MICROBIAL METAGENOMIC ANALYSIS OF PADDY FIELD SOIL OF BARIO, THE KELABIT HIGHLAND OF SARAWAK

SUPANG LAU MEI LING

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

MICROBIAL METAGENOMIC ANALYSIS OF PADDY FIELD SOIL OF BARIO, THE KELABIT HIGHLAND OF SARAWAK

SUPANG LAU MEI LING

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Supang Lau Mei Ling

Matric No: SGR110123

Name of Degree: Master of Science

Title of Thesis: Microbial Metagenomic Analysis of Paddy Field Soil of Bario, The Kelabit Highland of Sarawak

Field of Study: Microbiology (Biology & Biochemistry)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

MICROBIAL METAGENOMIC ANALYSIS OF PADDY FIELD SOIL OF BARIO, THE KELABIT HIGHLAND OF SARAWAK

ABSTRACT

Rice is known as one of the staple foods of the world particularly in Asian countries as the largest producer and consumer. The increasing demand and the declining supply of rice in Malaysia had always been an issue and that can only be solved with the imported rice from neighbouring countries. In Malaysia, one of the popular rice varieties comes from Bario, Sarawak and it is widely known as Bario rice. Bario rice is known as a unique form of delicacy in Malaysia. The farmers have been practicing traditional wet rice cultivation for centuries. This study aims to have an insight of the microbial community profile and the genes that involves in nutrient cycling in the paddy field soils obtained from Bario, Kelabit Highland of Sarawak using next-generation sequencing (NGS) approach. The DNA of the paddy field soils were extracted using Power Soil DNA Isolation Kit (MOBIO Laboratories, Inc) before subjected to NGS. The raw sequences obtained were trimmed and assembled using CLC Genomic Workbench 7.0. Assembled contigs were identified for the presence of prokaryotic components by BLASTN with GenBank 16S microbial database with e-value of $< 10^{-9}$. The data obtained from BLAST were analysed for taxonomic distribution using Metagenome Analyzer (MEGAN) version 5.2.3. The assembled contigs were annotated using Prokka, a software tool for rapid annotation of prokaryotic genomes. NGS data revealed immense diversity of ecologically important microbes present in the soil sample belongs to bacteria and archaea domains. The phylum Proteobacteria occupied the highest portion (57%) of the bacteria sequences in the F3F (Field 3F) paddy field soil sample followed by Euryarchaeota (12%), Actinobacteria (6%), Acidobacteria (3%), Verrucomicrobia (2%), Firmicutes (2%), Chloroflexi (2%) and Bacteroidetes (1%). Also, the most predominant microbial genera belong to *Geobacter* (17%), *Candidatus methanoregula* (7%), *Anaeromyxobacter*

(7%), *Methanosaeta* (3%), *Opitutus* (2%), *Burkholderia* (2%), *Syntrophus* (2%), *Bradyrhizobium* (2%) and *Pelobacter* (2%). There was a total of 413,191 genes annotated from Prokka and the genes such as *ntpJ, ctpV, nikR, narT, yydH, rip3, albF, narG, narY, narI, narV, cysO, zitB* and *zur* that involves in the nutrient cycling of the paddy field soil were identified. In conclusion, a total of 26 phylas, 41 microbial classes, 87 microbial orders, 177 microbial familiae and 388 microbial genera were found from this paddy field soil samples. Each of the microbial communities in the soil plays an important role in nutrient cycling.

Keywords: Bario rice; microbial community profile; genes; paddy field soil; NGS

ANALISIS METAGENOM MIKROB TANAH SAWAH PADI DI BARIO, TANAH TINGGI ORANG KELABIT DI SARAWAK

ABSTRAK

Beras merupakan salah satu makanan ruji dunia terutamanya di negara-negara Asia sebagai pengeluar dan pengguna terbesar. Permintaan yang semakin meningkat dan bekalan beras yang menurun di Malaysia sememangnya menjadi satu isu dan hanya boleh diselesaikan dengan mengimport beras dari negara-negara jiran. Di Malaysia, salah satu jenis varieti padi yang popular berasal dari Bario, Sarawak dan ia dikenali sebagai beras Bario. Beras Bario merupakan salah satu makanan istimewa yang jarang didapati di Malaysia. Petani-petani mengamalkan penanaman padi secara tradisional sejak berkurun lamanya. Kajian ini bertujuan untuk memberi gambaran tentang profil komuniti mikrob dan gen-gen yang terlibat dalam kitaran nutrien pada tanah sawah yang diperolehi dari Bario, Tanah Tinggi Orang Kelabit di Sarawak melalui pendekatan penjujukan generasi akan datang (NGS). DNA mikrob pada tanah sawah diekstrak menggunakan kit Power Soil DNA Isolation (MOBIO Laboratories, Inc) sebelum menggunakan NGS. Jujukan mentah yang diperolehi, dikemas dan dicantum menggunakan CLC Genomic Workbench 7.0. Kehadiran komponen prokaryotik dikenal pasti melalui kontig-kontig yang diperolehi dengan BLASTN dalam GenBank Mikrobiologi 16S dengan e-nilai $<10^{9}$. Data yang diperoleh daripada BLAST dianalisis untuk taburan taksonomi dengan menggunakan Metagenome Analyzer (MEGAN) versi 5.2.3. Kontig-kontig tersebut dianotasi menggunakan Prokka, alat perisian untuk penjelasan genom prokariotik dengan cepat. Data NGS menunjukan kepelbagaian mikrob yang ada di dalam sampel tanah sangat penting secara ekologi. Majoritinya adalah domain milik bakteria dan archaea. Filum Proteobakteria menduduki bahagian paling tinggi (57%) dari urutan bakteria dalam sampel tanah sawah F3F (Sawah 3F) diikuti oleh Euryarchaeota (12%), Actinobacteria (6%), Acidobacteria (3%), Verrucomicrobia (2%), Firmicutes 2%), Chloroflexi (2%) dan

Bacteroidetes (1%). Selain itu, genus mikrob yang paling utama adalah *Geobacter* (17%), *Candidatus metanoregula* (7%), *Anaeromyxobacter* (7%), *Methanosaeta* (3%), *Opitutus* (2%), *Burkholderia* (2%), *Syntrophus* (2%), *Bradyrhizobium* (2%) dan *Pelobacter* (2%). Sejumlah 413,191 gen telah dianotasi dari Prokka dan gen seperti *ntpJ, ctpV, nikR, narT, yydH, rip3, albF, narG, narY, narI, narV, cysO, zitB* dan *zur* yang terlibat dalam kitaran nutrien tanah sawah telah dikenalpasti. Kesimpulannya, terdapat sebanyak 26 filum, 41 kelas, 87 bangsa, 177 keluarga dan 388 marga mikrob yang telah ditemui dalam tanah sawah. Setiap mikrob tersebut memainkan peranan penting dalam kitaran nutrien.

Kata kunci: beras Bario; profil komuniti mikrob; gen; tanah sawah; NGS

ACKNOWLEDGEMENTS

First and foremost, I would like to give all the glory, honour and praise to the Lord Almighty for His unconditional love, guidance and blessings in completing my master research and thesis.

Secondly, my utmost gratitude goes to my supervisor Associate Professor Dr. Chan Kok Gan, FASc, FRSB, FMSA, for his guidance and never-ending encouragement throughout my master's research and thesis.

My deepest appreciation goes to University of Malaya for providing sufficient funding, high-end equipment as well as facilities throughout the commencement of my research through High Impact Research grant (UM.C/625/1/HIR/MOHE/CHAN/01, Grant No. A-000001-50001). Not forgetting, the Ministry of Higher Education for the SLAI scholarship awarded to me.

I would also like to thank my fellow lab mates, colleagues and friends for their countless assistance and moral support throughout this research and thesis. Last but not least, both my parents and family members for their unconditional love, relentless prayers and never-ending moral supports given to me in completing this master research and thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

-	Negative
-	To
#	Number of samples
%	Percentage
/	Or
<	Less than
× *g*	Gravitational force
×	Times
®	Registered
°C	Degree Celsius
µL	Microliter
µm	Micrometer
3’	Three prime
4WD	Four by Four
I	Roman numeral for number 1
III	Roman numeral for number 3
™	Trademark

ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
bp	Base pair
cm	Centimeter
Co	Cobalt
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
et al.	Latin word of "et alii" which means "and others"
EtBr	Ethidium bromide
etc	Latin word for Et cetera
Fe	Iron
g	Gram
GB	Gigabyte
GC	Guanine-Cytosine
gDNA	Genomic Deoxyribosenucleic acid
gen. et sp. nov.	Genus et Species Nova
ha	Hectare
HCl	Hydrochloric acid
ID	Identification
Inc.	Incorporation
Kb	Kilobase
Kg	Kilogram
L.	Latin
LS	Low sample
M	Molar
Mg^{2+}	Magnesium ion

min	Minutes
mL	Mililiter
mm	Milimeter
Mn	Manganese
Mo	Molybdenum
NCBI	National Centre for Biotechnology Information
ng	Nanogram
NGS	Next-generation sequencing
Ni	Nickel
nM	Nanomolar
PCR	Polymerase Chain Reaction
pH	Potential of hydrogen
psi	Pound per square inch
qPCR	Quantitative Polymerase Chain Reaction
RAST	Rapid Annotation Server
rDNA	Ribosomal DNA
RM	Ringgit Malaysia
rRNA	Ribosomal RNA
sp.	Species
spp.	Species plural
SRA	Short Read Archive
TBE	Tris-Borate-EDTA
V	Volt
v/v	Volume/Volume
w/w	Weight/Weight
w/v	Weight/ Volume

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Rice is known as one of the staple food of the world especially in most Asian countries including Malaysia. The increasing demand and the declining supply of rice in Malaysia had always been an issue and that can only be solved with the imported rice from neighbouring countries (Rajamoorthy et al., 2015). An adult Malaysian for instance consumed approximately two and half plate of rice per day in a study on food consumption pattern of the adult populations in Malaysia (Norimah et al., 2008). Fully dependent on imported rice will also give negative impact on the country's economy long term therefore the Malaysian government should encourage local rice production and help to boost yields of local rice production to cater the everyday needs of the people (Rajamoorthy et al., 2015).

Bario rice is considered as a premium aromatic rice in Malaysia. Bario rice has gained its popularity not only among Malaysians but tourists from all around the world that visited Ba'kelalan, Bario and Miri. Due to the low productivity of Bario rice, it is considered as a rare form of delicacy in Malaysia. Approximately 600 to 720 tonnes of Bario rice is produced in the Kelabit Highlands and only a portion of it is sold in Sarawak yearly and the market price for the authentic Bario rice in Sarawak can easily cost up to RM 10 per kg to RM18 per kg (Jiwan et al., 2015) and about Rm20 per kg to RM 28 per kg in the Peninsula Malaysia. Bario rice cultivation is unique compared to other rice cultivation in Malaysia because it is planted under flooded condition at a high altitude and cool climate. Due to the geographical area of Bario highland, the use of inorganic fertilizer could increase the cost of production of Bario rice. Therefore, most of the farmers still opt for traditional rice cultivation.

Hence, the overall objectives of this thesis were to study the microbial assemblage in the paddy field soil in Bario, the Kelabit Highland of Sarawak and to identify the microbial genes that are important for nutrient recycling.

The objectives of this research project include the following:

1. To sequence the metagenomic library obtained from F3F paddy field soils in Bario Highlands using Next-Generation Sequencing (NGS),
2. To study Bario paddy field soil microbial community profile,
3. To identify functional genes that involves in nutrient cycling.

CHAPTER 2: LITERATURE REVIEW

2.1 Bario, The Kelabit Highlands of Sarawak

The name Bario was coined from the Kelabit language "Bariew" which means "The valley of the wind" and is locally known as the Kelabit Highlands among the people in Sarawak. The main indigenous tribe that inhabited Bario are the Kelabits, one of the 27 indigenous tribes in Sarawak and they are one of the smallest ethnic groups in Sarawak. The estimation of their total population was put between 6,500 to 6,800 worldwide. However, the exact number was unknown till today. Many of the Kelabits especially the younger generation have migrated to the cities and other parts of the world to seek for better education and opportunities. Some, however, remained and continue to reside there, in the same manner as their forefathers did. The name Kelabit was incidentally coined by Charles Hose, the first Resident of Baram because he misheard the term Pa Labb'd (a place name) as Kelabit (Harrisson, 1959).

Bario lies on a plateau in the Kelabit Highland, southeast of Miri in the Fourth Division of Sarawak at 3500 feet above sea level. The plateau called Plain of Bah was surrounded by mountains which forms the Apo Duat range in the east while the Tamabu range to the south, north and west (Ismail & Din, 1998; Sheldon et al., 2013). The main rivers are the Dapur (known as Lubbun on older maps) and the Kelapang (or Baram). Dapur river flows near Bario and Pa' Umur whereas the Kelapang river flows through Pa' Main. Both of the rivers flow southward and will adjoin approximately 15 km south of Ramudu and ultimately steering into the Baram through a series of cascades in the southwestern corner of the highlands above Lio Matoh (Sheldon et al., 2013). Bario has a mild and cool climate with the temperature of 18 to 26°C and an annual rainfall of about 2,213 mm (Ismail & Din, 1998). Until World War II, the journey to Bario required about a month journey comprising of a long boat ride from Marudi to Lio Matoh (the highest

manoeuvrable point on the Baram river) and finally by foot (Harrisson, 1959). After the new airstrip was launched in April 1961 by the then governor of Sarawak, Sir Anthony Abell, Bario became more accessible to the coast and vice versa (Saging,1977). After such historical event, the trip to Bario mostly involve a flight from Miri to Bario and subsequently after a period of time, a road has been opened to link Bario and the coastal of Sarawak (Harrison, 1959; Sheldon et al., 2013). Now, Bario became more accessible to the world through air transport as well as logging roads. There are three flights daily from Miri (nearest city) to Bario operated by MASwings Sdn Bhd, a regional airline operating the Rural Air Services in East Malaysia. The carrier, which is a wholly-owned subsidiary of Malaysia Airlines, is part of the transportation services division of Malaysia Aviation Group Bhd.

The soils of Bario, the Kelabit highlands are derived from accreting and non-accreting alluvium which composed mainly of poorly drained clays, podzolic sands and 'climatogenic' organic soils (Seng et al., 1998).

2.2 Traditional Wet Paddy Cultivation

Wet paddy cultivation is quite a laborious work. It revolves around six major activities which are selecting the plot, clearing a plot, planting the paddy, minding the crop, harvesting and processing and storage. The elaborate activities are as follows (Bala, 1993).

a) Selecting the Plot

In the past, this is done every 5 to 7 years. The chosen land is usually beside a river and is selected by the community through the discussion among the heads of households in respective long house.

b) Clearing the Plot

This method involves clearing of bushes and undergrowth, felling of trees, drying of the land and burning. This method will also be done after the harvest. This is usually done as a cooperative activity by the long house community. They will move from one plot to another by turns on mutual basis. Burning of lands will take place on higher ground usually around mid-afternoon after the morning dew had dried and before the sunset where there was still light.

c) Planting the Paddy

Subsequent step is nursery preparation. The nursery plot is filled with water before the seeds are scattered in them. The time of planting will be decided by the elders in the community by looking at the phase of the moon. Planting of paddy is done by both men and women in which the men will make a hole in the ground with a stick while the women will insert 5 to 10 paddy seedlings in each of the hole.

d) Guarding the Paddy

This steps usually involves weeding. This is usually done by women. The plot of land was surrounded with bamboo fence to prevent wild animals as well as the buffaloes from entering the plot of land. However, as the harvesting season approaches, there were also birds to worry about and hence an elaborate methods of pest management have been developed.

e) Harvesting

The decision for harvesting was made by men but it was the women usually mother or daughter) who begin the ritual for harvesting. It was believed that the rice spirit prefers women for their gentle manner.

f) Processing and Storage

The storage is usually performed by an older woman because she is regarded as blessed in the community.

2.3 Traditional Wet Paddy Cultivation vs Modern Wet Paddy Cultivation

The local Kelabits have been practicing sustainable traditional wet paddy cultivation for centuries and in fact, they are popularly known for their cultivation of aromatic "Bario rice" which is regarded as a rare form of delicacy in Malaysia (Jiwan et al., 2015). Rice is not only staple food of the Kelabits but it is also a symbol of status, wealth and prestige (Janowski, 1988; Hew & Sharifah, 1998).

There is no mechanization adopted in growing and harvesting process of Bario rice as some of the farmers still practice traditional way of rice cultivation. The minimal fertilizer usage in the paddy cultivation was mainly because the farmers assumed that the soil was fertile by leaving paddy rice straws in the field to fallow for 4 to 5 months after harvesting process as an organic source of nutrient for the next planting season (Samuel et al., 2015). The decomposition and disintegration process between the rice straws, buffalo waste and urine that occurred within the soil will therefore provide the necessary nutrients for the incoming growth cycle of the rice paddy (Samuel et al., 2015). One traditional characteristic of paddy cultivation in Bario is the extensive use of buffaloes. A regular family would farm the same piece of land each year therefore naturally, one could foresee that the soil will lose its nutrients over time and hence its fertility also decreases after each planting season. To circumvent this, the farmers will leave the buffaloes in the field after harvesting to feed on dried paddy straws as well as to plough the soil. The buffaloes will not only feed and plough the lands but also provide source of organic manure which are ploughed back into the soil as well (Sanggin, 1998). The paddy field will be irrigated via drainage and irrigation system from running streams from the mountain and excess water from the upper paddy fields will be channelled to the lower paddy fields before being drained into the drain and finally into the river (Samuel et al., 2015).

In 2011, the traditional wet rice cultivation has been slowly integrated with mechanized wet rice cultivation before a fully mechanized system in wet rice cultivation is introduced. Bario Rice Industry Development is a project under the National Key Economic Area (NKEA)'s Entry Point Project (EPP) 11, was instigated at a cost of RM17 million. NKEA projects were introduced under the Government Transformation Programme (GTP). The development includes levelling of paddy fields, construction of seven irrigation dams with irrigation pipes to the fields, construction of farm roads, ploughing, planting and harvesting services for 200 ha of paddy fields in Bario, and the building of a drying and milling factory in Bario. Their aim was to increase the production of Bario rice. The first harvest from this project was marketed in the first quarter of 2014 (Aubrey, 2014). With the use of these modern machinery, the farmers are able to skip these laborious rice cultivation steps. In the recent years, as the machineries were used, buffaloes were rarely found in the field.

