EFFECT OF H₂ ON SOIL BACTERIAL COMMUNITY STRUCTURE AND GENE EXPRESSION

By
Yinan Zou

A Thesis Submitted to Saint Mary's University, Halifax, Nova Scotia in Partial

Fulfillment of the Requirements for the Degree of Master of Science in Applied Science

July, 2012, Halifax, Nova Scotia

Copyright Yinan Zou, 2012

Approved: Dr. Gefu Wang-Pruski,
External Examiner
Nova Scotia Agriculture College

Approved: Dr. Zhongmin Dong,
Senior Supervisor

Approved: Dr. Doug Strongman,
Supervisory Committee

Approved: Dr. Adam Piorko,
Supervisory Committee

Approved: Dr. Jeremy Lundholm,
Graduate Studies Representative

Date: August 14, 2012

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Canada

ABSTRACT

EFFECT OF H_2 ON SOIL BACTERIAL COMMUNITY STRUCTURE AND GENE EXPRESSION

By Yinan Zou

Previous studies have showed that H_2 evolved from HUP- legume nodules is oxidized by rhizospheric microorganisms and alters the soil microbial diversity in rhizosphere. This H_2 oxidation is linked with the CO_2 fixation and O_2 reduction activities. Metatranscriptome analysis in this study revealed that soil bacterial community was changed significantly by H_2. Populations of *β-proteobacteria* and *δ-proteobacteria* were stimulated by H_2 exposure whereas *Actinobacteria* was suppressed. Results also showed that genes related with CO_2 fixation and O_2 reduction were up-regulated after H_2 treatment. Several speculated pathways were generated for the H_2 metabolisms in the soil. The quantification analysis of hydrogenase, the key enzyme in H_2 oxidation in soil, was carried out using *HoxK* NiFe hydrogenase small subunit gene via real time PCR. The result showed that the gene copy number had a significant increase after H_2 treatment.

August 14, 2012

ACKNOWLEDGEMENTS

It would not have been possible to write this master thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Special appreciation goes to my supervisor, Dr Zhongmin Dong, for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experimental and thesis works have contributed to the success of this research. Not forgotten, my appreciation to my committee members, Prof. Doug Strongman and Prof. Adam Piorko for their support and knowledge regarding this topic.

I would like to express my appreciation to my external examiner, Prof. Gefu Wang-Pruski, for her instructive comments for my thesis writing and precious evaluation of my thesis during my thesis defense.

My acknowledgement also goes to all the technicians and office staffs of Department of Applied Science and Faculty of Graduate Studies and Research for their co-operations.

Sincere thanks to all my friends Xuan Yang, Bryan Flynn, Jin Wang, Henry Annan, Sarah Hall and others for their kindness and moral support during my study. Thanks for the friendship and memories.

Last but not least, my deepest gratitude goes to my beloved parents and my uncle's family for their endless love, prayers and encouragement. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. General introduction

1.1 Crop rotation

Crop rotation (also known as crop sequencing) is an agricultural practice of growing different crops in the same field in different seasons followed by certain sequences. It appeared in Europe in time of Roman Empire and was spread by Romans to European farmers as called "food, feed, and fallow" (Harris, 1995) and has been recorded in Qin and Han dynasties (221 B.C.-220 A.D.) in China (Gong et al., 2000). At beginning, crop rotation, which was firstly introduced in 16th century by a Frisian farmer called Hemmema (Slicher Van Bath, 1963), usually refers to two-field crop rotation (two-year crop rotation). It means only half of the field was in use and the rest stayed unused in the first year. In the next year, two halves were reversed. Then the system was developed to three-field crop rotation system which enlarged the rotation effect by the additional crops. Four-field rotation system came out in the early 16th century and was well received by European farmers by 1800 (Butt, 2002).

Crop rotation was beneficial because it firstly enabled farmers to rest the land and allowed nitrogen from atmosphere to active free-living bacteria, provided crop alternation (Grigg D, 1989) and reduced the effect of soil-borne diseases (Connor, 2011). Secondly, crop rotation system was able to take full advantage of soil capacity and made land thoroughly weeded (Andreae B, 1981).

Legumes are well-known as nitrogen contributors to a succeeding crop (Heichel G.H, 1987; Power J.F, 1989). Soil fertility can be significantly improved by legumes; it allows farmers to relatively reduce the application of chemical fertilizers. Because of the

ability of enhancing soil N fertility, legume crop rotation system is widely utilized in the world which serves as the rotation system. The influence of legume crop was reported to last for two or three years or even longer period for cotton production (Cooper, 1999; Rochester *et al.*, 2001; Hulugalle and Danielle, 2005). However, the crop growth promotion cannot be fully ascribed to nitrogen leftover from N_2 fixing legumes (Hesterman, 1986).

1.2 Nitrogen fixation and hydrogen production

1.2.1 Nitrogen fixation

Nitrogen, which is one of the elements in amino acid, is essential for life on Earth. For plants, although elemental nitrogen is absolutely necessary, they cannot absorb atmospheric nitrogen directly. Normally, nitrogen source can either be derived from N-fertilizer or come from atmospheric N_2 via biological nitrogen fixation process (BNF). Legumes can fix massive atmospheric nitrogen in their root nodules by bacterial symbionts (Giller, 2001). This biological nitrogen fixation is the greatest source of nitrogen inputs to agricultural system (David *et al.*, 2011). In BNF process, atmospheric N_2 is usually converted to ammonia and the general formula for BNF is

$$N_2 + 8H^+ + 8e^- \rightarrow 2NH_3 + H_2.$$
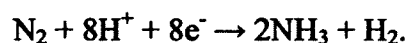
The key enzyme of this reaction is called nitrogenase. In addition, the nitrogen fixation process requires input of 16 equivalents of ATP molecules while one molecule of H_2 is produced per one N_2 molecule fixed.

1.2.2 Hydrogen production as by-product

During nitrogen fixation, hydrogen is produced as by-product (Wilson and Burris, 1947; Kamen and Gest, 1949; Hoch et al., 1960). It was discovered that the formation of H_2 is closely related to N_2 fixation (Phelps and Wilson, 1941). Later on H_2 production and evolution were found in soybean nodules (Hoch et al., 1957). It was estimated that 70% of the electrons passed from nitrogenase to nitrogen (Karel, 1976) while the rest went to hydrogen in BNF process (Hunt and Layzell, 1993, Ruiz-Argüeso et al., 1978). Each season, legume crop can produce 240,000 liters of H_2 each hectare while fixing 200 kg N ha^{-1} (Dong et al., 2003).

1.2.3 HUP status of legume root nodules

It was found that, in some pea root nodules, hydrogen was not evolved but was assimilated by nodules through uptake hydrogenase (HUP) activity (Dixon 1967). This activity appeared in legume root nodules which are named as HUP+. Some root nodules which lack the ability of taking up hydrogen produced by nitrogen fixation are known as HUP - nodules. Because of the capacity of taking up H_2, HUP + nodules usually release little hydrogen gas to rhizosphere, while HUP - nodules can deliver abundant H_2. The energy as H_2 produced from nodules equals to about 5% of total energy from crops' net photosynthesis (Dong and Layzell, 2001).

It was formerly believed that the H_2 loss from nodules was a disadvantage of HUP- compared to HUP+. The lost H_2 was utilized by neither plants themselves nor N_2 fixation bacteria in root nodules. La Favre and Focht (1983) found H_2 from nodules was quickly assimilated by soil organisms living in soil close to nodules. This activity led to

the increase of soil biomass in that area (Popelier *et al.*, 1985) and stimulation of O_2 reduction and CO_2 fixation (Dong and Layzell, 2001). The selection of HUP- symbiosis by nature and breeders suggested the H_2 evolved from nodules increases the soil fertility adjacent to nodules and eventually benefits the plant growth (Dong and Layzell, 2002). This improvement was carried out by H_2 oxidation activities around HUP- nodules by soil bacteria (McLearn and Dong, 2002). What is more, soil microbial community in this area is noted altered significantly by H_2 released from HUP- nodules (Zhang *et al.*, 2009); populations of several phyla from rhizosphere are rapidly stimulated by H_2, such as *Variovorax* and *Burkholderia* (Maimaiti *et al.*, 2007).

1.3 Plant-bacteria interaction

1.3.1 The rhizosphere

The rhizosphere is the narrow region of the soil habitat where plant roots are involved in numerous biological, chemical, and physical activities. Complex gradients are formed and released into this area by plants. The majority of those substances are organic and originally coming from plant photosynthetic and other processes. Population of microbes in this area usually live along with plant roots, their activities are cross-influenced by each other. Resources, such as organic carbon, mineral nutrients and water, are universal biological currencies exchanged and transformed in rhizosphere (Cardon and Whitbeach, 2007). Biological activity in rhizospheric area can also influence the patterns of community structure, ecosystem processes, and in patterns of soil development.

1.3.2 Plant-Growth-Promoting Rhizobacteria (PGPR)

In rhizobacteric area, it was found by Hiltner (1904) that the population of bacteria was richer compared to bulk soil. One possible reason is that plant roots secrete metabolites from 5-21% of plant carbon fixation as root exudates to benefit the rhizospheric bacteria (Marschner, 1995). In order to maximize the benefit for root environment, rhizobacteria have to compete with other rhizospheric microbes for nutrients and sites that can be occupied on the root (Lugtenberg and Kamilova, 2009). The biological metabolites in rhizosphere were reported as signaling compounds that are perceived by other bacteria or root cells of host plants (Van Loon and Bakker, 2003; Bais et al., 2004; Kiely et al., 2006).

Depending on different mechanisms, PGPR can be divided into several classes. *Rhizobium* and *Bradyrhizobium* were reported as biofertilizers who can form nodules on legume roots and convert atmospheric nitrogen into ammonia for plant growth (Spaink et al., 1998; Van Rhijn and Vanderleyden, 1995). Some pollutant-degrading rhizobacteria living on, or are close to, plant root that can use root secretion as major nutrient source are used as rhizoremediators to replace some pollutant-degrading bacteria that cannot degrade pollutants when they starve under wild condition (Kuiper et al., 2001). Some bacteria are able to produce plant hormone that stimulates plant growth. For example, Auxin-generating *P. fluorescens* WCS365 strain was reported leading to a significant increase in the root weight of radish, but showing no effect in cucumber, sweet pepper or tomato (Kamilova et al., 2006). The N_2- fixing bacterium *Azotobacter paspali* could produce plant growth factors such as IAA, gibberellins, and cytokinins, rather than sharing nitrogen fixation (Okon et al., 1998). PGPR containing

13

1-aminocyclopropane-1-carboxylate (ACC) deaminase were able to reduce plant ethylene level, therefore, helped plant growth and development. Such rhizobacteria could relieve plant stress from polyaromatic hydrogcarbons, from heavy meatls such as Ca^{2+} and Ni^{2+}, and from salt and draught (Glick *et al.*, 2007).

1.3.3 Hydrogen Oxidizing bacteria

It is well known that the addition of legumes in crop rotation can contribute to the significant increase in growth and yield. However, only 25% of crop plant growth can be attributed to application of N nutrition, the rest of effect still remained unknown (Bolton *et al.*, 1976). Recent studies showed that H_2 synthesized as by-product of N_2 fixation in root nodules may be partially responsible for the plant growth promotion. The soil H_2 uptake activity was noted being related with the soil microbial biomass (Popelier *et al.*, 1985). McLearn and Dong (2002) reported that the addition of antibiotics significantly reduced the H_2 uptake activity in soil, whereas fungicides didn't affect much. Therefore they concluded that the H_2 oxidizing bacteria should be responsible for soil uptake activity. Additionally, soil O_2 reduction and CO_2 fixation activities were also observed when H_2 was taking up in soil (Dong and Layzell, 2001).

The H_2 oxidizing bacteria (*Knallgas*-bacteira) are a group of chemolithoautotrophic bacteria defined by their capacity of decomposing H_2 in soil for energy. Calvin cycle is one of carbon metabolic pathways in H_2 oxidizing bacteria, which is similar with other photolithotrophic bacteria (Bowien and Schlegel, 1981). Energy and reducing power are used to activate hydrogenase which is the key enzyme of H_2 oxidation process.

Isolates of H_2 oxidizing bacteria were able to promote primary root elongation by 57-254%. The number and length of lateral roots and the amount of root hair also increased inoculated with H_2 oxidizing bacterial isolates (Maimaiti, 2007). H_2-oxidizing bacteria were considered responsible for legume plants growth promotion, although the mechanism of the plant-bacteria interaction between H_2-oxidizing bacteria and plants was still unknown.

1.4 Major molecular biological techniques for identification of plant-growth promoting bacteria

1.4.1 Fingerprint methods

The study of soil microbial communities was hampered by inability of classifying microorganisms, because over 99% of microorganisms in nature cannot be cultured in lab condition (Torsvik et al., 1990). In order to have a better understanding of soil microbial community in structure and function, techniques that can complement the traditional biological methods are required. The application of fingerprint techniques provides microbiologists an easy access to obtain an overall view of soil microbial community and diversity. These techniques for microbial community analyses are mainly based on PCR amplified DNA fragments. The common gene target for microbial community study is the small subunit ribosomal RNA gene (rDNA) because it widely exists in all bacteria on Earth and has been fully investigated for years. With the development of modern technology, several DNA fingerprint methods have been put into application for microbial community study, such as single-strand conformation polymorphism (SSCP),

denaturing and temperature gradient gel electrophoresis (DGGE/TGGE), restriction fragment length polymorphism (RFLP) and terminal restriction fragment length polymorphism (T-RFLP).

1.4.1.1 Single strand conformation polymorphism (SSCP)

Single-strand conformation polymorphism (SSCP) is a molecular method that has been widely used in mutation analysis and microbial diversity studies (Lee *et al.*, 1996; Schwieger and Tebbe, 1998). With this technique, PCR amplified DNA fragments, usually 16s rRNA gene with specific primer, can show different mobility on a non-denaturing gel based on folded conformation of single- stranded DNA. In order to exclude the possible reannealing during electrophoresis and multiple bands during separation patterns, phosphorylated primer before specific digestion of phosphorylated strands with lambda exonuclease or biotinylated primer after denaturation are applied (Lee *et al.*, 1996; Schwieger and Tebbe, 1998; Scheinert *et al.*, 1996). For the purpose of identifying the bacterial population in the community, SSCP –patterns needs to be hybridized with taxon specific probes and bands need to be excised and sequenced (Muyzer, 1999). In addition, fragment size for SSCP should range from 150 bp to 400 bp for optimal separation purpose.

1.4.1.2 Denaturing and temperature gradient gel electrophoresis (DGGE/TGGE)

Denaturing gradient gel electrophoresis (DGGE) is firstly introduced by Muyzer *et al* (1993) in microbial ecology study. DNA from a complex of microorganisms is amplified by PCR and separated in a polyacrylamide gel containing a linear gradient of

DNA denaturants (Muyzer *et al.*, 1998). DNA fragments with different sequences will separate at different position in the gel depending on melting behavior. The principle of DGGE is shared with temperature gradient gel electrophoresis (TGGE).

Both DGGE and TGGE are very popular molecular methods for microbial ecological studies and becoming regular in laboratories. They are widely applied in microbial community diversity analyses, differential gene expression in complex systems studies (Muyzer, 1999). The technique is reliable, reproducible, rapid and inexpensive. Compared with other fingerprinting techniques, such as terminal restriction fragment length polymorphism (T-RFLP), DGGE/TGGE is able to identify the community members by sequencing excised bands or by hybridization analysis using specific probes (Liu *et al.*, 1997). However, this method also has some limitations. For instance, DGGE/TGGE always has trouble with separation of relatively small DNA fragment and co-migration of DNA fragments with different sequences (Vallaeys *et al.*, 1997). In addition, the sensitivity of detecting rare community is always limited by hybridization analysis.

1.4.1.3 Restriction fragment length polymorphism (RFLP) and terminal restriction fragment length polymorphism (T-RFLP)

Restriction fragment length polymorphism (RFLP) is known as a PCR-based technique for multiple purposes. In microbiology study, it is used to characterize microbial communities by 16S rDNA fragments. After PCR amplification, 16S rDNA fragments are digested by restriction enzymes, separated according to the fragment length by gel electrophoresis. However, sometime overestimation of community

population occurs while using RFLP in micro biodiversity studies (Muyzer, 1999). In order to overcome the shortage, the fluorescently-labelled PCR product is applied to RFLP technique. Therefore, it is called terminal restriction fragment length polymorphism (T-RFLP) which shares the similar principle with RFLP technique. During amplification process in T-RFLP, all products are terminally fluorescently-labeled at the 5' end. After digestion by restriction enzymes, fluorescent terminal fragments with different length are generated and sequenced by DNA sequencer.

T-RFLP is more popular for routine analysis because it is more sensitive compared to DGGE/TGGE. It always comes with the phylogenetic assignment tool (PAT), the combination of these two methods can provide a rapid, automated approach for phylogenetic analysis for terminal restricted fragment (Kent *et al.*, 2003). In addition, because of its reproducibility and potential for automatisation, T-RFLP is more suitable for large complex sample scale (Smalla *et al.*, 2007; Osborne *et al.*, 2006). However, for all fingerprinting techniques, it is very hard to provide detailed information about microbial community. Therefore, more advanced method is required for microbial community studies.

1.4.2 PCR - Clone Library and Real-Time PCR

The PCR amplification, clone library build-up and sequencing are the common molecular tools to study the phylogenic diversity of microbial community containing specific genes in their genomes. In microbiology studies, 16s rRNA genes are usually used as targets for general micro-diversity research. For particular microbial community study, specific gene sequences that encode key compound for bacterial activity are

commonly chosen as PCR targets. Particular primer pairs will also be designed based on those sequences. PCR fragments will be inserted into bacteria via plasmids. Therefore, a large collection of bacteria containing PCR products is generated, which is also called clone library. The library can be used for further study such as DNA sequencing.

In order to improve the accuracy of this technique, the primer set and product size must be taken into consideration when designing experiments. Because longer PCR products may not perform appropriately for measurement of community structure compared to products with smaller size (Julie *et al.*, 2009). However, longer targets can also provide increased diversity of community, whereas shorter amplicons may require deeper sequencing effort because clone library may contain many more different types of sequences. Beside product size, Silvia *et al* (2006) also suggested PCR should be performed with combination of multiple replicate PCR amplifications, less amplification cycles and significant difference between denaturation and annealing temperature with shorter extension period.

Real Time PCR, also called qPCR, is a routine tool in molecular biology for quantitative gene analysis at DNA level. It is a continuous collection of fluorescent signal from polymerase chain reaction throughout cycles. Generally, qPCR follows the principle of regular PCR. The difference between qPCR and regular PCR is fluorescence and reference. Usually qPCR can be divided into two different types based on fluorescence: qPCR with non-specific fluorescent dye and with specific fluorescent probes. Fluorescence in the former will bind any double-stranded DNA fragment in the reaction whereas the later will only combine with the fragment that has a fluorescent reporter. For non-specific fluorescent dye, SYBR Green I is the most commonly used. SYBR Green I

can be detected at the wavelength of around 520 nm which can be reached in most real time PCR thermal cyclers. In addition, relative low cost and environmental friendly also make it suitable for many applications. Specific fluorescent probes are usually applied for detection purpose. qPCR also can be grouped as absolute qPCR and relative qPCR. The difference is that absolute quantification is using quantified amount of standard for standard calibration curve generation, but relative qPCR requires an appropriate internal reference as standard which is always a 'housekeeping' gene. Comparison of C_T values in absolute qPCR should be able to provide copy number or concentration of genes in the sample whereas relative qPCR could give a ratio of amount of initial target sequence between experimental and control samples. Compared to absolute qPCR protocol, a standard calibration curve is not necessary for relative qPCR.

1.4.3 Metatranscriptome analysis

In the past, all transcriptome analysis were based on three different commercially available techniques: Roche 454, Illumina and ABI SOLiD. However, these techniques are usually applied in eukaryotic system for identification purpose. Metatranscriptome analysis is a newly developed technique which is similar to those techniques above. However, metatranscriptome analysis is more advanced because it could handle prokaryotic RNA that is less stable than eukaryotic RNA. The technique appeared in 2008, Nagalakshmi (2008) and Wilhelm (2008) reported the RNA sequencing of whole microbial transcriptomes of yeasts *Saccharomyces cerevisae* and *Schizosaccharomyces pombe*. They both stated transcription of most non-repetitive sequence in yeasts and many detailed information like novel genes, introns and their boundaries, 3' and 5'

boundary mapping and more. However, the target is still from eukaryotes. In 2009, prokaryotic whole transcriptome analysis had been reported by many studies using tiling arrays or RNA sequencing technique. The first reviews of these studies appeared in 2009 and 2010 (Croucher *et al.*, 2009; van Vliet andWren, 2009; Sorek and Cossart, 2010; van Vliet, 2010).

Metatrancriptome analysis is attracting increasing attention in microbial ecosystem study. Comparing with eukaryotic RNA, working with prokaryotic RNA is always a great challenge. Bacterial RNA does not have a poly-A tail (Deutscher, 2003), which could prevent RNA from degradation. Therefore, it is impossible to isolate bacterial RNA using hybridization to immobilized poly-T. In addition, the majority of total RNA from bacteria is rRNA and tRNA (up to 80%) (Condon, 2007) and has a very short half-life. Hence, it is reasonable that transcriptome study of eukaryotic RNA is earlier described than the study on prokaryotes.

In metatranscriptome analysis on bacterial RNA, mRNA is collected from total RNA and then is interrupted into small fragment. Using these small pieces of mRNA as template, the second-strand cDNA is synthesized. After agarose gel electrophoresis, fragments are amplified by PCR process. At last, the cDNA library is generated using Illumina technique. The downstream application is used for multiple bioinformatics analysis for different purpose, such as function annotation, metabolic pathway, coding region prediction and gene expression difference.

With the rapid development of sequencing technology, there is no doubt that these techniques will provide us a deeper insight in transcriptional and post-transcriptional activities in prokaryotes in ecosystem. Now the microbial ecosystem is able to be

considered in a brand new aspect.

1.5 Objectives

My research consists of three different objectives: 1) Soil microbial community change induced by H_2 simulated root nodule environment. 2) The effect of nodule released H_2 on soil bacterial gene expression. 3) Hydrogenase gene activity induced by H_2 treatment.

Soil bacterial community study induced by H_2 gas has been discussed using different techniques such as Clone-library and Terminal restriction fragment length polymorphism (T-RFLP). However, results obtained from different techniques were always controversial in some phylum. Therefore, it is necessary to ultilize metatranscriptome analysis to detect the soil microbial community change induced by H_2 at the level of RNA.

