PHYLOGENY AND CHARACTERISATION OF

*nifH* GENE POOL FROM EX-TIN MINING SOIL

OF EASTLAKE, KAMPAR, PERAK

By

HAN SHING

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

April 2013
ABSTRACT

PHYLOGENY AND CHARACTERISATION OF
*nifH* GENE POOL FROM EX-TIN MINING SOIL
OF EASTLAKE, KAMPAR, PERAK

HAN SHING

Ex-tin mining land consists of nutrient-deficient soil. They are associated with high level of heavy metals and waste products of alluvial mining. This unfavorable land condition is not suitable for agriculture activities and is mostly abandoned. Nitrogen fixing microorganisms in soil has been reported to improve nutrient uptake, increase soil fertility status, rendering application in agriculture as potential inoculums for ex-tin mining rehabilitation. Thus, soil microbial community in ex-tin mining area was investigated to determine the diversity of diazotrophic communities. A soil sample was collected in ex-tin mining area within the rhizosphere of a plant in Eastlake of Kampar, Perak. Total microbial DNA was directly extracted from the soil and was subjected to nested PCR amplification using two pairs of degenerate primers. The secondary PCR amplicons were gel-purified, followed by ligation with pGEM®-T Easy Vector and subsequently transformed into competent *Escherichia coli* JM109 cells. Screening and selection of transformed cells with correct insert size were based on colony PCR. Recombinant plasmids were purified and sequenced, yielding a total of 24 partial *nifH* homologous gene sequences. BlastX analysis revealed high
similarities to *nifH* gene sequences corresponding to unculturable nitrogenase microorganisms. Molecular phylogenetic tree demonstrated that all the twenty-four NifH clones were clustered within the proteobacterial grouping of Cluster I, mainly within the subdivision of δ-Proteobacteria. This result indicated that low diversity of *nifH* gene pool was obtained in this study rather than phylogenetically diverse *nifH* sequences from a wide range diazotrophs.
I would like to take this opportunity to express my deep sense of gratitude to all those who have contributed significantly by sharing their knowledge and experience in the completion of this project work. First, I would like to extend my sincere gratitude to my research supervisor, Assistant Professor Dr Choo Quok Cheong for his patient assistance, valuable suggestions and relenting demeanour. Next, I would like to thank my research projects team members Fong Guang Yao, Gou Jau Ying, Kin Zi xian, and Kiek Qiu Wuen. Without their earnest advice and indefatigable support, this project would not be completed successfully within stipulated time. Last but not the least, I want to thank my dear family members whose blessings, inspiration and encouragement have resulted in the successful completion 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.

_____________________
HAN SHING
This project report entitled “PHYLOGENY AND CHARACTERISATION OF nifH GENE POOL FROM EX-TIN MINING SOIL OF EASTLAKE, KAMPAR, PERAK” was prepared by HAN SHING and submitted as partial fulfilment of the requirements for the degree of Bachelor of Science (Hons) in Biotechnology at Universiti Tunku Abdul Rahman.

Approved by,

____________________________
(Assist. Prof. Dr. CHOO QUOK CHEONG) Date: ……………………

Supervisor
Department of Biological Science
Faculty of Science
Universiti Tunku Abdul Rahman
PERMISSION SHEET

It is hereby certified that HAN SHING (ID No: 09ADB03718) has completed this final year project entitled “PHYLOGENY AND CHARACTERISATION OF nifH GENE POOL FROM EX-TIN MINING SOIL OF EASTLAKE, KAMPAR, PERAK” under supervision of Dr. Choo Quok Cheong (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.

Yours truly,

__________________

(HAN SHING)
TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENT iv
DECLARATION v
APPROVAL SHEET vi
PERMISSION SHEET vii
TABLE OF CONTENTS viii
LIST OF TABLES xi
LIST OF FIGURES xii
LIST OF ABBREVIATIONS xiii

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 4
2.1 The Importance of Biological Nitrogen Fixation 4
2.2 Characteristics of Nitrogenase Enzyme Complex 5
2.3 Physicochemical Properties of Ex-tin Mining Land 6
2.4 Soil and Microbial Diversity 7
   2.4.1 Microbial Diversity in Soil Environment 7
   2.4.2 The Effect of Soil Texture and Environmental Conditions on Microbial Diversity 8
   2.4.3 Limitation in Studying Soil Microbial Diversity 9
2.5 Methods to Assess Soil Microbial Diversity 10
   2.5.1 Cultivation-Dependent Method 10
   2.5.2 Cultivation-Independent Method 11
2.6 The Use of Molecular Approaches in Studying Diazotrophic Diversity 12
   2.6.1 Nested PCR Amplification of nifH Gene 12
2.6.2 \textit{nifH} Gene as Marker for Analysis of Diazotrophic Diversity