2.4 Factors Affecting the Success of Bario Rice Cultivation

According to Sanggin et al. (1998), several factors that may contribute to the success of Bario rice cultivation in Bario are the micro-climate, the transportation problem as well as shortage in labour.

a) Micro-climate

Bario, the Kelabit Highlands, located at higher elevation compared to other parts of Sarawak, has a moderately cool temperature. Sanggin et al. (1998) also stated that Bario rice been grown on experimental basis in other parts of Sarawak with slightly higher temperatures were neither the growth nor the taste of the rice were as good as the ones locally grown in Bario itself. The plateau also receives ample amount of sunshine and rainfall throughout the year, in which both are the basic necessities for successful rice cultivation. Irrigation water derived from the mountains using the gravity-fed pipe water

were channelled to the paddy fields. These basic requirements coupled with the moderately cool temperature may contribute to the success of paddy cultivation in Bario. (Sanggin et al., 1998).

b) Transportation

The primary reason why the farmers in Bario opt for using no chemical fertilizer was because of transportation. The exorbitant cost of transporting fertilizer to Bario was too much for the farmers in Bario. As mentioned earlier, the mode of transportation to Bario were by air and land (approximately 12 to 14 hours' drive from Miri using 4WD pickup truck). This adversity has basically forced the farmers in Bario to depends on buffalo manure rather than chemical fertilizers (Sanggin et al., 1998).

c) Labour

Paddy cultivation in Bario mainly depends on family labour. Some smaller families with bigger farms would have to hire additional labour from Indonesia or the local Penans. Besides that, most of the Kelabits actually migrated out of Bario to further their education as well as seek better job opportunities in major cities all around the world (Sanggin et al., 1998).

Therefore, all these contributed to the shortage of Bario rice supplies to meet the local demands as well as the demands from tourists from all over the world. Bario rice has gained its popularity not only among Malaysians but tourists from all around the world that visited Ba'kelalan, Bario and Miri. Approximately 600 to 720 tonnes of Bario rice is produced in the Kelabit Highlands and a portion of it is sold in Sarawak yearly and the market price for the authentic Bario rice can easily cost up to RM 10 per kg to RM18 per kg (Jiwan et al., 2015).

2.5 Bario Rice

There were many Bario rice varieties can be found in the market which includes Adan Halus, Adan Sederhana, Adan Merah, Adan Hitam, Adan Pulut, Adan Tuan, Adan/Bario Celum, Bario Pendek, Bario Banjal, Bario Sederhana, Bario Brunei, Bario Selepin, Bario Tinggi and Bario Tuan (Lee et al., 2014; Nicholas et al., 2014). However, in Bario, the most commercialize crop is the Adan Halus variety (*Oryza sativa* L. var. *adan halus*). Adan Sederhana is mostly planted in the lowland of Sarawak (Lee et al., 2014). Adan Halus and Bario Tuan rice variety are categorized as white rice while Bario Merah and Bario Celum is categorized as coloured rice with red and black bran layers respectively (Nicholas et al., 2014).

In a study by Nicholas et al. (2014), Bario Celum and Bario Tuan showed promising results as compared to MR 219. They have moderate glycemic index (GI) which plays significant roles in maintaining or slightly lowering blood sugars, low in fat content, high in crude fibre content, lower carbohydrate content and low energy content compared to the commercialized regular rice, MR 219 (Nicholas, 2014).

Paddy cultivation in Bario resembles nature farming as it is based on rain-fed with low fertilizer input (Jiwan et al., 2015). In the recent years, the farming in Bario has diverted from traditional to mechanized system but some farmers still maintain their traditional farming. About 40 to 60% of the total harvested Bario rice were traded while the remaining were kept for their own consumption (Jiwan et al., 2015).

On the 10th of March 2008, Bario Rice has been registered as product of Geographical Indication (GI) with the Malaysian Intellectual Property Organization (MyIPO) (Lee et al., 2015).

Adan rice (*Oryza sativa* L. var. *adan)* is famous for its soft and sticky texture, fine and elongated grains, mild aromas and ethereal taste (Kevin et al., 2007; Wong et al., 2009). This rice variety is planted once a year in August and harvested in February the following year. The maturation period of this variety ranges between 170 to 175 days after sowing with 130 cm to 135 cm average in height (Samuel et al., 2015). To date, the potential production for this variety ranges from 600 to 720 tonnes with the total area planted with Adan rice is less than 400 ha (Jiwan et al., 2015).

In a study by Nicholas et al. (2014), it is reported that Adan Halus variety is considered as a good source of protein, has low fat content, higher crude fibre content, lower carbohydrate content and low energy content compared to the regular rice, MR 219. This showed that Adan rice has great potential to be considered as health food. Food with low energy level contributes to healthy diet. In the same experiment, they also reported that Adan Halus has higher thiamine (vitamin B1) content compared to MR 219. This shows that Adan Halus is more nutritious as thiamine plays significant role in energy metabolism as compared to MR 219. Besides that, fresh Adan Halus was reported to have low amylose content of 10.48% which makes it firmer and stickier (Nicholas et al., 2013).

2.6 Microbial Ecology of Soil

Nemergut et al. (2013) defined a community as a cluster of hypothetical interacting species that co-exist in similar space and time (Nemergut et al., 2013). It is a study that seeks to investigate the structure of biological assemblages, their functional interaction as well as how the community structure changes over space and time (Konopka, 2009). Unlike macrobial communities, microbial study has the potential for quick turnover. Besides that, the researchers of microbial communities' studies are often caught in dilemma between if the sample taken were small enough to be relevant for

microorganisms but yet large enough to be relevant to ecosystem processes for instance 1 g of soil (Nemergut et al., 2013). The microbial communities in every ecological niche on Earth play very important roles in maintaining the well-being of the associated ecosystems. Due to the advancement of DNA sequencing and metagenomics in the recent years, the knowledge of organismal composition and metabolic functions of diverse communities in the ecosystem has tremendously increased. However, the understanding of the whole ecological systems of these diverse microbial communities was yet to catch up (Xiao et al., 2017).

Soil is a complex ecosystem which holds major reservoir of microbial genetic diversity (Curtis et al., 2002; Robe et al., 2003). Soil microbes play major roles in the ecosystems which include nutrient uptake, nitrogen and carbon cycles, as well as soil (van der Heijden et al., 2008). They are important regulators of plant productivity, particularly in nutrient-deficient ecosystems where plant symbionts are responsible for the acquisition of limiting nutrients (van der Heijden et al., 2008). Mycorrhizal fungi and nitrogen-fixing bacteria are accountable for 5 to 20% of all nitrogen in grasslands and savannah and 80% of nitrogen in temperate and boreal forests. Soil microbes also regulates plant productivity through mineralization of nutrients and it is estimated about 20,000 plant species are completely dependent on microbial symbionts for growth and survival (van der Heijden et al., 2008).

Hansel et al. (2008) investigated the composition and diversity of microbial communities and specific functional groups involved in key pathways in the geochemical cycling of nitrogen, iron and sulfur. The results showed that microbial communities and ammonia-oxidizing as well as Fe (III)-reducing communities varied greatly depending on carbon availability, water content and pH. In particular, the archaeal *Crenarchaeota* 16S

rDNA sequences share distribution and diversity similar to those of the *amoA*-based ammonia-oxidizing archaeal community, suggesting a role for the archaeal community in ammonia oxidation and thus, nitrogen cycling. The study highlights the complexity and heterogeneity of the soil microbial community structure and their metabolic potential (Hansel et al., 2008).

The rhizosphere is a region in the soil that contains diverse soil-borne microbes. It is an extremely active region in the soil ecosystem in which the microbes living in that region were influenced by the chemical release from the roots (Arjun & Harikrishnan, 2011). Commencing studies on the soil microbes and plants interaction in the rhizosphere are crucially important to comprehend the natural processes for instance carbon cycling, nutrient cycling and ecosystem functioning (Arjun & Harikrishnan, 2011). However, ecologists still encounter challenges in linking the diverse microbes in the rhizosphere and their role in the ecosystem. These microbes have diverse metabolic capabilities and were responsible in plant health, hence, knowledge on their community structure is crucial in understanding their individual roles in the natural ecosystem (Arjun & Harikrishnan, 2011). Buée et al. (2009) reported that the higher the diversity of microbes in the rhizosphere maybe due to the root activities such as the presence of high level of organic exudates in which it provides ideal ecological niches for the growth of microbes and hence, leading to increase number of microbes in the soil ecosystem (Buée et al., 2009). Some studies reported that the microbes in the rhizosphere are plant growth promoters arousing seedling development and growth (Dakora, 2003). The diverse microbial communities and their activities in the rhizospheres of paddy field soils affects the soil fertility as well as the nutrient cycling efficiency (Arjun & Harikrishnan, 2011).

2.7 Microbial Diversity of Paddy Field Soil

Rice (*Oryza sativa* L.) is known as one of the staple food of the world especially in most Asian countries including Malaysia. Paddy fields inhabited by diverse microorganism that play significant roles in the conservation of the soil quality as well as the paddy's health (Liesack et al., 2000). Bacteria, as the most abundant group of soil microbes, are dynamically involved in numerous biogeochemical processes of bulk and rhizosphere soils (Buée et al., 2009).

Aslam et al. (2013) did a study on the diversity of the bacterial community in the rice rhizosphere managed under conventional and no-tillage practices using culture-dependent approach. Their results claimed that the bacterial communities were different at certain growth stages in rice. In this study, they managed to culture 132 isolates of bacterial strains among which 60 strains came from conventional tillage soil while 72 strains from no-tillage soil. The most abundant bacterial phyla in conventional tillage soil were Proteobacteria (38%), followed by Actinobacteria (22%), Firmicutes (33%), Bacteroidetes (5%) and Acidobacteria (2%) whereas in the no-tillage soil, Proteobacteria (63%) were the most abundant bacterial phyla followed by Actinobacteria (24%), Firmicutes (6%) and Bacteroidetes (8%) during the four rice cultivation stages. Meanwhile, the most abundant bacterial families were Paenibacillaceae (15%), Bacillaceae (13%), Sphingomonadaceae (10%), Intrasporangiaceae (7%), Xanthomonadaceae (7%), Methylobacteriaceae (7%), Microbacteriaceae (5%), Caulobacteraceae (5%), Sphingobacteriaceae (3%), Micrococcaceae (3%), Burkholderiales incertae sedis (3%) and Staphylococcaceae (3%) while under no-tillage conditions, the most abundant families were Bradyrhizobiaceae (19%), Xanthomonadaceae (14%), Sphingomonadaceae (13%), Intrasporangiaceae (7%), Microbacteriaceae (6%), Bacillaceae (4%), Flavobacteriaceae (4%), Phyllobacteriaceae

(4%), Cellulomonadaceae (3%), Chitinophagaceae (3%), Mycobacteriaceae (3%) and Pseudomonadaceae (3%). They also reported that in both field, the number of Proteobacteria increased from the pre-sowing to the vegetative stage but remained constant until the ripening stage. But a different case for Actinobacteria, the number decreased gradually with cultivation time in both field. In the meantime, the Firmicutes group was affected by tillage practices and growth phase of rice in conventional tillage soils. Subsequently, in no-tillage soil, the Bacteroidetes group was found at pre-sowing, reproductive and ripening stages while in conventional tillage soils, they can be found at the pre-sowing stage (Aslam et al., 2013).

According to Ahn et al. (2012), in a study on the characterization of bacterial and archaeal communities in paddy field soils that was subjected to long-term fertilization practices, the most predominant bacterial phyla were Chloroflexi, Proteobacteria, and Actinobacteria (58 -76%) followed by Acidobacteria (4 - 7%), Firmicutes (2 - 6%), Bacteroidetes (2 - 7%), Planctomycetes (0.4 - 2%), and Gemmatimonadetes (1 - 2%). From the total sequences obtained from the paddy field soils, only 18 - 35% were able to be classified up to a genus level which demonstrates that the bacterial diversity in the paddy field soils are still mostly undiscovered. The phylum Chloroflexi which occupied the highest portion of the bacterial sequences in most of the soil samples at 21 - 33% were mostly conquered by Anaerolineaceae (54 - 70%) and Caldilineacea (8 - 14%) at a class level. The next most abundant phylum was the Proteobacteria (20 - 31%) and it can be divided into two functional groups namely chemoorganotrophs and chemolithotrophs. Chemoorganotrophs utilizes fermentation products for example fatty acids, alcohols or methane while chemolithotrophs utilizes reduced inorganic compounds for instance ammonia, sulphur or iron (II) for energy sources. In this study, they found both functional groups present in the paddy field soils. Under chemoorganotrophs group, *Pseudolabrys,*

Hyphomicrobium, Rhodobium, Methylocystis, Anaeromyxobacter, Desulfobacca, Geobacter, and Methylobacter were found while *Nitrosomonas, Thiobacillus*, and *Sideroxydans* were found under the chemolithotrophs group. Subsequently, the third highest abundance in the paddy field soils belong to the phylum Actinobacteria (5 - 23%). The most dominant genera under this phylum were *Arthrobacter, Marmoricola, Oryzihumus, Terrabacter, Nocardioides, Frankia,* and *Mycobacterium.* In the other hand, *Arthrobacter,* being the most dominant genus occupied 4 - 27% of the total sequences under Actinobacteria phylum. In the same study, the phylum distribution of archaeal communities in the paddy field soils showed that Crenarchaeota was the most predominant phyla occupying 67 - 90% followed by Thaumarchaeota (6 - 20%) and Euryarchaeota (3 - 19%). In addition to that, the most predominant class level under the phylum Euryarchaeota was Thermoplasmata (38 - 94%) and Methanomicrobia (6 - 56%).The most predominant genera of the class level Methanomicrobia was Methanosaeta (30 - 100%), Methanosarcina (0 - 17%) and Methanocella (0 - 60%) while in the phylum Thaumarchaeota, the most abundant genus was *Candidatus Nitrososphaera,* occupying 63 - 98% of the total sequences (Ahn et al., 2012).

In another study on the bacterial community variations in an alfalfa-rice crop rotation system using 16S rRNA gene 454-pyrosequencing by Lopes et al. (2014), they reported the presence of a total of 39 phyla with 19 of the phyla showed sequence abundance above 0.1%. Among the 19 phyla, the predominant group were Acidobacteria (32.4%) followed by Proteobacteria (26.3%), Chloroflexi (8.6%), Actinobacteria (7.5%), Bacteroidetes (7.3%) and Gemmatimonadetes (6.6%). They concluded that in September, rice-cropped soil showed lower diversity and lower relative abundance of rare OTUs than the uncropped soil, a relative increase in the abundance of Thermodesulfovibrionaceae was observed from April to September and finally in the fourth year of crop rotation, the

relative abundance of Acidobacteria and presumably anaerobic bacteria was significantly higher than in the third year compared to the higher abundance of presumably aerobic bacteria in the third than in the fourth year, mainly in April (Lopes et al., 2014).

In addition to that, a study by Xiao et al. (2016) on the analyses of microbial community of two contrasting soil cores (flooded paddy soils and dry corn field soil) contaminated by alimony and arsenic was conducted. They reported that a total of 44 phyla that were found with 8 most predominant phyla which accounted for more than 95% of the total 16S rRNA sequences obtained from Illumina Miseq platform. The 8 phyla were Proteobacteria (27.37%), Chloroflexi (23.97%), Acidobacteria (23.76), Nitrospirae (6%), Actinobacteria (5.55%), GAL15 (2.28%), Planctomycetes (1.97%), and AD3 (1.19%). While at class level, Ktedonobacteria consist of 18.76% of the total reads and followed by Deltaproteobacteria (10.83%), DA052 (7.94%), Alphaproteobacteria (7.74%), Betaproteobacteria (6.35%), Acidobacteria (6.05%), and Nitrospira (6.01%). The microbial richness and diversity as well as the alpha diversity indices differs significantly between the two soil cores (Xiao et al., 2017).

He et al. (2017), in a study on the bulk and rhizosphere soil bacteria communities in paddy fields under mixed heavy metal contamination using Illumina-based analysis, reported 15 bacterial phyla with the dominant phyla consist of more than 1% of the total community in each bulk and rhizosphere soil were Proteobacteria (25.5% - 38.9%), Actinobacteria (22.9% - 38.5%), Firmicutes (12.0% - 19.4%), Acidobacteria (4.5% - 10.7%), Gemmatimonadetes (2.3% - 6.5%), Chloroflexi (2.1% - 4.8%), Bacteroidetes (1.2% - 3.4%) and Nitrospirae (1.3% - 1.8%) while the less abundant bacterial phyla in both soil samples included Chlorobi, Verrucomicrobia, Spirochaetes, Elusimicrobia, Cyanobacteria and the candidate phyla OT1 and TM7. However, the predominant

bacterial phyla differed among the soil samples collected from different sites. For instance, the bulk and rhizosphere soils collected from Liantang village revealed that Actinobacteria being the most predominant bacteria phyla whereas in the bulk and rhizosphere soils collected from Fankou town showed higher diversity in Proteobacteria, Firmicutes, Gemmatimonadetes and Bacteroidetes. In addition, the rhizosphere soil from both Liantang village and Fankou town displayed higher relative abundance of Chloroflexi, Chlorobi and Spirochaetes than the bulk soils. They also reported a total of 25 orders were assigned in which the dominant orders in both bulk and rhizosphere soil were Actinomycetales, Bacillales, Clostridiales, Gaiellales, Rhizobiales, Myxococcales, Solirubrobacterales, Acidimicrobiales, Solibacterales, Syntrophobacterales, Rhodospirillales and Nitrospirales. In addition to that, the predominant genera in at least 3 paddy fields were reported to be *Bacillus, Clostridium, Rhodoplanes, Thiobacillus, Anaeromyxobacter* and *Candidatus Solibacter* while other genera that were found in all soils and were dominant in some samples were *Mycobacterium*, *Kaistobacter*, *Geobacter*, *Streptomyces*, *Tepidibacter*, *Phycicoccus*, *Nitrospira*, *Bradyrhizobium*, *Terracoccus*, *Anaerospora* and *Desulfosporosinus,* They also reported a rare order and genus in paddy field soil of Saprospirales, HOC36 and SC-I-84 and *Anaerospora* respectively (He et al., 2017).

2.8 Microbial Communities of Other Staple Food Farm Soil

There are many staple foods around the world for example wheat, maize (corn), millet, sorghum, soybean, root, tubers and many more. Among these, three of them namely wheat, maize (corn) and rice contribute about 60% of the world's food energy intake (FAO, 1995). Over centuries, many researches were done to study these valuable crops. Among the research, were the studies on the microbial communities in the natural ecosystem.