The effect of H_2 oxidizing bacteria on plant growth promotion had already been studied by field trial and root elongation test with H_2 oxidizing bacteria inoculation (Maimaiti *et al.*, 2007). In the meantime, several researches on genes in H_2 oxidizing bacteria that are indirectly involved with plant growth promotion had been conducted. However, it still remains the question that if there is any other gene activity which is missed by us. In addition, gas exchange activity in H_2 treatment has been monitored and a few data had been obtained, but it still needs more focus. Therefore, metatranscriptome analysis is the appropriate tool to study the whole transcriptome activity in our particular ecosystem and bring the possibility to uncover the myth of mechanism of H_2 oxidizing bacteria on plant growth promotion.

Previous researches showed the presence of H_2 oxidizing bacteria in soil adjacent to root nodules. Hydrogenase gene, the key gene of H_2 oxidation process, was select by us to investigate the abundant differential of hydrogenase genes in air and H_2 treated soil systems.

2. The Metatranscriptome Analysis of H_2 induced Soil Bacterial Community Structure Changes and Gene Expression Shifts

2.1 Introduction

2.1.1 H_2-dependent soil bacterial community change

It has been reported that H_2 uptake activity around root nodules is connected with microbial activities in that area, particularly with bacteria (Popelier et al., 1985; Dong and Dayzell, 2002). Those organisms that are able to oxidize H_2 are called H_2 oxidizing bacteria. Some of those bacteria had been characterized as well and three genera of who were successfully isolated by Maimaiti (2007) from H_2 treated soil and soil adjacent to HUP- legume root nodules. In addition, La Favre et al. (1983) indicated that H_2 uptake rate and microbial biomass declined as in area further away from root nodules. It suggested that H_2 released from root nodules might affect the microbial community structure in rhizosphere. Unfortunately, it is impossible to detect whole microbial community structure only by isolating methods, because over 99% of soil bacteria cannot be cultured on media (Torsvik et al., 1990). Hence, more advanced techniques are required for soil ecological purpose.

A similar study was conducted by Urich et al (2008), results showed that the bacterial community structure in bulk soil mainly consist of *Actinobacteria* and *Proteobacteria* as two largest phylum. In *Proteobacteria*, *Alphaproteobacteria* occupied the most population. *Burkholderiales* order was the largest group in *Betaproteobacteria*.

Stein et al (2005) demonstrated the success of utilizing fluorescence in situ

hybridization (FISH) and DNA staining techniques to study the soil bacterial community change induced by H_2. Their results demonstrated that population of *Betaproteobacteria*, *Gammaproteobacteria* and *Cytophaga Flavobacterium Bacteroides* were stimulated significantly by H_2 treatment.

The soil bacterial community structure change could also be monitored by using the combination of Terminal Restriction Fragment Length Polymorphism technique (T-RFLP) and phylogenetic assignment tool (PAT) (Zhang *et al.*, 2009). T-RFLP analysis was proved being capable of comparing bacterial community composition from multiple environments and the combination with PAT could extend the function by providing a more specific phylogenetic analysis (Kent *et al.*, 2008). Part of data from T-RFLP indicated the observation of *Variovorax, Burkholderia* and *Flavbacterium*, which was corresponded with Maimaiti's result (Maimaiti, 2007). The population of *Gammaproteobacteria* was induced in both soil samples from H_2 treatment system and field trial, whereas *Actinobacteria, Alphaproteobacteria* and *Firmicutes* were suppressed by H_2 significantly. However, all those methods were still based on DNA related analysis. In our study, we applied metatranscriptome analysis to detect the soil microbial community change at the level of RNA. Hence we were aiming to provide the bacterial community profile through a new perspective.

2.1.2 CO_2 fixation and O_2 reduction in H_2 oxidizing bacteria

H_2 oxidizing bacteria are chemolithoautotrophic, aerobic organisms that are able to use H_2 for aerobic respiration and reduction of CO_2 (Schink and Schlegel, 1978). Dong and Layzell (2001) demonstrated that CO_2 fixing and O_2 reducing activity were

significantly stimulated in highly concentrated H_2 environment. The O_2 concentration was reduced from 20.95% in air to 0.2% after 3-week H_2 treatment. Dong and Layzell (2001) also noticed that, when soil was treated with H_2, reducing power from H_2 eventually went to O_2 and CO_2 at the percentage of 60% and 40%, respectively. Previous study conducted by Simpson (1979) proved the presence of ribulose bisphosphate carbohylase/oxygenase (Rubisco), the key enzyme of Calvin Cycle, in free-living HUP+ rhizobia for CO_2 fixation. It catalyzes carboxylation of ribulose-1,5-bisphosphate (known as RuBP) in Calvin Cycle. The growth of *Rhodospirillum rubrum*, a genus of photosynthetic bacteria of the family Rhodospirillaceae, under H_2 and CO_2 condition in a completely synthetic medium was reported by John and Howard (1962). This species was reported containing the Rubisco as a dimer of large subunits (L2) (Tabita and McFadden, 1974). However, how H_2 from root nodules was involved with Calvin cycle in hydrogen oxidizing bacteria is still a myth.

2.2 *Materials & Methods*

2.2.1 Soil sample preparation

All soil samples in this study were collected from Annapolis Valley, Nova Scotia. Samples were mixed with sand (2:1 v/v) and were filled into 60 ml syringes for H_2 treatment.

All soil samples were divided into two set as H_2 treatment and air treatment. H_2 treatment was regarded as experimental group which was consistently treated with gas mixture of H_2 and air (1000 ppm), whereas air treated samples were consistently treated with room air as control (Fig. 1).

Figure 1 A simplified diagram of the H_2 treatment system (Dong and Layzell, 2001)
Both flasks were provided compressed room air by a clean air container. The left flask
was connected with a regular power supply which generated H_2 by a stable electric
current. This flask was considered as experimental group. The right flask didn't have H_2
supply; therefore it was treated as control group. All gas was mixed before it was
transported to soil: sand mixture. The concentration of H_2 in H_2 treatment was controlled
by stabilizing the electric current and flow rate at 100 ppm. The concentrated H_2 mixture
gas went through soil: sand mixture from V1 to V2 (atmosphere). In control group, the
air flew through V3 to V4 (atmosphere).

H₂ treatment

V2

V4

Soil:Sand
mixture

Air
treatment

V1

100 mM
Phosphoric Acid

V3

Clean
Compressed
Air

Regulated
Power
Supply
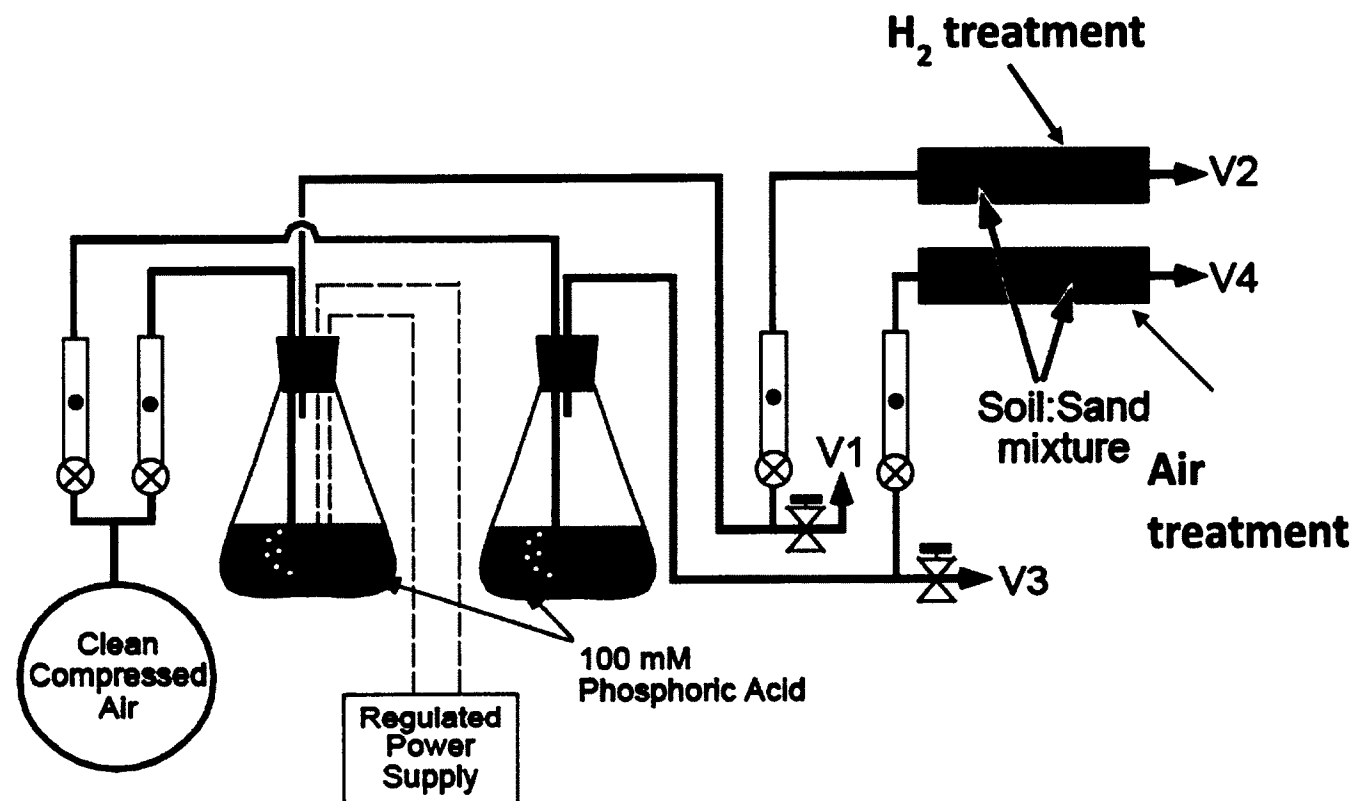
Figure 2 A simplified diagram of the hydrogen uptake capacity measurement system (He, 2010)

H_2 was produced in the flask by the regular power supply. The gas was mixed with room air and transported through the tube. Valves 1, 2, 3 and 4 are operated to make the sensor determine the concentration (ppm) of hydrogen in the mixed gas stream before and after passing the soil column.

Vent

H_2–air mixture

Air pump 2

switch

Clear compressed air

V3

V1

Air pump 1

Regular power supply

switch

V2

Computer analysis system

Voltage signal

V4

Voltage sensor

Vent
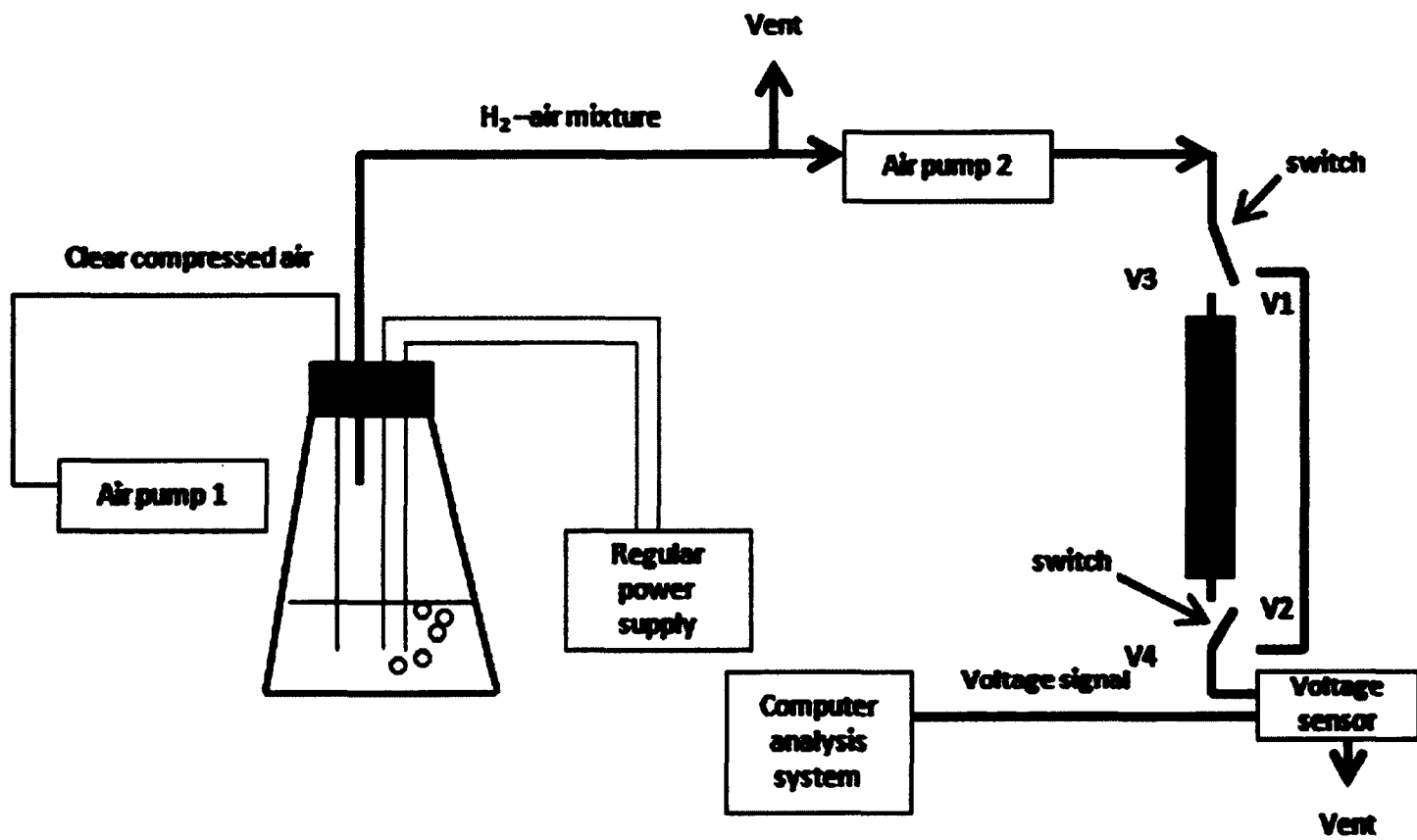
2.2.2 Measurement of H$_2$ uptake rate

H$_2$ uptake rate was measured throughout the hydrogen treatment using a H$_2$ sensor (Model S121, Quibit Systems Inc) and was recorded by the Data Logger program as described by Dong and Layzell (2001) (Fig. 2). The data was collected as voltage before (turning on V1 & V2 and turning off V3 & V4) and after (turning on V3 & V4 and turning off V1 & V2) the mixed gas stream passing the soil column. The amount of electrolytic hydrogen (Z: μmol/min) in the flask (Figure 2) was calculated using the following equation:

Z (μmol /min) = (3.00*104*C *Cu) / Av---①

C (Coulomb Constant): 6.24*1018 (A-1);

Cu (current of electrolysis): mA;

Av (Avogadro Constant): 6.02 *1023 (mol-1).

From Equation① the following equation was derived to calculate the concentration of electrolytic hydrogen in the mixed gas stream (H: ppm):

H (ppm) = [1.00*103*Z *GC*(273.15+T)]/(273.15*FR)-------------- ------②

Z (amount of electrolytic hydrogen per minute):   μmol/min;

GC (gas constant): 22.41 L/mol at 0 oC and 1 atmosphere pressure;

T (temperature): °C;   FR1 (Flow Rate One): ml/min.

A standard curve was generated to build up the relationship between voltage and hydrogen concentration (ppm) by setting a series of mixed gas streams with gradient hydrogen concentration (from 0.55-500 ppm). This is carried out by regulating the current of electrolysis and flow rate using measeuement system (Fig. 2). When V1 and V2 were turned on, H$_2$ –air mixture gas was going to the H$_2$ sensor directly, therefore the

sensor could record the voltage of H_2 –air mixture gas with known concentration. Based on Graphpad Prism 5.0, two standard curves of voltage versus hydrogen concentration (ppm) were generated:

Range 0.55 – 100 ppm: ppm=21.92*v/(2.101-v)-2.803, R^2=0.9962

Range 100 – 500 ppm: ppm=163.6*v/(4.273-v)-9.941, R^2=0.995

After 4 week H_2 treatment, the H_2 uptake rate in H_2 treated sample (343.63 nmol/h*g) was significantly higher than that in air treated soil sample (39.24 nmol/h*g). Then the soil sample was ready for further experiments.

2.2.3 RNA extraction and purification

Total RNA was collected by using RNA PowerSoil® Total RNA Isolation Kit (Catalog Number 12866-25) after 4 weeks H_2 treatment. The maximum amount of soil sample for each reaction is 1 gram. Soil samples were gathered immediately from both hydrogen and air treatment and were transferred into 15ml Bead Tube (provided). Then add 2.5 ml of Bead Solution (provided) to the Bead Tube and vortex to mix. After vortexing, add 0.25 ml of Solution SR1 (provided) to the Bead Tube and vortex to mix. Add 0.8 ml of Solution SR2 (provided) and vortex at maximum speed for 5 minutes. Remove the Bead Tube from Vortex Adapter and add 3.5 ml of phenol: chloroform: isoamyl alcohol (user supplied) and vortex to mix until the biphasic layer disappears. Place the Bead Tube on the Vortex Adapter and vortex at maximum speed for 10 minutes. Remove the Bead Tube from the Vortex Adapter and centrifuge at 2500 x g for 10 minutes at room temperature. After centrifugation, carefully transfer the upper aqueous phase to a clean 15 ml Collection Tube (provided). Add 1.5 ml of Solution SR3 (provided)

to the aqueous phase and vortex to mix. Then incubate tubes at 4°C for 10 minutes. Centrifuge incubated tubes at 2500 x g for 10 minutes at room temperature and transfer the supernatant, without disturbing the pellet, to a new 15 ml Collection Tube. Add 5 ml of Solution SR4 (provided) to the Collection Tube containing the supernatant and invert or vortex to mix. Then incubate tubes at room temperature for 30 minutes. Centrifuge tubes at 2500 x g for 30 minutes at room temperature after 30 min incubation. Discard the supernatant and invert the 15 ml Bead Collection Tube on paper tower for 5 minutes. Shake Solution SR5 to mix before use. Add 1 ml of Solution SR5 to the 15 ml Collection Tube and resuspend the pellet completely in the 45°C water bath for 10 minutes and vortex for several times. Prepare one RNA Capture Column (provided) while water bathing for each RNA Isolation sample. Remove the cap of a 15 ml Collection Tube and place the RNA Capture Column inside the tube. Add 2 ml of Solution SR5 to the RNA Capture Column and allow it to flow through the column and collect in the 15 ml Collection Tube. Add the RNA Isolation sample from water bathing onto the RNA Capture Column and allow it to flow through the column. Collect the flow through in the 15 ml Collection Tube. Wash the column with 1ml of Solution SR5. Allow it go gravity flow and collect the flow through in the 15 ml Collection Tube. Then transfer the RNA Capture Column to a new 15 ml Collection Tube. Shake Solution SR6 (provided) to mix and add 1 ml of Solution SR6 to the RNA Capture Column to elute the bound RNA into 15 ml Collection Tube. Allow SR6 to gravity flow into the 15 ml Collection Tube. Transfer the RNA to a 2.2 ml Collection Tube (provided) after elution and add 1 ml of Solution SR4. Invert tubes at least once to mix and incubate at minus 20°C for 10 minutes. Centrifuge the 2.2 ml tubes at 13000 x g for 15 minutes at room temperature to

pellet RNA. Decant the supernatant and invert the 2.2 ml Collection Tube onto a paper towel for 10 minutes until dry out. Resuspend the RNA in 100 μl of Solution SR7. Concentration of extracts was tested using Thermo Scientific NanoDrop 2000.

Add 4 Units of DNase enzyme and an appropriate amount of 10X DNase digestion buffer to RNA samples. Incubate tubes at 37°C for 45 minutes. Add one volume of phenol: chloroform: isoamyl alcohol and vortex. Incubate tubes at room temperature for 5minutes after vortexing. Then centrifuge tubes at 10000 x g for 5 minutes. Transfer the upper phase to a new tube. Add 1/10 volume of 5M NaCl and 2.5 volume of 100% ethanol into the tube and invert to mix. Incubate tubes at -20°C for 30 minutes and centrifuge at 10000 x g for 10 minutes. Decant supernatant, and add 1 ml of 70% ethanol and invert to mix. Incubate tubes at 4°C for 10 minutes and centrifuge at 10000 x g for 10 minutes. Discard the supernatant and air dry for appropriately 10 to 15 minutes. Resuspend the pellet in 50 ul of Sigma RNase free water. Quantity and quality of purified RNA sample were measured by Thermo Scientific NanoDrop 2000. 2% of agarose gel electrophoresis was carried out to test the existence of RNA and absence of DNA.

Total RNA extraction from soil was carried out twice from February, 2011 to July, 2011. The first extraction lasted for less than two months. Soil samples from treatment were collected each time and were used for extraction immediately. For each time of total RNA extraction, 40 ng of total RNA from H_2 treatment and 20 ng for air treatment were needed for requirement for metatranscriptome analysis. Considering the loss of total RNA during purification process, 60 ng of total RNA from H_2 treatment and 30 ng for air treatment were collected before purification. Generally speaking, the yield of total RNA

from H_2 treatment could reach about 2 ng from 1 g of soil sample and less than 1 ng from 1 g of air treated soil sample. Because of the limitation of equipments, RNA collection process lasted longer than we expected.

2.2.4 Shipping preparation for RNA samples

Purified RNA (H_2 treated soil total RNA sample was labeled as YZH, air treated soil total RNA sample was labeled as YZA) samples were sent to Beijing Genomic Institution (BGI) Hong Kong for metatranscriptome analysis. In order to maintain the RNA quality and avoid the degradation during the transportation, purified RNA samples were dried out using the vacuum chamber and stored as powder in GenTegra RNA Cluster Tubes (Catalog Number GTR3122). RNA samples were applied into cluster tubes from kit and dried down over night in vacuum chamber. Dried out samples were packed at room temperature and shipped via Fedex International shipment.