2.6.3. NifH Phylogenetic Analysis

2.6.4 Limitation of Molecular Approaches for the Study of Diazotrophic Diversity

3 MATERIALS AND METHODS

3.1 Materials

3.2 Soil Sampling

3.2.1 Study Area

3.2.2 Soil Sample Collection

3.3 Soil Physicochemical Properties Analysis

3.4 Total DNA Extraction

3.5 Nested PCR Amplification

3.5.1 \textit{nifH} Gene Amplification using Degenerated Primers

3.5.2 Nested PCR Amplification Parameter

3.5.3 Purification of PCR Products

3.6 Molecular Cloning of 360bp Target DNA and Colony PCR

3.6.1 Ligation of Purified DNA into Cloning Vector

3.6.2 Preparation of Competent Cells, \textit{Escherichia coli} (strain JM109)

3.6.3 Transformation

3.6.4 Screening for Recombinant Clones Using Colony PCR

3.7 Extraction of Recombinant Plasmid from Cell

3.8 DNA Sequencing and Phylogenetic Analysis

3.8.1 DNA Sequencing

3.8.2 Analysis of Nucleotide Sequences

3.8.3 Multiple Sequence Alignment (MSA)

3.8.4 Construction of Phylogenetic Tree

4 RESULTS

4.1 Analysis of Soil Physicochemical Properties

4.2 Total Soil DNA Extraction
4.3 Nested PCR and Gel Purification of PCR product 36
4.4 Transformation of E. coli JM109 37
4.5 Colony PCR 38
4.6 Recombinant Plasmid Extraction 39
4.7 BlastX Analysis 40
4.8 Multiple Sequence Alignment 43
4.9 Construction of Phylogenetic Tree 43

5 DISCUSSION 48
5.1 Soil Sampling 48
  5.1.1 Selection of Sampling Site 48
  5.1.2 Sampling Strategies 49
  5.1.3 Physicochemical Properties of Soil Sample 50
5.2 Total DNA Extraction 52
5.3 Nested PCR Amplification 53
5.4 Molecular Cloning of nifH Homologous Gene 54
5.5 BlastX Alignment of nifH sequences 56
5.6 Phylogenetic Tree Analysis 58
  5.6.1 Phylogeny of Group A Diazotrophs 61
  5.6.2 Phylogeny of Group B Diazotrophs 62
  5.6.3 Phylogeny of Group C Diazotrophs 63
  5.6.4 Phylogeny of Group D Diazotrophs 64
5.7 Current Limitation and Future Studies 66

6 CONCLUSION 68

REFERENCES 69
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Materials and source</td>
<td>19</td>
</tr>
<tr>
<td>3.2</td>
<td>Bacterial strains and plasmids</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>DNA sequences of each degenerated primers used in nested PCR amplification to screen <em>nifH</em> gene from soil sample</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>Parameters of nested PCR amplification</td>
<td>25</td>
</tr>
<tr>
<td>3.5</td>
<td>Material and volume used in ligation reaction mixture</td>
<td>26</td>
</tr>
<tr>
<td>3.6</td>
<td>Parameters for colony PCR amplification</td>
<td>28</td>
</tr>
<tr>
<td>3.7</td>
<td>DNA sequences of T7 and SP6 primers for colony PCR</td>
<td>29</td>
</tr>
<tr>
<td>3.8</td>
<td>Representatives of known diazotrophs retrieved from NCBI and their corresponding accession numbers</td>
<td>31</td>
</tr>
<tr>
<td>4.1</td>
<td>Physicochemical Properties of soil sample from sampling site</td>
<td>34</td>
</tr>
<tr>
<td>4.2</td>
<td>Purity and concentration of extracted total DNA from soil</td>
<td>36</td>
</tr>
<tr>
<td>4.3</td>
<td>Purity and concentration of gel purified secondary PCR product</td>
<td>37</td>
</tr>
<tr>
<td>4.4</td>
<td>List of recombinant plasmids and their corresponding amino acid identities of partial <em>nifH</em> sequences compared to that of the nearest known diazotrophs in protein database</td>
<td>41</td>
</tr>
<tr>
<td>4.5</td>
<td>List of recombinant plasmid and their corresponding translated amino acid sequences</td>
<td>44</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Representative NifH phylogeny</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>The location of soil sampling region</td>
<td>21</td>
</tr>
<tr>
<td>3.2</td>
<td>Diagram showing annealing position of each primer and expected size of the PCR products in nested PCR amplification</td>
<td>24</td>
</tr>
<tr>
<td>4.1</td>
<td>Soil texture triangle</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>Agarose gel image of total soil DNA, PCR negative control, primary PCR products, secondary PCR products and gel purified secondary PCR product</td>
<td>36</td>
</tr>
<tr>
<td>4.3</td>
<td>Plates showing the growth of <em>E. coli</em> JM109 cells transformed with ligation mixture and positive and negative control plates</td>
<td>38</td>
</tr>
<tr>
<td>4.4</td>
<td>Agarose gel image with representative of colony PCR products from white colonies</td>
<td>39</td>
</tr>
<tr>
<td>4.5</td>
<td>Agarose gel image with representative of extracted recombinant plasmids</td>
<td>40</td>
</tr>
<tr>
<td>4.6</td>
<td>NifH phylogeny obtained based on 59 NifH polypeptide sequences of known diazotrophs and 24 translated NifH polypeptide sequences from this study</td>
<td>47</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Absorbance</td>
<td></td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
<td></td>
</tr>
<tr>
<td>anf</td>
<td>Alternative nitrogen fixation gene</td>
<td></td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
<td></td>
</tr>
<tr>
<td>Blast</td>
<td>Basic Local Alignment Search Tool</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
<td></td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturing gradient gel electrophoresis</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
<td></td>
</tr>
<tr>
<td>ExPaSy</td>
<td>Expert Protein Analysis System</td>
<td></td>
</tr>
<tr>
<td>FRIM</td>
<td>Forest Research Institute Malaysia</td>
<td></td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside</td>
<td></td>
</tr>
<tr>
<td>MEGA</td>
<td>Molecular Evolutionary Genetics Analysis</td>
<td></td>
</tr>
<tr>
<td>MoFe</td>
<td>Iron molybdenum</td>
<td></td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Nitrogen</td>
<td></td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
<td></td>
</tr>
<tr>
<td>nif</td>
<td>Nitrogen fixation gene</td>
<td></td>
</tr>
<tr>
<td>NifH</td>
<td>Translated <em>nif</em>H* sequence</td>
<td></td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>T-RFLP</td>
<td>Terminal restriction fragment length polymorphism</td>
<td></td>
</tr>
<tr>
<td>vnf</td>
<td>Vanadium nitrogenase gene</td>
<td></td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-bromo-4-chloro-3-indolyl- β-D-thiogalactopyranoside</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Ex-tin mining land is defined as tracts of waste-lands made up of washed waste products of alluvial mining. These lands are also referred to as tin-tailing land which is known to have low water holding capacity, high hydraulic conductivity, low nutrient status and poor structural stability (Shamshuddin et al., 1986). In tin-tailing area, the availability of nitrogen limits plant productivity and this often affects plant community and their ecosystem processes.