In a study by Tian et al. (2017) on the changes in soil microbial communities after 10 years of winter wheat cultivation versus fallow in an organic-poor soil in the Loess Plateau of China, they reported that a total of 37 phyla and 746 genera was generated from the 16S rRNA gene sequences obtained from the soil samples. The most dominant phyla in all the soil samples reported to be Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. They represented an average of 82.5% of all the microbes' sequences in all the soil samples. The most predominant genera in all the soil samples which represent an average of 10.3% of all the bacterial sequences were *Bacillus*, *Bacteroides*, *Lactococcus*, and *Steroidobacter.* Among the three treated soil samples, *Bacteroides* found to be the most dominant genera in BF (continued bare fallow without fertilization or wheat cultivation) soil at 9.1% but significantly decreased in FW (fertilized winter wheat) and NF (continued natural fallow without fertilization or wheat cultivation) soils. However, no significant differences in the relative abundances of other dominant genera were found among the three treated soils (Tian et al., 2017).

In another study on the shifts in the microbial communities in soil, rhizosphere and roots of two major crop (maize and soybean) systems under elevated carbon dioxide and ozone conducted by Wang et al. (2017), they reported 320 OTUs less abundant soil microbes in the eO_3 (elevated ozone) soil in which hybrid maize was grown. From the total sequences, the most dominant phyla belonged to Actinobacteria (25%), followed by Proteobacteria (21.6%), Acidobacteria (15%), Chloroflexi (10.4%), Verrucomicrobia (7.8%) and Planctomycetes (5.6%). However, a total of 1158 OTUs were significantly enriched in the soil under eO_3 in which hybrid maize was planted. The phylum Proteobacteria was the most dominant phylum among these enriched soil OTUs at 27.6% followed by Actinobacteria at 23.7%, Acidobacteria at 15.5%, Chloroflexi at 6.7%, Verrucomicrobia at 6.4% and Planctomycetes at 6.9%. In the other hand, at family level,

the analysis revealed the presence of nitrifying bacteria such as Nitrososphaeraceae, Nitrospiraceae, Nocardioidaceae, and 0319-6A21 where the hybrid maize was planted. The OTUs of nitrogen fixing bacteria for instance Sphingomonadaceae, Rhizobiaceae, Termomonosporaceae, Micromonosporaceae, Streptomycetaceae and Bradyrhizobiaceae were significantly increased under eO_3 condition which further suggest the impact of eO_3 in nitrogen cycling in which the hybrid maize was grown. In the same research, they did a differential OTU abundance analysis to differentiate the OTUs between aCO_2 (ambient CO_2) and eCO_2 (elevated CO_2) in the rhizosphere of soybean and they revealed 496 OTUs were significantly enriched or depleted in the rhizosphere samples under eCO_2 conditions. The reported OTUs were members of 68 families in which was dominated by Comamonadaceae (aCO2: 9.9%, eCO2: 10.4%) followed by Sphingomonadaceae (aCO2: 6.0%, eCO2: 4.8%), Gaiellaceae (aCO2: 2.7%, eCO2: 2.4%) and finally, Streptomycetaceae (aCO2: 4.6%, eCO2: 4.1%) (Wang et al., 2017).

Donn et al. (2014), in a study on the evolution of bacterial communities in the wheat crop rhizosphere reported that the most dominant phyla in all the soils samples were Proteobacteria, Actinobacteria and Bacteroidetes. Proteobacteria particularly β-proteobacteria dominated TB (tightly bound to the rhizoplane after washing and root endophytes) soil at over 70% of the total sequences at the earliest sampling time, V1 while in LB (loosely bound after washed from the roots and attached soil) soil, they were composed of only 40% of the total sequences in LB soil. The LB soil were composed of higher proportions of Acidobacteria, Firmicutes and unclassified taxa. However, at later sampling stages, Actinobacteria was found to be more ubiquitous in TB soil than that of LB soil. In addition to that, Streptomycetaceae was the most dominant family in the TB than LB fraction. In TB fraction at V1 (vegetative stage year 1), Pseudomonadaceae were the most dominant family while Streptomycetaceae were found to be the most dominant

family at V2 (similar stage in year 2). The phyla Chloroflexi, Acidobacteria Group 1 and Nocardioidaceae was significantly reduced in the TB as compared to LB fraction. The dominance of Proteobacteria particularly members of the Oxalobacteraceae decreased with plant age. In this study, Oxalobacteraceae that dominated the new roots showed significant reduction on mature and senescing roots. This study also revealed that Pseudomonadaceae predominated in the early vegetative growth at year 1 (V1) than that of reproductive stage (R1). Likewise, Flavobacteriaceae were the predominant family of Bacteroidetes were greatly reduced in number on senescing roots (Sb) but on mature roots (R1), Sphingobacteriaceae consist of greater share of the LB community (Donn et al., 2014).

2.9 Importance of Microbes in the Soil Ecosystem

Soil microorganism comprises of bacteria, actinomycetes, fungi, protozoa and algae. They play vital roles in supporting their host in the natural ecosystem by symbiotic relationship. There are three common groups of soil microorganisms can be found in the rhizospheres namely beneficial, pathogenic microorganisms and commensal. Their interaction and competition for plant nutrition explains the soil suppressive ability against insects and pathogens (Berendsen et al., 2012; Lakshmanan et al., 2014).

The key drivers that influence the diversity of rhizhospheric microbial includes soil types, plant host genotype and agronomy practices (Lakshmanan et al., 2014; Philippot et al., 2013). Both biotic for example host variety, genotypes, growth stages, root vicinity and root structure, and abiotic factors for instance soil pH, temperature, seasonal changes and presence of rhizospheric exudates will act as chemical indicator for microorganism and hence, influence their community structure as well as function within

the complex plant and root microbiomes (Berendsen et al., 2012; Berg & Smalla, 2009; Dennis et al., 2010; Lakshmanan et al., 2014; Philippot et al., 2013).

Actinomycetales, gram-positive bacteria responsible for the production of many natural antimicrobial drug compounds found in the market for instance streptomycin, actinomycin and streptothricin (Moore et al., 2009; Ofaim et al., 2017; Zhou et al., 2012). Besides producing antibiotics, bacteria from this order also involved in a pathway related to the degradation of fluorobenzoate and compounds of polychlorinated biphenyls (PCBs) pollutant family and the pathways that were influenced by the removal includes lipids, terepnoids and plant induced secondary metabolites categories (Ofaim et al., 2017). Another taxanomic group in the root environment that play key roles in the rhizosphere was the Rhizobiales (Berg & Smalla, 2009; Ofaim et al., 2017). Rhizobiales in the plant roots involved in the metabolism of linoleic acid and geraniol associated pathways in which both compounds were both plant exudates that are used as carbon sources in the root's rhizosphere (Folman et al., 2001; Ofaim et al., 2017; Owen et al., 2018).

In addition, the bacterial order Sphingomonadales in the plant roots were involved in the phenylpropanoid and flavonoid-related pathways (Ofaim et al., 2017). The role of plant exudates for instance flavonoids, organic acids and carbohydrates works as determinants of the microbes' community structure in the root rhizosphere (Narasimhan et al., 2003; Ofaim et al., 2017; Ofek-lalzar et al., 2014; Schulz et al., 2007). Matsumura et al. (2015) reported that bisphenol degradation in the root was influenced by the removal of Actinomycetales, Pseudomonadales and Burkholderiale instead of Rhizobiales (Matsumura et al., 2015; Ofaim et al., 2017). In the other hand, some bacterial taxonomic group plays important roles in the metabolism of potential regulators in plant-microbe interactions. For instance, Pseudomonadales taxonomic group was reported to

exclusively affects the production of B- group vitamins for instance vitamin B6 in the rhizosphere (Marek-Kozaczuk & Skorupska, 2001; Ofaim et al.,2017).

The genera *Clostridium, Thiobacillus, Anaeromyxobacter, Geobacter* and *Desulfosporosinus* play key-role in iron and iron cycling in paddy field soils while the genus Rhodoplanes may play crucial role in improving soil fertility (Sun et al., 2015). Most of the species belonging to the Chloroflexi, Anaerolineaceae and Caldilineacea at class level are strictly anaerobes and involved in sugar and polysaccharides fermentation by converting it into organic acids and hydrogen (Grégoire et al., 2011; Podosokorskaya et al., 2013). Hence, the Chloroflexi-associated bacteria are theoretically known as the primary bacterial degraders of polysaccharides in the anoxic zones of paddy field soils (Ahn et al., 2012).

Furthermore, the phylum Proteobacteria are divided into two functional group namely chemoorganotrophs and chemolithotrophs. The former group uses fermentation products such as fatty acids, alcohols or methane while the latter group uses reduced inorganic compounds for instance ammonia, sulphur or iron (II) for energy sources. Examples of genera under the former group are *Pseudolabrys, Hyphomicrobium, Rhodobium, Methylocystis, Anaeromyxobacter, Desulfobacca, Geobacter,* and *Methylobacter* while examples of genera under the latter group are *Nitrosomonas, Thiobacillus,* and *Sideroxydans.* These bacterial groups are generally not fermentative as they utilize oxygen, nitrite, nitrate, sulphate or iron (III) as electron acceptor, hence they are expected to be highly active in zones where both either organic or inorganic electron donors and external electron acceptors are found (Ahn et al., 2012).

Next, the examples of genera under the phylum Actinobacteria are *Arthrobacter, Marmoricola, Oryzihumus, Terrabacter, Nocardioides, Frankia,* and *Mycobacterium*. The species belonging to these genera are generally known to be aerophilic or microaerophilic in which they are responsible for the degradation of organic matter in the oxic zones of paddy field soils (Ahn et al., 2012). The high relative abundance of the genus Arthrobacter are due to their ability to resist starvation and dryness besides their nutritional adaptability (Jones & Keddie, 2006).

Besides that, the archaeal group was reported to be anaerobic or facultatively anaerobic. The species of the genus *Thermoplasmata* (Phylum Euryarchaeota) were also reported to be extremely acidophilic in which they grow well at pH less than 2 (Reysenbach, 2001; Golyshina et al., 2009) but there are some environmental species known as WCHD3- 02 were obtained from different sources as claimed such as marine sediment, anaerobic digester, rumen, cattle compost and deep subsurface groundwater. In the other hand, the genera *Methanosaeta* and *Methanosarcina* of phylum Methanomicrobia was reported to be acetotrophic methanogens and cosmopolitan in the paddy field soil ecosystems (Conrad et al., 2006).

Methanocella spp*.,* a hydrogenotrophic methanogens, isolated from paddy field soil (Sakai et al., 2008, 2010) were previously known as Rice Cluster I. They are responsible for methane production in the rhizosphere of paddy field soils (Conrad et al., 2006). Pester et al. (2011) and Brochier-Armanet et al. (2012) reported that all the isolates associated with Thaumarchaeota showed their ability to oxidize ammonia aerobically (Brochier-Armanet et al., 2012; Pester et al., 2011). *Nitrosomonas* or *Nitorosococcus* (genus level), an ammonia-oxidizing bacterial groups, was responsible in the nitrification in the paddy field soils (Ahn et al., 2012).

2.10 Metagenomics

Metagenomics can be defined as the technique to study genetic material recovered directly from environmental samples. It is a culture-independent analysis of the whole genetic composition in a sample unlike genomics, which analyse the genomic DNA from a single organism or cell (Gilbert & Dupont, 2011; Handelsman, 2004; Mirete et al., 2016). It explores the composition of functional gene of microorganism communities and provides better understanding and pictures in comparison to 16S rRNA gene which are often based on the diversity of a single gene (Handelsman, 2004; Nesme et al., 2016; Thomas et al., 2012). Over a decade, metagenomics has emerged as a powerful tool used to provide access to the "unculturable" majority of microbial communities by gaining access to the physiology and genetics of the uncultured microorganisms (Handelsman, 2004; Lorenz et al., 2002; Nesme et al., 2016; Whitman et al., 1998).

Metagenomics were able to provide genetic information on potentially novel enzymes or biocatalysts, genomic linkages between function and phylogeny for uncultured organisms, and evolutionary profiles of community function and structure (Thomas et al., 2012). The rapid and cost-efficient high-throughput next-generation sequencing (NGS) has accelerated the progress of sequence-based metagenomics and it can be proven by the increasing amount of metagenome shotgun sequence datasets in a decade (Thomas et al., 2012). High-throughput sequencing platforms with increased capacity, assist in characterization and quantification of soil microbial diversity (Nesme et al., 2016).

Metagenomics has been widely used as a standard tool used for many laboratories and ecologist, the same manner as 16S rRNA gene fingerprinting methods used to describe microbial community profile (Gilbert & Dupont, 2011; Thomas et al., 2012). In

addition, shotgun metagenomics is a prevailing, high-resolution method which enable the study of microbial diversity *in situ*.

Environmental metagenomic libraries has shown to be great mining resources for new microbial enzymes and antibiotics with promising application in biotechnology, medicine as well as industry (Riesenfeld et al., 2004; Rondon et al., 2000). The vast majority of the biosphere's genetic and metabolic diversity contains a staggering number of yet uncharacterized microbial genomes (Pace, 1997; Torsvik et al., 2002). The advancement of metagenomics was able to reveal information about uncultured bacteria circumventing the culture-dependent methods (Lorenz et al., 2002).

However, microbial ecologists faced challenges in interpreting big data to further understanding the relationship between soil microbes and the soil ecosystem generated from the current sequencing technologies (Nesme et al., 2016). The definitive goal of metagenomics is to offer a descriptive and finally prognostic taxonomic and metabolic model of an ecosystem (Gilbert & Dupont, 2011).

2.11 Soil Functional Metagenomics

Soil metagenomic studies are rapidly increasing over the years however these studies are prone to preconceptions and limitation such as in cell lysis, nucleic acid extraction, sequencing mistakes and are limited to consistent quantification and annotation of sequenced genes (Lombard et al., 2011; Prosser, 2015). Nevertheless, these limitations are unimportant hence the full coverage on soil metagenome is not usually achieved for instance in a study, they reported that even with 300 GB of data, the full coverage of the soil diversity was not achieved (Howe et al., 2014; Prosser, 2015).

Traditional methods such as PCR or microarray could not identify majority of unknown genes belonging to environmental microorganisms. Hence, with the use of functional metagenomics approach, the exploration of novel resistance genes was made possible again by circumventing the culture-dependent method and sequence bias (Allen et al., 2009; Gibson et al., 2015; Sommer et al., 2009; Vercammen et al., 2013; Wang et al., 2017). With this approach, novel resistance genes from the sequence obtained from diverse environmental resistome were revealed (Allen et al., 2009; Gibson et al., 2015; Wang et al., 2017).

Soil microbes comprises of many genes that encodes for biological processes such as nitrification, denitrification, ammonia oxidation and many more within them. Besides, a single enzyme may involve in several physiological pathways and participate in the functionality of an ecosystem or may have multiple functions (Prosser, 2015). For instance, *nirK* genes, which encodes for the enzyme nitrite reductase, responsible for reducing nitrite to nitric oxide, participate in nitrite oxidation, ammonia oxidation, anaerobic ammonia oxidation and denitrification. Furthermore, the gene *amoA*, encodes for the enzyme ammonia monooxygenase, involve in ammonia and methane oxidation (Arp & Stein, 2003; Prosser, 2015).

Wang et al. (2017) discovered different classes of tetracycline resistance genes from soil metagenomes in China through functional metagenomic approach. Their results verified the potential of discovering vast novel resistance genes from soil microorganisms originate from pristine environment (Davies & Davies, 2010; Martínez et al., 2015; Wang et al., 2017). In another study by Hjort et al. (2014), they discovered a novel chitinase which was formerly characterized as suppressive to phytopathogens that show some antifungal activity obtained from a Swedish field soil. This was the first reported active

chitinase found, produced and purified via metagenomic approach. The discovery of this novel chitinase has huge potential as environmentally friendly alternatives to noxious chemical-derived fungicide against crop fungal pathogens (Hjort et al., 2014).

Souza et al. (2018) discovered 42,631 hydrolases which belongs to five classes from four shotgun metagenomes namely no-tillage (NT), conventional tillage (CT), crop succession such as soybean/wheat (CS) and crop rotation of soybean/maize/wheat/lupine/oat (CR) derived from southern Brazil. The abundance of hydrolases increased five-fold in NT soils compared to CT soils. Besides that, they also discovered other important genes such as lipases, laccases, cellulases, proteases, amylases and pectinases from the four metagenomes using metagenomic approaches (Souza et al., 2018).

Furthermore, in another study by Goethem et al. (2018), they manage to identify some antibiotic resistance gene from Antarctic soils using metagenomic approach. Contigs from assembled shotgun metagenome were able to efficiently used to access antibiotic resistance genes from environmental resistomes (D'Costa et al., 2006). They reported 177 resistance genes that are resistant to natural antibiotics. Their hypothesis on the antibiotic resistance genes found in this study was derived from antibiotic-producing species was true and was backed with the existence of antibiotic biosynthesis genes present in most phyla encoding resistance (Goethem et al., 2018).

Metagenomic datasets obtained from shotgun sequencing were able to reveal biosynthetic and metabolic pathways of microbes in the natural ecosystem. It helps to further understands the relationship between the microbes and the host itself. For example, in a study by Lu et al. (2018), on the microbial communities found in the

rhizosphere of wheat, barley and two rice varieties at seedling stage. They reported a high level of degradation pathways in wheat and barley such as limonene, glycan and pinene which may be associated with the digestion of root exudates by the rhizosphere microbes. In the meantime, they also observe the lower abundance functions in 10 metabolism pathways and 9 biosynthesis pathways in which half were associated with amino acid metabolism. The results suggest that the rhizosphere microbes may have abundant amino acids from the root exudates and therefore reduce the biosynthesis of amino acid. They concluded that from the four crops tested at seedling stage, the microbes that exhibit specific features especially in promoting plant growth was evidently present in the soils (Lu et al., 2018).

2.12 NGS

The study of microbial diversity and function metagenomes were simplified with NGS application by evading the need to culture fastidious bacteria which are often unculturable under standard laboratory condition and media (Ismail et al., 2017; Torsvik & Øvreås, 2002; Torsvik et al., 2002).

The next-generation sequencing platforms such as Illumina/Solexa, Roche/454, HeliScope/Helicos BioSciences and Life/APG are fast and reasonably priced compared to the traditional Sanger's dideoxy sequencing of cloned amplicons (Metzker, 2010). The application of NGS technologies on soil biodiversity in various ecosytems has been increasing rapidly for example in grasslands, agricultural lands, forest lands, desert lands and Artic and Antarctic soils (Mardis, 2008; Nielsen & Wall, 2013; Orgiazzi et al., 2015). Shotgun metagenomic sequencing prevails over high-throughput 16S rRNA amplicon sequencing by omitting the use of standards primers to detect rRNA genes which are often untraceable (Brown et al., 2015).