2.2.5 Metatranscriptome analysis

2.2.5.1 Pipeline of experiments

Beads with Oligo(dT) were used to remove poly(A) mRNA from eukaryote after total RNA was collected. rRNAs from prokaryocytes were isolated by kits. Fragmentation buffer was added for interrupting mRNA to short fragments. Taking these short fragments as templates, random hexamer-primer was used to synthesize the first-strand cDNA. The second-strand cDNA was synthesized using buffer, dNTPs, RNaseH and DNA polymerase I, respectively. Short fragments were purified with QiaQuick PCR extraction kit and resolved with EB buffer for end reparation and adding

poly (A). After that, the short fragments were connected with sequencing adapters. And, after the agarose gel electrophoresis, the suitable fragments were selected for the PCR amplification as templates. At last, the library could be sequenced using Illumina HiSeq™ 2000 (Fig. 1)

Figure 3 Experiment pipeline of transcriptome analysis

38
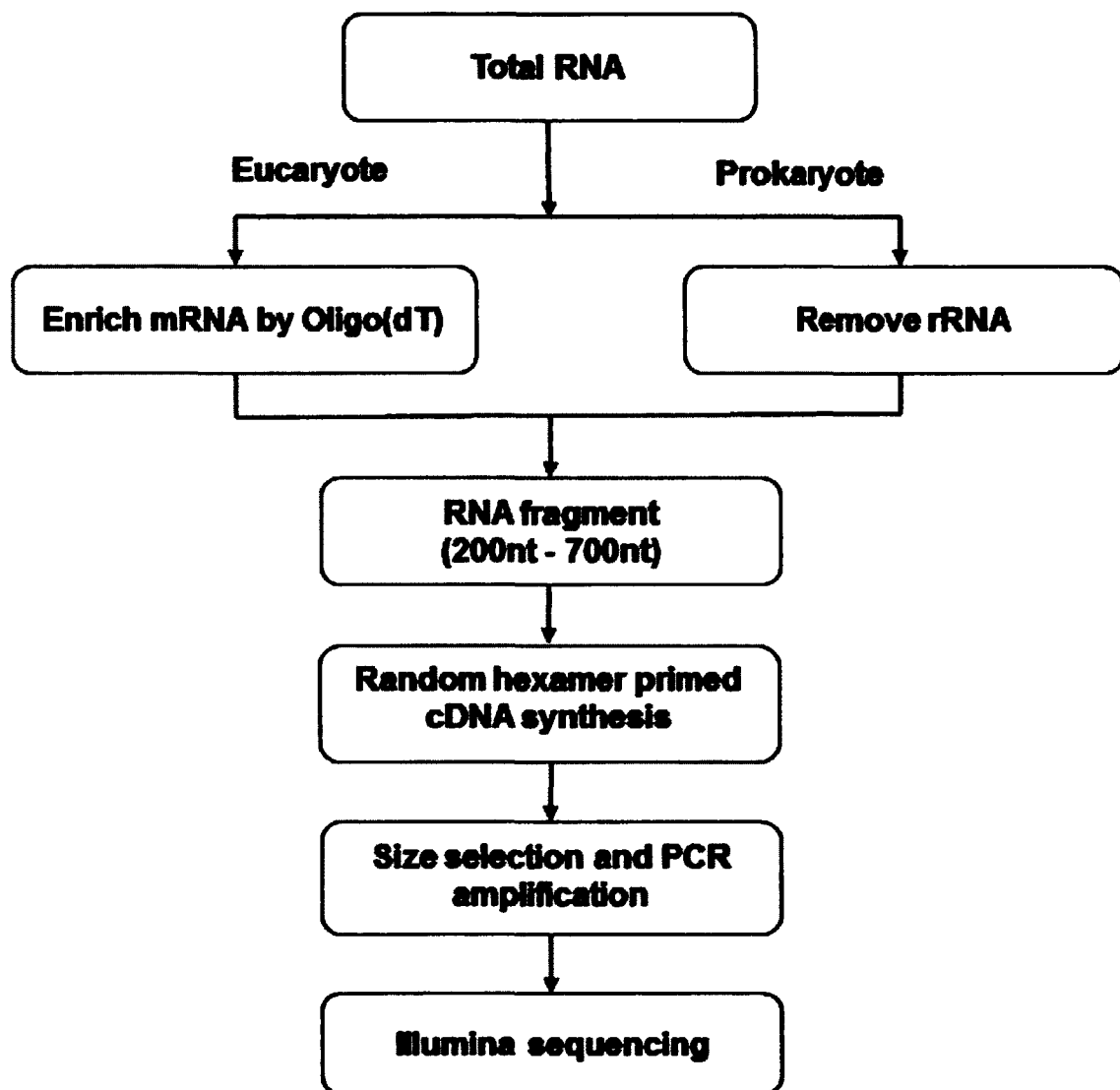
2.2.5.2 Pipeline of Bioinformatics Analysis

2.2.5.2.1 Output statistics

Image data output from sequencing machine was transformed by base calling into sequence data, which is called raw data or raw reads and stored in fastq format. Raw reads produced from sequencing machines contained dirty reads which contain adapters, unknown or low quality bases. These data can negatively affect following bioinformatics analysis. Therefore, dirty raw reads (reads with adaptors, unknown nucleotides larger than 10% and low quality) were discarded. The following analysis was based on clean reads, which were generated by filtering raw reads (Fig. 4).

2.2.5.2.2 Assemble

Transcriptome de novo assembly was carried out with short reads assembling program - SOAPdenovo. SOAPdenovo firstly combined reads with certain length of overlap to form longer fragments without N, which are called contigs. Then the reads were mapped back to contigs; with paired-end reads it was able to detect contigs from the same transcript as well as the distances between these contigs. Next, SOAPdenovo connects the contigs using N to represent unknown sequences between each two contigs, and then Scaffolds were made. Paired-end reads were used again for gap filling of scaffolds to get sequences with least Ns and cannot be extended on either end. Such sequences were defined as Unigenes. When multiple samples from a same species were sequenced, Unigenes from each sample's assembly could be taken into further process of sequence splicing and redundancy removing with sequence clustering software to acquire non-redundant Unigenes as long as possible. In the final step, blastx alignment (evalue <

0.00001) between Unigenes and protein databases like nr, Swiss-Prot, KEGG and COG is performed, and the best aligning results were used to decide sequence direction of Unigenes. If results of different databases conflict with each other, a priority order of nr, Swiss-Prot, KEGG and COG should be followed when deciding sequence direction of Unigenes. When a Unigene happens to be unaligned to none of the above databases, a software named ESTScan would be introduced to decide its sequence direction. For Unigenes with sequence directions, their sequences from 5' end to 3' end was provided; for those without any direction their sequences from assembly software were provided.

2.2.5.2.3 *Community structure analysis*

To estimate the community structure of each sample, all of the pair-ended reads aligned to Ribosomal Database Project (RDP) were overlapped to be longer tags, and the length of overlap region was at least 10bp. Total tags were overlapped for air treated sample and H_2 treated sample. For each sample, about 130 thousand tags were aligned against SILVA 106 again with BLASTN, the E value was set to be 1e-5 and only first 50 hits were used to determine taxonomy for each tag based on LCA (last common ancestor) principle.

Figure 4 Pipeline of Bioinformatics Analysis

```
                    ┌─────────────────┐
                    │    Raw reads    │
                    └─────────────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │   Clean reads   │
                    └─────────────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │    Unigenes     │
                    └─────────────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │ Gene Annotation │
                    └─────────────────┘
```

| Unigene Expression Annotation | Protein Coding Region Prediction (CDS) | ◄── | Unigene Function Annotation |

| Unigene Expression Difference Analysis | | Unigene GO Classification and Pathway Analysis |

| GO enrichment analysis | ◄── | ──► | Pathway enrichment analysis |
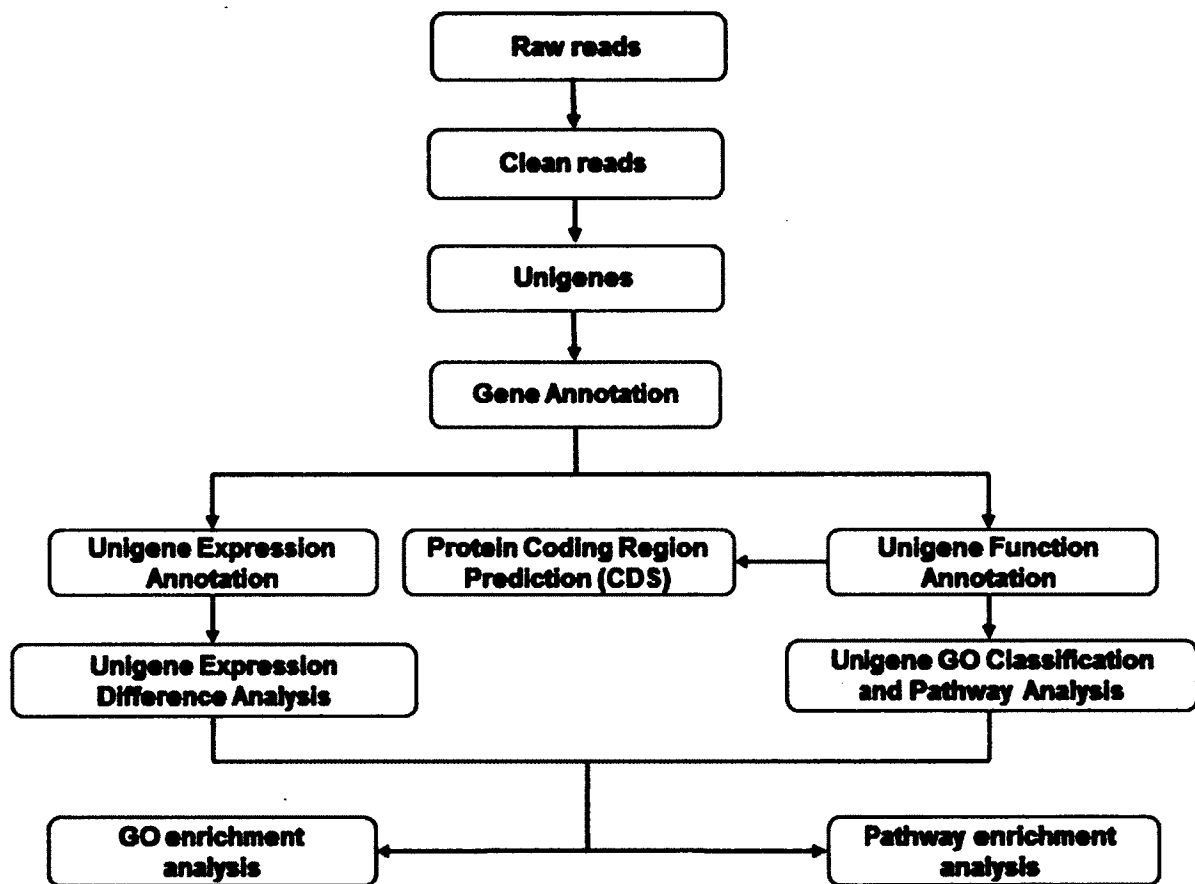
2.2.5.2.4 Unigene Function Annotation

Unigene annotation provided information of expression and functional annotation of Unigene. Information of functional annotation gave protein functional annotation, COG functional annotation and Gene Ontology (GO) functional annotation of Unigenes.

Unigene sequences were firstly aligned by blastx to protein databases like nr, Swiss-Prot, KEGG and COG (e-value<0.00001), retrieving proteins with the highest sequence similarity with the given Unigenes along with their protein functional annotations.

KEGG database contains systematic analysis of inner-cell metabolic pathways and functions of gene products. It helps studying complicated biological behaviors of genes. With KEGG annotation, we can get Pathway annotation of Unigenes could be achieved.

COG is a database where orthologous gene products are classified. Every protein in COG is assumed to evolve from an ancestor protein, and the whole database is built on coding proteins with complete genome as well as system evolution relationships of bacteria, algae and eukaryotic creatures. Unigenes were aligned to COG database to predict and classify possible functions of Unigenes.

2.2.5.2.5 Unigene GO Classification

Gene Ontology (GO) is an international standardized gene functional classification system which offers a dynamic-updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component and biological process. The basic unit of GO is GO-term. Every GO-term belongs to a type of

ontology.

With nr annotation, Blast2GO program was applied to get GO annotation of Unigenes. Blast2GO has been cited by other articles for more than 150 times and is widely recognized as GO annotation software. After getting GO annotation for every Unigene, WEGO software was used to do GO functional classification for all Unigenes and to understand the distribution of gene functions of the species from the macro level.

2.2.5.2.6   Unigene Metabolic Pathway Analysis

KEGG is a database that is able to anaylze gene product during metabolism process and related gene function in the cellular processes. With the help of KEGG database, we can further study genes' biological complex behaviors, and by KEGG annotation we can get pathway annotation for Unigenes.

2.2.5.2.7   Protein Coding Region Prediction (CDS)

Unigenes were firstly aligned by blastx (evalue<0.00001) to protein databases in the priority order of nr, Swiss-Prot, KEGG and COG. That is, we first align unigenes to nr, then Swiss-prot, then KEGG, and finally COG. Unigenes aligned to a higher priority database would not be aligned to lower priority database. The alignments ended when all alignments are finished. Proteins with highest ranks in blast results were taken to decide the coding region sequences of Unigenes, then the coding region sequences were translated into amino sequences with the standard codon table. So both the nucleotide sequences (5' - 3') and amino sequences of the Unigene coding region were acquired. Unigenes that cannot be aligned to any database were scanned by ESTScan, producing nucleotide sequence (5'-3') direction and amino sequence of the predicted coding region.

2.2.5.2.8 Unigene Expression Difference analysis

The calculation of Unigene expression used RPKM method (Reads Per kb per Million reads) (Mortazavi *et al.*, 2008), the formula is shown below:

$$RPKM = \frac{10^6 \times C}{NL/10^3}$$

Set RPKM to be the expression of Unigene A, and C to be number of reads that uniquely aligned to Unigene A, N to be total number of reads that uniquely aligned to all Unigenes, and L to be the base number in the CDS of Unigene A. The RPKM method is able to eliminate the influence of different gene length and sequencing level on the calculation of gene expression. Therefore the calculated gene expression can be directly used for comparing the difference of gene expression between samples.

A rigorous algorithm was used to identify differentially expressed genes between two samples.

The null hypothesis and alternative hypothesis to identify expressed genes between two samples are defined as following:

H0: a gene has same expression level in two samples

H1: a gene has different expression level in two samples

X is denoted as number of reads that can uniquely map to gene A. For each transcript representing a small fraction of the library, p(x) will closely follow the Poisson distribution.

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad (\lambda \text{ is the real transcripts of the gene})$$

The total clean read number of the sample 1 is N1, and total clean read number of

sample 2 is N2; gene A holds x reads in sample1 and y reads in sample 2. The probability of gene A expressed equally between two samples can be calculated with the following formula:

$$p(i|x) = \left(\frac{N_2}{N_1}\right)^i \frac{(x+i)!}{x!i!(1+\frac{N_2}{N_1})^{(x+i+1)}}$$

FDR (False Discovery Rate) control is a statistical method used in multiple hypotheses testing to correct for p-value. In practical terms, the FDR is the expected false discovery rate; for example, if 1000 observations were experimentally predicted to be different, and a maximum FDR for these observation was 0.1, then 100 out of these observations would be expected to be false discovered. FDR was used with the ratio of RPKMs of the two samples at the same time.

The smaller the FDR is and the larger the ratio is, the larger difference of the expression level between the two samples is. In our analysis, we choose those with FDR ≤0.001 and ratio larger than 2 differentially expressed unigenes (DEG) were then carried out into GO functional analysis and KEGG Pathway analysis.

2.2.5.2.9 *Pathway enrichment analysis*

This analysis is divided into two parts: Gene Ontology (GO) functional analysis (molecular function, cellular component and biological process) and KEGG Pathway Analysis.

First, mapping all differentially expression genes to each term of Gene Ontology database (http://www.geneontology.org/) and calculating the gene numbers each GO term has. We get a gene list and gene numbers for every certain GO term, then using

hypergeometric test to find significantly enriched GO terms in DEGs comparing to the genome background.

Different genes usually cooperate with each other to exercise their biological functions. Pathway-based analysis helps to further understand genes biological functions. KEGG is the major public pathway-related database. Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing with the whole genome background. After multiple testing correction, we chose pathways with Q value$\leq$0.05 are significantly enriched in DEGs. With Pathway significantly enrichment, we can test biochemical pathways and signal transduction pathways which DEGs take part in.

2.3 Results

2.3.1 Soil bacteria community change compared between treated and air soil sample

The soil bacterial community structure was detected using the method mentioned above. Results were collected and grouped into three hierarchies: phylum, class and order. In air treated soil sample, Actinobacteria occupied the largest phylum (26.60%) followed by Proteobacteria (26.03%), Acidobacteria (10.44%) and Planctomycetes (8.70%) (Fig. 5). The majority of Proteobacteria consisted of four classes: Alphaproteobacteria (10.44%), Deltaproteobacteria (7.94%), Betaproteobacteria (4.86%) and Gammaproteobacteria (2.42%) (Fig. 6). In Actinobacteria (26.60%), Actinobacteridae occupied 15.97%, whereas Rubrobacteridae and Acidimicrobidae took 5.77% and 3.85%

respectively. In *Alphaproteobacteria* (10.44%) (Fig. 7), the population in this class was mainly *Rhizobiales* (5.65%), the rest were *Rhodospirillales* (2.11%) and *Sphingomonadales* (1.91%). *Betaproteobacteria* (4.86%) in air treated soil were composed of *Burkholderiales* (2.32%) and *Nitrosomonadales* (1.18%). In *Gammaproteobacteria*, *Xanthomonadales* (2.13%) was the overwhelming majority. *Deltaproteobacteria* was mainly composed of *Myxococcales* (5.80%).

In H_2 treated soil sample, *Proteobacteria* occupies almost half (45.35%) of the microbial population classified by phylum (Fig. 8). The second largest phylum in H_2 treated soil was *Actinobacteria* (13.20%) followed by *Acidobacteria* (9.32%), *Planctomycetes* (8.98%) and *Chloroflexi*(4.38%). *Betaproteobacteria* was the largest group (17.75%) in *Proteobacteria*. *Deltaproteobacteria* occupied 16.65% of total microbial population. *Alphaproteobacteria* and *Gammaproteobacteria* took 8.80% and 1.71% respectively (Fig. 9).   In the class of *Actinobacteria* (13.20%), the major order was *Actinobacteridae* (9.17%) followed by *Rubrobacteridae* (2.19%) and *Acidimicrobidae* (1.45%) (Fig. 10). *Burkholderiales* from *Betaproteobacteria* class was still the largest group in the whole community; the other order from the same class was *Nitrosomonadales* (1.37%). In *Alphaproteobacteria* (8.80%), the majority was *Rhizobiales* (5.72%) followed by *Rhodospirillales* (1.36%). *Xanthomonadales* (1.45%) was still the major order in *Gammaproteobacteria* class (1.71%). *Deltaproteobacteria* class was mainly composed of *Myxococcales* (13.99%).

Figure 5 Microbial community structure in air treated soil (Phylum)

Microbial community structure in air treated soil (phylum)

Legend:
- Actinobacteria
- Proteobacteria
- Acidobacteria
- Planctomycetes
- Chloroflexi
- Gemmatimonadetes
- Verrucomicrobia
- Firmicutes
- NA
- Bacteroidetes
- Nitrospirae
- Crenarchaeota

Labels: 1.99%, 1.42%, 1.24%, 2.69%, 2.76%, 2.89%
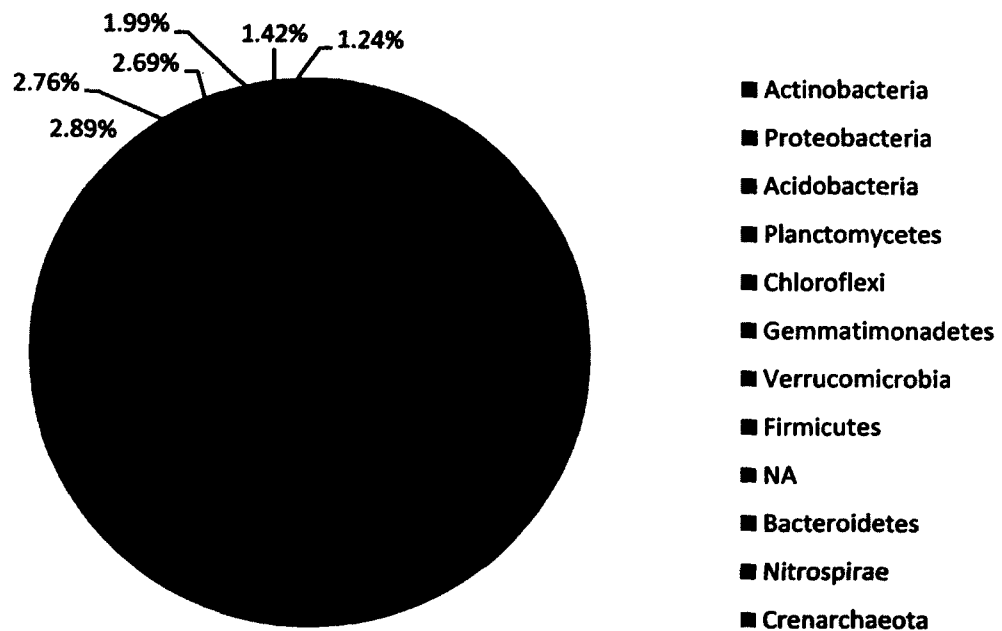
Figure 6 Microbial community structure in air treated soil (class)

Microbial community structure in air treated soil (class)

- Actinobacteria
- Acidobacteria
- Alphaproteobacteria
- Deltaproteobacteria
- Planctomycetacia
- Betaproteobacteria
- Gemmatimonadetes
- Phycisphaerae
- Gammaproteobacteria
- OPB35_soil_group
- KD4-96
- Sphingobacteria
- Bacilli
- Nitrospira
- Clostridia
- Soil_Crenarchaeotic_Group(SCG)
- TK10