Many ways can be used to solve the nutrient deficiency problem in ex-tin mining land. Biological nitrogen fixation by nitrogen fixing microorganism is essential for maintaining fertility in many ecosystems. In natural ecological restoration of mining wastelands, biological nitrogen fixation plays a vital role in introducing nitrogen nutrient into the land to provide support for vegetation growth (Zhan and Sun 2011). Biological nitrogen fixation is also one of the most important biological processes on Earth. The major types of biological nitrogen fixation include symbiotic and free-living nitrogen fixation. It is a key process to reduce atmospheric nitrogen to form ammonia, the form of nitrogen nutrient that can be utilized by living organism for the synthesis of many bioorganic compounds (Cheng 2008). This process is catalysed by nitrogenase enzyme complex which is employed by diverse groups of diazotrophs in different ecosystems (Poly et al., 2001a; Elbeltagy and Ando 2008). Diverse groups of nitrogen fixing microorganisms are able to use the renewable source of energy to fix atmospheric nitrogen under mild condition, such as normal temperature and pressure.
Nitrogen, the fifth most abundant element on Earth is essential for the synthesis of nucleic acids, amino acids and proteins. Despite the importance of nitrogen and their abundance in the atmosphere, nitrogen gas is highly inert and nonreactive, as two nitrogen elements are bound together by a triple bond, making it extremely difficult to react with other elements. Hence, fixed inorganic nitrogen such as nitrate (NO$_3^-$) and ammonium (NH$_4^+$) ions are the limiting factor for primary productivity in both marine and terrestrial ecosystems. Although ammonium ion is the major nitrogen source and intermediate in many metabolic reactions, it can lead to toxicity in plants when cultured on ammonium ion as the sole nitrogen source (Britto and Kronzucker 2002). Majority of plants absorb inorganic nitrate from soil while ammonium is used less by plants due to their extreme toxicity in high concentration (Pidwirny 2006).

The nitrogenase enzyme complex consists of two conserved proteins: MoFe protein encoded by $nifD$ and $nifK$ genes; and the Fe protein encoded by the $nifH$ gene (Rosado et al., 1998). The $nifH$ gene is one of the oldest existing functional genes in the history of gene evolution rendering their high conservation among nitrogen fixing microorganisms (Rosado et al., 1998; Singh et al., 2010). Thus, the highly conserved feature of $nifH$ gene makes it an ideal molecular marker to study and determine the potential nitrogen fixing community in various environments.

Free living prokaryotes are able to fix atmospheric nitrogen without any association with any host. They are ubiquitous in soil, phylogenetically and physiologically diverse (Burgmann et al., 2004). Diazotrophs from phototrophs, autotrophs,
chemotrophs, and heterotrophs can perform nitrogen fixation. Cultivation-based strategies had severe limitations to study diversity of soil diazotrophs due to the physiological diversity of diazotrophs and the fact that many of the prokaryotes in ecosystem are not always culturable (Burgmann et al., 2004). Therefore, culture-independent method such as molecular-based approaches have been developed and successfully applied to determine diversity of diazotrophs in different soil systems. These studies include forest soil, pasture, agricultural soil, wetland soil, and rhizospheres. Other ecological environments where culture-independent studies of nitrogen fixation was done include marine plankton (Zehr et al., 1998), microbial mats, and aggregates (Olson et al., 1999), salt marsh grasses (Bagwell et al., 1998), terrestrial soils (Rosado et al., 1998), hydrothermal vents (Mehta et al., 2003) and rhizospheres of rice (Ueda et al., 1995).

The objectives of this study include:

- To determine the physicochemical properties of ex-tin mining soil sample of Eastlake, Kampar.
- To determine the diversity of culture-independent diazotrophs within the sampling region.
- To determine novel *nifH* homologous sequences from non-culturable diazotrophs
- To elucidate the distribution and structure of phylogenetic tree of nitrogen fixation community.
CHAPTER 2

LITERATURE REVIEW

2.1 The Importance of Biological Nitrogen Fixation

Nitrogen is an essential element for life and the limiting factor in primary production within natural terrestrial and marine ecosystems (Vitousek and Howarth 1991). Primary producers consume large amounts of nitrogen as compared to other nutrients, a very costly process in term of energy obtained and used. Although 78% of the Earth’s atmosphere is nitrogen gas, plants are unable to directly utilize and assimilate the nitrogen gas from the atmosphere (Danso and Eskew 1984).