Big-scale sequencing technologies allows deeper study into layers of microbial communities and are essential in bestowing unbiased view of the phylogenetic composition and functional diversity of environment microbial communities (Zwolinski, 2007). The ability of big-scale sequencing techniques to generate billions of reads at relatively lower cost with high speed is useful in many application especially whole-genome sequencing, metagenomics, metatranscriptomics and proteogenomics and the sequencers' performance significantly improves over the years (Glenn, 2011; Liu et al., 2012). High-throughput sequencer such as Illumina had shown promising and speedy method in analysing the diversity of soil microbes and their community structure (He et al., 2017; Myrold et al., 2013; Rincon-florez et al., 2013).

The advantages of Illumina Hiseq 200 is the output generated can reach up to 600 GB per run and could finish in a duration of 8 days. They have predicted that it could reach 1 TB/run in the future. In comparison to 454/Roche and SOLiD sequencing platforms, Illumina Hiseq 2000 and Illumina Miseq is the cheapest in sequencing with $0.02/million bases in terms of reagents. Thousands of samples could be run simultaneously with mutiplexing incorporated in P5/P7 primers and adapters. Illumina Hiseq 2000 sequencer, with the aid of Truseq v3 reagents and associated softwares, was enhanced on the high GC sequencing (Liu et al., 2012).

2.13 Bioinformatic Tools

DNA or genome annotation is the process of identifying the location of the genes and classifying their functions in the genome sequence (Richardson & Watson, 2013). There were many online annotation servers available for free for instance NCBI, RAST and xBASE2. The NCBI provides a Prokaryotic Genomes Automatic Annotation Pipeline service via email with a short turn-around time (Stewart et al., 2009). RAST, on the other

hand, is a server in the web known for annotating bacterial and archaeal genomes with turn-around results in less than a day (Aziz et al., 2008) while xBASE2 , the most rapid among the three pipelines, have turn-around time in a matter of hours (Chaudhuri et al., 2008). However, Prokka is another available software tool that can be installed on any Unix system using command line and it organizes the existing tools to achieve a rich and more reliable annotation of genomic bacterial sequences (Seemann, 2014). However, Prokka depends on external prediction tools features to classify the coordinates of genomic sequence in contigs. The external tools used are listed in Table 2.1 with all of the tools, except for Prodigal, provide coordinates and appropriate labels to describe the features. Prokka was designed to be both rapid and accurate tool.

Table 2.1: Bioinformatic prediction tools used by Prokka (Seemann, 2014)

Bioinformatic Tools	**Predicted Features**	**References**
Prodigal	Coding sequence (CDS)	(Hyatt et al., 2010)
RNAmmer	Ribosomal RNA genes (rRNA)	(Lagesen et al., 2007)
Aragorn	Transfer RNA genes	(Laslett & Canback, 2004)
SignalP	Signal leader peptides	(Petersen et al., 2011)
Infernal	Non-coding RNA	(Nawrocki & Eddy, 2013)

CHAPTER 3: MATERIALS AND METHODS

3.1 Equipment and Instruments

Instruments used during the course of this study included -20ºC freezer (Liebherr, UK); 2100 Bioanalyzer (Agilent Technologies, USA); 4ºC chiller (Thermo Scientific, USA); -80ºC freezer (Gaia Science, Singapore); agarose gel electrophoresis (AGE) (Biorad, USA); autoclave machine (Hirayama, USA); centrifuge machine (Eppendorf, North America); eco-spin microcentrifuge (Elmi, Latvia); fume hood (Esco Technologies, USA); gel documentary image analyzer (UVP, USA); high performance UV transilluminator (UVP, USA); HiSeq 2000 next generation DNA platform (Illumina, USA); ice maker (Nuove Tecnologie Del Freddo, Italy); laminar flow cabinet (Esco, Technologies, USA); magnetic stirrer hot plate (Labmart, USA); milli-Q® integral water purification system (Merck, Germany); nanodrop spectrophotometer (Thermo Scientific, USA); pH meter (Sartorius, Germany); polymerase chain reaction (PCR) T100 thermal cycler (Biorad, USA); Qubit® 2.0 fluorometer (Invitrogen, USA); thermomixer (Eppendorf, North America); vortex mixer (Core Life Sciences, CA); Eco qPCR (Illumina, USA) and weighing machine (Sartorius, Germany).

Equipment used in this study included disposable petri dishes; laboratory glassware (Schott's bottles, conical flasks, volumetric flasks, measuring cylinder, beaker); syringe (Terumo, USA); syringe filter (0.22 µm pore size) (Sartorius, Germany); polypropylene tubes (15 mL and 50 mL); pipettes (Eppendorf, North America), pipette tips (Eppendorf, North America) and microtiter plate (6-well and 96-well).

3.2 Chemical Reagents

All the chemical reagents used in this study are of analytical grade purchased from Bio-Rad Laboratories Ltd., U.S.A.; Promega Ltd, U.S.A.; Invitrogen Corp., U.S.A.; Epicentre, U.S.A. and Illumina, Inc., U.S.A.

3.3 Buffer Solutions

Preparation of in-house solutions and buffers used in this study were sterilized by autoclaving at 121 ˚C, at 15 psi for 20 min. Heat sensitive solutions were filter sterilized with sterile 0.2 μm pore size nitrose cellulose membrane (Sartorius, Germany).

3.3.1 Tris-HCl Buffer (1 M, pH 8.0)

Briefly, 60.5 g of Tris base was weighed and dissolved in 400 mL of distilled water. The pH was adjusted to 8.0 by adding approximately 21 mL of concentrated hydrochloric acid (HCl) to the solution and finally distilled water was added to make the final volume of 500 mL.

3.3.2 Ethylenediaminetetraacetic Acid (EDTA) (0.5 M. pH 8.0)

Briefly, 93.05 g of EDTA were weighed and dissolved in 400 mL of distilled water. The pH was adjusted to 8.0 by adding sodium hydroxide pellet and finally distilled water was added to make the final volume of 500 mL.

3.3.3 10× Tris Borate Ethylenediaminetetraacetic Acid (TBE) Buffer

Briefly, 108 g of Tris and 55 g of Boric acid were weighed and dissolved in 800 mL of distilled water using a magnetic stirrer. Then 40 mL of 0.5 M EDTA (pH 8.0) was added to the mixture and finally distilled water was added to make the final volume of 1 litre.

3.4 DNA Marker

DNA ladder markers used in this study were GeneRuler^{TM} 1 kb DNA ladder purchased from Fermentas International Inc., Canada.

3.5 F3F Sample Collection

A total of 100 soil samples were randomly collected on June, 2014 from various points in a paddy plot at 0-10 cm depth by using sterile 50 mL Eppendorf conical tubes. The samples were kept frozen at -20°C upon reaching to the laboratory. The GPS location was at N 03° 44.799', E 115° 26.912' with the elevation of 3517 feet above sea level.

3.6 Genomic DNA Extraction

The soil metagenome DNA extraction was performed using Power Soil DNA Isolation Kit (MOBIO Laboratories, Inc) following the manufacturer's protocol with some modifications. Approximately 0.25 g of soil was added to the PowerBead tubes and was gently vortexed to mix. Then, 60 µL of Solution C1 was added and the tubes were inverted several times or vortexed briefly. The PowerBead tubes were secured horizontally on a flat-bed vortex pad with tape and vortexed at maximum speed for 1 hour. The tubes were later centrifuged at 10,000 × *g* for 30 seconds at room temperature (25°C). After that, the supernatant approximately 400 to 500 µL were transferred to a clean 2 mL collection tube. Next, 250 µL of Solution C2 was added and the tubes were vortexed for 5 seconds before incubation at 4 ºC for 5 minutes. Upon completion of the incubation period, the tubes were centrifuged at room temperature for 1 minute at 10,000 × *g*. About 600 µL of supernatant were carefully transferred into a clean 2 mL Collection Tube by avoiding the pellet. A volume of 200 µL of Solution C3 was added to the supernatant and was vortexed briefly before incubation at 4 ºC for 5 minutes. The tubes were subsequently centrifuged at room temperature for 1 minute at 10,000 × *g*. Then, 750

µL of supernatant were carefully transferred into a clean 2 mL Collection Tube by avoiding the pellet. Later, 1200 µL of Solution C4 were added to the supernatant and were vortexed for 5 seconds. Approximately 675 µL were loaded into each spin filters and were centrifuged at 10,000 × *g* for 1 minute at room temperature. The flow through was discarded and an additional 675 µL of the supernatant were added to the spin filter and centrifuged at 10,000 × *g* for 1 minute at room temperature. The process was repeated until finish. Next, 500 µL of Solution C5 were added to the spin filter and were centrifuged at room temperature for 30 seconds at 10,000 × *g* and the flow through were discarded. The washing step using Solution C5 was repeated again and completed this washing step by centrifuging again at room temperature for 1 minute at 10,000 × *g*. The spin filters were carefully transferred into clean 2 mL collection tubes. Finally, 30 µL of Solution C6 were added to the center of the white filter membrane and were incubated for 1 minute before centrifugation at room temperature for 30 seconds at 10, 000 × *g*. The final step was repeated again before the spin filters were discarded and the DNA was kept at -20 ºC.

3.7 Quantification and Qualification of Genomic DNA

The quality of the extracted DNA was checked on NanoDrop 2000c (Thermo Scientific, USA) and quantified on Qubit 2.0 (Life 19 Technologies, USA) fluorometer.

3.8 Agarose Gel Electrophoresis (AGE)

For agarose electrophoresis, 1 % (w/w) of agarose gel was prepared, put into the gel electrophoresis tank and 1× TBE was added until the TBE fully covered the gel in the tank. Finally, 3 µL of the genomic DNA extracted and 2 µL of 1 kb DNA ladder marker was loaded into 1 % (w/v) ethidium bromide (EtBr) stained agarose gel prepared. The

electrophoresis was done at 70 V for 1 hour and visualized on UVP ultraviolet trans-illuminator upon completion.

3.9 DNA Library Preparation

The following step was outsourced for sequencing to third party company available. Sequence libraries were generated using the TruSeq DNA Sample Preparation Kits (Illumina, USA) according to the manufacturer's guidelines. The Low Sample (LS) protocol was used and the procedures were as follows.

3.9.1 Fragmentation of Genomic DNA

The extracted gDNA was quantified using *Qubit* 2.0 (Life 19 Technologies, USA) fluorometer and the gDNA samples were normalized to a final volume of 55µL at 20 ng/µL per well of the PCR plate. The DNA was then sheared using Covaris S220 (Thermo Fisher Scientific, USA).

3.9.2 Performing End Repair

The End Repair procedure was performed according to manufacturer's protocol. The protocol can be summarised as follows. After the End Repair Mix was added to the fragmented DNA, it will be incubated in pre-heated thermal cycler at 30°C for 30 minutes. After incubation, the AMPure XP Beads and PCR grade water will be added to make a diluted beads mixture according to the formula below.

AMPure XP Beads: # of samples ×160 µL × 0.85= µL AMPure XP Beads

PCR grade water: # of samples ×160 µL × 0.15= µL PCR grade water

After the mixture is ready, it will be added to the incubated End Repair Mix. The mixture is then incubated again at room temperature for 15 minutes before placing it on the magnetic stand. The mixture was left on the magnetic stand for 15 minutes at room

temperature or until the liquid appeared to be clear. The supernatant was then removed and the process of adding diluted AMPure XP Beads was repeated once again. With the mixture on the magnetic stand, freshly prepared 80% ethanol was added to the sample without disturbing the beads. Then, the mixture was incubated at room temperature for 30 seconds before all of the supernatant was removed and discarded carefully so the beads were not disturbed. The washing step using 80% ethanol was repeated twice and was let to stand at room temperature for 15 minutes to dry and before removing it from the magnetic stand. The dry pellet was resuspended with Resuspension buffer and was incubated at room temperature for 2 minutes. Then, the mixture was placed on the magnetic stand at room temperature for 5 minutes or until the liquid appeared to be clear. Finally, the clear supernatant was transferred to the new 0.3 mL PCR plate.

3.9.3 Adenylate 3' Ends

Thawed A-tailing control and thawed A-tailing mix was added and mixed well into each well of the 0.3 mL PCR plate. The plate was placed on the pre-heated thermal cycler and incubated at 37°C for 30 minutes. Finally, it was removed from the thermal cycler and then immediately preceded with Adapter Ligation step.

3.9.4 Ligation of Adapters

The thawed Ligation control and Ligation Mix were added to each well of the plate respectively. Then, the thawed DNA adapter index was added to each well of the plate and the entire volume were gently pipetted 10 times to mix thoroughly. The plate was incubated on the preheated thermal cycler with the lid closed at 30 °C for 10 minutes. Then, Stop Ligation buffer was added to each well of the plate to inactivate the ligation and the entire volume was pipetted up and down 10 times to mix thoroughly. The mixed AMPure XP beads were added to each of the well of the plate and mixed prior to the plate

incubation at room temperature for 15 minutes. The process using AMPure XP beads and 80% ethanol was repeated before being air dried at room temperature for 15 minutes. The dried pellet was then resuspended with Resuspension buffer and mixed thoroughly. The plate was incubated at room temperature for 2 minutes before it was placed on the magnetic stand at room temperature for 5 minutes or until the liquid appeared clear. Then, the clear supernatant from each well were transferred to the corresponding well of the new 0.3 mL PCR plate. The same procedure was repeated again until the clear supernatant was transferred from each well of the plate to the corresponding well of the new 0.3 mL PCR plate labelled with the PCR barcode.

3.9.5 Enrichment of DNA Fragments

Thawed PCR master mix was added to each well of the PCR plate and mixed thoroughly by pipetting. The PCR plate was placed in the closed pre-programmed thermal cycler and PCR was selected to amplify the plate. The similar step using AMPure XP beads and freshly prepared 80% ethanol was repeated again. The air-dried pellet in each well was resuspended with Resuspension buffer and mixed thoroughly by pipetting. The PCR plate was incubated at room temperature for 2 minutes before the PCR plate was placed on the magnetic stand at room temperature for another 5 minutes or until the liquid appeared clear. Then, the clear supernatant was transferred from each well of the PCR plate to the corresponding well of the new 0.3 mL PCR plate.

3.9.6 Library Validation

The quantification of the sample library was performed using Kapa Biosystems Library Quantification kit (Illumina) and Eco qPCR (Illumina, USA) system according to the Illumina Sequencing Library qPCR Quantification guide. In the other hand, the qualification of the sample library was performed using a High Sensitivity DNA chip on

a Bioanalyzer 2100 (Agilent, USA) which was used to verify the size of the PCR enriched fragments as well as to check the template size distribution.

3.9.7 Normalization and Libraries Pooling

The sample library was transferred from each well of the plate to the corresponding well of the new MIDI plate. The concentration of the sample library in each well was normalized to 10 nM using Tris-Cl 10 nM, pH 8.5 with 0.1% (v/v) of Tween 20. The entire normalized sample library was mixed gently by pipetting.

3.10 NGS of Soil Metagenomic DNA

A pre-run wash was performed on the cBot and a protocol was selected from the cBot. The prepared samples were then loaded to the cBot and Start button was selected and the Run Status screen will be opened and the run began.

3.11 Sequence Analysis of Extracted Metagenomic DNA

Raw output data generated was retrieved from the company and personally analysed. The raw output data were processed and trimmed using CLC Genomic Workbench 7.0. Sequences less than 15 bp were discarded. The minimum contig length was set as 250 bp. Assembled contigs were identified for the presence of prokaryotic components by BLASTN with GenBank 16S Microbial database with e-value of $< 10^{-9}$.

3.12 Taxanomic Assignments of Paddy field Soil Microbiome

The data obtained from BLAST were analysed for taxonomic distribution using Metagenome Analyzer (MEGAN) version 5.2.3 (Huson et al., 2011). The microbial diversity was calculated based on the number of assigned reads. The taxanomic

classification summaries obtained from MEGAN was subjected to Community-Analyzer tool (Kuntal et al., 2013) for analysis.

3.13 Gene Prediction of Paddy Field Soil Metagenome

The assembled data exported from CLC Genomic Workbench 7.0 were the subjected to Prokka pipelines to annotate bacterial, archaeal and viral genomes for gene prediction.

3.14 Nucleotide Sequences Accession Number

The paddy field soil metagenomic raw reads were deposited in the NCBI Short Read Archive with accession number SRA174292 under study accession SRP044070.

CHAPTER 4: RESULTS

4.1 Metagenomic DNA Quality and Quantity

The quality and quantity of the extracted metagenomic DNA were checked prior to sequencing. Table 4.1 shows the metagenomic DNA extracted from the paddy field soil of the Kelabit Highlands.

Table 4.1: Quality and quantity of extracted metagenomic DNA

Sample	$A_{260/280}$	$A_{260/230}$	Quantity (ng/µL)
F3F	1.81	1.39	26.4

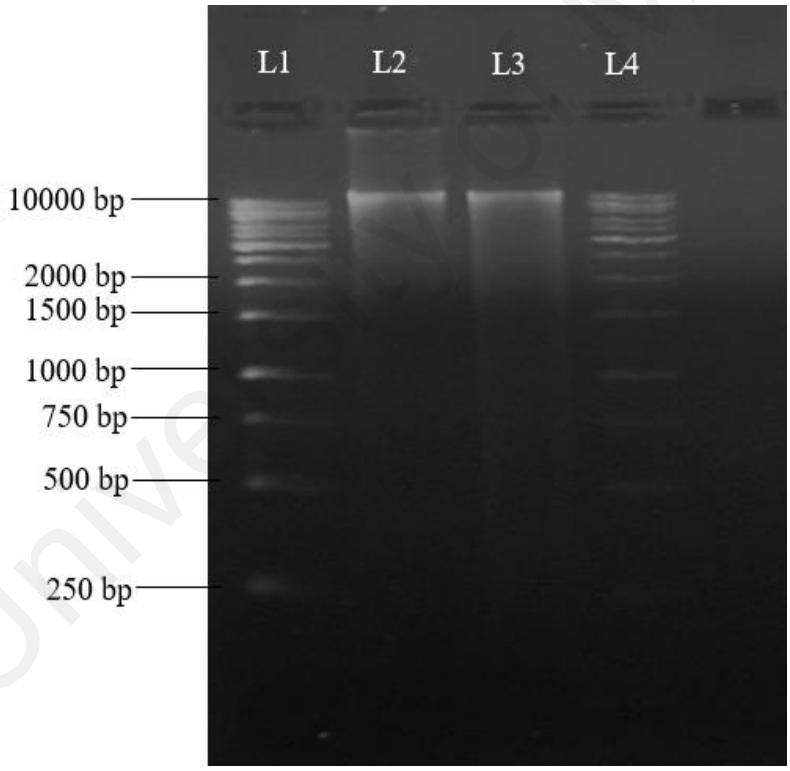

Figure 4.1: Agarose gel electrophoresis of the extracted DNA from F3F paddy field soil sample. L1: 1 kb DNA ladder, L2-L3: Extracted metagenomic DNA, L4: 1kb DNA ladder.