1.43% 1.26% 1.23% 1.20%
1.42%
1.47% 1.46%
1.64%
2.42%
2.79%
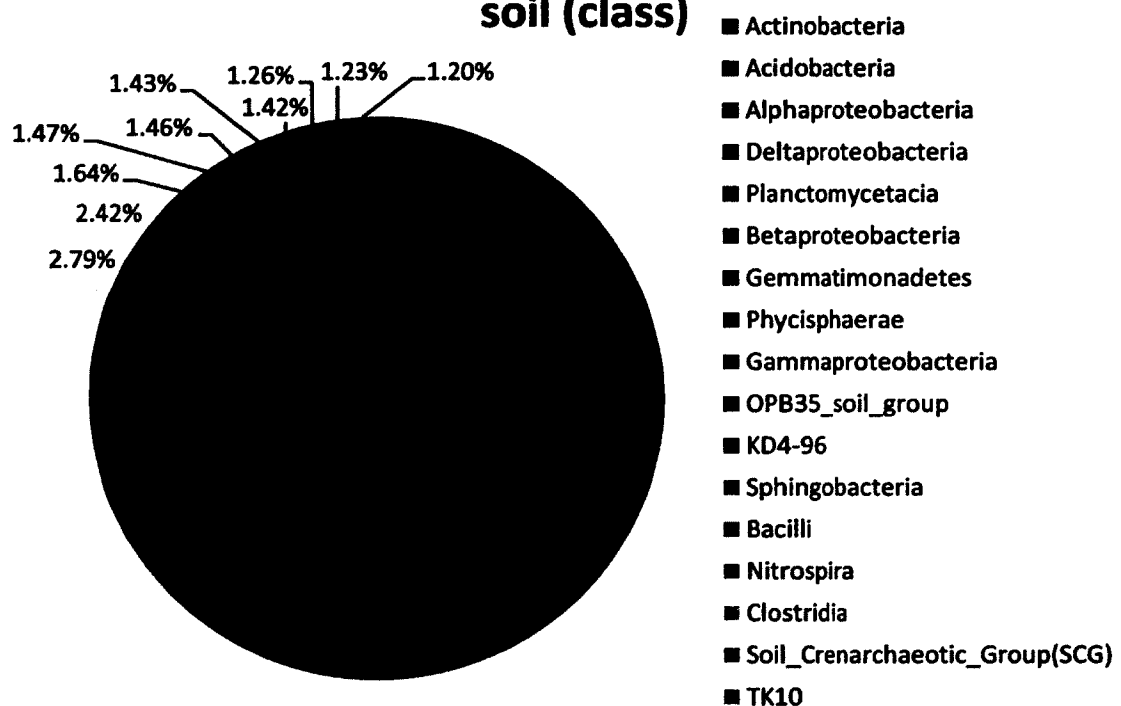
Figure 7 Microbial community structure in air treated soil (order)

Microbial community structure in air treated soil (order)

1.43% 1.25% 1.18% 1.11%
1.42%
1.91% 1.46%
2.11%
2.13%
2.32%

- Actinobacteridae
- Myxococcales
- Rubrobacteridae
- Rhizobiales
- Planctomycetales
- DA023
- Gemmatimonadales
- Acidimicrobidae
- WD2101_soil_group
- Burkholderiales
- Xanthomonadales
- Rhodospirillales
- Sphingomonadales
- Sphingobacteriales
- Bacillales
- Nitrospirales
- Clostridiales
- Nitrosomonadales
- Candidatus_Solibacter
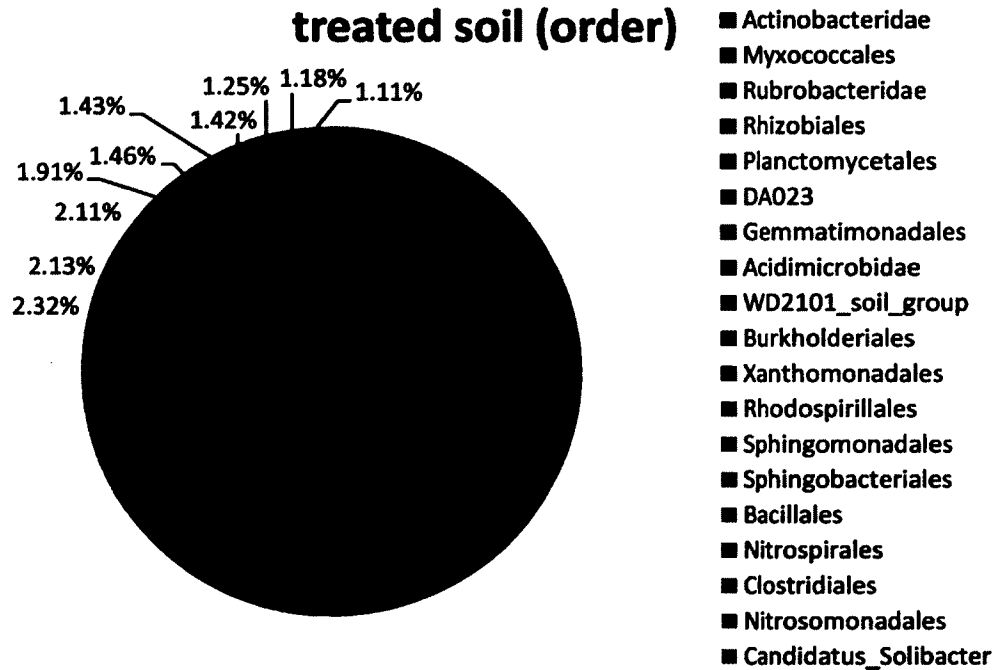
Figure 8 Microbial community structure in H₂ treated soil (Phylum)

Microbial community structure in H$_2$ treated soil (Phylum)

2.38% 2.15%

3.60% 2.85%

4.28%

4.38%

- ■ Proteobacteria
- ■ Actinobacteria
- ■ Acidobacteria
- ■ Planctomycetes
- ■ Chloroflexi
- ■ Verrucomicrobia
- ■ NA
- ■ Firmicutes
- ■ Bacteroidetes
- ■ Gemmatimonadetes
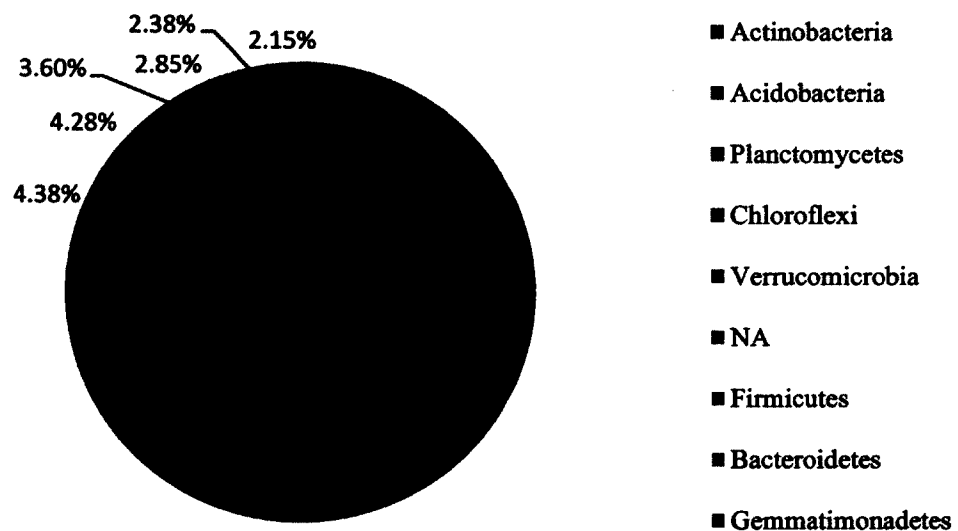
Figure 9 Microbial community structure in H$_2$ treated soil (Class)

Microbial community structure in H_2 treated soil (Class)

1.71%
1.77%
2.15%
2.28%
1.34%
1.30%
1.13%
1.02%

- Betaproteobacteria
- Deltaproteobacteria
- Actinobacteria
- Alphaproteobacteria
- Acidobacteria
- Planctomycetacia
- OPB35_soil_group
- Gemmatimonadetes
- Bacilli
- Gammaproteobacteria
- Spartobacteria
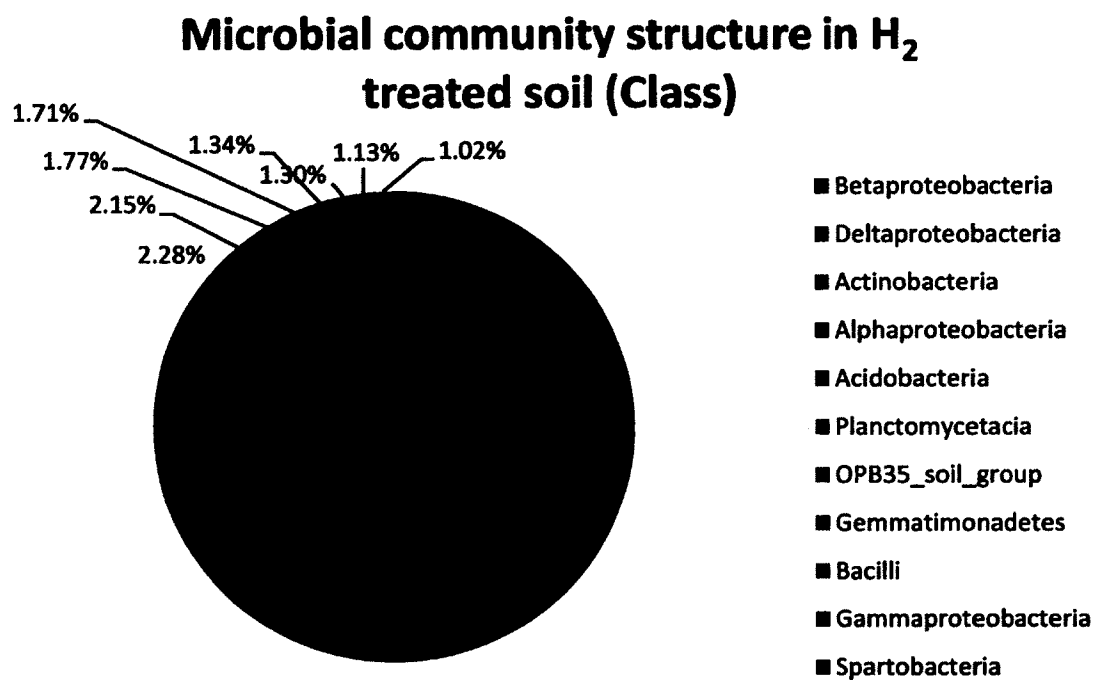
Figure 10 Microbial community structure in H$_2$ treated soil (order)

Microbial community structure in H$_2$ treated soil (order)

1.34% 1.30% 1.09% 1.01%
1.36%
1.37%
1.45%
1.45%
1.76%
2.02%

- Burkholderiales
- Myxococcales
- Actinobacteridae
- Planctomycetales
- Rhizobiales
- DA023
- Rubrobacteridae
- Gemmatimonadales
- Bacillales
- Acidimicrobidae
- Xanthomonadales
- Nitrosomonadales
- Rhodospirillales
- Chthoniobacterales
- Sphingobacteriales
- Candidatus_Solibacter
- Clostridiales
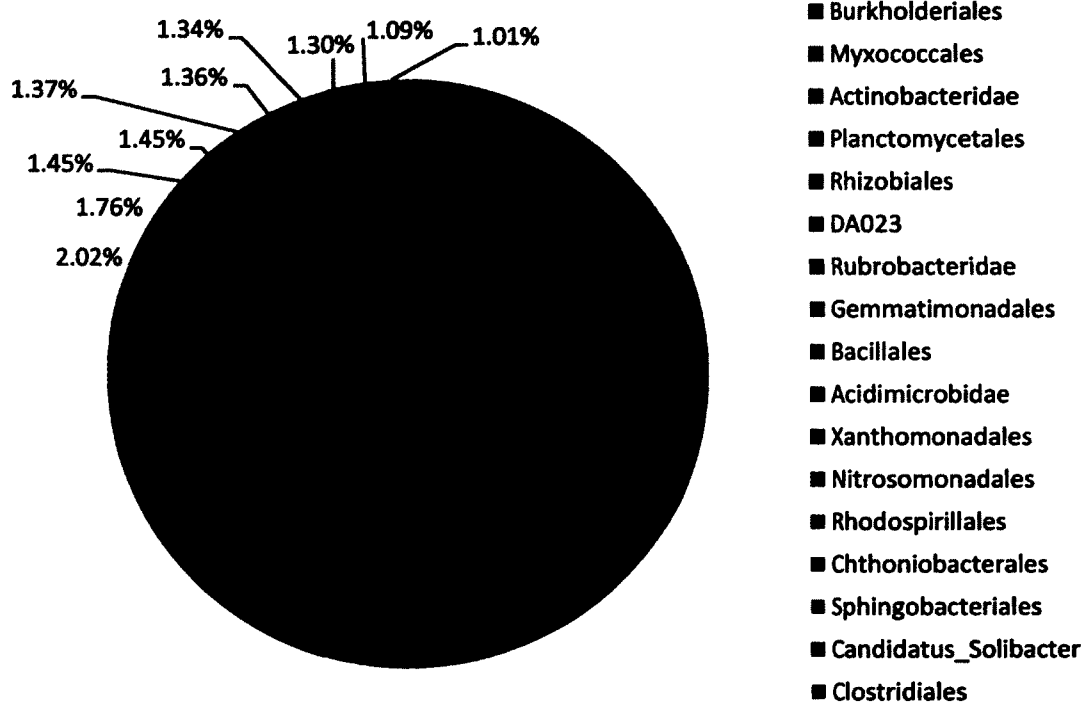
2.3.2 Gene expression analysis

All RNA samples were filtered through the protocol. Clean reads were selected after filtering (Table. 1). 10907 pieces of unigenes were aligned from all clean reads. 10694 pieces of unigenes were sequenced, annotated with multiple online databases. After annotation, 4024 pieces of total unigenes were successfully matched up through all databases. Unigene expression difference analysis was carried out among all sequenced unigenes. 5445 pieces of unigenes were up-regulated after H_2 treatment, whereas 5249 pieces were down-regulated. After filtering by picking up all Differentially Expressed Genes (DEGs) within the range of FDR$\leq$0.001 AND |log2Ratio|$\geq$1, 1262 pieces of unigenes were considered as up-regulated, whereas 2142 pieces of unigenes were down-regulated and 7290 pieces were considered as non-DEGs (Fig. 11). The percentage of differentially expressed unigenes in total unigenes was relatively higher than usual. That is because the amount of H_2 provide to soil in our treatment was higher than the amount that legume root nodule can release to rhizosphere. Therefore, we can ensure that all bacteria in our soil could be fully treated by H_2 and their gene expression differences could be maximized as well. However, regardless of the restricted condition of DEG. We combined the annotation file with unigene expression difference file containing over 10000 pieces of unigenes together for further analysis. In the new file, 2463 pieces of unigenes were up-regulated and 1561 pieces were down-regulated.

In our research, unigenes that were up-regulated after H_2 treatment were paid the most attention, because it was more likely that those expression differences were induced by H_2. We did compare some unigenes that had the most significant decrease of gene

expression in H_2 treated soil sample (Table. 2), but most of those products cannot even be identified through annotation. Hence, it seems to be harder to seek direct evidence of the relationship between H_2 and bacterial metabolism through down regulated genes.

All annotated unigenes were classified by GO function and divided into three ontologies: molecular function, cellular component and biological process (Fig. 12)

According to the GO function classification, metabolic process (12.85% of total unigenes) and cellular process (10.05%) from biological process, cell (13.91%) and cell part (12.75%) from cellular component and binding (14.30%) and catalytic activity (13.43%) occupied the most unigenes in both samples.

Table 1 Output statistics of sequencing for both YZH and YZA RNA samples

Samples	Total Reads (M)	Total Nucleotides (Gb)	Q20 percentage	N percentage	GC percentage +	No duplication (M)
YZA	24.33	2.43	95.90%	0.01%	55.84%	17.68
YZH	24.38	2.44	96.34%	0.01%	53.88%	16.66

* Total Nucleotides = Total Reads1 x Read1 size + Total Reads2 x Read2 size; Total Reads and Total Nucleotides are actually clean reads and clean nucleotides; Q20 percentage is proportion of nucleotides with quality value larger than 20; N percentage is proportion of unknown nucleotides in clean reads; GC percentage is proportion of guanidine and cytosine nucleotides among total nucleotides; No duplication is clean reads without duplication

Figure 11 Distribution of comparison of all differentially expressed genes
(FDR$\leq$0.001 AND |log2Ratio|$\geq$1)

Expression Level YZA vs YZH

FDR ≤ 0.001 AND |log2Ratio| ≥ 1
■ up-regulated genes
■ down-regulated
■ Not DEGs
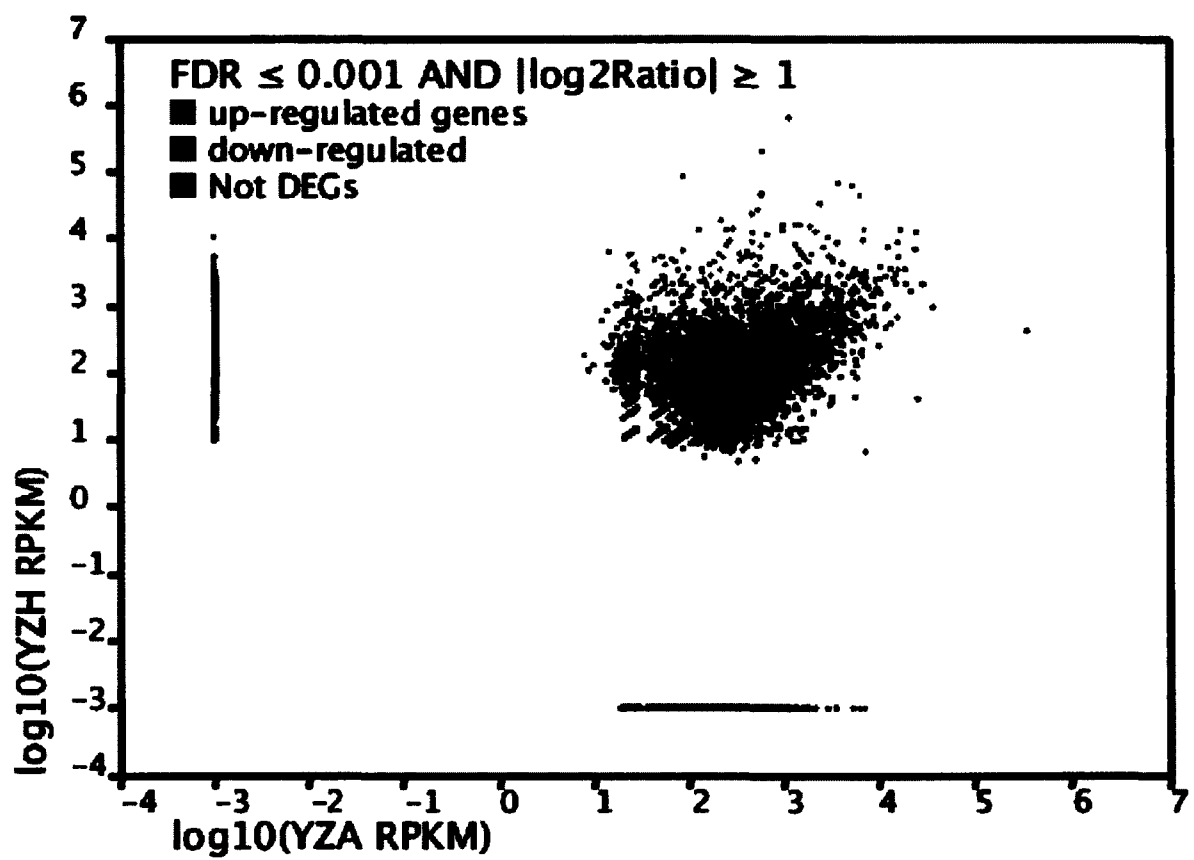
Table 2 A table of 20 unigenes that have the most significant decrease of gene expression in H₂ treated soil sample

Product Name	Average expression difference (RPKM)
conserved domain protein	20105.9931
minus strand conserved hypothetical protein	17281.9326
RecName: Full=Protein TAR1	17430.1591
conserved hypothetical protein	12437.6439
conserved hypothetical protein	12231.5773
hypothetical protein	10877.8780
minus strand hypothetical protein	9357.2686
minus strand conserved hypothetical protein	7069.7687
minus strand hypothetical protein DR_0254	7049.1886
minus strand hypothetical protein AcavDRAFT_4806	6825.1772
RNA-dependent RNA polymerase	6679.2828
senescence-associated protein	6673.7277
minus strand conserved hypothetical protein	6492.5951
minus strand hypothetical protein ACLA_028940	6163.4525
minus strand conserved hypothetical protein	5936.2723
minus strand LOW QUALITY PROTEIN: retrotransposon protein	5898.1397
LOW QUALITY PROTEIN: conserved hypothetical protein	5316.7500
minus strand cell wall-associated hydrolase	5012.1396
minus strand RNA-dependent RNA polymerase	4687.7705
minus strand LOW QUALITY PROTEIN: conserved hypothetical protein	4675.4713

Figure 12 GO classification of differentially expressed unigene

Percent of genes

biological regulation
cell proliferation
cellular component organization or biogenesis
cellular process
developmental process
establishment of localization
localization
locomotion
metabolic process
multi-organism process
multicellular organismal process
regulation of biological process
response to stimulus
signaling
signaling process

cell
cell part
extracellular region
extracellular region part
macromolecular complex
membrane-enclosed lumen
organelle
organelle part
virion
virion part

binding
catalytic activity
molecular transducer activity
protein binding transcription factor activity
transcription regulator activity
transporter activity

biological_process

cellular_component

molecular_function

Number of genes

YZA-vs-YZH

2.3.2.1 Carbon metabolism

Unigenes encoding ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco), fructose-bisphosphate aldolase, ribulose-phosphate 3-epimerase, glyceraldehyde 3-phosphate dehydrogenase, fructose-1,6-bisphosphatase, transketolase and phosphoheptose isomerase enzymes were found being up-regulated. Those proteins were believed being involved with Calvin–Benson-Bassham cycle (Calvin cycle). 35 pieces of Rubisco genes were found more active in H_2 treated soil compared to control. The most significant difference of gene expression that related with carbon metabolism is coming from Rubisco, which has 6945.707 more reads per kb per million reads (RPKM) in H_2 treat sample than that in control (Table. 3, Fig. 13).

2.3.2.2 Hydrogen metabolism

Hydrogenase gene, the key enzyme that catalyzes the H_2 oxidation reaction, was found in 35 different unigenes and has a cumulative RMPM value of 3669.396. Two other genes encoding potential electron receptors from H_2, ferredoxin—NADP (+) reductasewere and NADH dehydrogenase, were observed. The expression difference in cumulation of NADH dehydrogenase (1996.779 RPKM) was lower than that of hydrogenase genes. But, ferredoxin—NADP (+) reductasewere didn't share the similar differential (89.0881 RPKM) with NADH dehydrogenase genes (Table. 4, Fig. 14).

2.3.2.3 Electron transport chain

Several unigenes encoding proteins that potentially associated with prokaryotic respiration chain are stimulated by H_2 treatment (Table. 5, Fig. 15). Cytochrome c and

cytochrome c oxidase are most active proteins in this group. Others are relatively quiet but still express more in H_2 treated soil sample.