Biological nitrogen fixation is responsible for 65% of annual fixation while industrial processes only represent 25% (Singh et al., 2010). There are two types of biological nitrogen fixation and they are free-living and symbiotic nitrogen fixation. One species of symbiotic nitrogen fixers can fix up to 600Kg nitrogen per hectare per year during growing season (Wagner 2012). The estimated input of nitrogen fixed by free-living nitrogen fixers can achieve 60Kg nitrogen per hectare per year in natural environment (Zhan and Sun 2011).

Industrial processes for manufacturing of nitrogen fertilizer heavily rely on natural fossil fuel (Bohlool et al., 1992). Substantial quantities of hydrogen obtained from natural gas and of energy are maintained at high temperature and pressure condition to allow nitrogen to react with hydrogen for the synthesis of ammonia through Haber-Bosch process. Thus, excessive use of nitrogen fertilizer in agriculture could further
accelerate the depletion of natural and non-renewable energy resources used in fertilizer production (Bohlool et al., 1992). In addition, excessive use of fertilizers will cause impact on human health and eutrophication of rivers and lakes. Hence, in order to substitute for nitrogen fertilizer, biological nitrogen fixation can help to maintain nitrogen reservoir in environment to achieve high crop production and yield (Peoples and Craswell 1992).

2.2 Characteristics of Nitrogenase Enzyme Complex

Biological nitrogen fixation is catalyzed by the nitrogenase enzyme complex. This biological process can be found in a number of prokaryotic microorganisms such as cyanobacteria, symbiotic rhizobia, and free-living Azotobacter, Klebsiella, and Clostridium (Yun and Szalay 1984). Nitrogenase is a metalloenzyme composed of two protein components called dinitrogenase reductase (Fe protein) and dinitrogenase (MoFe protein) (Mylona et al., 1996). Fe protein is homodimer with molecular weight of 64000Da (Sorlie et al., 2001) and is encoded by nifH gene. The MoFe protein is $\alpha_2\beta_2$ heterotetramer with molecular weight of 240000Da (Sorlie et al., 2001) and is encoded by nifD and nifK genes. The nifH gene is highly conserved among all diazotrophic organisms, thereby making it an ideal biomarker to determine the diversity of diazotrophs in natural environment (Deslippe and Egger 2006). The MoFe and Fe protein can be damaged irreversibly by oxygen by inappropriate oxidation of metalloclusters, leading to degradation of functional three dimension protein structure (Lery et al., 2010). Alternative nitrogenases are nitrogenases which lack Mo in MoFe protein (Bishop et al., 1980). They include vanadium nitrogenases and ferric
nitrogenases. These alternative nitrogenases contain different biochemical properties and polypeptide sequences from Mo-nitrogenase (Chien and Zinder 1994).

The equation for biological nitrogen fixation can be represented as the following:

\[ \text{N}_2 + 8\text{H}^+ + 8\text{e}^- + 16\text{ATP} \rightarrow 2\text{NH}_3 + \text{H}_2 + 16\text{ADP} + 16\text{P}_i \]

It indicates that biological nitrogen fixation is very costly in term of chemical energy requirement by consuming 8 ATP to produce only one mole of NH\(_3\). The atmospheric nitrogen is composed of two nitrogen atoms joined by a triple covalent bond which makes the molecule become highly inert and nonreactive. Therefore, nitrogenase requires high energy to catalyse the breaking of the bond and addition of three hydrogen atoms to one nitrogen atom to form ammonia.

2.3 Physicochemical Properties of Ex-tin Mining Land

Tin mining activities in Malaya begun about 150 years ago, resulting in large areas of barren land called tin tailings. These tin-tailing wastelands estimated to be about 250000 hectares, representing almost two percent of total land area of Peninsular Malaysia (Majid et al., 1994). Ex-tin mining land consists of two fractions known as sand tailing and slime tailing (Majid et al., 1994). Sand tailing consists of coarse and fine sand while slime tailing consists mainly of very fine soils and mineral deposits such as silt and clay. Mine wasteland usually comprises of bare stripped area, loose soil piles, waste rock and overburden surfaces, subsided land areas and other degraded land by mining facilities (Sheoran et al., 2010). The mining activities disrupted soil components such as soil horizons and structure, soil microbe populations, and crucial nutrient cycles resulted in destruction of existing vegetation and soil profile. Reliance
on natural succession to restore sand tin tailings without any human aid can take a very long time during which the tin tailings remain economically barren (Nurtjahya et al., 2009).

Low organic matter, low nitrogen, phosphorus, and potassium nutrient contents, low or high pH, and high content of heavy metals such as cadmium, manganese, and aluminum compounds are the chemical properties of an ex-tin mining land (Ang 1994). Eighty percent of this wasteland has little or no organic matter, predominant coarse materials that cause excessive drainage, low water and nutrient retention capacities and high surface temperature (Mokhtaruddin and Norhayati 1995). The cation-exchange capacity of all tin mining lands is very poor (Nurtjahya et al., 2009). Therefore, tin mining land is extremely difficult to be reclaimed by afforestation and extra efforts are required to improve the fertility of the tin tailing before tree planting.

2.4 Soil and Microbial Diversity

2.4.1 Microbial Diversity in Soil Environment

Microbial diversity is defined as genetic diversity within species; diversity of bacterial and fungal species in microbial communities; and ecological diversity in community structure, complexity of interactions, number of trophic levels, and guilds (Nannipieri et al., 2003). In molecular-ecological term, it is defined as the number and distribution of different DNA sequence types present in community in a given habitat (Garbeva et al., 2004). Majority of plants and animals usually compete among their communities for resources and spacing. However, lack of competition in any microbial communities within soil surface as many soil microorganisms are spatially inhabited
most of the time (Zhou et al., 1997). Soil microorganisms contact among each other in microhabitats for very short time through formation of water bridges between soil particles and aggregates immediately after raining (Nannipieri et al., 2003).

