Metatranscriptome analysis of microbial communities in rice microcosms

Dissertation

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

(Dr. rer. nat)

dem Fachbereich Biologie

der Philipps-Universität Marburg/Lahn

vorgelegt von

Yongkyu Kim

aus Seoul, Südkorea

Marburg an der Lahn | 2012

Metatranscriptome analysis of microbial communities in rice microcosms

Doctoral thesis

Submitted in the fulfillment of the requirements for a doctoral degree
"Doktorgrad der Naturwissenschaften (Dr. rer.nat.)"
to the faculty of biology – Philipps-Universität Marburg

by

Yongkyu Kim

From Seoul, South Korea

Marburg/Lahn | 2012

The research for the completion of this work was carried out from October 2008 to April 2012 in the Department of Biogeochemistry at the Max Planck Institute for Terrestrial Microbiology under the supervision of PD. Dr. Werner Liesack.

Thesis was accepted to the Dean, Faculty of Biology, Philipps-Universität Marburg

on:

First reviewer: PD. Dr. Werner Liesack
Second reviewer: Prof. Dr. Martin Thanbichler

Date of oral examination:

Publication

The following papers were published by the date of submission of the present thesis:

1) Mettel C[¥], **Kim Y[¥]**, Shrestha PM & Liesack W (2010) Extraction of mRNA from Soil. *Applied and Environmental Microbiology* **76:** 5995-6000
[¥] These authors contributed equally to this work

2) Kulichevskaya IS, Serkebaeva YM, **Kim Y**, Rijpstra WIC, Damsté JSS, Liesack W & Dedysh SN (2012) Telmatocola sphagniphila gen. nov., sp. nov., a Novel Dendriform Planctomycete from Northern Wetlands. *Frontiers in Microbiology* **3:** 1-9

Dedicate to my wife and family

Table of Contents

Metatranscriptomics is a state-of-the-art technique to elucidate the functional activities of microbial communities, but its application is still limited to marine microbial assemblages. In my PhD project, we established the complete approach of soil metatranscriptomics, involving RNA extraction, cDNA library preparation by random priming, 454-pyrosequencing, and bioinformatic data analysis. The approach was tested on microbial communities in the oxic surface layer and the anoxic bulk soil of flooded rice paddy soil microcosms.

Total RNA was recovered in high integrity and purity by low-pH phenol extraction (pH 5.0) followed by Q-Sepharose column chromatography. We were able to enrich mRNA sequences up to 50-70% in the metatranscriptome libraries using Ribo-Zero™ rRNA removal kit (Meta-Bacteria). All the 454 reads obtained were preprocessed prior to data analysis to minimize sequence ambiguities. A total of 10,000 SSU-ribotags (total RNA) were analyzed to elucidate community composition in the oxic and anoxic zones at three different incubation time points (25, 45 and 90 days after transplantation of rice seedlings). Additionally, about 45,000 and 12,000 mRNA-tags (enriched mRNA) were obtained for the analysis of functional activities in, respectively, the oxic and anoxic zone of 90-day-old rice microcosms.

SSU-ribotag data analysis revealed no major temporal changes in community composition except for *Geobacter, Clostridia* and methanogens in the anoxic bulk soil. However, the taxonomic composition of microbial communities was clearly distinct between the oxic and anoxic zones, with cyanobacteria being the dominant group in the surface layer. Although mRNA-tags related to basic cellular functions were most abundant in both mRNA datasets, the expression of specific functions in response to different oxygen conditions was observed such as, for example, methane oxidation in the oxic zone and methanogenesis in the anoxic zone.

Our metatranscriptomic approach provides a means to analyze the composition and functional gene expression of complex soil microbial communities while avoiding the limitations of PCR-based approaches.

Metatranskriptomik ist eine ‘state-of-the-art’ Technik um die funktionellen Aktivitäten mikrobieller Gemeinschaften aufzuklären. Die Anwendung hat sich bis auf wenige Ausnahmen bis jetzt auf marine Lebensräume fokussiert. In meiner Doktorarbeit haben wir den gesamten Ablauf der Metatranskriptomik, einschließlich RNA Extraktion, Vorbereiten von cDNA Bibliotheken mittels Zufalls-priming, 454 Pyrosequenzierung und bioinformatische Analyse für terrestrische Lebensräume etabliert. Der gesamte, entwickelte Ablauf wurde bezogen auf mikrobielle Gemeinschaften in oxischer- und anoxischer Erde von Reisbodenmikrokosmen umfassend getestet.

Gesamt-RNA von hoher Integrität und Reinheit konnte bei einem pH Wert von 5 mittels Phenolextraktion kombiniert mit Q-Sepharose Säulenchromatographie gewonnen werden. Durch Anwendung des Ribo-Zero ™ rRNA removal kits (Meta-Bacteria) war es uns möglich mRNA bis zu einem Grad von 50-70% in unseren Metatranskriptombibliotheken anzureichern. Alle erhaltenen 454 reads sind vor der eigentlichen Datenverarbeitung vorverarbeitet worden um potentielle Unklarheiten auszuräumen. Insgesamt 10,000 SSU-ribotags (Gesamt-RNA) sind analysiert worden um die Gemeinschaftsstruktur in den oxischen und anoxischen Zonen an drei unterschiedlichen Inkubationszeitpunkten (25, 45 und 90 Tage, nach dem Pflanzen von Reiskeimlingen) aufklären zu können. Zusätzlich konnten 45,000 beziehungsweise 12,000 mRNA-tags (angereicherte mRNA) für die Analyse funktioneller Aktivitäten, bezogen auf die oxische und anoxische Zone 90 Tage alter Reismikrokosmen gewonnen werden.