2.3.2.4   Other genes

Nitrite reductase is the enzyme that was observed with increase of gene copies in H_2 treated soil samples compared to air treated soil (He, 2010). In our study, the gene expression result agrees with the conclusion by He (Table. 6, Fig. 16). However, the fluency (2) is relatively low which means there are only two species of bacteria being more active in the expression of nitrite reductase genes.

There are a few genes with names that are hardly functionally recognized, but their expression activities are too significant to ignore, such as Tar1p gene, hypothetical protein CaO19.6835, hypothetical protein AnaeK_3670 and hypothetical protein TSTA_040370 (Table. 6, Fig. 16). The expression difference of hypothetical protein CaO19.6835 gene in accumulation was extremely higher (109362.958 RPKM) than any other gene found via metatranscriptome analysis.

Table 3 A table of frequency of occurrences and expression difference of genes
related to carbon metabolism in H_2 treated sample compared to air treated
soil sample

Product Name	Frequency of occurrence	Cumulative reads in air treated sample (RPKM)	Cumulative reads in H₂ treated sample (RPKM)	Cumulative expression difference (RPKM)	Average expression difference (RPKM)
Fructose-bisphosphate aldolase	6	0	686.1115	686.1115	114.3519
Transketolase	11	0	1504.6555	1504.6555	136.7869
Fructose-1,6-bisphosphatase	5	0	1011.8780	1011.8780	202.3755

Figure 13 Frequency of occurrences and cumulative expression difference of genes related tocarbon metabolism in H_2 treated sample compared to air treated soil sample

Frequency of occurences and expression difference of genes related with carbon metabolsm in H₂ treated sample compared to air treated soil sample

Frequency of occurrence ■   **Cumulative expression difference (RPKM)** ■
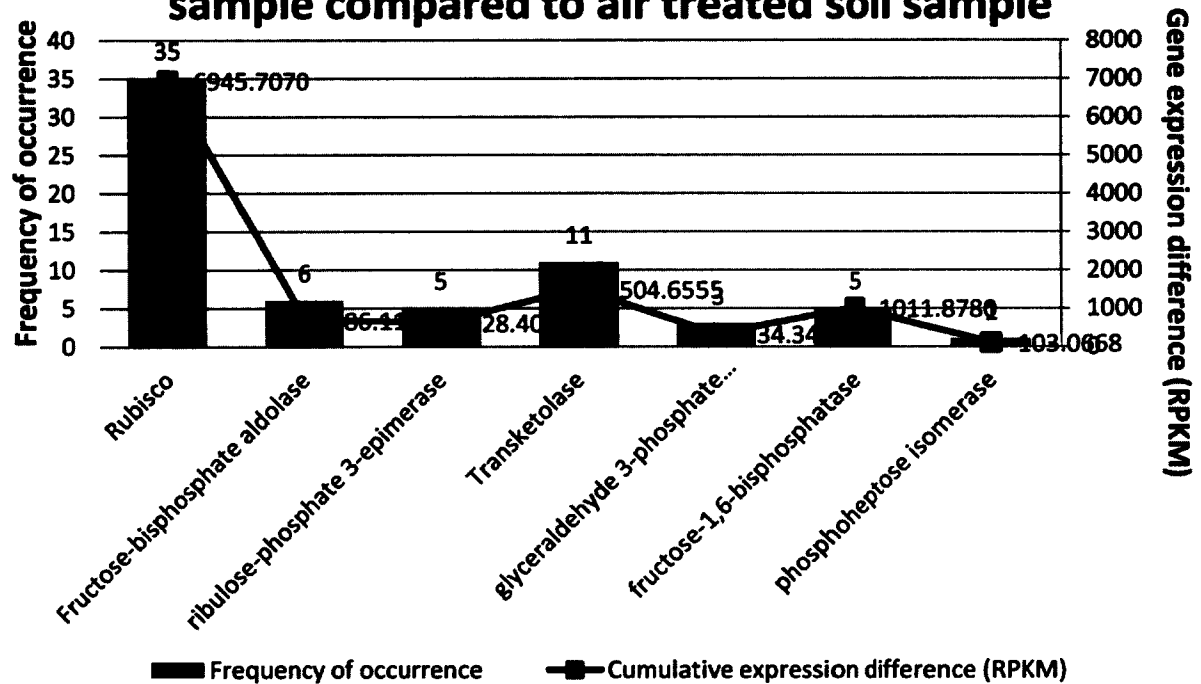
Table 4 A table of frequency of occurrences and expression difference of genes
related to hydrogen metabolism in H_2 treated sample compared to air treated
soil sample

Product Name	Frequency of occurrence	Reads in air treated sample (RPKM)	Reads in H_2 treated sample (RPKM)	Cumulative expression difference (RPKM)	Average expression difference (RPKM)
Ferredoxin—NADP(+) reductase	2	0	89.0881	89.0881	44.5441

Figure 14 Frequency of occurrences and cumulative expression difference of genes
related with hydrogen metabolism in H_2 treated sample compared to air
treated soil sample

Frequency of occurences and cumulativeexpression difference of genes related with hydrogen metabolsm in H_2 treated sample compared to air treated soil sample
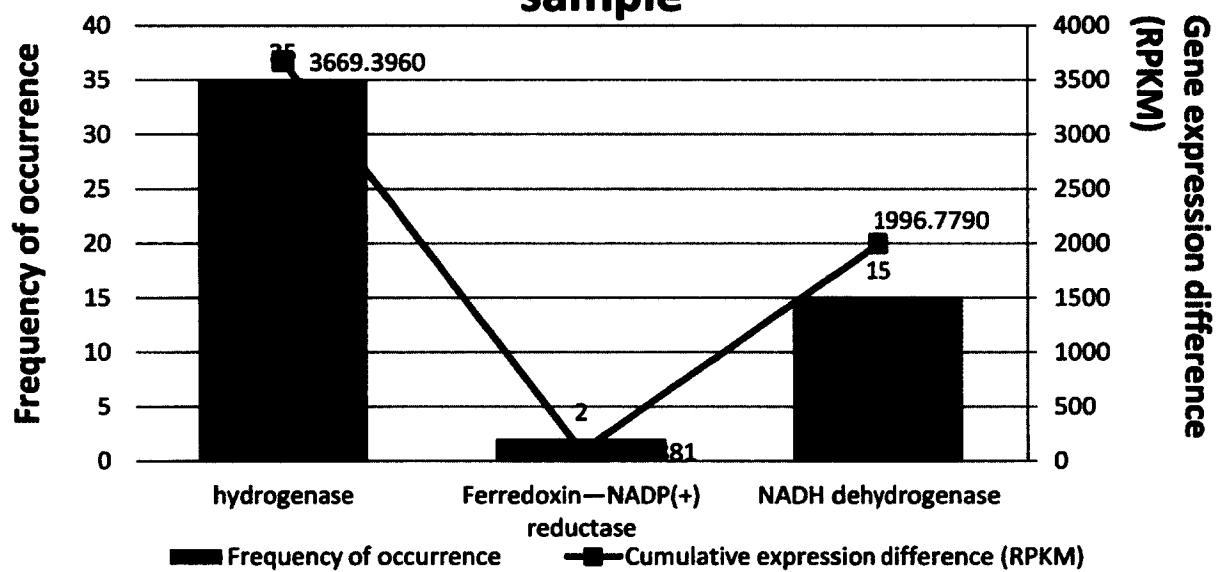
Table 5 A table of frequency of occurrences and expression difference of genes related with electron transport in H_2 treated sample compared to air treated soil sample

Product Name	Frequency of occurrence	Reads in air treated sample (RPKM)	Reads in H₂ treated sample (RPKM)	Cumulative expression difference (RPKM)	Average expression difference (RPKM)
Hydrogen:quinone oxidoreductase	3	0	210.9056	210.9056	70.3019
Electron transferring flavoprotein dehydrogenase	2	0	126.0715	126.0715	63.0358
NADH ubiquinone oxidoreductase	5	0	461.4256	461.4256	92.3851

Figure 15 Frequency of occurrences and cumulative expression difference of genes
related with electron transport in H_2 treated sample compared to air
treated soil sample

Frequency of occurences and cumulative expression difference of genes related with electron transportation in H₂ treated sample compared to air treated soil sample
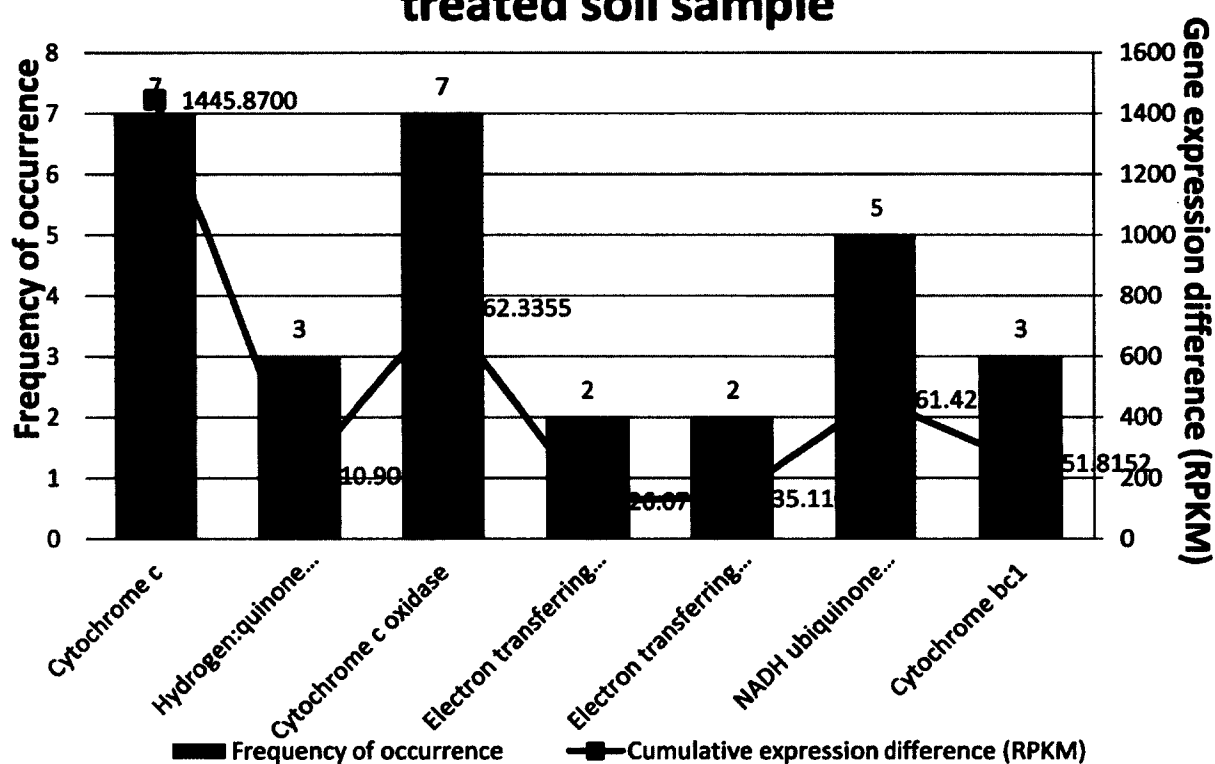
Table 6 A table of frequency of occurrences and expression difference of genes
remained unknown or related with plant growth promotion in H_2 treated
sample compared to air treated soil sample

Product Name	Frequency of occurrence	Reads in air treated sample (RPKM)	Reads in H₂ treated sample (RPKM)	Cumulative expression difference (RPKM)	Average expression difference (RPKM)
Hypothetical protein AnaeK_3670	36	0	3739.8180	3739.8180	103.8838
Hypothetical protein CaO19.6835	14	7442.2420	116805.2000	109362.9580	7811.6411

Figure 16 Frequency of occurrences and cumulative expression difference of genes remained unknown or related with plant growth promotion in H_2 treated sample compared to air treated soil sample

Frequency of occurences and cumulative expression difference of genes remained unknown or related with plant growth promotion in H_2 treated sample compared to air treated soil sample
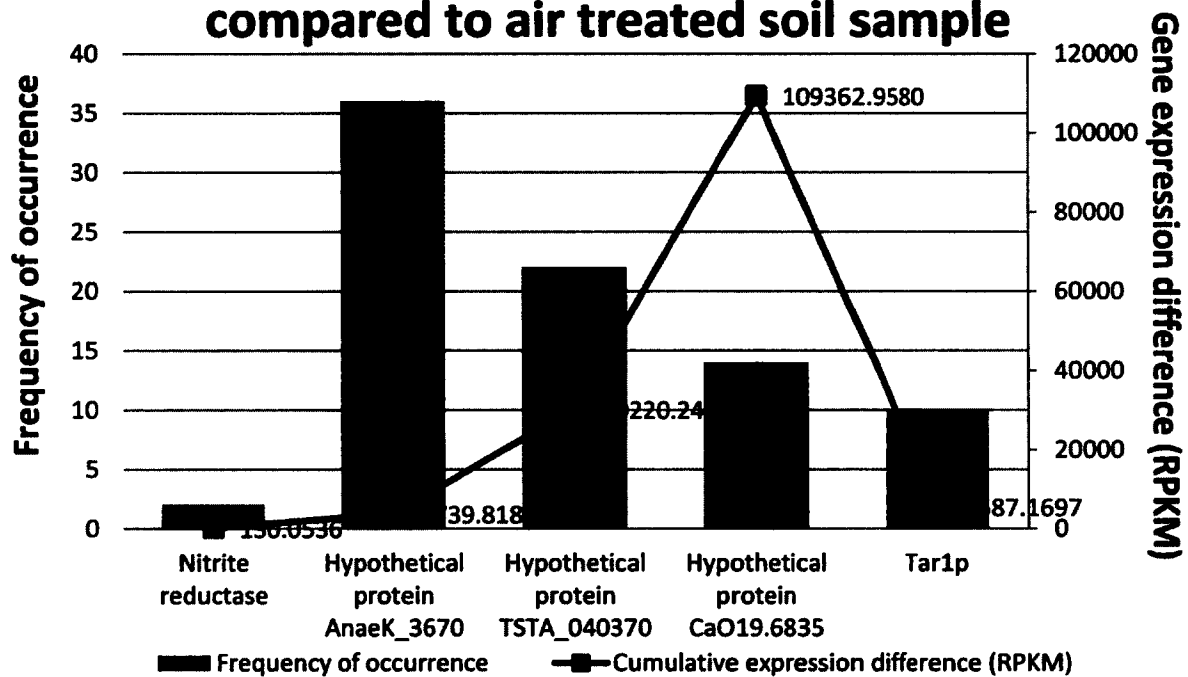
2.4 Discussion

2.4.1 Soil bacteria community change compared between H_2 and air treated soil samples

It was reported by Flynn (2010) that soil bacteria composition was always changing even after H_2 uptake rate reached its peak. In our study, instead of taking soil samples from the same date of treatment, we collected soil within a longer period to ensure the result of soil bacterial community change reflected an overall variability of the total bacterial population between H_2 treated and air treated soil.

After 4 week treatment, population of *Proteobacteria* was stimulated from 26.03% to 45.35% in phylum, whereas *Actinobacteria* and *Chloroflexi* were suppressed by H_2 treatment (Fig. 17). In *Proteobacteria* group, population of *Betaproteobacteria* and *Deltaproteobacteria* were enriched after H_2 treatment, but *Acidobacteria*, *Alphaproteobacteria* and *Gammaproteobacteria* experienced the decline of population (Fig. 18). *Burkholderiales* is the most significantly increased group in *Betaproteobacteria* (4.9% in air treatment, 17.8% in H_2 treatment). The other order, *Nitrosomonadales*, didn't change much (1.2% in air treatment, 1.4% in H_2 treatment). In *Deltaproteobacteria* class, the population enrichment is mostly due to *Myxococcales* (5.8% in air treatment, 14% in H_2 treatment) (Fig. 19).

According to results conducted by T-RFLP method using phylogenetic assignment tool (PAT) from previous study (He, 2010); H_2 treatment was able to induce the enrichment of *Proteobacteria*, particularly in classes of *Betaproteobacteria* and

Gammaproteobacteria. But the *Alphaproteobacteria* community was dramatically

impacted by H_2 to half size before the treatment. In addition, *Deltaproteobacteria* was

Figure 17 Soil bacterial community change induced by H_2 treatment (phylum)

Soil bacterial community change induced by H₂ treatment (phylum)
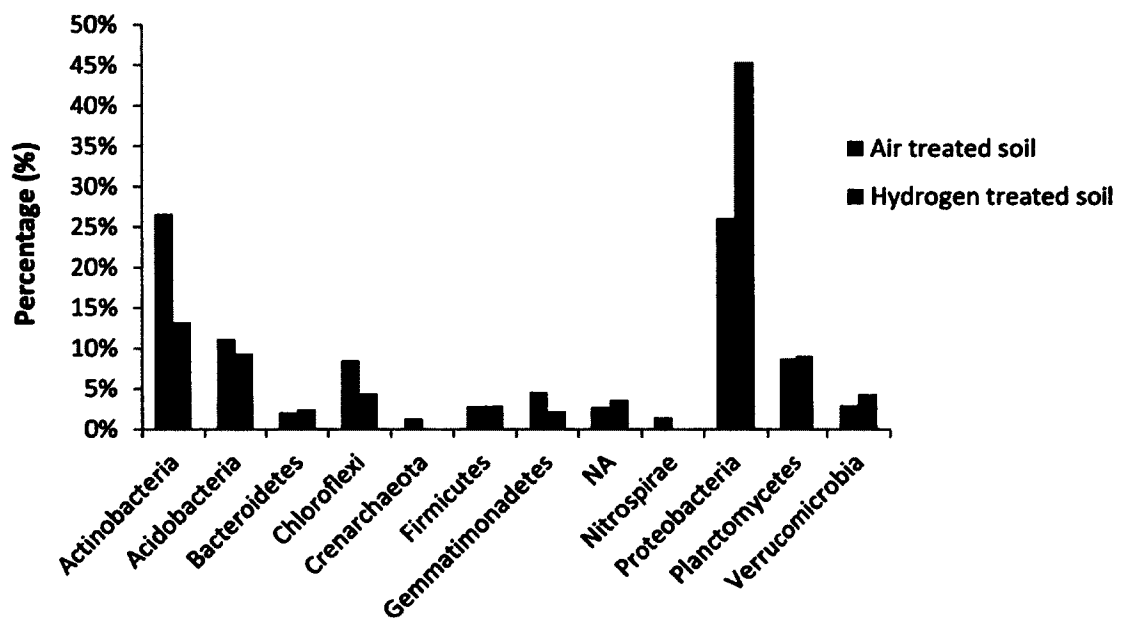
Figure 18 Soil bacterial community change induced by H₂ treatment (class)

Soil bacterial community change induced by H$_2$ treatment (order)

Percentage (%)

- Air treated soil
- Hydrogen treated soil

Acidobacteria
Actinobacteria
Alphaproteobacteria
Bacilli
Betaproteobacteria
Clostridia
Deltaproteobacteria
Gammaproteobacteria
Gemmatimonadetes
KD4-96
Nitrospira
OPB35_soil_group
Phycisphaerae
Planctomycetacia
Spartobacteria
Soil_Crenarchaeotic_G...
Sphingobacteria
TK10

Figure 19 Soil bacterial community change induced by H_2 treatment (order)

Soil bacterial community change induced by H₂ treatment (class)

■ Air treated soil   ■ Hydrogen treated soil
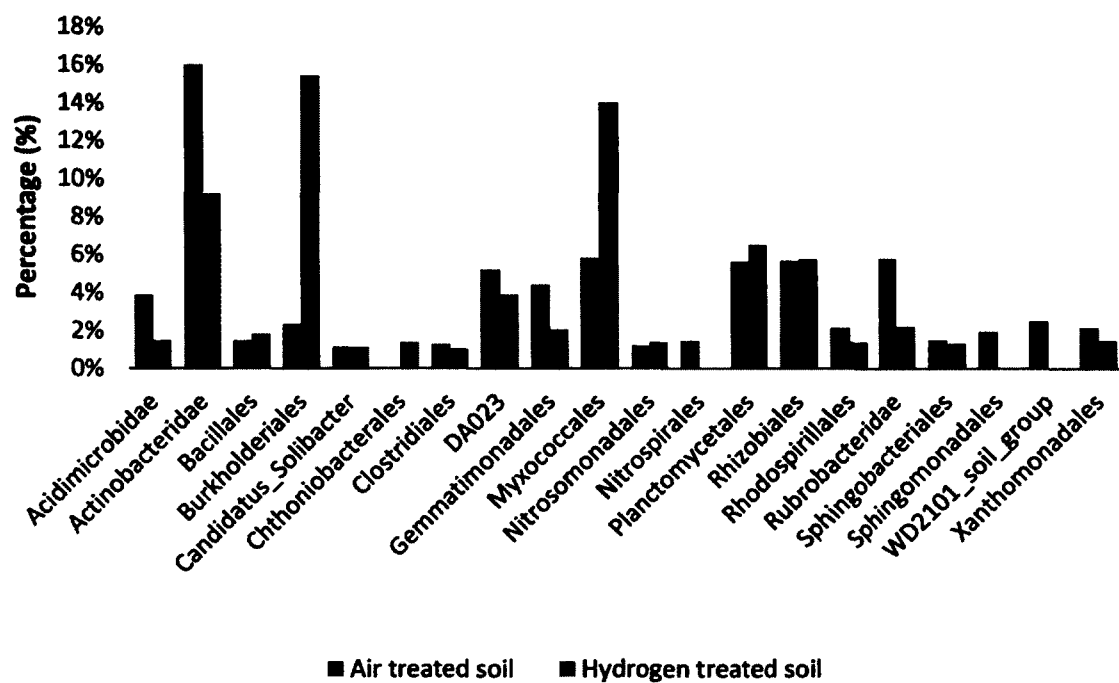
not even detected by T-RFLP and PAT. In our study, our results conducted from metatranscriptome analysis agreed with microbial community profiles of *Alphaproteobacteria* and *Betaproteobacteria*. Results from both methods revealed dramatic increase of population of *Burkholderiales* from *Betaproteobacteria*, which experienced almost 7 folds increase after H_2 treatment which is also supported by Maimaiti (2007). Inoculation test of *Burkholderiales* species also showed positive effects on plant growth. In *Deltaproteobacteria*, the population explosion should be ascribed to *Myxococcales*, which had been detected by T-RFLP and PAT analysis. Here we confirm that *Burkholderiales* order from *Betaproteobacteria* is one of the microbial factors of plant growth promotion and *Myxococcales* order from *Deltaproteobacteria* may have the same potential but needs further study.