Microbial diversity can be observed on soil surface and decreases when soil level is deeper as the content of organic nutrients is higher on soil surface (Nannipieri et al., 2003). About 6000 different bacterial genomes per gram of soil were measured and this indicated that the diversity of microbial population is large (Torsvik et al., 1996). Microorganisms can be found inhibiting in soil ubiquitously and other microbial species only can be discovered in particular location. If microorganisms are ubiquitous and no differences exist among different soil types throughout the Earth, they probably can be dispersed by water current, wind, and animals (Nannipieri et al., 2003). On the other hand, if they are geographically unique and have a restricted distribution, the microbial diversity in soil can be huge globally (Nannipieri et al., 2003).

2.4.2 The Effect of Soil Texture and Environmental Conditions on Microbial Diversity

The diversity and abundance of soil microbes can be affected by constraints in their environment such as availability of nutrients, pH, moisture content, soil texture as well as artificial disturbances such as agriculture, pesticide and pollution (Kohler et al., 2011). Two main determinants of soil microbial community structure are plant type and soil type (Meliani et al., 2012). Soil texture influences spatial and temporal patterns of water retention which determines number of soil microhabitats in unsaturated soil and in turn influencing microbial diversity. A finer-textured soil has
more water-filled pores while more coarse soil has more isolated water films which result in a larger number of hydrated and isolated microhabitats available for microbial colonization (Chau et al., 2011). This can in turn increase the potential for multiple microbial species to coexist in close proximity without direct competition for natural resources to survive and reproduction, thereby leading to higher microbial diversity (Carson et al., 2010; Chau et al., 2011;).

Different plant species have different biochemical composition. Vast diversity of vegetation will change the production and range of organic compounds in detritus (dead leaves and roots) which is resource availability that constraining and controlling microbial community (Zak et al., 2003). Soil is dynamic system comprised of environmental factors such as nutrients, pH, water content, temperature, pressure, spatial relationships and ionic composition which can vary over time. Therefore, microbial communities will fluctuate over time in the dynamic soil microhabitats as the microbes have to adapt variable situations in environment and their growth can be affected (Nannipieri et al., 2003).

2.4.3. Limitation in Studying Soil Microbial Diversity

Heterogeneous soil contains many microhabitats that are suitable for microbial growth which result in bacteria highly aggregated in soil existing in clumps or hot spots (Kirk et al., 2004). Replicates of 1-5g of soil sample are usually used to study microbial diversity and then make conclusion about the microbial community. This approach could cause problems such as innate heterogeneity of soil and spatial distribution of microorganisms (Trevors 1998). Microbial diversity and population size can be
underestimated resulting in high variability between replicates and low statistical power (Klironomos et al., 1999). Microbial species in such small scale of soil sample can cause biased result and detection of dominant microbial populations. To overcome the problem of spatial heterogeneity and distribution, it is suggested that smaller scale with large number of soil samples is more preferential to study microbial diversity in soil habitats (Grundmann and Gourbiere 1999).

Problem of culturing soil microorganisms is arising as large phenotypic and genetic diversity found in soil microbial species (Kirk et al., 2004). Only a limited number of microbial species can be recovered from soil by traditional culture-based techniques, resulting in underestimation of the amount and the significance of nitrogen fixer (Martensson et al., 2009). Cell lysis, nucleic acid extraction and PCR-based diversity studies that associated with molecular-based approach can cause biasness and thus, studying microbial diversity in soil can be problematic (Kirk et al., 2004).

2.5 Methods to Assess Soil Microbial Diversity

2.5.1 Cultivation-Dependent Method

Traditionally, analysis of soil microbial communities is dependent on culturing techniques using a variety of culture media designed to obtain pure culture of different microbial species live in natural soil (Hill et al., 2000). Different prokaryotic microorganisms have different growth requirements resulted from different physiological properties and this will prevent their simultaneous cultivation (Widmer et al., 1999). Culture-base methods are very time consuming and limited to study soil microbial diversity as only small fraction of microorganisms in soil are culturable.
(Janssen et al., 2002). It is estimated that less than 0.1% of microorganism found in typical agricultural soils are culturable using current culture media formulation (Torsvik et al., 1990; Hill et al., 2000). The nifH sequences of nitrogenase can be divided into 49 different subgroups, yet twenty-two of the subgroups are not cultivatable and many of the remaining subgroups contain only one or few members that have been cultivated successfully in laboratory (Buckley et al., 2007).

2.5.2. Cultivation-Independent Method

Because the physiological variety of nitrogen fixing communities and many prokaryotes are non-cultivatable, culture-independent methods are applied to solve the limitation of culture-dependent method. Culture-independent method uses molecular techniques which have been developed and successfully applied to study diazotrophic communities in many soil ecosystems (Burgmann et al., 2004). A number of molecular approaches include DNA reassociation, DNA-DNA and mRNA:DNA hybridization, DNA cloning and sequencing, and PCR-based methods (Kirk et al., 2004).