4.2 NGS of F3F Paddy field Soil

A total of 33.6 GB of raw reads generated from the F3F paddy field soil. The quality reads after trimming was 25.9 GB of sequences. The average length of sequence generated after quality trim was 89.9 bp. Table 4.2 shows the summary of NGS data statistics of the F3F soil sample. The assembled sequences of the F3F paddy field soil generated a total of 1,382,225 contigs of more than 250 bp in length, with N50 of 540 bp.

Table 4.2: Summary of Hiseq sequencing data

Sample	F3F
No. of sequences generated (bp)	24,318,655,838
No. of reads generated	275,118,914
No. of quality sequences generated (bp)	1,070,788,506
No. of quality reads generated	272,038,684
Avg length of quality reads (bp)	101.0
No. of contigs[1]	1,382,225

[1] Calculated based on contigs of more than 250 bp

4.3 NGS Analysis of F3F Paddy field Soil

4.3.1 Microbial Taxanomic Distribution of F3F Paddy field Soil

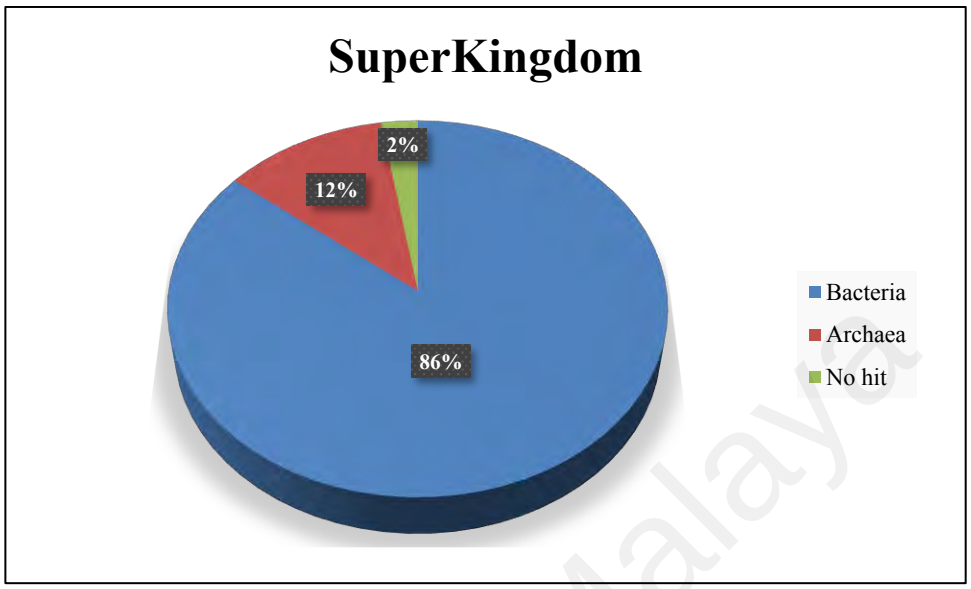

Figure 4.2: Microbial diversity of F3F paddy field soil at superkingdom level. Percentage of relative abundance to domain Bacteria (86%) and Archaea (12%). The other 2% was assigned under No hit (APPENDIX A).

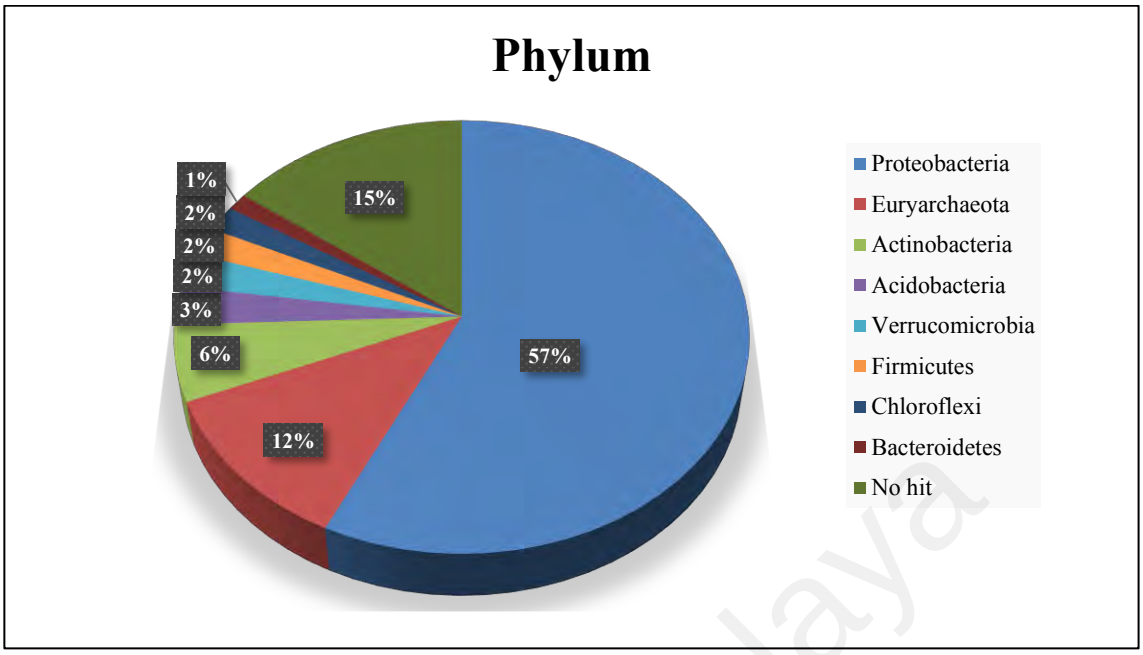

Figure 4.3: Microbial diversity of F3F paddy field soil at phylum level

A total of 26 phyla have been identified. The top 8 most dominant phyla in the soil are the Proteobacteria (57%), Euryarchaeota (12%), Actinobacteria (6%), Acidobacteria (3%), Verrucomicrobia (2%), Firmicutes (2%), Chloroflexi (2%) and Bacteroidetes (1%). The other remaining 15% was categorized under No hit (Refer to APPENDIX B).

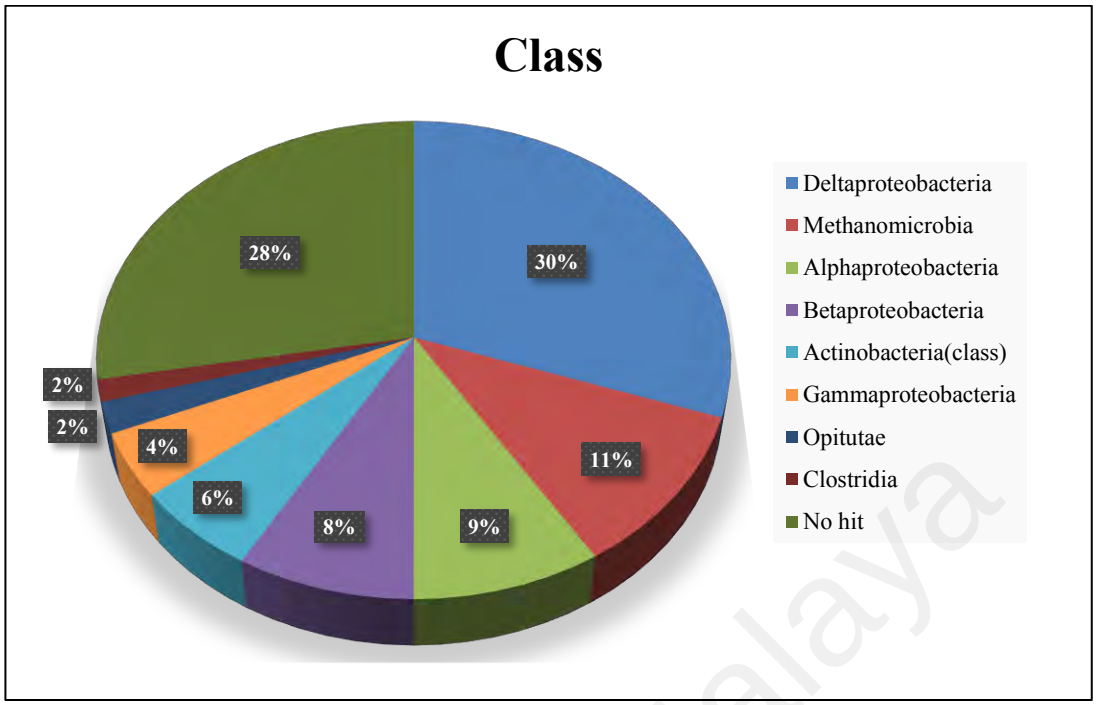

Figure 4.4: Microbial diversity of F3F paddy field soil at class level

As for the class-level rank, 41 class-level groups of microbes were found. The 8 most dominant Class were the Deltaproteobacteria (30%), Methanomicrobia (11%), Alphaproteobacteria (9%), Betaproteobacteria (8%), Actinobacteria (6%), Gammaproteobacteria (4%), Opitutae (2%) and Clostridia (2%). The other remaining 28% was categorized under No hit (Refer to APPENDIX C).

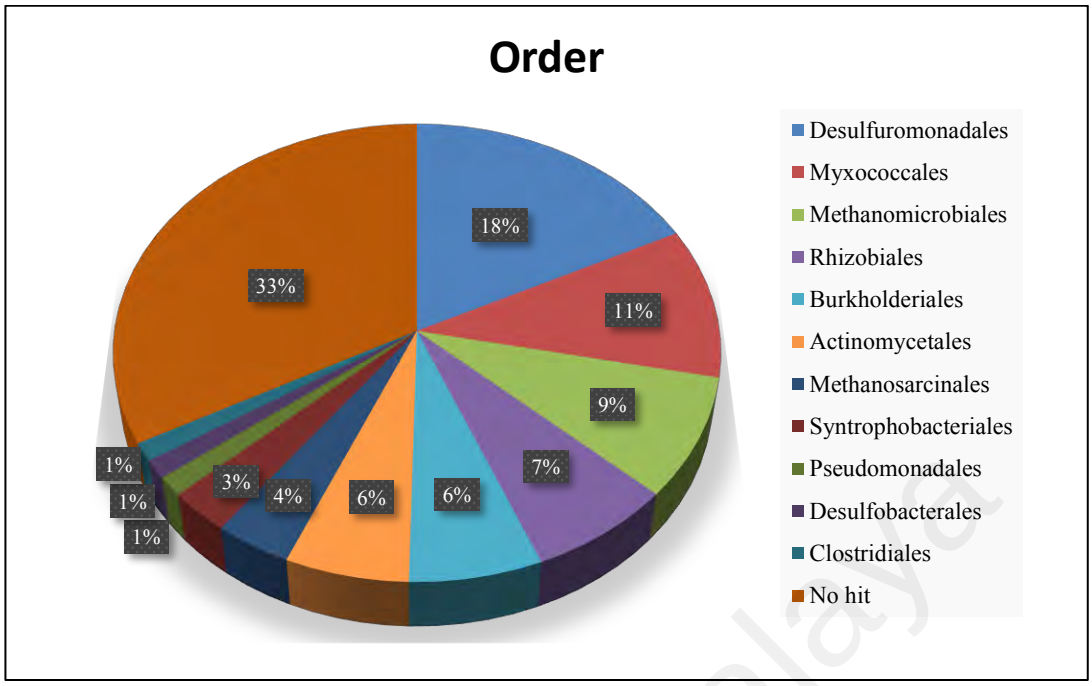

Figure 4.5: Microbial diversity of paddy field soil at order level

Meanwhile, in order level, a total of 87 Order were assigned in which 11 of them were found to be the most dominant among all. They are Desulfuromonadales (18%), Myxococcales (11%), Methanomicrobiales (9%), Rhizobiales (7%), Burkholderiales (6%), Actinomycetales (6%), Methanosarcinales (4%), Syntrophobacteriales (3%), Pseudomonadales (1%), Desulfobacterales (1%) and Clostridiales (1%). The other remaining 33% was categorized under No hit (Refer to APPENDIX D).

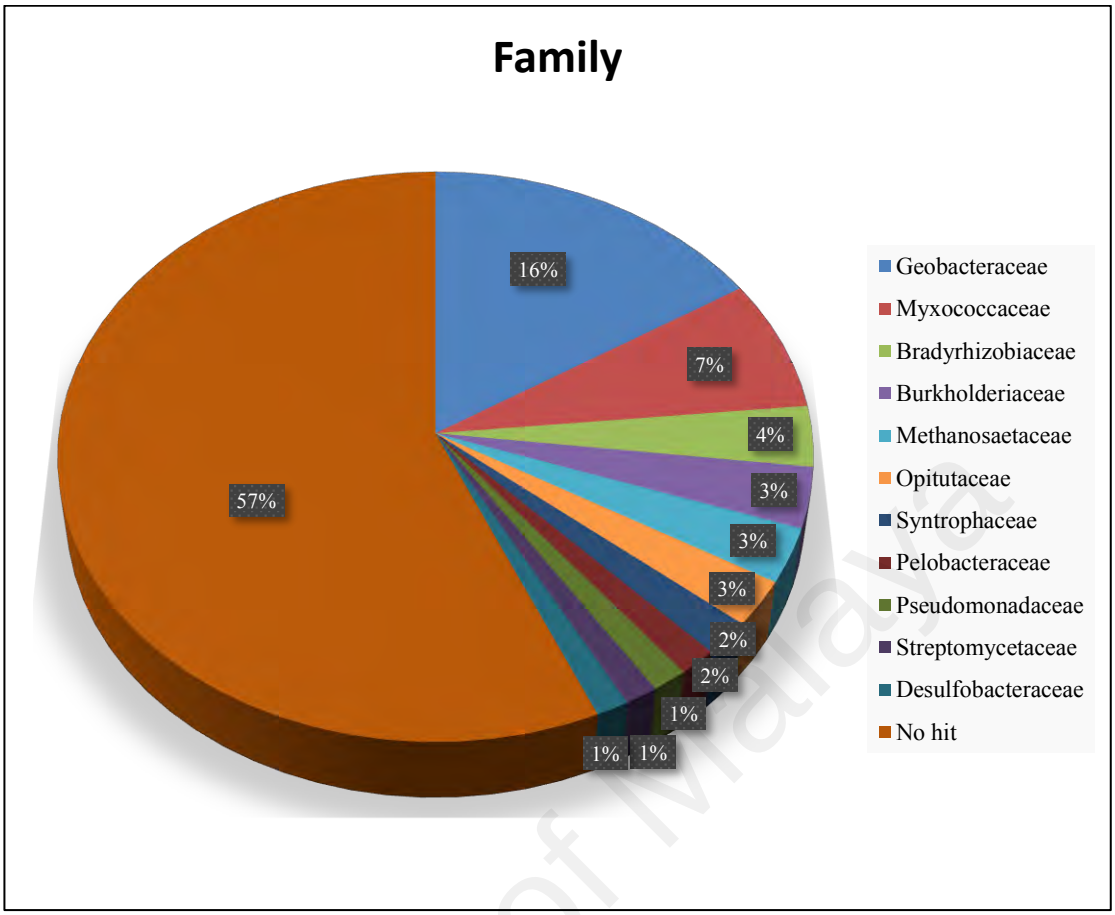

Figure 4.6: Microbial diversity of F3F paddy field soil at family level

In the family level, a total of 177 Family of microbes can be found in the soil. Out of the 177 family, 11 family were the most abundant in the soil. They are Geobacteraceae (16%), Myxococcaceae (7%), Bradyrhizobiaceae (4%), Burkholderiaceae (3%), Methanosaetaceae (3%), Opitutaceae (3%), Syntrophaceae (2%), Pelobacteraceae (2%), Pseudomonadaceae (1%), Streptomycetaceae (1%) and Desulfobacteraceae (1%). The other remaining 57% was categorized under No hit (Refer to APPENDIX E).

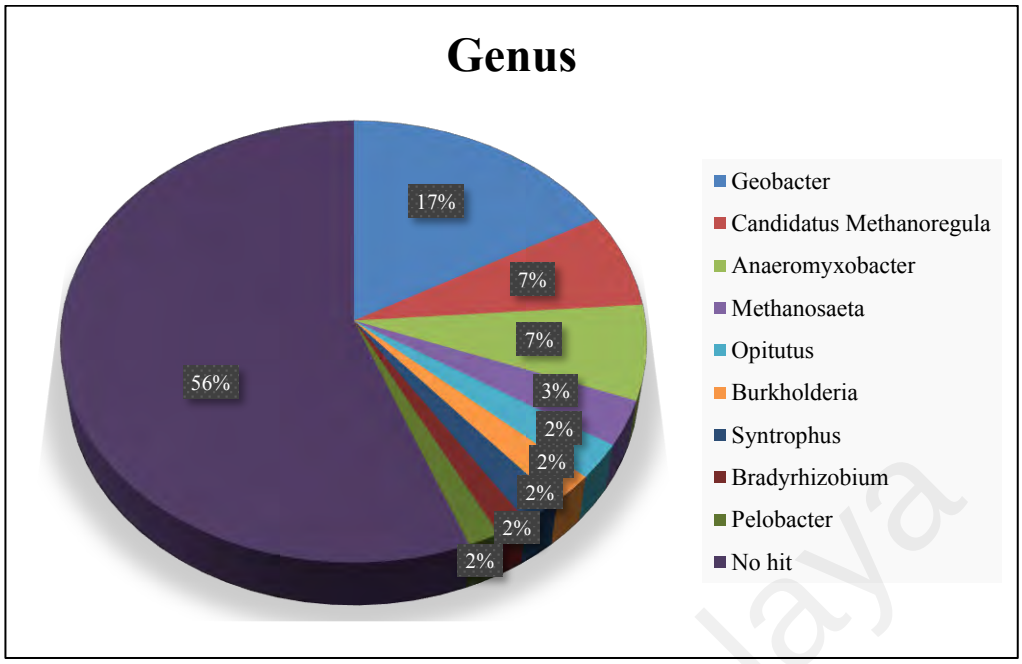

Figure 4.7: Microbial diversity of F3F paddy field soil at genus level

In the other hand, a total of 388 microbial genera can be found from the analysis. Out of these, 9 microbial genera were found to be most dominant in the soil. They are *Geobacter* (17%), *Candidatus methanoregula* (7%), *Anaeromyxobacter* (7%), *Methanosaeta* (3%), *Opitutus* (2%), *Burkholderia* (2%), *Syntrophus* (2%), *Bradyrhizobium* (2%) and *Pelobacter* (2%). The other remaining 56% was categorized under No hit. (Refer to APPENDIX F).