It is important to mention that, in our study, all those changes of microbial community structure were obtained based on community composition. It only reflected the percentage of total population that one certain group of bacteria occupied in whole bacterial community in our soil. The result was unable to express the absolute change of population of any group of bacteria. Considering results from H_2 treated soil sample were obtained from a higher quantity of total RNA compared to the quantity from air treated soil sample. Here we suggest that our result of soil microbial community change was relative and the absolute change of each group of bacteria might be opposite to our results.

2.4.2 Gene expression analysis

2.4.2.1 Hydrogen related carbon metabolism (CO_2 assimilation) and electron transport chain (O_2 reduction)

Eight groups of genes that encode enzymes participating Calvin cycle pathways have been found very active in H_2 treated soil sample. Flynn (2010) also presented the significant stimulation of Rubisco gene after 4 week H_2 treatment. Genes encoding enzymes involved with other bacterial CO_2 assimilation pathways, such as reverse Krebs Cycle and Wood–Ljungdahl pathway, haven't been found any increase of gene expression in H_2 treated soil sample. Here we confirm that Calvin cycle is the major and probably the only pathway for CO_2 assimilation in H_2 oxidizing bacteria.

Although it has been clearly proved that Calvin cycle is the major pathway in H_2 oxidizing bacteria, the relationship between H_2 and Calvin cycle is still unknown. Usually H_2 acts as electron donors in microbial metabolism, therefore the only way that H_2 gets involved with CO_2 fixation, especially Calvin cycle, is providing electrons. Generally speaking, reduced nicotinamide adenine dinucleotide phosphate (NADPH) is the only direct electron donor in Calvin cycle. However, there is no direct evidence proving that H_2 can transfer electrons directly to NADP(+) in Calvin cycle or through an electron intermediate. We assume that both possibilities can happen in our soil sample. In our study, all hydrogenase genes, as the enzyme of first step reaction in H_2 related electron transport, were annotated by their chemical structure, it is impossible to determine the function of hydrogenase and potential electron acceptor *via* enzyme names. One possibility of how H_2 gets involved in Calvin cycle is that the electrons from H_2 are

transferred immediately to NADP(+) by hydrogenase (hydrogen dehydrogenase (NADP) or hydrogen: NADPH+ oxidoreductase). However, we prefer the hypothesis that H_2 added to soil in our study transfers electrons through electron intermediates. We observed that ferredoxin—NADP(+) reductase gene was stimulated by H_2 in soil, and this is the enzyme which transfer electrons to NADP(+) from reduced ferredoxin. Furthermore, among all types of hydrogenase proteins, there is one type catalyzing conversion of H_2 and oxidized ferredoxin to H+ and reduced ferredoxin, which can connect H_2 and NADPH reasonably in our study. Therefore, we think it is more likely that electrons from H_2 are transferred to reduced ferredoxin by membrane bound hydrogenase as an electron intermediate and are passed to NADP(+) by ferredoxin—NADP(+) reductase (Fig. 20). Our data showed that NADH dehydrogenase was up regulated by H_2 treatment. This implied that activities with NADH performed very active in H_2 treated soil. For Calvin Cycle in bacteria, there is a chance that NADH may be involved with the production of organic compounds at the level of hexoses (Lengeler et al., 2009). This possibility needs to be confirmed in the future.

The relationship between H_2 and respiration chain (O_2 reduction) seems more obvious than that between H_2 and CO_2 fixation. It has been reported that, in *Bradyrhizobium Japonicum*, ubiquinone (UQ) is participating the transfer of electrons from H_2 (Eisbrenner and Evans, 1982). Furthermore, cytochromes *o* and *aa3* as terminal oxidases, cytochromes *b* and *c* types as electrons carriers, and UQ as a hydrogen carrier are also identified in the same species. This is also the case for aerobic H_2 oxidizing bacteria (Evan et al., 1987). However, in *Xanthobacter autotrophicus*, cytochrome *b* acts as proximal electron acceptor for hydrogenase *in vivo* (Schink, 1982). Based on result

from metatranscriptome analysis, we hypothesize that electrons from H_2 can be passed to O_2 *via* multiple electron transport pathways. One of the pathway is that H_2 pass electrons to NAD(+) via hydrogenase at the first step. NAD(+) becomes NADH when it receives two electrons from H_2 and transfer electrons to ubiquinone. Electrons from ubiquinone are transferred by cytochrome *bc1* to cytochrome *c* and eventually reach O_2 via cytochrome *c* oxidase. However, previous studies suggested that NADH was not involved with the H_2 oxidizing bacterial respiration chain (Bernard and Schlegel, 1974; O'Brian and Maier, 1982; Ferber *et al.*, 1995). Furthermore, one type purified heterodimeric hydrogenase from *B. japonicum* was demonstrated having capacity of H_2-ubiquinone oxidoreductase, which indicates the possibility that ubiquinone can receive electrons from H_2 by hydrogenase. Therefore we present another potential pathway of H_2-electron transport chain in H_2 oxidizing bacteria. Electrons from H_2 are immediately transferred to ubiquinone by hydrogenase and flow to O_2 *via* cytochrome *bc1*, cytochrme *c* and cytochrome oxidase (Fig. 21). Despite of two potential pathways mentioned above, there probably is the third pathway in H_2 oxidizing bacteria. Genes that encode electron-transferring flavoprotein (ETF) and electron-transferring flavoprotein dehydrogenase (ETF dehydrogenase) are observed more active in H_2 treated soil sample. ETF dehydrogenase functions as transferring electrons from ETF to ubiquinone. It was reported by Ghisla and Thorpe (2004) that ETF and ETF dehydrogenase were participating respiratory chain. However, we lack the evidence of presence of any enzyme that enables the electron transport between H_2 and ETF. Hence, this pathway needs further study.

2.4.2.2 Other genes and potential of new genes

Nitrite reductase is the enzyme that catalyzes nitrite reduction in denitrification process. Previous study conducted by He (2010) demonstrated that the gene copies of soil after H_2 treatment increased significantly, which was concluded that most H_2-oxidizing bacteria may contain denitrification gene or they are denitrifiers to produce N_2O. In our study, the nitrite reductase gene expression difference was observed between H_2 treated and air treated soil sample. This provided the evidence that H_2 is able to induce the biological denitrification process. However, only two pieces of nitrite reductase gene were found more active in H_2 treated soil and the gene expression difference was not that significant. More specific study of RNA level of this particular gene needs to be conducted in order to get a better understanding of these results.

Figure 20 Potential H$_2$-dependent CO$_2$ fixation in H$_2$ oxidizing bacteria

Potential H$_2$-dependent CO$_2$ fixation in H$_2$ oxidizing bacteria
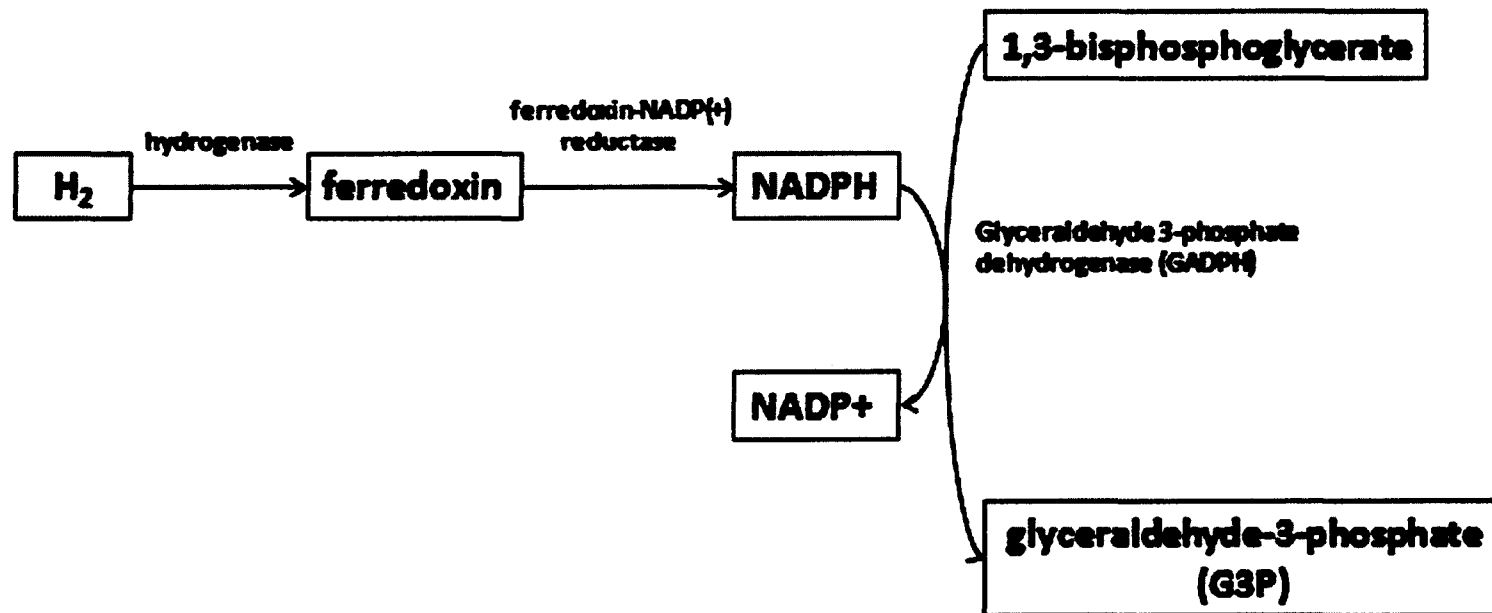
Reduction phase of Calvin Cycle

Figure 21 potential H_2-dependent electron transport chain in H_2 oxidizing bacteria

Potential H$_2$-dependent electron transportation chain in H$_2$ oxidizing bacteria
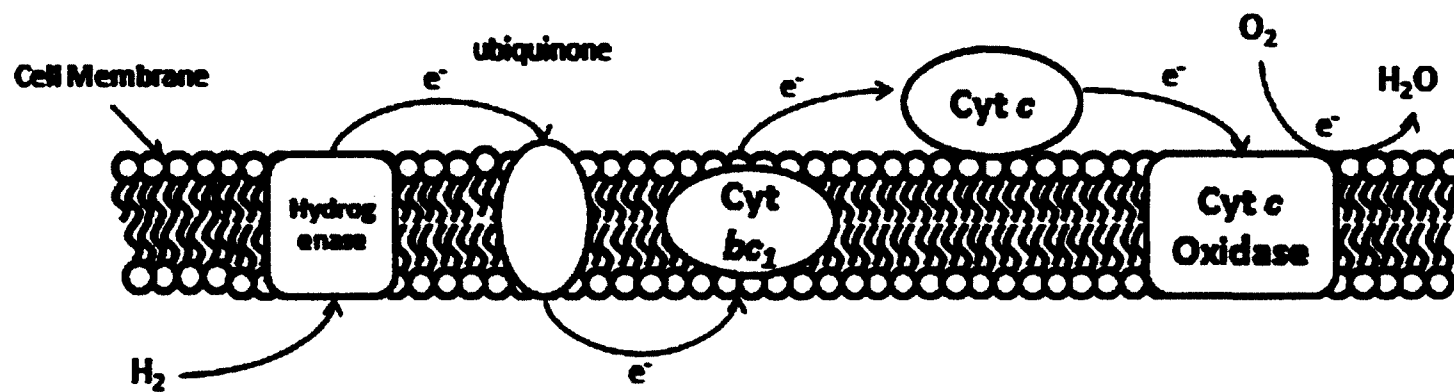
Despite of genes associated with bacterial CO_2 fixation and O_2 reduction, according to the result of differentially expressed gene analysis, many other genes were extremely active after H_2 treatment, but they were annotated with hypothetical protein or names that cannot be identified, such as hypothetical protein AnaeK_3670, hypothetical protein TSTA_040370, hypothetical protein CaO19.6835 and Tar1p. Some of those genes had showed the most significant difference at the gene expression level between two samples. We assume that induction of gene expression about those genes after H_2 treatment should have a close connection with increase of population of H_2 oxidizing bacteria. Unfortunately, their roles in biological activities in H_2 oxidizing bacteria are still unknown. Further study needs to focus on those gene activities.

3. Hydrogenase gene study

3.1 Introduction

Legumes are plants that can fix N_2 because of the mutualistic symbiotic relationship with rhizobia found in root nodules of legume. Well-known legumes include alfalfa, clover, peas, beans, lentils, lupins, mesquite, carob, soy, and peanuts. Rhizobia are soil bacteria that have the capacity of fixing N_2 from the air and convert it into other nitrogen-containing derivatives which can be absorbed by legumes. This ability of N_2 fixation has been used in crop rotation to enhance both growth and yield for millennia.

In recent years, it was found that H_2 is produced as a byproduct during the N_2 fixation process (Simpson and Burris, 1984). In addition, H_2 can diffuse out of some nodules which lack the capacity of H_2 uptake (HUP-) in rhizobia (Uratsu et al., 1982). Conrad and Seiler (1979) indicated that H_2 produced in nodules was consumed by the "soil". Later on, it was reported that the activity of H_2 oxidation in soil was accompanied with the increase of microbial biomass in soil (Popelier et al., 1985). McLearn and Dong (2002) demonstrated that autoclaving could remove the ability of H_2 uptake of the "soil". In addition, antibiotics (penicillin with streptomycin) were reported having more significant capacity of inhibiting H_2 uptake ability compared to fungicide (benomyl) in H_2-treated soil. Hence, the evidence suggests that it is the H_2-oxidizing bacteria that take up the H_2 produced from biological N_2 fixation.

There are variety species of H_2 oxidizing bacteria in the soil, most of them are aerobic. Although some of those bacteria had been well studied, there are still many of those remained to be determined (Bowien and Schlegel, 1981). In taxonomy studies

using molecular methods, 16s rRNA gene is normally considered as a good target gene for cloning and sequence analysis which is available to study unknown bacteria. In my study, due to the phylogenetic diversity of aerobic H_2-oxidizing bacteria, the hydrogenase genes could be another choice for DNA cloning and quantitative analysis. Hydrogenases are metalloproteins which are located in either the cytoplasm or the periplasm of bacteria. Hydrogenase can catalyze the reversible heterolytic cleavage of molecular hydrogen:

$$H_2 \leftrightarrow 2H^+ + 2e^-.$$

Usually hydrogenase can be classified into one of three types based on the chemical clusters and metal atoms contained in the center of enzyme (Vignais and Colbeau 2004). NiFe hydrogenase consists of Fe-S clusters and NiFe complex. FeFe hydrogenase contains the same clusters as NiFe hydrogenase but FeFe complex in the center (Vignais and Colbeau 2004). The third type of hydrogenase is called "Fe-S clusters free" hydrogenase since it is dependent on a Fe cofactor instead of Fe-S clusters, which is obviously different from those previous ones (Vignais and Colbeau 2004). Cammack (1999) showed that NiFe hydrogenase is responsible for the H_2 uptaking process, while the "Fe-S clusters free" hydrogenase plays a crucial role in the H_2 evolution process. To isolate and amplify the hydrogenase gene, DNA primers (a strand of nucleic acid) should be added to the target gene so that the DNA polymerase is able to replicate at the 3'-end of the primer and copy the opposite strand. So far, many primers have been designed and successfully applied to the study the H_2-oxidizing bacteria community (Leitao *et al.*, 2006).

In general, NiFe hydrogenase and FeFe hydrogenase are two major families that differ functionally. FeFe hydrogenase is involved with H_2 production, but NiFe

hydrogenase tends to be more involved with H_2 oxidation.

NiFe hydrogenase is composed of a large subunit (LSU) encoded by *HoxG* gene, a small subunit (SSU) encoded by *HoxK* gene and 10 iron atoms and 1 atom of nickel (Lorenz *et al.*, 1989). The molecular masses of both SSU and LSU are approximate 30 kDa and 60kDa, respectively (Przybyla *et al.*, 1992; Reeve and Beckler, 1990). The presence of NiFe hydrogenase has been reported in many bacterial species *Bradyrhizobium japonicum* (Arp, 1985), *Rhodobacter capsulatus* (Seefeldt *et al.*, 1987), *Azotobacter vinelandii* (Seefeldt and Arp, 1986), *Thiocapsa roseopersicina* (Gogotov, 1978), *Alcaligenes eutrophus* (Schink and Schlegel, 1979), *Desulfovibrio baculatus* (He *et al.*, 1989), and *Desulfovibrio gigas* (Hatchikian *et al.*, 1978). However there is no research conducted on hydrogenase(s) from microorganisms living in proximity to legume root nodules. In our study, we used molecular methods (PCR and real time PCR) to detect hydrogenase activity in such microorganisms.

3.2 Materials & Methods

3.2.1 Soil treatment

Soil used for hydrogenase gene study was the same as the soil for metatranscriptome analysis. All soil was collected from Annapolis Valley, Nova Scotia. Samples were mixed with sand (2:1 v/v) and were filled into 60 ml syringes.

All soil samples were divided into two sets as H_2 treatment and air treatment. H_2 treatment was regarded as experimental group which was consistently treated with gas mixture of H_2 and air (1000ppm), whereas air treated samples were consistently treated

with room air as control (Fig. 1).

The hydrogen uptake rate was measured throughout the hydrogen treatment using a H_2 sensor (Model S121, Quibit Systems Inc) and recorded by the Data Logger program as described by Dong and Layzell (2001) (Fig. 2). The data was collected as voltage before (turning on V1 & V2 and turning off V3 & V4) and after (turning on V3 & V4 and turning off V1 & V2) the mixed gas stream passing the soil column.

The standard curve was generated to establish the relationship between voltage and hydrogen concentration (ppm) by setting a series of mixed gas streams with gradient hydrogen concentration (from 0.55-500 ppm). This is carried out by regulating the current of electrolysis and flow rate using measuring system (Fig. 2). When V1 and V2 were turned on, H_2 –air mixture gas was going to the H_2 sensor directly, therefore the sensor could record the voltage of H_2 –air mixture gas with known concentration. Based on Graphpad Prism 5.0, two standard curves of voltage versus hydrogen concentration (ppm) were generated:

Range 0.55 – 100 ppm: ppm=21.92*v/(2.101-v)-2.803, R^2=0.9962

Range 100 – 500 ppm: ppm=163.6*v/(4.273-v)-9.941, R^2=0.995

After 4 week H_2 treatment, the H_2 uptake rate in H_2 treated sample (343.63 nmol/h*g) was significantly higher than that in air treated soil sample (39.24 nmol/h*g). Then the soil sample was ready for further experiments.

3.2.2 Total DNA extraction from soil samples

DNA was isolated from both treatments using MoBio PowerSoil® DNA Isolation Kit and MoBio UltraCleanTM DNA Purification Kit.

For MoBio PowerSoil® DNA Isolation Kit, 0.25 grams of soil was added into 2ml Bead Solution tube for maximum yields. After vortexing, Solution C1 with the volumn of 60 µl was added followed by vortexing at maximum speed for 10 minutes. All supernatant was transferred to a new 2 ml Collection Tube with the addition of 250 µl of Solution C2 and was incubated at 4°C for 5 minutes. The solution was centrifuged for 1 minute at 10000 x g. All supernatant was transferred to a new clean 2 ml Collection Tube. The solution was centrifuged at the same speed at room temperature for 1 minute and was transferred to another 2 ml Collection Tube. Solution C4 was added with the volume of 1200 µl. The mixed solution was centrifuged at the same speed for the same period separately due to the maximum volume of centrifuging tube provided. After centrifugation, all supernatant was kept and mixed with 500 µl of Solution C5 followed with centrifugation for 30 seconds at 10000 x g. The tube was centrifuged again for 1 minute after the removal of flow through. Then the spin filter in the centrifuging tube was carefully transferred to a new 2 ml Collection Tube followed with the addition of 100 µl of Solution C6. After centrifugation at room temperature for 30 seconds at 10000 x g and removal of spin filter, DNA was in the tube and ready for purification.

The protocol for DNA extraction using MoBio UltraCleanTM DNA Purification Kit was similar with that for MoBio PowerSoil® DNA Isolation Kit except that the addition of Solution S1 was followed with 200 µl of Solution IRS (Inhibitor Removal

Solution).

3.2.3 Primer design

Primer for hydrogenase gene study was designed based on hydrogenase small subunit *HoxK* gene sequences downloaded from NCBI Genbank. All sequences were aligned together using ClustalX 2.0.12 and the phylogenetic relationship was viewed by Geneious 5.4.3. All newly designed primer stats were reviewed according to National Institute of Standard and Technology online Primer Tools. The primer order was scheduled and produced with help of Eurofins MWG Operon company.

3.2.4 Amplification for Cloning

PCR condition was as follow: 3 minutes at 95°C for denaturation, 35 cycles of 30s at 95°C, 30s for annealing for (58 °C for primer set 1 and 2, 63 °C for primer set 3), 30s at 72°C for extension, and a final cycle, 10 minutes at 72°C Multiple PCR reactions were pooled together to increase final concentration. PCR products were purified with the QIAquick® PCR purification kit (QIAGEN Inc. Mississauga, ON). 25µl PCR system included: 1mM primers, , 2.5 µl of 10× Buffer, 2.5 µl of 2mM dNTP, 0.2 µl of 5U/ µl Taq enzyme (UBI Life Sciences Ltd. SK, Canada)), 1 µl of DNA template and Sigma® water to complete the 25µl volume.

3.2.5 Cloning

The amplicons were ligated into the pGEM®-T Easy Vectors with pGEM®-T Easy Vectors System (Promega Corporation, Madison, WI) followed by the same protocols above. The products were transformed to JM109 cells and screened for positive

colonies.

Plasmid DNA samples extracted from positive colonies were sent for sequencing by Macrogen. Sequencing results were provided by Macrogen and checked manually with primers, size and enzyme cutting site.

3.2.6 Plasmid Linearization

Before real-time PCR, the plasmid DNA should be linearized. 1μg of each sample was digested by 60U SalI at 37°C for 2 hours. Afterward, the reaction was stopped by using QIAquick® Gel Extraction Kit (QIAGEN Inc. Mississauga, ON). The products were checked by 1% agarose gel, ethidium bromide stain.

3.2.7 Generation of Standard Curve

Real-time PCR was carried out in ABI Prism 7000 (Applied Biosystem, Foster City, CA). The reaction system contained: 0.5mM primers, 1μl linearized plasmid DNA, 10 μl of Green-2-Go qPCR Mastermix (Bio Basic Inc, Markham Ontario, Ca) and Sigma® water was added to complete the 20 μl reaction system. The condition of real-time PCR was similar with regular PCR. The differences were one additional step, 2 minutes at 50°C before denaturation, and after 10 minutes extension, there was one more association stage.