In recent years, nifH gene in environmental samples has been widely studied by using molecular methods such as clone library analysis and PCR-based methods which can provide a more complete picture of diazotrophic community in various environmental samples (Zhan and Sun 2011). Molecular based methods require DNA or RNA sample to be directly extracted from environmental sample without in vitro cultivation and isolation (Sorensen et al., 2009). However, harsh extraction methods such as bead
beating can shear the nucleic acids and thus create problems in subsequent PCR detection (Wintzingerode et al., 1997).

2.6 The Use of Molecular Approaches in Studying Diazotrophic Diversity

2.6.1 Nested PCR Amplification of nifH Gene

Previous studies on nifH gene diversity were based on PCR amplification using degenerated primers (Zehr et al., 1998; Poly et al., 2001b; Rosch et al., 2002). Prior knowledge of target nifH sequence is required to design and construct PCR primers to target and amplify the nifH gene from environmental samples. In order to achieve universal amplification successfully, the PCR primers are designed based on highly conserved amino acid sequences of the nifH gene. Although the nifH genes are highly conserved at amino acid level, nifH genes may not be homologous at DNA level due to high degree of degeneracy of genetic code (Zehr and McReynolds 1989). A nested PCR protocol is developed to achieve specific amplification of nifH gene from bulk soil DNA using highly degenerate oligonucleotide primer sets (Widmer et al., 1999). This protocol is successfully applied to amplify the nifH gene from different soil samples and from diverse groups of organisms (Shaffer et al., 2000; Burke et al., 2002).

Use of highly sensitive amplification procedures of nested PCR can eliminate problems in samples that either have few nitrogen-fixing microorganisms or have high level of inhibitory compounds so that these samples do not inhibit the PCR reaction (Zehr and Turner 2001). High purity of PCR primers is essential to avoid multiple amplification products in any PCR experiments. One-step nifH PCR reaction has
potential to amplify non-specific products of approximately the same size as \textit{nifH} gene which can be detected by cloning and sequencing the PCR amplicon (Zehr and Turner 2001). However, nested PCR has great ability to help in reducing amplification of non-specific products.

Nested PCR has been used widely to detect and identify potential nitrogen fixing microorganisms in total DNA extract from different ecosystems due to its high sensitivity. However, it was found that PCR reagents such as \textit{Taq} DNA polymerase, buffers, dNTPs, and primers supplied might contains trace amounts of \textit{nifH}-containing genomic DNA, resulting in producing unwanted amplification products (Zehr et al., 2003b; Goto et al., 2005; Bostrom et al., 2007). Restriction digestion, ultrafiltration, DNase treatment, and UV exposure of PCR reagent can reduce the contamination, but complete elimination cannot achieve and decreasing of amplification efficiency can be resulted (Zehr et al., 2003b; Goto et al., 2005). Thus, the studies of gene diversity from environmental sample should take precautions regarding the contamination to ensure the clone libraries generated from PCR amplicons are not the results of contaminated PCR reagents.

\subsection*{2.6.2 \textit{nifH} Gene as Marker for Analysis of Diazotrophic Diversity}

The complete nucleotide sequence of \textit{nifH} gene is 870bp long and encodes a polypeptide of 290 amino acids (Normand and Bousquet 1989). \textit{nifH} gene is highly conserved among diazotrophs and NifH phylogenetic tree is very similar with 16S rRNA phylogenetic tree (Mehta et al., 2003). However, nitrogen fixing microorganisms are phylogenetically diverse and it is difficult to assess the
distribution and diversity of nitrogen fixing microorganisms solely based on 16S rRNA phylogenetic diversity studies (Mehta et al., 2003). Since nitrogen fixation is exhibited by phylogenetically heterogeneous groups of prokaryotes, detection of a marker gene that is unique and is required for nitrogen fixation is possible to provide a method to determine nitrogen fixation activity in environment.

The nifH, nifD and nifK gene have been shown to be homologous in all diazotrophs and contain highly conserved sequences, reflecting a strict structural requirements of nitrogenase for proper catalytic function (Hennecke et al., 1985). However, nifH gene is selected as a marker to study diazotrophic diversity as it can provide the best phylogenetic resolution (Hennecke et al., 1985; Normand & Bousquet, 1989; Hirsch et al., 1995). nifH gene remains the most thoroughly studied with an extensive collection of sequences obtained from both cultured and uncultivated organisms from multiple environments (Zehr et al., 2003a).

The phylogenetic information of the nifH gene has been used frequently to analyze uncultivable diazotrophs in different ecosystems including wetland soils (Chelius and Lepo 1999; Piceno et al., 1999), forest soils (Widmer et al., 1999; Shaffer et al., 2000), pasture and agricultural soils (Poly et al., 2001b), rhizospheres (Ueda et al., 1995; Lovell et al., 2000), estuaries (Affourtit et al., 2001), marine (Zehr and McReynolds 1989; Zehr et al., 1996; Zehr et al., 1998), freshwater (Zani et al., 2000) and even in termite guts (Ohkuma et al., 1996; Noda et al., 1999; Ohkuma et al., 1999).
2.6.3. NifH Phylogenetic Analysis

Phylogenetic analysis is important to determine the evolutionary relationships among species of a gene or among loci of a multigene family (Bos and Posada 2005). *nifH* gene is a good choice to be applied for identifying and detecting unknown diazotrophs (Zehr and Capone 1996). The phylogeny of *nifH* gene can be used to identify phylogenetically group unknown diazotrophs based on their *nifH* gene sequences. Although *nifH* gene has been used in phylogenetic studies, discrepancies in the aspect of taxonomic placement are observed when topologies of NifH and 16S rRNA trees are compared (Young 1992). Lateral gene transfer might cause the existence of such discrepancies (Normand and Bousquet 1989).

