTATATCTCCTGAGCTTTGC
CGTCGCGTATCGGCATCGCGAAACCGCCGCACACGACATCTCCGAGA
ACGTCATAGAATACATAATCAAGGCCGATCTTTTCAGAGTAAGCTCCA
AGCTGCTCAAGCAGGTTGATTGAAGCGATTATGCCTCTGCCGGCGCAG
CCCACACCCGGTTCCGGCCCTCCGGATTCAACGCAACGGGTGTTCTTG
AATCCCGGTTTAACCACGTCCTCAAGGTCAAGGTCTTCGCCTTCCATG
CGAAGCGTATCGAGCACTGTTTTCTGGGACAGGCCTCCAAGCAGAAGT
CTGGTAGAGTCCGCTTTAGGGTCGCAG

>1st_BASE_537425_pTKY_1412_8_SP6

GGGGCCATCATTTCCCCGGATACGACTATGTATATCTCCTGAGCTTTGC
CGTCGCGTATCGGCATCGCGAAACCGCCGCACACGACATCTCCGAGA
ACGTCATAGAATACATAATCAAGGCCGATCTTTTCAGAGTAAGCTCCA
AGCTGCTCAAGCAGGTTGATTGAAGTGATTATGCCTCTGTCGGCGCAG
CCCACACCCGGTTCCGGCCCTCCGGATTCAACGCAACGGGTGTTCTTG
AATCCCGGTTTAACCACGTCCTCAAGGTCAAGGTCTTCGCCTTCCATG
CGAAGCGTATCGAGCACTGTTTTCTGGGACAGGCCTCCAAGCAGAAGT
CTGGTAGAGTCCGCCTTAGGATCACAG

>1st_BASE_537424_pTKY_1412_2_SP6

GAGGCCATCATCTCGCCGGAGCAGACGATGTAGATCTCCTCAGCCTTG
TTCTCCCGGATCGGCATGGCAAACCCGCCGCAGACCACGTCGCCCAAG
ACGTCGTAGAAGACGAAGTCCAGGTCCGGGGTGTAGGCGCCGTTCTCT
TCCAGGAAGTTGATGGCGGTGATGACCCCCCGGCCGGCACAGCCGAC
GCCAGGCTCCGGGCCGCCCGACTCGACGCATTTCACGTCGCCGTAACC
GACTTTGAGGACGTCATCCAGTTCCAGGTCCTCGACCGTGCCCAGTTC
CCGCACCAGGTCCATGACCGTGGACTGGGCCTTGGCGTGGAGGATCA
GACGGGTGGAGTCGGCCTTGGGATCGCAG

>1st_BASE_529273_pTKY0212_SP6

GGGGCCATCATTTCGCCCGAGGTCACGATGTAGATCTCCTTGGCCTTG
CCTTCCCGGATCGGCATGGCAAAACCGCCGCAGACCACGTCGCCGAG
CACGTCGTAGGACACAAAATCTGTGTCGTCGCCGTACGCGCCGTTCTC
TTCCAGGAAATTGATGGCCGTGATCACGCCGCGGCCCGCGCATCCCAC
GCCCGGCTCGGGACCGCCCGACTCCGCGCACTTGATGCTCTTGAAGCC
CTGCAGTAGCACTTCATCCAGTTCCAGGTCCTCCACGGTGCCCTTTTCC
CGGGCCATGTCCATGACGGTATTCTGCATCTTGCGATTAAGAATGAGC
CGCGTCGAATCCGCCTTGGGGTCACAG

Appendix D

D.1 Multiple Sequence Alignment of *Nif*H Sequence

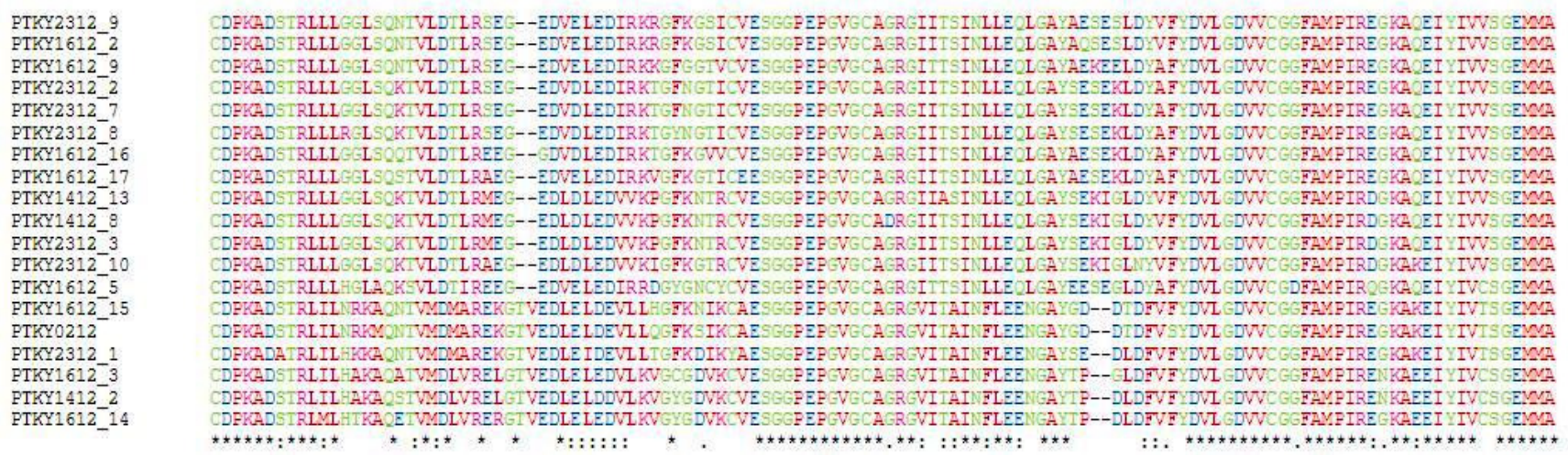

Figure 4: Multiple sequence alignment of 19 *Nif*H sequences. Amino acids are designated using single-letter abbreviations. Gaps are indicated using hyphens. Degree of conversation is specified using the asterisk, colon and period symbols – (*) indicates a single, fully conserved residue; (:) indicates conservative substitutions; (.) indicates semiconservative substitutions.