3.2.8 Quantification of *HoxK* in soil samples

DNA sample was diluted to 1ng/μl and added into the 96-well plate (Applied Biosystem, Foster City, CA) with the volume of 1 μl as tempelate, 0.5mM primers 10 μl of Green-2-Go qPCR Mastermix (Bio Basic Inc, Markham Ontario, Ca) and fill with

Sigma® water to 20 μl as reaction volume. Inhibition effect was determined by adding 10^4 copies of the plasmid DNA containing PCR products. Statistics were analyzed using Graphpad Prism 5.0.

3.3   Results

3.3.1 Primer design for *HoxK* gene

Three pairs of specific primers for hydrogenase *HoxK* gene were successfully designed (Table 7) and tested in both DNA samples from H_2 treated and air treated soil. Primer set 1 was firstly designed (Fig. 22). However, in order to quantify gene copies using the same pair of primers, the product size needs to be reduced to around 200 bp. Primer set 2 and 3 were designed based on the gene sequence amplified by PCR with primer set 1 (Table 7). Primer set 2 with a smaller size of product was used for further study (Fig. 23).

3.3.2 Sequencing results of plasmids with target genes

Two clones with *HoxK* gene obtained by PCR with primer set 1 and set 2 were sequenced separately. Both clones were identified as hydrogenase small subunit from *Oligotropha carboxidovorans* OM5. Binding sites of each set of primers were indicated (Fig. 24). Binding site of primer set 1containing primer HoxKF1 and primer HoxKR2 was highlighted as yellow. The sequences stained with green color were the binding area of set 2 (primer RTF90 and primer RTR283). Because set 3 shared the same reverse primer, the binding site of set 3 was colored as primer RTF67 in blue and primer RTR283 in yellow at the 3' end. Product size of primer set 1, 2 and 3 were 489 bp, 213 bp and

estimated 234 bp, respectively.

Table 7 Primer sequences used to amplify fragments from *HoxK* gene in the H$_2$ oxidation pathway

Prime r set	Primer Name	Sequence 5'-3'	Length (bp)	Product
Set 1	HoxKF1	CACGGYCTBGARTGYACCTG	489	hydrogenase small subunit [*Oligotropha carboxidovorans* OM5]
	HoxKR2	GGGCASCCVGGBACYTTGA		
Set 2	RTF90	AGATTATGATGACACGCTGA	234	hydrogenase small subunit [*Oligotropha carboxidovorans* OM5]
	RTR283	CCATGAGATGATCGCCTT		

**Figure 22 Primer test using primer HoxKF1 and primer HoxFR2. Air: DNA
extracted from air treated soil sample; H$_2$: DNA from H$_2$ treated soil
sample**
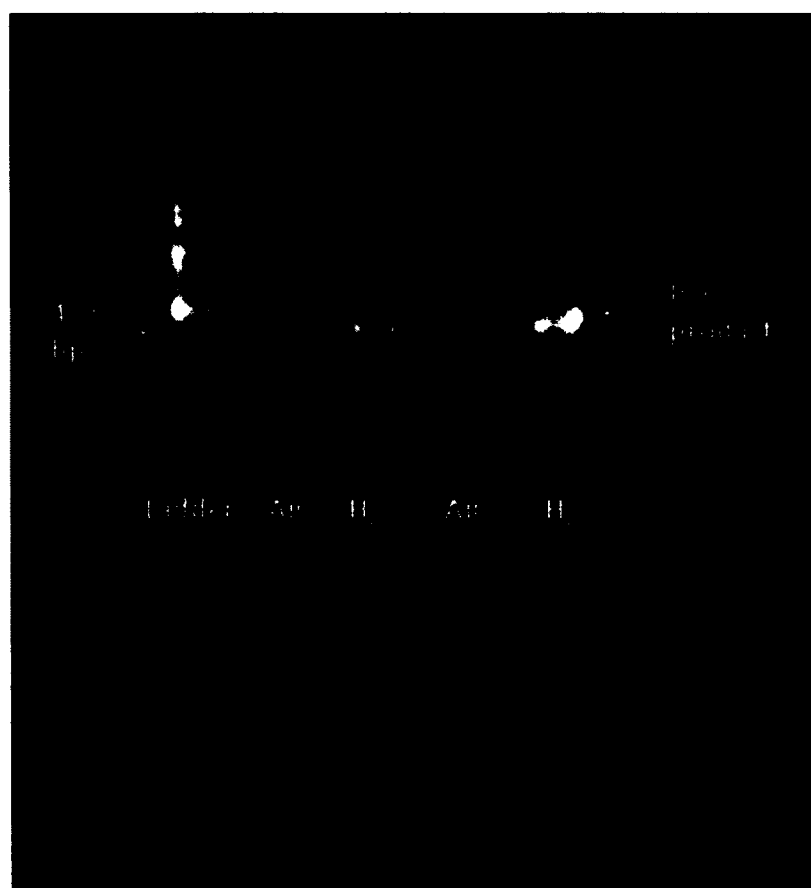
Figure 23 Primer test using primer RTF90 and primer RTR283. Air: DNA extracted from air treated soil sample; H_2: DNA from H_2 treated soil sample; P: plasmid DNA containing *HoxK* gene; N: pure water.

200 bp → PCR product

Ladder Air H₂ P N H₂ Air P
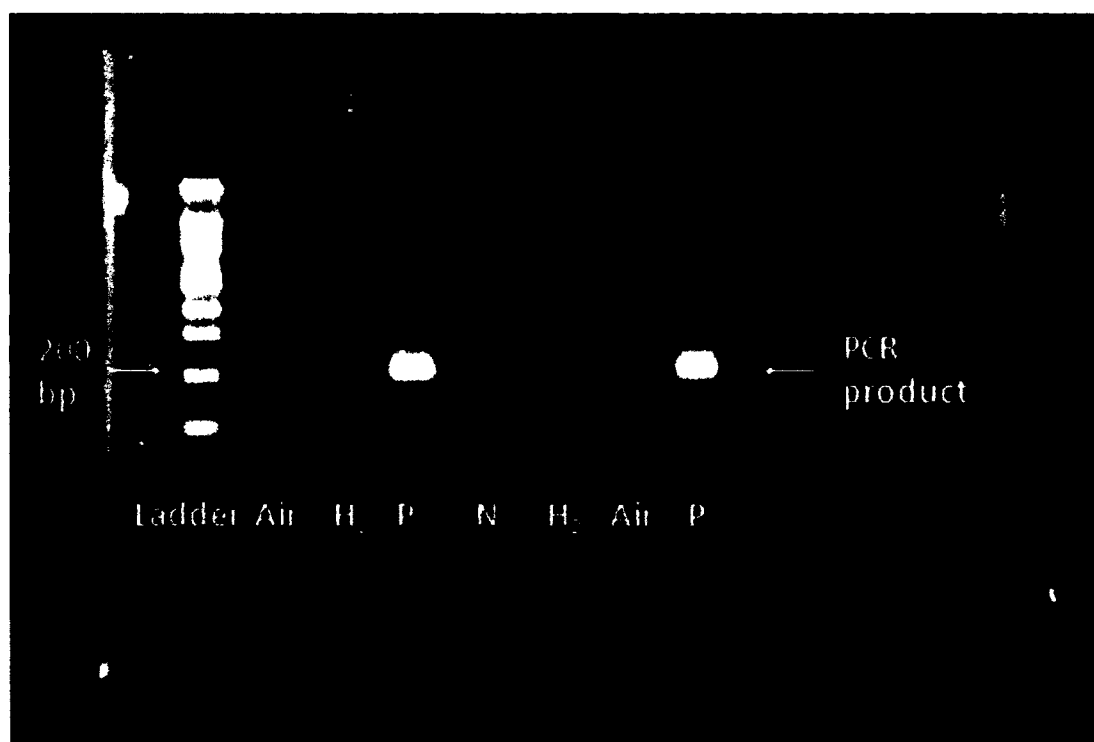
Figure 24 Primer binding sites for primer set1 (primer HoxKF1 and primer HoxR2), set2 (primer RTF67 and primer 283) and set3 (primer RTF90 and primer RTR283)

5'-CACGGTCTGGAGTGTACCTGCTGCTCGGAGTCCTTTATCCGCTCTGCACATCCACTGG
TGAAAGAC██████████████████████CGCT████████████████████TGGCCGCCGC
GGGACATCAGGCGGAAGCTGCGCTCGCAGATACCATCGAACGCTACAAGGGCAATTATATT
CTGGCCGTGGAAGGCAATCCGCCGCTCAACGAAGCTGGCATGTTCTGCATCATCGGTGGT
AAGCCTTTTGTCGATCAGCTTCGTTACGCCGCCAAACATGCC██████████████████G
GCTCCTGCGCCAGCCATGGTTGCGTGCAGGCAGCGCGGCCGAATCCAACCCGCGCAACCC
CGGTGCATCAGGTCATCACCGACAAGCCGATCATCAAAGTCCCCGGCTGCCCAATCACTA
GTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACG**CGTTGGATGC**
ATAGCTTGA-3'

3.3.3 Performance of standard curve

Real time PCR for hydrogenase *HoxK* gene using primer RTF90 and primer RTR283 were carried out 6 times and standard curves were generated by liberalized plasmids DNA contain PCR product amplified by primer RTF90 and primer RTR283 (Fig. 25, 26) using 3 DNA samples extracted from the same soil treatment but by two different commercial kits (PowerSoil® DNA Isolation Kit and UltraCleanTM DNA Purification Kit). The detect limitation is 10^2 copies. The *HoxK* gene copy numbers of each real time PCR set was calculated out using its own standard curve.

3.3.4 Quantification of *HoxK* gene in soil sample

With the standard curve for each set of real time PCR, the copy number (copies per gram soil) of two samples, H (H_2 treated sample) and A (air treated sample) were detected. Each sample was diluted to 1 ng/µl for a better performance, because the inhibition can be diluted as the DNA sample. The absolute number of gene copies in each sample varied all the time (Table 8). The size of standard error for all 6 sets for both samples was relatively bigger.

The copy number of *HoxK* gene in both H_2 treated and air treated samples varied significantly in each single set of real time PCR (Fig. 27). The inhibition efficiency could not be detected because the copy numbers of template with plasmid DNA in all 6 sets were undetermined. In addition, copy numbers of *HoxK* gene in all sets varied a lot. Thus, it will be less convincible to use any of those absolute number as the real copy number for either of the treatment. As a replacement, we calculated the ratio of *HoxK gene* increased in H_2 treated soil samples compared to air treated samples. According to those

ratios, we found after the H_2 treatment, the copy number of hydrogenase *HoxK* gene number in *O. carboxidovorans* OM5 raised 1.14 fold to 28.04 times (Table 8). The p-value calculated using Wilcoxon matched pairs test is 0.0156 which indicated that the difference of gene copy in H_2 and air treated soil sample was significant. The ratio from set 6 was extremely higher than any other ratios; one explanation for such a high value could be that there is error happened during PCR preparation. Because ratio from set 5 that shared the same DNA samples with set 6 was close to others. Thus, regardless of the ratio from set 6, we concluded that the *HoxK* gene copy number increased 1.14 to 4.19 times after H_2 treatment. Considering the inhibition of PCR technique itself, we believe that the real ratio can be different.

Figure 25 Standard curves of real time PCR set 1 using RTF90 and RTR283
primers and DNA extracted by PowerSoil® DNA Isolation Kit

Standar curve for RT-PCR1

Chart with y-axis "Cycle threshld" (values 2, 7, 12, 17, 22, 27, 32) and x-axis "Log starting HoxK copy numbers" (values 3, 4, 5, 6, 7, 8, 9, 10). Legend: ◆ RT-PCR1
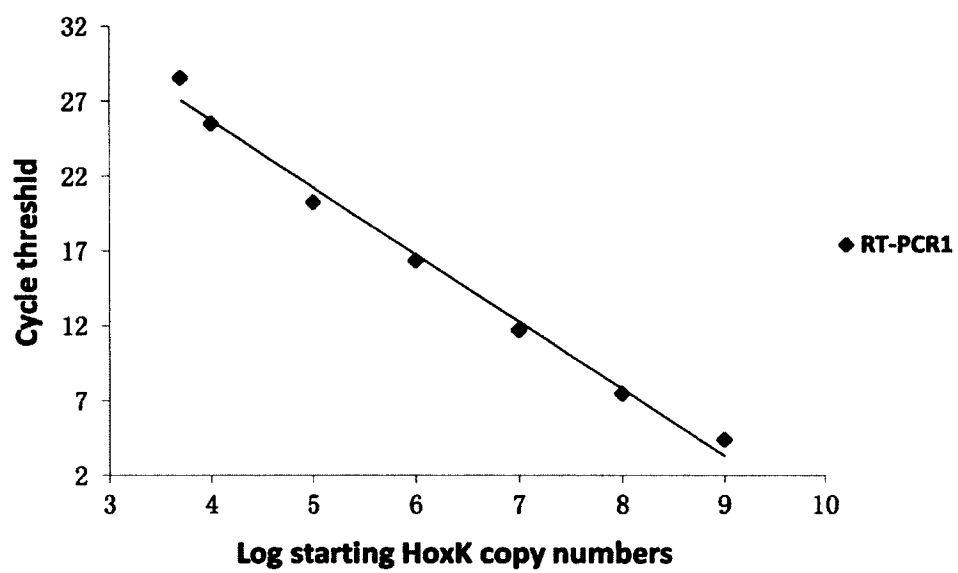
Figure 26 Standard curves of real time PCR set 2 to 6 using RTF90 and RTR283 primers and DNA extracted by PowerSoil® DNA Isolation Kit

Standard Curve for real time PCR set 2 to 6

Y-axis: Cycle threshld

X-axis: Log starting *HoxK* copy numbers

Legend:
- ◆ RT-PCR2
- ■ RT-PCR3
- ▲ RT-PCR4
- ✕ RT-PCR5
- ✳ RT-PCR6
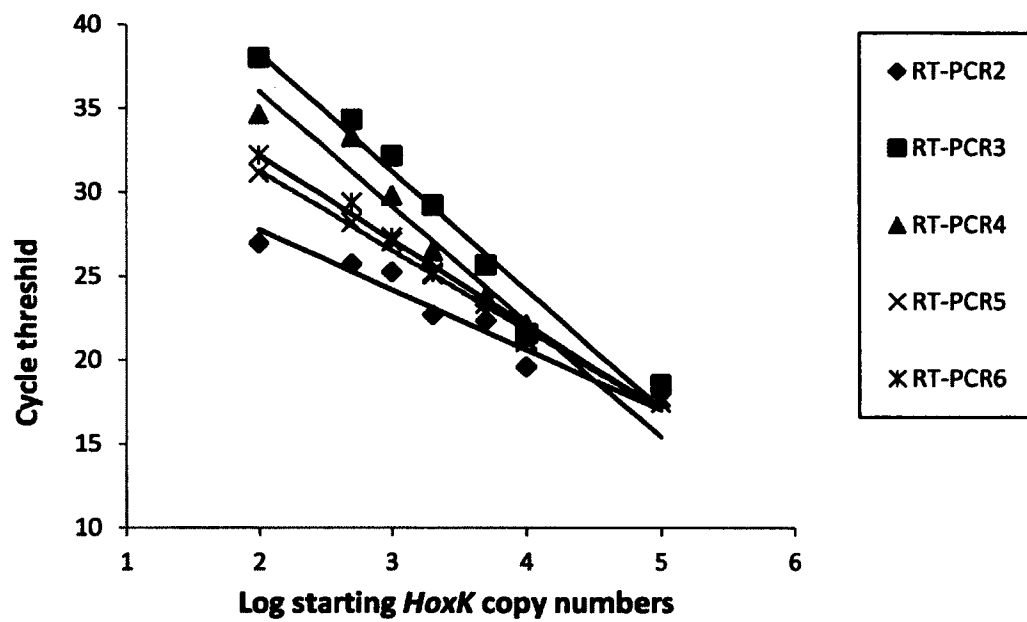
Table 8 Copy number of *HoxK* gene (copies per gram soil) detected by real time PCR using primer RTF90 and primer RTR283

Gene copy (copies per gram soil)	DNA extracted by PowerSoil kit		DNA extracted by UlatrClean kit		DNA extracted by UlatrClean kit		Average *HoxK* gene copy number	Standard Error	P-value (Wilcoxon matched pairs test)
	RT-PCR1	RT-PCR2	RT-PCR3	RT-PCR4	RT-PCR5	RT-PCR6			
H₂ treated sample	4.82E+05	1.20E+07	4.90E+06	1.64E+06	2.92E+05	1.01E+06	3.39E+06	1853472.171	0.0156
Air treated sample	1.61E+05	5.62E+06	1.76E+06	4.79E+05	5.63E+04	3.47E+04	1.35E+06	893266.8767	
Ratio(H₂/Air)	1.99	1.14	1.79	2.43	4.19	28.04	1.507357315		

Figure 27 Average *HoxK* gene copy numbers from real time PCR set 1 to 6 in H_2 and air treated soil samples

Average *HoxK* gene copy number in H$_2$ and air treated soil samples
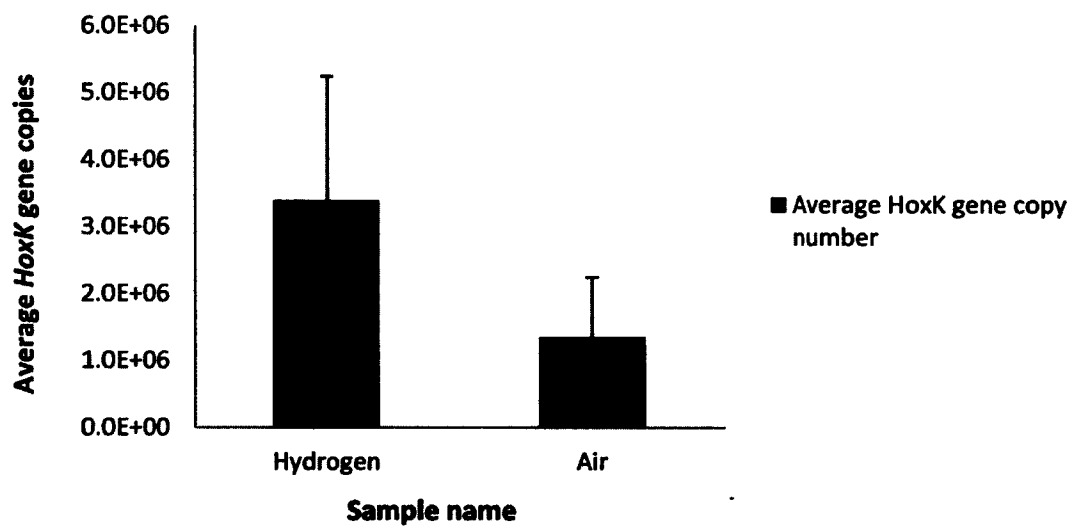
3.4   *Discussion*

Quantification of bacteria capable of hydrogen oxidation is important for a better understanding of H_2 oxidizing activity and in the rhizosphere adjacent to nodule root nodules. In this study, a real-time PCR assay was carried out to quantify the hydrogenase using the *HoxK* gene encoding NiFe hydrogenase small subunit. The results showed that the gene copies of soil after H_2 treatment increased significantly.

3.4.1 H_2 fertilization and quantification of *HoxK* gene

H_2 was observed released by legume root nodules by scientists years ago. The effect of H_2 left over by nodules on plant growth had been studied as well at the plant physiology level. However the mechanism of H_2 fertilization at molecular biology level still remained unclear. That is why we chose hydrogenase as the target to understand this myth.

H_2 can act as original electron donors in many microbiological metabolisms. The key enzyme of all those pathways is the hydrogenase which catalyzes the first step of all reactions. In general, hydrogenase protein consists of two parts: a larger subunit (HoxK) and a small subunit (HoxG). The hoxK part has an essential role in hydrogenase catalysis (Sayavedra and Arp, 1993). In our study, the primer for hydrogenase study was designed based on NiFe hydrogenase small subunit (HoxK) gene. The result from regular PCR method indicated that the newly designed primer pair (RTF90 and RTR283) could amplify a product that has a length of 213 base pairs. Later on the result from real time PCR demonstrated that more *HoxK* gene copies were observed in H_2 treated sample rather than that in control. This observation provided the direct evidence that H_2 in soil

can enrich the population of H_2 oxidizing bacteria containing NiFe hydrogenase small subunit in soil. The relative difference of *HoxK* gene copies in H_2 and air treated samples was significant. Because our primers designed based on NiFe hydrogenase *hoxK* genes resulted in only one PCR product matched with hydrogenase sequence of *Oligotropha carboxidovorans* OM5 from *Alphaproteobacteria*, we believe the total amount of all hydrogenase gene in soil can be much more different compared between H_2 treated soil sample and control.

3.4.2 Limitations for hydrogenase *HoxK* gene study

One of the limitations for our hydrogenase gene study is that, we could find a pair primers that are able to amplify hydroganase genes from multiple species, which can reflect the truth of hydrogenase in H_2 oxidizing bacteria in soil more precisely. As He (2010) and our metatranscriptome analysis presented, the major population of bacteria induced by H_2 treatment was *Burkholderiales* and *Myxococcales* from *betaproteobacteria* and *deltaproteobacteria* respectively. Unfortunately, neither of speices belonging to those orders was observed.

Another limitation of hydrogenase *HoxK* gene study is the inhibition during amplification process. The inhibition of PCR amplification, especially humic acid or humic substances co-extracted with DNA could strongly impact the gene quantification analysis by inhibiting the *Taq* enzyme (Porteous and Armstrong, 1991). The dilution of DNA sample can relatively reduce the influence of inhibition. In our study, we used 1 µl of DNA solution (relatively 5ng/µl) in 25 µl reaction for regular PCR and 1µl of diluted DNA solution (1 ng/µl) in 25 µl reaction for real time PCR. To calculate the amplification

efficiency, we added estimated copies of plasmids containing *HoxK* gene fragments into each sample as reference. However, most of times, the copy number of samples with plasmids could not even be detected. Thus, our absolute copy number of *HoxK* gene in each set of real time PCR lacked reliability. This is the reason why we only calculated the increasing degrees in H_2 treated soil sample compared to control group.

3.4.3 Future work on hydrogenase study

As all limitations mentioned above, the future work of hydrogenase study should focus on primer design based on sequences from *Burkholderiales* and *Myxococcales* according to our metatranscriptome result. In addition, the detection of certain gene copy number cannot imply the gene activity under certain circumstances. Thus, in order to understand the how active the hydrogenase gene is in H_2 treatment, reverse transcription PCR for hydrogenase gene needs to be carried out.