There are essentially four major clusters in NifH phylogenetic tree and all *nifH* genes can be classified into one of these clusters (Chien and Zinder 1994; Chien and Zinder 1996). Cluster I belong to the most conventional eubacterial MoFe nitrogenases from cyanobacteria, proteobacteria (α, β, and γ), and γ-proteovacterial *vnfH*. Cluster II contains eubacterial alternative nitrogenases lacking Mo and several methanogen nitrogenases. Cluster III consists of Mo nitrogenase (*nif-1*) genes from gram-positive eubacterium *Clostridium pasteurianum* and the *nif-2* genes from the archaeon *Methanosarcina barkeri*. Lastly, Cluster IV is highly divergent and consists of predicted gene products which are likely to serve in functions other than nitrogen fixation. These organisms are solely from methanogen, such as *Methanococcus voltae*. Figure 2.1 shows a phylogeny of NifH polypeptide sequences with Cluster I-IV based on the *nifH* sequences in the database.
Figure 2.1: Representative NifH phylogeny. Tree showing phylogeny of NifH polypeptide sequences, constructed by the neighbor-joining method adapted from Choo et al., (2003). Graphic representation of the tree was made using NJPlot software. The database accession numbers are indicated after the abbreviations. Cluster I to IV assignments are described elsewhere. The data was analysed with 100 bootstrap values. The values presented above the nodes are the bootstrap values generated. Bootstrap values below 50% are not shown. The scale bar represents 0.02 substitution per site. Abbreviations: Abr, Azospirillum brasilense; Afa, Alcaligenes faecalis; Asp, Nostoc sp. strain PCC7120; Avi, Azotobacter vinelandii; Bsp, Bradyrhizobium sp. strain ANU289; Bja, Bradyrhizobium japonicum; Cpa, Clostridium pasteurianum; Csp, Cyanothece sp. strain ATCC 51142; Fal-ArI3, Frankia alni strain ArI3; Fsp-EUIK1, Frankia sp. strain EUIK1; Fsp-FaC1, Frankia sp. strain FaC1; Fsp, Fischerella sp. strain UTEX1931; Gdi, Gluconacetobacter diazotrophicus; Hse, Herbaspirillum seropedicae; Kpn, Klebsiella pneumoniae; Mba, Methanosarcina barkeri; Mma, Methanococcosmaripaludis; Mth(H), Methanothermococcus thermolithotrophicus; Mmar, Methanothermococcus marburgensis strain Marburg; Mth, Methanothermobacter marburgensis strain Marburg; Paz, Paenibacillus azotoxicans; Pbo, Plectonema boryanum; Ret, Rhizobium etli; Rle, Rhizobium leguminosarum; Rsp, Rhizobium sp. strain NGR234; Sme, Sinorhizobium meliloti; Tfe, Acidithiobacillus ferrooxidans; Tsp, Trichodesmium sp. strain IMS101.
2.6.4. Limitation of Molecular Approaches for the Study of Diazotrophic Diversity

Although molecular techniques have been generally used to prevent the problem arise from cultivation-based methods, there are some limitations in studying soil microbial diversity. The first step in molecular method is cell extraction from soil component. If the extraction approach is too gentle, Gram-negative bacterial cells would be lysed but not Gram-positive bacterial cells. If the extraction approach is too harsh, both Gram-negative and Gram-positive bacteria can be lysed but their DNA will become fragmented (Kirk et al., 2004). Fragmented DNA will contribute to the formation of chimeric PCR products (Liesack et al., 1991). Lysis efficiency in cell extraction varies between microbial groups, and between spores and mycelia influencing the DNA extraction yield and subsequent purification of DNA (Prosser 2002).

DNA extraction may also lead to biasness in studying soil microbial diversity. Harsh DNA extraction approach such as bead beating can cause DNA to be sheared or fragmented by bead beater although it can produce a significant amount of DNA (Leff et al., 1995). In addition, various biotic and abiotic components such as organic matter or humic acid in environmental sample are inhibitory substances which can be co-extracted and may interfere with subsequent PCR analysis (Wintzingerode et al., 1997; Kirk et al., 2004).

*Taq* DNA polymerase has been known to have an intrinsic error rate during DNA synthesis in PCR reaction due to lacking of 3’-5’ exonuclease (proofreading) activity
(Lawyer et al., 1993). The observed error frequencies for Taq DNA polymerase is ranged from one error per 290 nucleotides \((3 \times 10^3)\) to one error per 5411 nucleotides \((2 \times 10^4)\) (Eckert and Kunkel 1991). Single-base substitution errors rate produced by Taq DNA polymerase is 1 for each 9000 nucleotides synthesized (Tindall and Kunkel 1988; Eckert and Kunkel 1990).