4.3.2 Genome Annotation of F3F Paddy field Soil

The list of genes that were annotated using Prokka software were compiled in a table in APPENDIX G. Table 4.3 shows the summary of genes that were involved in the nutrient cycling.

Table 4.3: Summary of genes involved in nutrient cycling

Gene	Product
nrgA	Ammonium transporter NrgA
cbiM	Cobalt transport protein CbiM
cbiN	Cobalt transport protein CbiN
cbiQ	Cobalt transport protein CbiQ
nirQ	Denitrification regulatory protein NirQ
nirT	Denitrification system component NirT
nfuA	Fe/S biogenesis protein NfuA
nifB	FeMo cofactor biosynthesis protein NifB
fur	Ferric uptake regulation protein
nikMN	Fused nickel transport protein NikMN
ntcA	Global nitrogen regulator
hoxN	High-affinity nickel transport protein
iscA	Iron-binding protein IscA
cooF	Iron-sulfur protein
corC	Magnesium and cobalt efflux protein CorC
corA	Magnesium transport protein CorA
mgtE	Magnesium transporter MgtE
mntB	Manganese transport system membrane protein MntB
modB	Molybdenum transport system permease protein ModB
mopA	Molybdenum-pterin-binding protein MopA
mopII	Molybdenum-pterin-binding protein 2
cnrA	Nickel and cobalt resistance protein CnrA
nikO	Nickel import ATP-binding protein NikO
nikQ	Nickel transport protein NikQ
nikB	Nickel transport system permease protein NikB
nifU	NifU-like protein

Table 4.3, continued

Gene	Product
iscU	NifU-like protein
napA	Nitrate reductase
narB	Nitrate reductase
nasA	Nitrate reductase
narJ	Nitrate reductase molybdenum cofactor assembly chaperone NarJ
narX	Nitrate/nitrite sensor protein NarX
nasA	Nitrate transporter
narK	Nitrate/nitrite transporter NarK
narT	Putative nitrate transporter NarT
ntrC	Nitrogen assimilation regulatory protein
vnfA	Nitrogen fixation protein VnfA
fixK	Nitrogen fixation regulation protein FixK
nifL1	Nitrogen fixation regulatory protein
glnG	Nitrogen regulation protein NR(I)
glnL	Nitrogen regulation protein NR(II)
ptsN	Nitrogen regulatory protein
glnB	Nitrogen regulatory protein P-II
glnK	Nitrogen regulatory protein P-II 2
nifH	Nitrogenase iron protein
nifH1	Nitrogenase iron protein 1
nifD	Nitrogenase molybdenum-iron protein alpha chain
nifK	Nitrogenase molybdenum -iron protein beta chain
nifK1	Nitrogenase molybdenum -iron protein beta chain
vnfD	Nitrogenase vanadium-iron protein alpha chain
vnfK	Nitrogenase vanadium-iron protein beta chain
nifW	Nitrogenase-stabilizing/ protective protein NifW
ntpJ	Potassium/ sodium uptake protein NtpJ
ctpV	Putative copper-exporting P-type ATPase V
nikR	Putative nickel-responsive regulator
narT	Putative nitrate transporter NarT

Table 4.3, continued

Gene	Product
ntpJ	Potassium/ sodium uptake protein NtpJ
ctpV	Putative copper-exporting P-type ATPase V
nikR	Putative nickel-responsive regulator
narT	Putative nitrate transporter NarT
yydH	Putative peptide zinc metalloprotease protein YydH
rip3	Putative zinc metalloprotease Rip3
albF	Putative zinc protease AlbF
narG	Respiratory nitrate reductase 1 alpha chain
narY	Respiratory nitrate reductase 2 beta chain
narI	Respiratory nitrate reductase 1 gamma chain
narV	Respiratory nitrate reductase 2 gamma chain
cysO	sulfur carrier protein CysO
zitB	Zinc transporter ZitB
zur	Zinc uptake regulation protein

CHAPTER 5: DISCUSSION

5.1 Microbial Diversity and Its Benefit To The Ecosystem

The taxonomic composition of microbial in paddy field soil varies from soil to soil and are dependent on the environment (Rokunuzzaman et al., 2016). Among the factors that influence the microbial structure in the soil ecosystem are pH (Tripathi et al., 2012), interference (Blasiak et al., 2014), physicochemical properties of the soil (Andrew et al., 2012; Liliensiek et al., 2012; Mitchell et al., 2010) and vegetation types (Mitchell et al., 2010). Paddy field soil microbes plays vital roles in improving soil quality, root forms as well as rice growth (Liesack et al., 2000).

From the results, bacterial domain occupied the most in comparison to archaeal domain in the F3F soil metagenome (Figure 4.2). The No hit category present in all the taxonomic distribution (Figure 4.2 to Figure 4.7) was due to the limited range of 16S microbes in the 16S microbial database used in the BLAST. The increasing relative abundance under the No hit category over the taxonomic classification could also suggest that the paddy field soil might contain a substantial amount of unidentified microbes that are yet to be discovered.

In a study on the microbial communities in the paddy field soil in Korea by Ahn et al. (2012), they reported the most predominant bacteria phyla (in orderly manner) found was Chloroflexi, Proteobacteria, Actinobacteria, Acidobacteria, Firmicutes and Bacteroidetes while in archaeal communities, the predominant phyla was Thaumarchaeota and Euryarchaeota. In another separate study on the microbial communities composition in paddy field soil irrigated with acid mine waste water in China by Sun et al. (2015), they reported the most predominant phyla in orderly manner was found to be Proteobacteria, Chloroflexi, Acidobacteria, Nitrospirae, Bacteroidetes,

Verrucomicrobia and Planctomycetes. Hence, suggested that the factors that affects the microbial community structure in the soil were pH, nitrate and sulfate content (Sun et al., 2015). In this study, the most predominant phyla (in orderly manner) was found to be Proteobacteria, Euryarchaeota, Actinobacteria, Acidobacteria, Verrucomicrobia, Firmicutes, Chloroflexi and Bacteroidetes (Figure 4.3).

The phylum Proteobacteria occupied the highest portion of the bacteria sequences in the F3F paddy field soil sample. Proteobacterial taxa can be divided into two functional groups based on their known physiology: chemoorganotrophs and chemolithotrophs (Ahn et al., 2012). Chemoorganotrophs utilizes fermentation products such as fatty acids, alcohols or methane while chemolithotrophs utilizes reduced inorganic compounds such as ammonia, sulfur or iron (II) for energy sources (Ahn et al., 2012). These bacterial groups are supposed to be active in zones where both external electron acceptors and with either organic or inorganic electron donors are present because they utilizes oxygen, nitrite, nitrate, sulfate or iron (III) as an electron acceptor and are generally not fermentative (Ahn et al., 2012).

The most predominant genus in this study was *Geobacter* (Figure 4.7). *Geobacter* communities are commonly known for their ability to reduce metals as they are often found most abundant in subsurface where Fe (III) is abundant and available for microorganisms (Lovley et al., 2004). Thus, suggested that *Geobacter* may plays a significant role in iron reduction in paddy soils (Sun et al., 2015). *Anaeromyxobacter* populations are another commonly found microorganism in paddy field soils (Sun et al. 2015). A member of *Anaeromyxobacter* genera known as *Anaeromyxobacter dehalogenans* is able to utilize a wide range of electron acceptors, including halogenated phenols, soluble and insoluble Fe(III), nitrate, oxygen, fumarate, uranium(VI),

selenium(IV) and arsenate (He & Sanford 2003; He & Yao 2010; Kudo et al. 2013; Sanford et al., 2002; Wu et al., 2006). On the other hand, the genus *Syntrophus* is a member of Syntrophobacteriales (3%) at order level and Syntrophaceae (2%) at family level (Figure 4.5 and 4.6). Members of *Syntrophus* genera are capable of of degrading aromatic compounds in syntrophic association with hydrogen-using microorganisms, metabolizes benzoate in coculture with hydrogen-using microorganisms and crotonate in pure culture, syntrophically metabolizes benzoate, gentisate, and 3-phenylpropionate, metabolize a variety of fatty acids in syntrophic association with hydrogen users (Mcinerney et al., 2008). The genus *Pelobacter* belongs to the same order as the genus *Geobacter* but is a member of Pelobacteraceae (2%) family (Figures 4.5 and 4.6). *Pelobacter* spp. which are predominantly found in sediments and sludge were also known to be dissimilatory iron(III) reducers and alcohol oxidizer (Lovley et al., 2004; Mcinerney et al., 2008).

The phylum Verrucomicrobia was mostly dominated by *Opitutus* (2%) at genus level and it belongs to the class and family of Opitutae (2%) and Opitutaceae (3%) respectively (Figures 4.4 and 4.6). The relative abundance of *Verrucomicrobia* and *Bacteroidia* increased during reproductive plant growth may be due to their ecophysiology, which is playing a role in carbon degradation (Breidenbach & Conrad, 2015; Rui et al., 2009; Sugano et al., 2005; Tanahashi et al., 2005). However, *Verrucomicrobia* were less detected in the uncultivated fields in comparison with rice and maize fields (Breidenbach & Conrad, 2015). *Bacteroidia* play major role in decomposition of rice plant residue (Breidenbach & Conrad, 2015). Chin et al. (1999,2001) reported that *Opitutus terrae,* a member of *Verrucomicrobia*, isolated from a paddy field were potential polysaccharolytic and saccharolytic and capable of hydrogen production (Chin et al., 1999, 2001).

Meanwhile, the genus *Burkholderia* belongs to the class, order and family of Betaproteobacteria (8%), Burkholderiales (6%) and Burkholderiaceae (3%) respectively (Figures 4.4, 4.5 and 4.6). Most *Burkholderia* species were notably known as nitrogen-fixing soil bacteria and displayed non-pathogenic interaction with plants but it promotes plant growth (Coenye & Vandamme, 2003; Souza et al., 2012; Suárez-moreno et al., 2012; Woli et al., 2017). The fact that *Burkholderia* species have the potential to promote plant growth is important for its use in agriculture (Souza et al., 2012; Woli et al., 2017). Studies have shown that soil innoculated with *Burkholderia brasiliense* and *Burkholderia vietnamiensis* resulted in 42-64 % significant increase in the growth of rice plants (Baldani et al., 2000). Further more, endophytes of *Burkholderia* presence in the Brazillian rice roots, stem and leaves were able to fix 31% of nitrogen for the rice plants (Baldani et al., 2000). The presence of *Burkholderia* in the paddy field soil may indicate that the soil is fertile as it acts a nitrogen-fixing bacteria and hence, promote the growth of rice plants.

Finally, the genus *Bradyrhizobium* belongs to the class, order and family of Alphaproteobacteria (9%), Rhizobiales (7%) and Bradyrhizobiaceae (4%) respectively (Figure 4.4, 4.5 and 4.6). Photosynthetic *Bradyrhizobium* has been reported as a natural endophyte of wild rice plants (Chaintreuil et al., 2000). Chaintreuil et al. (2000) reported that inoculation of paddy field soils with *Bradyrhizobium* sp. ORS278 were able to increase shoot growth and grain yield of rice thus showed its potential ability to increase rice production (Chaintreuil et al., 2000). Endophytic rhizobia such as bacteria of the genera *Rhizobium, Bradyrhizobium, Sinorhizobium, Allorhizobium, Mesorhizobium, and Azorhizobium*, are significantly important in both the environment and agriculture because they are responsible for most of the atmospheric nitrogen fixed on earth through

their symbioses with legumes and thus, may also have great potential to improve rice production (Ikeda et al., 2014).

The second most abundant phylum was Eurycharcheota (12%) which belongs to archaeal communities (Figure 4.3). The phylum Eurycharcheota was dominated by *Candidatus methanoregula* (7%) and *Methanosaeta* (3%) at genus level (Figure 4.7). The genus *Candidatus methanoregula* belongs to Methanomicrobia (11%) at class level which is the second most abundant class in the soil sample (Figure 4.4) and Methanomicrobiales (9%) at order level which is the third most abundant order in the soil sample (Figure 4.5). The genus *Methanosaeta* comes from the same class as *Candidatus methanoregula* but it belongs to the order and family of Methanosarcinales (4%) and Methanosaetaceae (3%) respectively (Figures 4.5 and 4.6). The genus *Methanosaeta* is an acetotrophic methanogen. Satomi et al. (2006) suggested that Methanosaeta, a filamentous acetate-utilizing methanogens isolated is universal in paddy fields soils and it plays major role in methane production from acetate (Izukami et al., 2006). The emission of methane gas, a greenhouse gas, to the atmosphere is rather harmful to the environment. The presence of this methanogens in the soil indicate that they are involved in the emission this greenhouse gas to the environment.

There are many microbial biofertilizers can be found in the market. They were invented and effectively used in agriculture (Bhardwaj et al., 2014). Subtituting harmful agrochemicals such as chemical fertilizers and pesticides with natural-occuring beneficial microbes could ameliorate the plant nutrition and grant protection against biotic (pathogens and pests) and abiotic (climatic change and pollution) stresses (Lakshmanan et al., 2014). There were many documentation on microbes segregration and application

as major replacement for agrochemicals for plant protection (Bhattacharyya & Jha, 2012; Doornbos & Loon, 2012; Lakshmanan et al., 2014).

5.2 Functional Gene Mining

A total of 413,191 genes were annoted using Prokka pipeline. Out of the 413,191 genes, 824 genes were found to be associated with nutrient cycling.

5.2.1 Nutrient Utilization in Bacteria

Nitrogen is a major components of the chlorophyll and chlorophyll is needed for photosynthesis process in plants. In agriculture sector, nitrogen is crucial in promoting plant growth and improve grain yields. Unfortunately, plant do not self-synthesize nitrogen (Masclaux-Daubresse et al., 2010) and required nitrogen supply from other sources including microbes from soil. As a result, the symbiosis relationship between plant and soil microbes has been extensively studied over the past decades. The word 'symbiosis' relatively means the mutual benefits gained between both party involved and the most popular plant-microbe symbiosis relationship are the nitrogen fixation. The nitrogen-fixing bacteria will fix nitrogen from the environment and the plants will gain access to the nitrogen fixed while the nitrogen-fixing bacteria gain access to the plant-derived carbon sources (Desbrosses & Stougaard, 2011). The genes *nirQ, nirT, ntcA, napA, narB, nasA, narJ, narX, narK, narT, ntrC, vnfA, fixK, nifL1, glnG, glnL, ptsN, glnB, glnK, nifH, nifH1, nifD, nifK, nifK1, vnfD,vnfK, nifW, narG, narY, narI and narV* were found to be involved in nitrogen cycling in the soil. The enzymes engaged in denitrification process are nitrate and nitrite were encoded by *nar/nap* and *nir* genes respectively (Olivares et al., 2013; van Spanning et al., 2007). A crucial nitrogen supply signalling component in both Gram-negative and Gram-positive bacteria is demonstrated by the *Escherichia coli* PII protein, encoded by *glnB* (Jack et al., 2001). This small

uridylated and non-uridylated forms of PII protein will communicate with other regulatory proteins that governs either glutamine synthetase activity or phosphorylation state of NtrC before transmitting the information concerning the nitrogen level of the cell to these interaction partners (Merrick & Edwards, 1995). Besides that, this protein were able to uridylate on a tyrosine residue under nitrogen limitation condition (Detsch & Stülke, 2003; Ninfa & Atkinson, 2000). PII type signal transduction protein GlnB and GlnK which control the activities of the membrane transport proteins and a transcription factor regulates nitrogen metabolism (Jack et al., 2001).

Ammonium is known to be one of the major and most prefered source of nitrogen for bacteria (Detsch & Stülke, 2003). One of gene that specializes in ammonium transfer is nrgA which can be found in the paddy field soil. NrgA is a membrane protein which is required for the transport and utilization of ammonium at low concentrations (Detsch & Stülke, 2003; Wray et al., 1994). Ardin et al. (2014) reported *nrgA* and *glnB* are co-transcribed as a single operon and *glnB* gene was found to be the regulator of the *nrgA* gene. These ammonium transporters are responsible for the movement of ammonium ions across the cell membrane which is fundamental for nitrogen metabolism in bacteria (Ardin et al., 2014; Khademi et al., 2004; Mcdonald & Dietrich, 2011).

Sulfur, similar to carbon and nitrogen, plays important role in plant metabolism and plant development. Sulfur carrier protein is encoded by *cysO* gene can be found in the metagenomic library (Table 4.3). Cysteine (Cys) is a building block of amino acids in proteins and act as the sulfur precursor of a myriad of sulfur compounds that plays pivotal role in cell metabolism and survival (Droux, 2004). Cys is the first organic-reduced sulfur compound and act as a sulfur donor for GSH synthesis,which is a molecule involved in

part of the redox control in the cell and later used as a reservoir and transporter for sulfur in plants (Droux, 2004; Foyer et al., 2001; Noctor & Gomez, 2002).

The whole mechanism of nickel and cobalt uptake in most bacteria and archaea is still unclear although Ni- and Co- containing enzymes are crucial to methanogens for energy metabolism and anabolism (Rodionov et al., 2006). Rodionov et al. (2006) reported that the most widespread uptake system of the two metal are the CbiMNQO-type transporters and proposed that "Cbi" and "Nik" were designated for systems related to cobalt and nickel homeostasis respectively. The *cbiM, cbiQ* and *cbiO* genes were foumd to be close to bacterial urease genes and shown to be of great significance in urease activity in cells grown under the scarcity of nickel (Chen & Burne, 2003; Gilmour et al., 2001). Regardless of the importance of nickel and cobalt for bacterial metabolism, their uptake must be regulated to circumvent toxic effect to the bacteria. For instance, in *Escherichia coli,* excess nickel is circumvented by the NikR repressor which will binds to the promoter region of nikABCDE operon in the presence of nickel (Pina et al., 1999; Rodionov et al., 2006). The CbiMNQO transporter systems of both *Salmonella enterica* serovar Typhimurium and *Rhodobacter capsulatus* in *E. coli* shows high preference for cobalt ions while the Nik(MN)QO system of *R.capsulatus* in which the M and N components are fused together to form single protein has high-affinity towards nickel (Rodionov et al., 2006). The *hoxN* gene is found present in each nickel-containing hydrogenase gene cluster and HoxN transporter from *Alcaligenes eutrophus* is the most widely studied among other nickel transporter (Komeda et al., 1997). In *A. eutrophus*, the activity of the HoxN-mediated transport system was surpressed by cobalt ion (Komeda et al., 1997).