The result of real time PCR can be enhanced by optimizing the preparation stage. All soil samples for regular and real time PCR are better to be collected as much as possible at one time and store in -80 °C for future study. Therefore, the difference among each set of amplification will be minimized. Dilution test and calculation of efficiency are also needed for each set of amplification, especially for real time PCR.

4. General discussion

Our objectives in this study were to understand the H_2 metabolism related gene activities and obtain a better view of microbial community structure changes in H_2 environment which is similar with the rhizosphere around legume root nodules. According to previous studies, CO_2 metabolism, especially CO_2 assimilation, and bacterial respiration chain were strongly connected with H_2 metabolism in bacteria that are capable of oxidizing atmospheric H_2. Our results indicated that during the H_2 treatment, many genes that encode proteins which are associated with bacterial electron transportation and CO_2 fixation, particularly Calvin cycle, have a much higher level of gene expression. Combining the result with NiFe hydrogenase *HoxK* gene study, we confirmed that hydrogenase gene not only contains a much higher quantity after H_2 treatment, but also is more active than that in air treatment. However, researches designed for specific genes need to be conducted for further confirmation about the involvement of H_2 in both O_2 reduction and CO_2 fixation process. Although genes that are induced by H_2 treatment are more likely connected directly with bacterial metabolisms which we are interested about, those experienced down regulation of gene expression during the treatment also need more investigation.

For hydrogenase study, a better primer pair that is able to cover more species of H_2 oxidizing bacteria, particularly *Burkholderiales* and *Myxococcales*, is needed. The soil sample collection protocol and preparation of PCR protocol need to be improved.

In terms of soil microbial community study, our result suggested agreed with

results conducted by others methods that *Protebacteria* experienced the most significant increase of population after H_2 treatment. However, there are a few controversies about other phylum between different methods.

5. References

Andreae, B. 1981. Farming, Development and Space: A world Agricultural Geography, translated from German by Howard F. Gregor. Walter De Gruyter Inc. Berlin, Germany.

Arp, D. J. 1985. *Rhizobium japonicum* hydrogenase: purification to homogeneity from soybean nodules, and molecular characterization. *Archives of Biochemistry and Biophysics*. 237:504-512.

Bais, H. P., Park, S.-W., Weir, T. L., Callaway, R. M., and Vivanco, J. M. 2004. How plants communicate using the underground information superhighway. *Trends in Plant Science*, 9: 26–32.

Bernard, U. and Schlegel, H.G. 1974. NADH- and H_2-oxidation in hydrogen bacteria studied by respiratory chain inhibitors. *Archives of Microbiology*. 95: 39-44.

Bolton, E. F., Dirks, V. A., and Aylesworth, W. A. 1976. Some effects of alfalfa, fertilizer and lime on corn yield in rotations on clay soil during a range of seasonal moisture conditions. *Canadian Journal of Soil Science*. 56: 21–25.

Bowien, B. and Schlegel, HG. 1981. Physiology and biochemistry of aerobic. *Annual Review of Microbiology*. 35:405-52

Butt, J.J. 2002. Daily life in the age of Charlemagne. Library of Congress Cataloging-in-Publication Data, Westport. UK

Cammack, R. 1999. Hydrogenase Sophistication. *Nature*. 397:214-5.

Cardon, Z. G. and Whitbeck, J. L. 2007. The rhizosphere: An ecological perspective. Elsevier Academic Press. Burlington, MA.

Cooper, J. L. 1999. A grower survey of rotations used in the New South Wales Cotton Industry. *Australian Journal of Experimental Agriculture*, 39: 743-755.

Condon, C. 2007. Maturation and degradation of RNA in bacteria. *Current Opinion in Microbiology* 10: 271–278.

Conrad, R., and Seiler, W., 1979. Field measurements of hydrogen evolution by nitrogen fixing bacteria. *Soil Biology & Biochemistry*. 11: 689-690.

Croucher, N.J., Fookes, M.C., Perkins, T.T., Turner, D.J., Marguerat, S.B., Keane, T., et al. 2009. A simple method for directional transcriptome sequencing using Illumina technology. *Nucleic Acids Research* 37: e148.

Connor, D. J., Loomis, R. S. and Cassman, K. G. 2011. Crop Ecology: Productivity and Management in Agricultural Systems. 2nd Edition, Cambridge.

Deutscher, M. P. 2003. Degradation of stable RNA in bacteria. *The Journal of Biological Chemistry* 278: 45041–45044.

Dixon, R.O.D. 1967. Hydrogen uptake and exchange by pea root nodules. *Annals of Botany*. 31: 179-188

Dong, Z., and Layzell, D. B. 2001. H_2 oxidation, O_2 uptake and CO_2 fixation in hydrogen treated soils. *Plant and Soil* 229, 1–12.

Dong, Z. and Layzell, D. B. 2002. Why do legume nodules evolve H_2 gas? Nitrogen Fixation: Global Perspectives : Proceedings of the 13th International Congress on Nitrogen Fixation, Hamilton, Ontario, Canada.

Dong, Z., Wu, L., Kettlewell, B., Caldwell, C.D. and Layzell, D.B. 2003. Hydrogen fertilization of soils – is this a benefit of legumes in rotation? *Plant, Cell and Environment*. 26: 1875-1879

Eisbrenner, G. and Evans, H. J. 1982. Carriers in the electron transport from molecular hydrogen to oxygen in *Rhizobium japonicum* bacteroids. *Journal of Bacteriology.* 149: 1005-1012.

Evans, H. J., Harker, A. R., Papen, H., Russell, S. A., Hanus, F.J. and Zuber, M. 1987. Physiology, biochemistry, and genetics of the uptake hydrogenase in Rhizobia. *Annual Review of Microbiology.* 41: 335–361.

Ferber, D.M., Moy, B. and Maier, R.J. 1995. *Bradhyrhizobium japonicum* hydrogen-ubiquinone oxidoreductase activity: quinine specificity, inhibition by quinine analogs, and evidence for separate sites of electron acceptor reactivity. *Biochimica et Biophysica Acta.* 1229: 334-346.

Flynn, B. 2011. Quantification of bacterial Rubisco genes and their expression in soil adjacent to H_2 releasing legume nodules (thesis).

Glick, B. R., Cheng, Z., Czarny, J. and Duan, J. 2007. Promotion of plant growth by ACC deaminase-producing soil bacteria. *European Journal of Plant Pathology.* 110: 329-339.

Grigg D. 1989. English Agriculture: an historical perspective. Blackwell Pub (April 1990). Birmingham, US.

Giller, K.E. 2001. Nitrogen Fixation in Tropical Cropping Systems. *CAB International*, Wallingford, UK, 423 pp.

Ghisla, S. and Thorpe, C. 2004. Acyl-CoA dehydrogenases. A mechanistic overview. *European Journal of Biochemistry.* 271: 494–508.

Gogotov, I. N. 1978. Relationships in hydrogen metabolism between hydrogenase and nitrogenase in phototropic bacteria. *Biochimie* 60:267-276.

Gong, Z., Lin, P., Chen, J. and Hu, X. 2000. Classical farming systems of China. In: nature farming and microbial applications. Xu H., Parr J.F., Umemura H. (ed.) p .14. *Food Products Press. NY, USA.*

Harris, H.C. 1995. Long-term trials on soil and crop management at ICARDA. In: Soil management: experimental basis for sustainability and environmental quality. Lal R. and Stewart B.A.(eds.) p.447. CRC Press, Boca Raton, Florida.

Hatchikian, L. C., Bruschi, M. and LeGall, J. 1978. Characterization of the periplasmic hydrogenase from *Desulfovibno gigas*. *Biochemical and Biophysical Research Communications.* 82:451-461.

He, S. H., Teixeira, M., Legall, J., Patil, D. S., Moura, I., Moura, J. J. G., DerVartanian, D. V., Huynh, B. H. and Peck, H. D. J. 1989. EPR studies with ^{77}Se enriched (NiFeSe) hydrogenase of *Desulfovibrio baculatus*. Evidence for a selenium ligand to the active-site nickel. *The Journal of Biological Chemistry*. 264:2678-2682.

He, X. 2010. Effect of h₂ on soil bacterial community structure and soil denitrifiers population (Thesis).

Heichel, G.H. 1987. Legumes as source of nitrogen in conservation tillage systems. 29-34. In J.F. Power (ed.) The role of legumes in conservation tillage systems. Proc. Natl. Conserv. Tillage Conf., Athens, GA. 27-29 April. Soil Conservation Society of America. Ankeny, IA.

Hesterman, O. B., Sheaffer, C. C., Barnes, D. K., Lueschen, W. E. and Ford, J. H. 1986. Alfalfa dry matter and nitrogen production and fertilizer nitrogen response in legume-corn rotations. *Agronomy Journal*. 78, 19–23.

Hiltner, L. 1904. Über neuere Erfahrungen und problem auf dem Gebiete der Bodenbakteriologie unter plant symbiots. *Nature Reviews Microbiology*. 2:43-56.

Hoch, G. E., Little, H. N. and Burris, R. H. 1957. Hydrogen evolution from soybean root nodules. *Nature (London)* 179:430-431.

Hoch, G. E., Schneider, K. C. and Burrism, R. H. 1960. Hydrogen evolution and exchange and conversion of N₂O to N₂ by soybean root nodules. *Biochimica et Biophysica Acta* 37:273-279.

Hulugalle, N. R. and Daniells, I.G. 2005. Permanent beds in Australian cotton production systems. *Australian Journal of Soil Research*, 39(2): 317-328.

Hunt, S. and Layzell, D.B. 1993. Gas exchange of legume nodules and the regulation of nitrogenase activity. *Annual Review of Plant Physiology and Plant Molecular Biology*. 44: 483-511

John, G. O and Howard, G. 1962. Hydrogen photosynthesis and alternative metabolic pathways in photosynthetic bacteria. *Bacteriological Reviews*. 26(1): 51-63.

Julie, A. H., Hilary, G. M., Susan, M. H., Phillip, R. N., Mitchell, L. S. and David, B. M. W. 2009. Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Enviromental Microbiology*. 11(5): 1292-1302.

Kamen, M. D. and Gest, H. 1949. Evidence for a nitrogenase system in the photosynthetic bacterium Rhodospirillum rubrum. *Science*. 109(2840):560

Kamilova, F., Kravchenko, LV., Shaposhnikov, AI., Makarova, N and Lugtenberg, BJJ. 2006. Effects of the tomato pathogen *Fusarium oxysporum* f. sp. *Radicis-lycopersici* and of the biocontrol bacterium *Pseudomonas fluorescenc* WCS365 on the composition of organic acids and sugars in tomato root exudates. *Molecular Plant-Microbe Interactions*. 19: 1121-1126.

Karel, R.S. and Harold, J.E., 1976. Hydrogen evolution: A major factor affecting the efficiency of nitrogen fixation in nodulated symbionts. *Proceedings of the*

National Academy of Sciences USA. 73:1207-1211.

Kent, A. D., Smith, D. J., Benson, J. B. and Triplett, E. W. 2003. Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. *American Society for Microbiology.* 69: 6768-6776.

Kiely, P. D., Haynes, J. M., Higgins, C. H., Franks, A., Mark, G. L., Morrissey, J. P., and O'Gara, F. 2006. Exploiting new systems-based strategies to elucidate plant-bacterial interactions in the rhizosphere. *Microbial Ecology,* 51: 257-266.

Kuiper, I., Bloemberg, G. V., and Lugtenberg, B. J. J. 2001. Selection of a plant-bacterium pair as a novel tool for rhizostimulation of polycyclic aromatic hydrocarbon-degrading bacteria. *Molecular Plant-Microbe Interactions.* 14: 1197-1205.

La Favre, J. S. and Focht, D. D. 1983. Conservation in soil of H_2 liberated from N_2 fixation by Hup– nodules. *Applied and Environmental Microbiology.* 46: 304–311.

Lee, D. H., Zo, Y. G. and Kim, S. J. 1996. Non-radioactive method to study genetic profiles of natural bacterial communities by single-stranded-conformation polymorphism. *Applied and Environmental Microbiology* 62: 3112-3120

Leitão, E., Oxelfelt, F., Oliveira, P., Moradas-Ferreira, P. and Tamagnini, P. 2005. Analysis of the *hupSL* operon of the non-heterocystous cyanobacterium *Lyngbya majuscula* CCAP 1446/4: Regulation of transcription and expression under light-dark regimen. *Applied Environmental Microbiology.* 71:4567-76.

Lengeler, J., Drews, G. and Schlegel, H. 2009. Biology of the Prokaryotes. Wiley-Blackwell; 1 edition.

Liu, W. T., Marsh. T. L., Cheng, H and Forney. L. J. 1997. Characterization of microbiaol diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology.* 63: 4516-4522.

Lorenz, B., Schneider, K. Kratzin, H. and Schlegel, H. G. 1989. Immunological comparison of subunits isolated from various hydrogenases of aerobic hydrogen bacteria. *Biochimica et Biophysica Acta.* 995:1-9.

Lugtenberg, B. and Kamilova, F. 2009. Plant-Growth-Promoting Rhizobacteria. *Annual Review of Microbiology.* 63: 541-556.

Maimaiti, J., Zhang, Y., Yang, J., Cen, Y. P., Layzell, D., Peoples, M., and Dong, Z. 2007. Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environmental Microbiology.* 9: 435-444.

Marschner, H. 1995. Mineral Nutrition of Higher Plants. London: Academic. 2nd ed.

McLearn, N. and Dong, Z. 2002. Microbial nature of the hydrogen-oxidizing agent in hydrogen-treated soil. *Biology and Fertility of Soils* 35:465-469.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7): 621-8.

Muyzer, G., de Waal, E. C. and Uitterlinden, A. G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes encoding for 16S rRNA. *Applied and Environmental Microbiology* 59: 695-700

Muyzer, G., Brinkhoff, T., Nübel, U., Santegoeds, C., Schäfer, H., and Wawer, C. 1997. Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. p. 1–27. In A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.) *Molecular Microbial Ecology Manual*, vol. 3.4.4. Kluwer Academic Publishers, Dordrecht, Netherlands.

Muyzer, G. 1999. Genetic fingerprinting of microbial communities: present status and future perspectives In C. R. Bell, M. Brylinsky, and P. Johnson-Green (ed.), Microbial biosystems: new frontiers. Proceedings of the 8th International Symposium on Microbial Ecology. Atlantic Canada Society for Microbial Ecology, Halifax, Canada.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344–1349.

O'Brian M. R. and Maier R. J. 1985. Role of ubiquinone in hydrogen-dependent electron transport in *Rhizobium japonicum*. *Journal of Bacteriology*. 161: 775-777.

Okon, Y., Bloemberg, G. V., and Lugtenberg, B. J. J. 1998. Biotechnology of biofertilization and phytostimulation. In *Agricultural Biotechnology*, ed. A, Altman. 327-349. New York: Marcel Dekker.

Osborne, C. A., Rees, G. N., Bernstein, Y. and Janssen, P. H. 2006. New threshold and confidence estimates for terminal restriction fragment length polymorphism analysis of complex bacterial communities. *Applied and Environmental Microbiology* 72:1270– 1278.

Phelps, A. S. and Wilson, P. W. 1941. Occurrence of Hydrogenase in Nitrogen Fixing Organisms. *Experimental Biology and Medicine* 47, 473-476.

Porteous, L. A., and Armstrong, J. L. 1991. Recovery of bulk DNA from soil by a rapid, small-scale extraction method. *Current Microbiology*. 22(6): 345-348.

Power, J.F. (editor). 1987. The role of legumes in conservation tillage systems. Soil Conservation Society of America, Ankeny, Iowa. 29-34.

Rochester, I. J., Peoples, M. B., Hulugalle, N. R.., Gault, R. R. and Constable, G. A.. 2001. Using legumes to enhance nitrogen fertility and improve soil condition in cotton cropping systems. *Field Crops Research*, 70(1): 27-41.

Popelier, F., Liessens, J., and Verstraete, W. 1985. Soil H_2-uptake in relation to soil properties and rhizobial H_2-production. *Plant and soil*. 85:85-96.

Przybyla, A. E., J. Robbins, N. Menon, and H. D. J. Peck. 1992. Structure-function relationships among the nickel-containing hydrogenases. *FEMS Microbiology Reviews.* 88:109-135.

Reeve, J. N., and G. S. Beckler. 1990. Conservation of primary structure in prokaryotic hydrogenases. *FEMS Microbiology Reviews.* 87:419-424.

Ruiz-Argüeso, T., Hanus, J. and Evans, H. J.. 1978. Hydrogen production and uptake by pea nodules as affected by strains of *Rhizobium japonicum. Archives of Microbiology.* 116:113-118.

Sayavedra L.A. and Arp D.J. 1993. In *Azotobacter vinelandii* hydrogenase, substitution of serine for the cysteine residues at positions 62, 65, 294, and 297 in the small (HoxK) subunit affects H_2 oxidation. *Journal of Bacteriology.* 175: 3414-3421.

Scheinert, P., Kruse, R., Ullmann, U., Söller, R. and Krupp, G. 1996. Molecular differentiation of bacteria by PCR amplification of the 16S-23S rRNA spacer. *J Microbiol Methods* 26: 103-117.

Schink, B. and Schlegel, H. 1978. Hydrogen metabolism in aerobic hydrogen-oxidizing bacteria. *Biochimie.* 60: 297-305.

Schink, B. and Schlegel, H. 1979. The membrane-bound hydrogenase of *Alcaligenes eutrophus.* I. Solubilization, purification, and biochemical properties. *Biochimica et Biophysica Acta* 567:315-324.

Schink, B. 1982. Isolation of a hydrogenase-cytochrome *b* complex from cytoplasmic membranes of *Xanthobacer autotrophicum* GZ 29. *FEMS Microbiology Letters.* 13: 289-293.

Schwieger, F. and Tebbe, C. C. 1998. A new approach to utilize PCR-Single-Strand Conformation Polymorphism for 16S rRNA-based microbial community analysis. *Applied and Environmental Microbiology.* 64: 4870-4876.

Seefeldt, L. C. and Arp, D. J. 1986. Purification to homogeneity of *Azotobacter vinelandii* hydrogenase; a nickel and iron containing a α dimer. *Biochimie* 68:25-34.

Seefeldt, L. C., McCollum, L. C., Doyle, C. M. and Arp, D. J. 1987. Immunological and molecular evidence for a membranebound, dimeric hydrogenase in *Rhodopseudomonas capsulata. Biochimica et Biophysica Acta.* 914:299-303.

Silvia, G. A., Ramahi, S., Vanja, K. and Martin, F. P. 2006. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Applied and Environmental Microbiology.* 71(12): 8966-8969.

Simpson, F. B., Maier, R. J. and Evans, H. J. 1979. Hydrogen-stimulated CO_2 fixation and coordinat induction of hydrogenase and ribulose-bisphosphate carboxylase in a H_2-uptake positive strain of *Rhizobium japonicum. Archives of Microbiology.* 123:1-8.

Simpson, F. B., Burris, R. H., 1984. A nitrogen pressure of 50 atmospheres does not

prevent evolution of hydrogen by nitrogenase. *Science*. 224: 1095-1097.

Slicher Van Bath, B. H. 1963. The agrarian history of Western Europe A.D. 500-1850. Edward Arnold (Publishers) LTD. London, UK.

Smalla K., Oros-Sichler, M., Milling, A., Heuer, H., Baumgarte, S., Becker, R., Neuber, G., Kropf, S., Ulrich, A. and Tebbe, C. C. 2007. Bacterial diversity of soils assessed by DGGE, T-RFLP and SSCP fingerprints of PCR-amplified 16S rRNA gene fragments Do the different methods provide similar results. *Journal of Microbiological Methods* 69:470-479.

Sorek, R. and Cossart, P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* 11: 9-16.

Spaink, H. P., Kondorosi, A. and Hooykaas, P. J. J. 1998. *The Rhizobiaeceae*. Dordrecht, The Netherlands: Kluwer Acad.

Stein, S., Selesi, D., Schilling, R. and Patties, I. 2005. Microbial activity and bacterial composition of H_2-treated soils with net CO_2 fixation. *Soil Biology & Biochemistry*. 37:1938-1945.

Tabita, F. R. and McFadden, B. A. 1974. D-Ribulose 1,5-diphosphate carboxylase from *Rhodospiritlum rubrum*. II. Quaternary structure, composition, catalytic, and immunological properties. *The Journal of Biological Chemistry*. 249: 3459-3464.

Torsvik, V., Coksoyr, J. and Daae, F. L. 1990. High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology* 56: 782-787.

Uratsu, S. L., Keyser, H. H., Weber, D. F. and Lim, S. T. 1982. Hydrogen uptake (HUP) activity of *Rhizobium japonicum* from major US soybean production areas. *Crop Science*. 22: 600-602.

Urich, T., Lanzen, A., Qi, J., Huson, D., Schleper, C. and Schuster, S. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE*. 3(6)

Vallaeys, T., Topp, E., Muyzer, G., Macheret, V., Laguerre, G. and Soulas, G. 1997. Evaluation of denaturing gradient gel electrophoresis in the detection of 16S rDNA sequence variation in rhizobia and methanotrophs. *FEMS Microbiology Ecology*. 24: 279-285.

Van Loon, L. C., and Bakker, P. A. H. M. 2003. Signalling in rhizobacteria-plant interactions. In H. De Kroon & E. J. W. Visser (Eds.), *Root ecology*, (*Ecological studies* Vol. 168, pp. 297-330). Berlin, Heidelberg: Springer-Verlag.

Van Rhijn, P., and Vanderleyden, J. 1995. The *Rhizobium*-plant symbiosis. *Microbiological Reviews*. 59: 124-142.

van Vliet, A. H., and Wren, B. W. 2009. New levels of sophistication in the transcriptional landscape of bacteria. *Genome Biology* 10: 233. doi:10.1186/gb-2009-10-8-233.

van Vliet, A. H. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiology Letters*. 302: 1-7.

Vignais, P. M. and Colbeau, A. 2004. Molecular biology of microbial hydrogenases. *Current Issues in Molecular Biology.* 6:159–188.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V. and Goodhead, I., Penkett, C. J., Rogers, J. and Bähler, J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453: 1239–1243.

Wilson, P. W. and Burris, R. H. 1947. The mechanism of biological nitrogen fixation. *Bacterial Rev.* 11(1): 41-73.

Zhang, Y., He, X and Dong, Z. 2009. Effect of hydrogen on soil bacterial community structure in two soils as determined by terminal restriction fragment length polymorphism. *Plant Soil.* 320: 295-305.