PCR amplification may cause biasness caused by two factors when applied to mixtures of closely related genes in DNA sample (Widmer et al, 1999). The selective amplification of certain sequences from a mixture of different sequences causes misrepresentation of gene abundance in DNA sample (Suzuki and Giovannoni 1996). The potential for PCR chimera formation during PCR reaction will result in nonfunctional pseudogene or artifacts (Liesack et al., 1991; Ohkuma et al., 1996; Widmer, et al., 1998).
CHAPTER 3

MATERIALS AND METHODS

3.1 Materials

Table 3.1: Materials and Source

<table>
<thead>
<tr>
<th>Materials</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeneRuler 100bp DNA ladder, GeneRuler 1kb DNA ladder, 6X DNA loading</td>
<td>Fermentas</td>
</tr>
<tr>
<td>dye, Isopropyl-beta-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-</td>
<td></td>
</tr>
<tr>
<td>indolyl-beta-D-galacto-pyranoside (X-Gal), 10X PCR buffer, Magnesium</td>
<td></td>
</tr>
<tr>
<td>chloride (MgCl₂), Taq Polymerase, GeneJET™ Gel Extraction Kit, GeneJET™</td>
<td></td>
</tr>
<tr>
<td>Plasmid Miniprep kit</td>
<td></td>
</tr>
<tr>
<td>2X Rapid Ligation, T4 DNA Ligase, pGEM-T Easy Vector, dNTPs mix, T7</td>
<td>Promega</td>
</tr>
<tr>
<td>promoter primers, SP6 promoter primers</td>
<td></td>
</tr>
<tr>
<td>PowerSoil™ DNA Isolation Kit</td>
<td>Mo BIO Laboratories Inc</td>
</tr>
<tr>
<td>Agarose</td>
<td>Vivantis</td>
</tr>
<tr>
<td>LB broth, LB agar</td>
<td>Pronadisa</td>
</tr>
</tbody>
</table>
CHAPTER 4

RESULTS

4.1 Analysis of Soil Physicochemical Properties

The analysis of soil physicochemical properties was tabulated in Table 4.1. The dry pH of the soil sample was 6.53 which are slightly acidic in term of chemical properties. The percentages of total nitrogen and organic carbon were 0.19 and 1.39 respectively. This result indicated that the sampling site shown a relatively low nutrient status which could affect effective plant growth and soil microbial communities. For the soil texture determination, the soil particle composition showed high tendency toward large particle size. Sand predominates at 85% with 69% of coarse sand and 16% of find sand, followed by 13% of clay and 4% of silt. Thus, the soil sample was classified as loamy coarse sand in soil texture class as analysed using soil texture calculator available at http://www.soils.usda.gov/technical/aids/investigations/texture/. The soil texture triangle was shown in Figure 4.1.

Table 4.1: Physicochemical Properties of soil sample from sampling site

<table>
<thead>
<tr>
<th>Physicochemical Properties</th>
<th>Soil Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry pH</td>
<td>6.53</td>
</tr>
<tr>
<td>Total nitrogen (%)</td>
<td>0.19</td>
</tr>
<tr>
<td>Total Organic Carbon (%)</td>
<td>1.39</td>
</tr>
<tr>
<td>Coarse Sand (%)</td>
<td>69</td>
</tr>
<tr>
<td>Fine Sand (%)</td>
<td>16</td>
</tr>
<tr>
<td>Silt (%)</td>
<td>4</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>13</td>
</tr>
<tr>
<td>Texture</td>
<td>Loamy Coarse Sand</td>
</tr>
</tbody>
</table>
5.1 Soil Sampling

5.1.1 Selection of Sampling Site

In this study, an isolated site of ex-tin mining land was selected for soil collection. This kind of land may only occupied by a limited number of microorganisms as compared to undisturbed land because of nutrient deficiency and extreme environmental condition. Low amount of organic matter in soil can limit the number and activity of soil bacteria (Heydarnezhad et al., 2012). Nitrogen fixation and vegetation can be affected by mineral nutrient deficiency in soil (Baligar and Duncan 1990). As shown in Figure 3.1, a few plant species were growing within the vicinity of sampling site and the land consists mostly of sand and little clay. The total nitrogen and organic carbon percentage in ex-tin mining land within vicinity of University Tunku Abdul Rahman were less than 1.5% shown by Tan (2011), indicating low fertility status. Since nitrogen element is one of the macronutrients for plant growth and biological nitrogen fixation happen in such poor land, it is possible to reveal some potential diazotrophic diversity in the sampling area.

Soil sampling is normally conducted within of plant rhizosphere which has higher possibility of harbouring large microbial communities (Ramos et al., 2000; Tan 2011). This is because organic nutrients may be exuded by plant root which provide the driving force for the development of active microbial population within the rhizosphere of plant roots (Watanabe and Roger 1984; Whipps 2001). In addition,
CHAPTER 6

CONCLUSION

A total of twenty-four partial *nifH* sequences were determined from the soil sample collected within the rhizosphere of vegetation of ex-tin mining land of University Tunku Abdul Rahman. NifH phylogeny revealed that all the translated NifH clones were grouped within the Proteobacterial grouping, mainly within the subdivision of Delta proteobacteria of Cluster I. This result indicated low diversity of diazotrophs within the study area was obtained instead of a broad range diversity of diazotrophs. Several NifH clones from the soil sample were shown to be closely related with microbial species which possess valuable application in bioremediation such as metal reducers and dehalogenation. However, the complete diversity picture of diazotrophs in the sampling site remains to be undefined and thereby, further studies is required.

Although only limited partial *nifH* sequences were obtained from this study, nevertheless, the discovery of twenty-four *nifH* novel sequences has added entry to the ever expanding the database of *nifH* sequences. This, in turn, will be beneficial in providing gene sequence references for similar studies and serve to complete the missing information on genome context.
REFERENCES


Barcellos, F.G. et al., 2007. Evidence of Horizontal Transfer of Symbiotic Genes from a Bradyrhizobium japonicum Inoculant Strain to Indigenous Diazotrophs Sinorhizobium (Ensifer) fredii and Bradyrhizobium elkanii in a Brazilian Savannah Soil. Applied and Environmental Microbiology, 73(8), pp. 2635-2643.