Molybdenum (Mo) is one of the trace elements found in the soil and it can exist in several oxidation states ranging from zero to VI, in which VI is the most commonly found in the soil (Kaiser et al., 2005). Besides that, Mo also serves as a cofactor for a number of enzymes involved in the metabolism of carbon, nitrogen and sulfur (Delgado et al., 2006). Molybdate (a more stable form of Mo) uptake system has been characterized in *E.coli*. in which the integration of molybdate into the cell was mediated by a high-affinity ABCtype transport system encoded by the *modABC* genes in which the *modB* gene present in the paddy field soil is the transmembrane component of the permease (Self et al., 2001).

Manganese (Mn) is essential for growth and survival of most living organisms. In bacteria, Mn plays a pivotal role in bacterial signal transduction. The *mntB* gene that was found in the paddy field soil is found to be a transmembrane protein (Jakubovics & Jenkinson, 2017).

In both Bacteria and Archaea, CorA appeared to be ubiquitous. CorA is absent in few species with the smallest genomes among many microbial genomes available where MgtE appears to take its place (Kehres et al., 1998) . CorA's vast distribution indicates that it is the primary Mg^{2+} transporter in this two Kingdoms and hence, CorA is Earth's most abundant Mg^{2+} transporter. Mg^{2+} transporters are unique members of known classes of transport protein (Grubbs & Maguire, 1987; Kehres et al., 1998).

Prokaryotes particularly bacteria will regulate their iron metabolism depending on the iron availability (Andrews et al., 2003). In *E. coli*, for instance, the regulation is mediated by the ferric-uptake regulator protein known as Fur (Hantke, 2001b). Fur protein will regulate the concentration of intracellular free iron in the iron storage system

(Hantke, 2001a). The iron acquisition and consumption will be imbalance such that the free iron levels become excessive due to the absence of Fur (Andrews et al., 2003).

5.3 Future Work

The current work is based on the soil samples collected during fallow period. Therefore, further studies should be done by collecting different soil samples from different plant growth stages and analysed using this approach, to have a comparative data to track the changes in microbial diversity in the soil.

Besides that, the culture-dependent method should also be done to do a comparative study with culture-independent metagenomics. Transcriptomic studies on the paddy field soil can also be done to study gene expression of valuable bacteria.

Future study on the functional analysis of the microbial communities in the paddy field soil obtained from Bario Highland is needed to have better understanding of their ecological roles and their relationship with the host. The study on the genes found from this preliminary study will also be conducted. The future trend of soil metagenomic is metaphenomic study (Jansson & Hofmockel, 2018). It is a study on the expressed product encoded in the metagenome and environment.

CHAPTER 6: CONCLUSION

This study revealed high abundance of crucially important microorganisms in Bario paddy field soil that plays major roles in nutrient cycling. A total of 26 phylas, 41 microbial classes, 87 microbial orders, 177 microbial familiae and 388 microbial genera were found in the paddy field soils. Each microbial genera that were present in the soil sample were major key players in both nitrogen fixation and carbon cycling, iron reducing, act as plant growth promoter and many other benefits to its ecosystem as reported by previous study.

There were a total of 413,191 genes annotated using Prokka pipeline from the paddy field soils. However, only 824 genes found were associated with nutrient cycling. The remaining genes found in the paddy field soil may have significant roles in other field for instance in the biotechnology, pharmaceutical and etc.

This is the first report on the microbial diversity of paddy field soil of Bario, the Kelabit highland of Sarawak using NGS metagenomic approach. Hence, this preliminary study will paved the way for further study on the microbial diversity and genes found in the soil.

REFERENCES

Ahn, J., Song, J., Kim, B., Kim, M., Joa, J., & Weon, H. (2012). Characterization of the bacterial and archaeal communities in rice field soils subjected to long-term fertilization practices. *The Journal of Microbiology*, *50*(5), 754-765.

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *International Society for Microbial Ecology*, *3*(2), 243-251.

Andrew, D. R., Fitak, R. R., Munguia-vega, A., Racolta, A., Martinson, V. G., & Dontsova, K. (2012). Abiotic factors shape microbial diversity in Sonoran desert soils. *Applied and Environmental Microbiology*, *78*(21), 7527-7537.

Andrews, S., Robinson, A. K., & Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiology Reviews*, *27*(2-3), 215-237.

Ardin, A. C., Fujita, K., Nagayama, K., Takashima, Y., Nomura, R., Nakano, K., … Matsumoto-nakano, M. (2014). Identification and functional analysis of an ammonium transporter in *Streptococcus* mutans. *PLoS ONE*, *9*(9), 1-10.

Arjun, J. K., & Harikrishnan, K. (2011). Metagenomic analysis of bacterial diversity in the rice rhizosphere soil microbiome. *Journal of Biotechnology, Bioinformatics and Bioengineering*, *1*(3), 361-367.

Arp, D. J., & Stein, L. Y. (2003). Metabolism of inorganic N compounds by ammonia-oxidizing bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, *38*(6), 471-495.

Aslam, Z., Yasir, M., Yoon, H. S., Jeon, C. O., & Chung, Y. R. (2013). Diversity of the bacterial community in the rice rhizosphere managed under conventional and no-tillage practices. *Journal of Microbiology*, *51*(6), 747-756.

Aubrey, S. (2014, January 16). NKEA first harvest of Bario rice on sale soon. *Borneo post online*, Retrieved from http://www.theborneopost.com/2014/01/16/nkea-first-harvest-of-bario-rice-on-sale-soon/.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., … Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*(75), 1-15.

Berendsen, R. L., Pieterse, C. M. J., & Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends in Plant Science*, *17*(8), 478-486.

Berg, G., & Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiology Ecology*, *68*(1), 1-13.

Bhardwaj, D., Ansari, M. W., Sahoo, R. K., & Tuteja, N. (2014). Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. *BMC Microbial Cell Factories*, *13*(66), 1-10.

Bhattacharyya, P. N., & Jha, D. K. (2012). Plant growth-promoting rhizobacteria (PGPR): Emergence in agriculture. *World Journal of Microbiology and Biotechnology*, *28*(4), 1327-1350.

Blasiak, L. C., Schmidt, A. W., Andriamiarinoro, H., Mulaw, T., Rasolomampianina, R., Applequist, W. L., … Hill, R. T. (2014). Bacterial communities in Malagasy soils with differing levels of disturbance affecting botanical diversity. *PLoS ONE*, *9*(1), 1-11.

Breidenbach, B., & Conrad, R. (2015). Seasonal dynamics of bacterial and archaeal methanogenic communities in flooded rice fields and effect of drainage. *Frontiers in Microbiology*, *5*, 1-16.

Brochier-Armanet, C., Gribaldo, S., & Forterre, P. (2012). Spotlight on the Thaumarchaeota. *International Society for Microbial Ecology*, *6*(2), 227-230.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., … Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain bacteria. *Nature*, *523*(7559), 208-225.

Buée, M., de Boer, W., Martin, F., van Overbeek, L., & Jurkevitch, E. (2009). The rhizosphere zoo: An overview of plant-associated communities of microorganisms, including phages, bacteria, archaea, and fungi, and of some of their structuring factors. *Plant and Soil*, *321*(1–2), 189-212.

Chaintreuil, C., Giraud, E., Prin, Y., Lorquin, J., Bâ, A., Gillis, M., … Ba, A. (2000). Photosynthetic Bradyrhizobia are natural endophytes of the African wild rice *Oryza breviligulata*. *Applied and Environmental Microbiology*, *66*(12), 5437-5447.

Chaudhuri, R. R., Loman, N. J., Snyder, L. A. S., Bailey, C. M., Stekel, D. J., & Pallen, M. J. (2008). xBASE2: A comprehensive resource for comparative bacterial genomics. *Nucleic Acids Research*, *36*, D543-D546.

Chen, Y. M., & Burne, R. A. (2003). Identification and characterization of the nickel uptake system for urease biogenesis in *Streptococcus salivarius* 57.I, *Journal of Bacteriology, 185*(23), 6773-6779.

Chin, K., Hahn, D., Hengstmann, U. L. F., & Liesack, W. (1999). Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. *Applied and Environmental Microbiology, 65*(11), 5042-5049.

Chin, K., Liesack, W., & Janssen, P. H. (2001). *Opitutus terrae* gen. nov., sp. nov., to accommodate novel strains of the division ' Verrucomicrobia ' isolated from rice paddy soil. *International Journal of Systematic and Evolutionary Microbiology, 51*(Pt 6), 1965-1968.

Coenye, T., & Vandamme, P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environmental Microbiology*, *5*(9), 719-729.

Conrad, R., Erkel, C., & Liesack, W. (2006). Rice Cluster I methanogens, an important group of Archaea producing greenhouse gas in soil. *Current Opinion in Biotechnology*, *17*(3), 262-267.

Curtis, T. P., Sloan, W. T., & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences*, *99*(16), 10494-10499.

D'Costa, V. M., McGrann, K. M., Hughes, D. W., & Wright, G. D. (2006). Sampling the antibiotic resistome. *Science*, *311*(5759), 374-378.

Dakora, F. D. (2003). Defining new roles for plant and rhizobial molecules in sole and mixed plant cultures involving symbiotic legumes. *New Phytologist*, *158*(1), 39-49.

Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance, *Microbiology and Molecular Biology Reviews, 74*(3), 417-433.

Delgado, M. J., Tresierra-Ayala, A., Talbi, C., & Bedmar, E. J. (2006). Functional characterization of the *Bradyrhizobium japonicum modA* and *modB* genes involved in molybdenum transport. *Microbiology*, *152*(Pt 1), 199-207.

Dennis, P. G., Miller, A. J., & Hirsch, P. R. (2010). Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiology Ecology*, *72*(3), 313-327.

Desbrosses, G. J., & Stougaard, J. (2011). Root nodulation : A paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host & Microbe*, *10*(4), 348-358.

Detsch, C., & Stülke, J. (2003). Ammonium utilization in *Bacillus subtilis* : Transport and regulatory functions of *NrgA* and *NrgB*. *Microbiology*, *149*(Pt 11), 3289-3297.

Baldani, V. L. D., Baldani, J. I., & Döbereiner, J. (2000). Inoculation of rice plants with the endophytic diazotrophs *Herbaspirillum seropedicae* and *Burkholderia* spp. *Biology and Fertility of Soils*, *30*(5-6), 485-491.

Donn, S., Kirkegaard, J. A., Perera, G., Richardson, A. E., & Watt, M. (2014). Evolution of bacterial communities in the wheat crop rhizosphere. *Environmental Microbiology*, *17*(3), 610-621.

Doornbos, R. F., & Loon, L. C. V. (2012). Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agronomy for Sustainable Development*, *32*(1), 227-243.

Droux, M. (2004). Sulfur assimilation and the role of sulfur in plant metabolism : A survey. *Photosynthesis Research*, *79*(3), 331-348.

Folman, L. B., Postma, J., & Veen, J. A. V. (2001). Communities at different root locations and plant. *Microbial Ecology, 42*, 586-597.

Foyer, C. H., Theodoulou, F. L., & Delrot, S. (2001). The functions of inter- and intracellular glutathione transport systems in plants. *Trends in Plant Science*, *6*(10), 486-492.

Gibson, M. K., Forsberg, K. J., & Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *International Society for Microbial Ecology*, *9*(1), 207-216.

Gilbert, J. A., & Dupont, C. L. (2011). Microbial metagenomics : Beyond the genome. *Annual Review of Marine Science*, *3*, 347-371.

Gilmour, H. D., Bosse, J. T., & MacInnes, J. I. (2001). Novel genes affecting urease activity in *Actinobacillus pleuropneumoniae*. *Journal of Bacteriology,183*(4), 1242-1247.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, *11*(5), 759-769.

Goethem, M. W. V., Pierneef, R., Bezuidt, O. K. I., Peer, Y. V. D., Cowan, D. A., & Makhalanyane, T. P. (2018). A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome*, *6*(1), 40.

Golyshina, O. V, Yakimov, M. M., Lünsdorf, H., Ferrer, M., Nimtz, M., Timmis, K. N., … Golyshin, P. N. (2009). *Acidiplasma aeolicum* gen. nov., sp. nov., a euryarchaeon of the family Ferroplasmaceae isolated from a hydrothermal pool , and transfer of *Ferroplasma cupricumulans* to *Acidiplasma cupricumulans* comb. nov. *International Journal of Systematic and Evolutionary Microbiology, 59*(Pt 11), 2815-2823.

Grégoire, P., Fardeau, M., Joseph, M., Guasco, S., Hamaide, F., Biasutti, S., … Ollivier, B. (2011). Isolation and characterization of *Thermanaerothrix daxensis* gen. nov., sp. nov., a thermophilic anaerobic bacterium pertaining to the phylum "Chloroflexi", isolated from a deep hot aquifer in the Aquitaine Basin. *Systematic and Applied Microbiology*, *34*(7), 494-497.

Grubbs, R. D., & Maguire, M. E. (1987). Magnesium as a regulatory cation: Criteria and evaluation. *Magnesium*, *6*(3), 113-127.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, *68*(4), 669-685.

Hansel, C. M., Fendorf, S., Jardine, P. M., & Francis, C. A. (2008). Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Applied and Environmental Microbiology*, *74*(5), 1620-1633.

Hantke, K. (2001a). Bacterial zinc transporters and regulators. *BioMetals*, *14*(3-4), 239-249.

Hantke, K. (2001b). Iron and metal regulation in bacteria. *Current Opinion in Microbiology*, *4*, 172-177.

He, H., Li, W., Yu, R., & Ye, Z. (2017). Illumina-based analysis of bulk and rhizosphere soil bacterial communities in paddy fields under mixed heavy. *Pedosphere*, *27*(3), 569-578.

He, Q., & Sanford, R. A. (2003). Characterization of Fe (III) reduction by chlororespiring *Anaeromxyobacter dehalogenans*. *Applied and Environmental Microbiology, 69*(5), 2712-2718.

He, Q., & Yao, K. (2010). Microbial reduction of selenium oxyanions by *Anaeromyxobacter dehalogenans*. *Bioresource Technology*, *101*(10), 3760-3764.

Hjort, K., Presti, I., Elväng, A., Marinelli, F., & Sjöling, S. (2014). Bacterial chitinase with phytopathogen control capacity from suppressive soil revealed by functional metagenomics. *Applied Microbiology and Biotechnology*, *98*(6), 2819-2828.

Howe, A. C., Jansson, J. K., Malfatti, S. A., Tringe, S. G., Tiedge, J. M., & Brown, C. T. (2014). Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences*, *111*(16), 4904-4909.

Huson, D., Mitra, S., & Ruscheweyh, H. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, *21*(9), 1552-1560.

Hyatt, D., Chen, G., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal : Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(119), 1-11.

Ikeda, S., Sasaki, K., Okubo, T., Yamashita, A., Terasawa, K., Bao, Z., … Minamisawa, K. (2014). Low nitrogen fertilization adapts rice root microbiome to low nutrient environment by changing biogeochemical functions. *Microbes and Environments*, *29*(1), 50-59.

Ismail, G., & Din, L. (1998). *Bario: The Kelabit highlands of Sarawak : A scientific journey through borneo*. Malaysia: Pelanduk Publications.

Ismail, Z., Sam, C., Yin, W., & Chan, K. (2017). Tropical mangrove swamp metagenome reveals unusual abundance of ecologically important microbes. *Current Science*, *112*(8), 2-7.

Izukami, S. M., Akeda, K. T., Kada, S. A., & Ujita, T. F. (2006). Isolation and characteristics of Methanosaeta in paddy field soils. *Bioscience, Biotechnology & Biochemistry*, *70*(4), 828-835.

Jack, R., Arcondé, T., & Merrick, M. (2001). P II signal transduction proteins , pivotal players in microbial nitrogen control. *Microbiology and Molecular Biology Reviews*, *65*(1), 80-105.

Jakubovics, N. S., & Jenkinson, H. F. (2017). Out of the iron age : New insights into the critical role of manganese homeostasis in bacteria. *Microbiology*, *147*(Pt 7), 1709-1718.

Jansson, J. K., & Hofmockel, K. S. (2018). The soil microbiome — from metagenomics to metaphenomics. *Current Opinion in Microbiology*, *43*, 162-168.

Jiwan, M., Henrita, S., Petrus, B., Abdullah, N. A. P., Osumanu, H. A., Wan Asrina, W. Y., … Mustapha, M. (2015, September). *Traditional natural farming system in the production of Bario rice (Adan rice) by Lun Bawang community in the highland of borneo, Sarawak, east Malaysia and potential for using biochar, paddy straw and buffalo dung bokashi.* Paper presented at Kuala Lumpur International Agriculture, Forestry and Plantation Conference (KLIAF 2015).

Kaiser, B. N., Gridley, K. L., Brady, J. N., Phillips, T., & Tyerman, S. D. (2005). The role of molybdenum in agricultural plant production. *Annals of Botany*, *96*(5), 745-754.

Kehres, D. G., Lawyer, C. H., & Maguire, M. E. (1998). The *CorA* magnesium transporter gene family. *Microbial & Comparative Genomics*, *3*(3), 151-169.

Kevin, M. T. S., Ahmed, O. H., Asrina, W. Y. W., Rajan, A., & Ahzam, M. (2007). Towards growing bario rice on lowland soils: A preliminary nitrogen and potassium fertilization trial. *American Journal of Agricultural and Biological Science*, *2*(2), 99-105.

Khademi, S., Iii, J. O. C., Remis, J., Robles-colmenares, Y., Miercke, L. J. W., & Stroud, R. M. (2004). Mechanism of ammonia transport by Amt/MEP/Rh: Structure of AmtB at 1.35Å. *Science*, *305*(5690), 1587-1595.

Komeda, H., Kobayashi, M., & Shimizu, S. (1997). A novel transporter involved in cobalt uptake. *Proceedings of National Academy of Sciences of The United States of America*, *94*(1), 36-41.

Konopka, A. (2009). What is microbial community ecology? *International Society for Microbial Ecology*, *3*, 1223-1230.

Kudo, K., Yamaguchi, N., Makino, T., Ohtsuka, T., Kimura, K., Tao Dong, D., & Amachi, S. (2013). Release of arsenic from soil by a novel dissimilatory arsenate-reducing bacterium, *Anaeromyxobacter* sp. strain PSR-1. *Applied Environment Microbiology*, *79*(15), 4635-4642.

Kuntal, B. K., Ghosh, T. S., & Mande, S. S. (2013). Community-Analyzer: A platform for visualizing and comparing microbial community structure across microbiomes. *Genomics*, *102*(4), 409-418.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H., Rognes, T., & Ussery, D. W. (2007). RNAmmer : Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100-3108.

Lakshmanan, V., Selvaraj, G., & Bais, H. P. (2014). Functional soil microbiome : Belowground solutions to an aboveground problem. *Plant Physiology*, *166*(2), 689-700.

Laslett, D., & Canback, B. (2004). ARAGORN , a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*(1), 11-16.

Lee, H. H., Bong, C. F. J., Loh, S. P., Sarbini, S. R., & Yiu, P. H. (2014). Genotypic , grain morphological and locality variation in rice phytate content and phytase activity. *Emirates Journal of Food and Agriculture*, *26*(10), 844-852.

Lee, H. H., Loh, S. P., Bong, C. F. J., Sarbini, S. R., & Yiu, P. H. (2015). Impact of phytic acid on nutrient bioaccessibility and antioxidant properties of dehusked rice. *Journal of Food Science and Technology, 52*(12), 7806-7816.

Liesack, W., Schnell, S., & Revsbech, N. P. (2000). Microbiology of flooded rice paddies. *FEMS Microbiology Reviews*, *24*(5), 625-645.

Liliensiek, A.-K., Thakuria, D., & Clipson, N. (2012). Influences of plant species composition, fertilisation and Lolium perenne ingression on soil microbial community structure in three Irish grasslands. *Microbial Ecology*, *63*(3), 509-521.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., … Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, *2012*, 1-11.

Lombard, N., Prestat, E., van Elsas, J. D., & Simonet, P. (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiology Ecology*, *78*(1), 31-49.

Lopes, A. R., Manaia, C. M., & Nunes, O. C. (2014). Bacterial community variations in an alfalfa-rice rotation system revealed by 16S rRNA gene 454-pyrosequencing. *FEMS Microbiology Ecology*, *87*(3), 650-663.

Lorenz, P., Liebeton, K., Niehaus, F., & Eck, J. (2002). Screening for novel enzymes for biocatalytic processes: Accessing the metagenome as a resource of novel functional sequence space. *Current Opinion in Biotechnology*, *13*(6), 572-577.

Lovley, D. R., Holmes, D. E., & Nevin, K. P. (2004). Dissimilatory Fe(III) and Mn(IV) reduction. *Advances in Microbial Physiology*, *49*, 219-286.

Lu, T., Ke, M., Peijnenburg, W. J. G. M., Zhu, Y., Zhang, M., Sun, L., … Qian, H. (2018). Investigation of rhizospheric microbial communities in wheat, barley, and two rice varieties at the seedling stage. *Journal of Agricultural and Food Chemistry*, *66*(11), 2645-2653.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, *24*(3), 133-141.

Martínez, J. L., Coque, T. M., & Baquero, F. (2015). What is a resistance gene? ranking risk in resistomes. *Nature Reviews Microbiology*, *13*(2), 116-123.

Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., & Suzuki, A. (2010). Nitrogen uptake, assimilation and remobilization in plants: Challenges for sustainable and productive agriculture. *Annals of Botany*, *105*(7), 1141-1157.

Matsumura, Y., Akahira-moriya, A., & Sasaki-mori, M. (2015). Bioremediation of bisphenol-A polluted soil by *Sphingomonas bisphenolicum* AO1 and the microbial community existing in the soil. *Biocontrol Science*, *20*(1), 35-42.

Mcdonald, T. R., & Dietrich, F. S. (2011). Multiple horizontal gene transfers of ammonium transporters/ammonia permeases from prokaryotes to eukaryotes : Toward a new functional and evolutionary classification. *Molecular Biology and Evolution, 29*(1), 51-60.

Mcinerney, M. J., Struchtemeyer, C. G., Sieber, J., Mouttaki, H., Stams, A. J. M., Schink, B., … Gunsalus, R. P. (2008). Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. *Annals of the New York Academy of Sciences*, *1125*, 58-72.

Merrick, M. J., & Edwards, R. A. (1995). Nitrogen control in bacteria. *Microbiological*, *59*(4), 604-622.

Metzker, M. L. (2010). Sequencing technologies- the next generation. *Nature Review Genetics*, *11*(1), 31-46.

Mirete, S., Morgante, V., & González-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*, *38*, 143-149.

Mitchell, R. J., Hester, A. J., Campbell, C. D., Chapman, S. J., Cameron, C. M., Hewison, R. L., & Potts, J. M. (2010). Is vegetation composition or soil chemistry the best predictor of the soil microbial community? *Plant and Soil*, *333*(1-2), 417-430.

Moore, B. S., Nett, M., & Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the *Actinomycetes*. *Natural Product Reports*, *26*(11), 1353-1508.

Myrold, D. D., Zeglin, L. H., & Jansson, J. K. (2013). The potential of metagenomic approaches for understanding soil microbial processes. *Soil Science Society of America Journal*, *78*(1), 3-10.

Narasimhan, K., Basheer, C., Bajic, V. B., & Swarup, S. (2003). Enhancement of plant-microbe interactions using a rhizosphere metabolomics-driven polychlorinated biphenyls. *Plant Physiology*, *132*(1), 146-153.

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933-2935.

Nemergut, D. R., Schmidt, S. K., Fukami, T., Neill, S. P. O., Bilinski, T. M., Stanish, L. F., … Wickey, P. (2013). Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews, 77*(3), 342-356.

Nesme, J., Achouak, W., Agathos, S. N., Bailey, M., Baldrian, P., Brunel, D., … Simonet, P. (2016). Back to the future of soil metagenomics. *Frontiers in Microbiology*, *7*, 73.

Nicholas, D., Hazila, K. K., Chua, H. P. & Rosniyana, A. (2013). Effects of aging on physico-chemical properties , nutritional compositions and cooking characteristics of Bario rice. *Journal of Tropical Agriculture and Food Science, 41*(2), 239-248.

Nielsen, U. N., & Wall, D. H. (2013). The future of soil invertebrate communities in polar regions: Different climate change responses in the Arctic and Antarctic? *Ecology Letters*, *16*(3), 409-419.

Ninfa, A. J., & Atkinson, M. R. (2000). PII signal transduction proteins. *Trends in Microbiology*, *8*(4), 172-179.

Noctor, G., & Gomez, L. (2002). Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signalling. *Journal of Experimental Botany*, *53*(372), 1283-1304.

Norimah, A. K., Safiah, M., Jamal, K., Siti, H., Zuhaida, H., Rohida, S., … Azmi, M. Y. (2008). Food consumption patterns: Findings from the Malaysian adult nutrition survey. *Malaysian Journal of Nutrition*, *14*(1), 25-39.

Ofaim, S., Ofek-lalzar, M., Sela, N., Jinag, J., Kashi, Y., & Charles, T. C. (2017). Analysis of microbial functions in the rhizosphere using a metabolic-network based framework for metagenomics interpretation. *Frontiers in Microbiology*, *8*, 1606.

Ofek-lalzar, M., Sela, N., Goldman-voronov, M., Green, S. J., Hadar, Y., & Minz, D. (2014). Niche and host-associated functional signatures of the root surface microbiome. *Nature Communications*, *5*, 4950.

Olivares, J., Bedmar, E. J., & Sanjuán, J. (2013). Biological nitrogen fixation in the context of global change. *Molecular Plant-Microbe Interactions*, *26*(5), 486-494.

Orgiazzi, A., Dunbar, M. B., Panagos, P., de Groot, G. A., & Lemanceau, P. (2015). Soil biodiversity and DNA barcodes: Opportunities and challenges. *Soil Biology and Biochemistry*, *80*, 244-250.

Owen, S. M., Clark, S., Pompe, M., & Semple, K. T. (2018). Biogenic volatile organic compounds as potential carbon sources for microbial communities in soil from the rhizosphere of *Populus tremula*. *FEMS Microbiology Letters, 268*(1), 34-39.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, *276*(5313), 734-740.

Pester, M., Schleper, C., & Wagner, M. (2011). The Thaumarchaeota : An emerging view of their phylogeny and ecophysiology. *Current Opinion in Microbiology*, *14*(3), 300-306.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785-786.

Philippot, L., Raaijmakers, J. M., Lemanceau, P., & Putten, W. H. van der. (2013). Going back to the roots : The microbial ecology of the rhizosphere. *Nature Reviews Microbiology*, *11*(11), 789-799.

Pina, K. D. E., Desjardin, V., Giordano, G., & Wu, L. (1999). Isolation and characterization of the *nikR* gene encoding a nickel-responsive regulator in *Escherichia coli*. *Journal of Bacteriology*, *181*(2), 670-674.

Podosokorskaya, O. A., Bonch-Osmolovskaya, E. A., Novikov, A. A., Kolganova, T. V, & Kublanov, I. V. (2013). *Ornatilinea apprima* gen. nov., sp. nov., a cellulolytic representative of the class Anaerolineae. *International Journal of Systematic and Evolutionary Microbiology*, *63*(Pt 1), 86-92.

Prosser, J. I. (2015). Dispersing misconceptions and identifying opportunities for the use of "omics" in soil microbial ecology. *Nature Reviews Microbiology*, *13*(7), 439-446.

Rajamoorthy, Y., Rahim, K. A., & Munusamy, S. (2015). Rice industry in Malaysia: Challenges, policies and implications. *Procedia Economics and Finance*, *31*(15), 861-867.

Richardson, E. J., & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*, *14*(1), 1-12.

Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: Genomic analysis of microbial communities. *Annual Review of Genetics*, *38*(1), 525-552.

Rincon-florez, V. A., Carvalhais, L. C., & Schenk, P. M. (2013). Culture-independent molecular tools for soil and rhizosphere microbiology. *Diversity*, *5*(3), 581-612.

Robe, P., Nalin, R., Capellano, C., Vogel, T. M., & Simonet, P. (2003). Extraction of DNA from soil. *European Journal of Soil Biology*, *39*(4), 183-190.

Rodionov, D. A., Hebbeln, P., Gelfand, M. S., & Eitinger, T. (2006). Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters : Evidence for a novel group of ATP-binding cassette transporters. *Journal of Bacteriology*, *188*(1), 317-327.

Rokunuzzaman, M., Ueda, Y., Li, C., Tanaka, S., & Ohnishi, K. (2016). Effects of land use changes from paddy fields on soil bacterial communities in a hilly and mountainous Aarea. *Microbes and Environments*, *31*(2), 160-164.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., … Goodman, R. M. (2000). Cloning the soil metagenome : A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, *66*(6), 2541-2547.

Rui, J., Peng, J., & Lu, Y. (2009). Succession of bacterial populations during plant residue decomposition in rice field soil. *Applied and Environmental Microbiology*, *75*(14), 4879-4886.

Sakai, S., Conrad, R., Liesack, W., & Imachi, H. (2010). *Methanocella arvoryzae* sp. nov., a hydrogenotrophic methanogen isolated from rice field soil. *International Journal of Systematic and Evolutionary Microbiology*, *60*(Pt 12), 2918-2923.

Sakai, S., Imachi, H., Hanada, S., Ohashi, A., Harada, H., & Kamagata, Y. (2008). *Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon , the first isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order Methanocellales ord. nov. *International Journal of Systematic and Evolutionary Microbiology*, *58*(Pt 4), 929-936.

Samuel, H., Jiwan, M., Abdullah, N. A. ., Bulan, P., Wan Yahya, W. A., Ahmed, O. H., … Muhammad, A. R. (2015). Seed quality, physical properties and proximate compositions of adan rice. *International Journal of Agriculture, Forestry and Plantation, 1*, 54-58.

Sanford, R. A., Cole, J. R., & Tiedje, J. M. (2002). Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic *Myxobacterium*. *Applied and Environmental Microbiology*, *68*(2), 893-900.

Schulz, S., Dickschat, J. S., & Schulz, S. (2007). Bacterial volatiles : The smell of small organisms. *Natural Product Reports*, *24*(4), 814-842.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068-2069.

Self, W. T., Grunden, A. M., Hasona, A., & Shanmugam, K. T. (2001). Molybdate transport. *Research in Microbiology*, *152*(3), 311-321.

Sheldon, F. H., Brown, C. E., Rahman, M. A., Tay, G. K., & Moyle, R. G. (2013). Ornithology of the Kelabit highlands of Sarawak, Malaysia. *Raffles Bulletin of Zoology*, *61*(2), 827-853.

Sommer, M. O. A., Dantas, G., & Church, G. M. (2009). Functional characterization of the antibiotic resistance reservoir in the human. *Science*, *325*(5944), 1128-1131.

Souza, A. D., Silva, A., Brandt, E. D., Pinnock, E., Vandamme, P., Maria, F., & Moreira, D. S. (2012). Diazotrophic *Burkholderia* species isolated from the amazon region exhibit phenotypical , functional and genetic diversity. *Systematic and Applied Microbiology*, *35*(4), 253-262.

Souza, R. C., Cantão, M. E., Nogueira, M. A., Vasconcelos, A. T. R., & Hungria, M. (2018). Outstanding impact of soil tillage on the abundance of soil hydrolases revealed by a metagenomic approach. *Brazilian Journal of Microbiology*, doi: 10.1016/j.bjm.2018.03.001.

van Spanning, R. J. M., Richardson, D. J., & Ferguson, S. J. (2007). Chapter 1 - Introduction to the biochemistry and molecular biology of denitrification *biology of the nitrogen cycle* (pp. 3-20). Amsterdam: Elsevier.

Food and Agriculture Organization of the United Nations [FAO] (1995). *Dimensions of need-Staple foods: What do people eat?* Rome: FAO Retrieved from http://www.fao.org/docrep/u8480e/u8480e07.htm

Stewart, A. C., Osborne, B., & Read, T. D. (2009). DIYA: A bacterial annotation pipeline for any genomics lab. *Bioinformatics*, *25*(7), 962-963.

Suárez-moreno, Z. R., Caballero-mellado, J., & Coutinho, B. G. (2012). Common features of environmental and potentially beneficial plant-associated *Burkholderia*. *Microbial Ecology*, *63*(2), 249-266.

Sugano, A., Tsuchimoto, H., Tun, C. C., Asakawa, S., & Kimura, M. (2005). Succession and phylogenetic profile of eubacterial communities in rice straw incorporated into a rice field: Estimation by PCR-DGGE analysis. *Soil Science & Plant Nutrition*, *51*(1), 51-60.

Sun, M., Xiao, T., Ning, Z., & Xiao, E. (2015). Microbial community analysis in rice paddy soils irrigated by acid mine drainage contaminated water. *Applied Microbiology and Biotechnology*, *99*(6), 2911-2922.

Tanahashi, T., Murase, J., Matsuya, K., Hayashi, M., Kimura, M., & Asakawa, S. (2005). Bacterial communities responsible for the decomposition of rice straw compost in a Japanese rice paddy field estimated by DGGE analysis of amplified 16S rDNA and 16S rRNA fragments. *Soil Science & Plant Nutrition*, *51*(3), 351-360.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - A guide from sampling to data analysis. *Microbial Informatics and Experimentation*, *2*(3), 1-12.

Tian, H., Wang, H., Hui, X., Wang, Z., Drijber, R. A., & Liu, J. (2017). Changes in soil microbial communities after 10 years of winter wheat cultivation versus fallow in an organic-poor soil in the loess plateau of China. *PLoS ONE*, *12*(9), 1-15.

Torsvik, V., & Øvreås, L. (2002). Microbial diversity and function in soil: From genes to ecosystems. *Current Opinion in Microbiology*, *5*(3), 240-245.

Torsvik, V., Øvreås, L., & Thingstad, T. F. (2002). Prokaryotic diversity — magnitude , dynamics, and controlling factors. *Science*, *296*(5570), 1064-1066.

Tripathi, B. M., Kim, M., Singh, D., Lee-Cruz, L., Lai-Hoe, A., Ainuddin, A. N., … Adams, J. M. (2012). Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microbial Ecology*, *64*(2), 474-484.

van der Heijden, M. G. A., Bardgett, R. D., & van Straalen, N. M. (2008). The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, *11*(3), 296-310.

Vercammen, K., Garcia-Armisen, T., Goeders, N., van Melderen, L., Bodilis, J., & Cornelis, P. (2013). Identification of a metagenomic gene cluster containing a new class A beta-lactamase and toxin-antitoxin systems. *MicrobiologyOpen*, *2*(4), 674-683.

Wang, P., Marsh, E. L., Ainsworth, E. A., Leakey, A. D. B., Sheflin, A. M., & Schachtman, D. P. (2017). Shifts in microbial communities in soil , rhizosphere and roots of two major crop systems under elevated CO_2 and O_3. *Scientific Reports*, *7*(1), 15019.

Wang, S., Gao, X., Gao, Y., Li, Y., Cao, M., Xi, Z., … Feng, Z. (2017). Tetracycline resistance genes identified from distinct soil environments in China by functional metagenomics. *Frontiers in Microbiology*, *8*, 1406.

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, *95*(12), 6578-6583.

Woli, A., Ku, A., Zielenkiewicz, U., & Banach, A. (2017). Metagenomic analysis of some potential nitrogen-fixing bacteria in arable soils at different formation processes. *Soil Microbiology*, *73*(1), 162-176.

Wong, S. C., Yiu, P. H., Bong, S. T. W., Lee, H. H., Neoh, P. N. P., & Rajan, A. (2009). Analysis of Sarawak Bario rice diversity using microsatellite markers. *American Journal of Agricultural and Biological Science*, *4*(4), 298-304.

Wray, L. V, Atkinson, M. R., & Fisher, S. H. (1994). The nitrogen-regulated *Bacillus subtilis nrgAB* operon encodes a membrane protein and a protein highly similar to the *Escherichia coli glnB*-encoded P_{II} Protein. *American Society for Microbiology*, *176*(1), 108-114.

Wu, Q., Sanford, R. A., & Lo, F. E. (2006). Uranium (VI) Reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology*, *72*(5), 3608-3614.

Xiao, E., Krumins, V., Xiao, T., Dong, Y., & Tang, S. (2017). Depth-resolved microbial community analyses in two contrasting soil cores contaminated by antimony and arsenic. *Environmental Pollution*, *221*, 244-255.

Xiao, Y., Angulo, M. T., Friedman, J., Waldor, M. K., & Weiss, S. T. (2017). Mapping the ecological networks of microbial communities. *Nature Communications*, *8*(1), (2402).

Zhou, Z., Gu, J., Li, Y., & Wang, Y. (2012). Genome plasticity and systems evolution in *Streptomyces*. *BMC Bioinformatics*, *13*(Suppl 10), S8.

Zwolinski, M. D. (2007). DNA sequencing: Strategies for soil microbiology. *Soil Science Society of America Journal*, *71*(2), 592-600.