Die Analyse der SSU-ribotags hat keine gravierenden Veränderungen der mikrobiellen Gemeinschaften gezeigt, mit Ausnahme von *Geobacter*, Clostridien und Methanogenen in der anoxischen Zone. Nichtsdestotrotz konnte belegt werden das klare taxonomische Unterschiede zwischen der oxischen und anoxischen Zone vorliegen, wobei Cyanobakterien die dominierende Gruppe der oxischen Zone darstellen. Obwohl mRNA-tags bezogen auf grundlegende zelluläre Funktionen das Gros in beiden Datensätzen ausmachten, konnten spezifische Funktionen in Abhängigkeit zur Verfügbarkeit von Sauerstoff gezeigt werden. So zum Beispiel

Methanoxidierung in der oxischen Zone und Methanogenese unter anoxischen Bedingungen.

Der von uns etablierte Ansatz der Metatranskriptomik stellt ein nützliches Mittel dar um die Zusammensetznung und funktionelle Genexpression von komplexen, terrestrischen mikrobiellen Gemeinschaften untersuchen zu können, ohne den Beschränkungen PCR-basierter Techniken zu unterliegen.

1. Introduction

Microorganisms are found everywhere in the biosphere, including soil, water bodies and sediments, but also in symbiotic interactions with plants and animals. They may even be detected in aerosols in the atmosphere. Besides, they exist at extreme environmental conditions such as, for example, high pressure, salt concentration or temperature. The total number of prokaryotic cells on earth was estimated to be 4-6 × 10^{30} cells. The prokaryotic biomass includes 350-550 Pg of C, 85-130 Pg of N and 9-14 Pg of P (Whitman *et al*, 1998). The total amount of prokaryotic carbon is comparable to that of total plant biomass (Hogan, 2009). As the most diverse group of organisms on earth, microbial growth and activities greatly influence the cycling and transformation of nutrients that make up living systems, involving, for example, fixation of gases (C & N), decomposition of organic matter or conversion of inorganic substances to simple molecules that can be exploited by other organisms.

1.1. Microorganisms exist in complex communities

Most microorganisms in nature live under either nutrient- or energy-limiting conditions. In order to survive in infertile situations, they exist in complex communities, interacting with physiologically different microorganisms in contiguous environments (Newman & Banfield, 2002). They communicate in spatial and temporal heterogeneity, primarily via the secretion and receipt of products of their own metabolic activities or extracellular signaling molecules (Surette *et al*, 1999). These signaling mechanisms enable microorganisms to detect changes in their environment, construct mutually beneficial associations with other organisms, gain advantages over competitors, and communicate with their host (Fenchel, 2002). For instance, fermenting bacteria produce hydrogen as a metabolite of carbohydrate metabolism. Hydrogen is then oxidized and converted into CH_4 by methanogens with CO_2 as an electron acceptor. The consumption of hydrogen by methanogens makes the complete fermentation of carbohydrates thermodynamically favorable. Understanding of microbe-microbe

and microbe-environment interactions remains one of the great challenges in microbial ecology.

One of the possibilities to study such interactions is cultivation. However, it is estimated that from most environments, less than 1% of microorganisms are cultivable under laboratory conditions (Rondon *et al*, 1999). Microbial function and activity in complex communities can differ from the individual behavior in laboratory culture (Schink, 2000). For example, *E.coli* is the most studied bacterial species, used for the development of genetic tools and as a model system for multiple research foci such as, for example, bacterial cell growth, and biochemical behavior and structure. However, the ecological role and niche of *E. coli* is not yet fully understood (van Elsas *et al*, 2011).

Therefore, the *in situ* identification of community members is one of the major challenges in microbial ecology, addressed in particular by molecular ecology methods. Diversity and structure of microbial communities are examined using various techniques including direct counting (Bloem, 1995; Weinbauer *et al*, 1998), phospholipid fatty acid (PLFA) analysis (Zelles & Bai, 1994; Yu & Harch, 2001), fluorescent *in situ* hybridization (FISH) (Christensen *et al*, 1999), molecular fingerprinting such as amplified rDNA restriction analysis (ARDRA) (Gich *et al*, 2000), terminal restriction fragment length polymorphism (T-RFLP) analysis (Horz *et al*, 2000) and denaturing/temperature gradient gel electrophoresis (DGGE/TGGE) (Heuer *et al*, 2001), and cloning and comparative sequence analysis of phylogenetic and functional marker genes. All these approaches can provide information on the diversity, structure and compositional change of microbial communities. They may also provide insights into a particular functional potential when functional markers such as *amoA*, *pmoA* or *mcrA* are used for analysis. However, they cannot provide global information on functional activity and changes therein, which occur in response to environmental cues and may be related to, for example, nutrient uptake, energy flow, or degradation of substrates.

Moreover, the distribution of metabolic pathways does not necessarily correspond to the microbial rRNA gene-based phylogeny, presumably due to lateral gene transfer (Pace, 1997). Therefore, species identification often does not allow firm conclusions about community function. Collectively, there is a strong need for alternative approaches that allow community-wide analysis of functional gene contents (metagenomics), functional gene transcription (metatranscriptomics), and functional protein production (metaproteomics) (**Figure 1.1**).

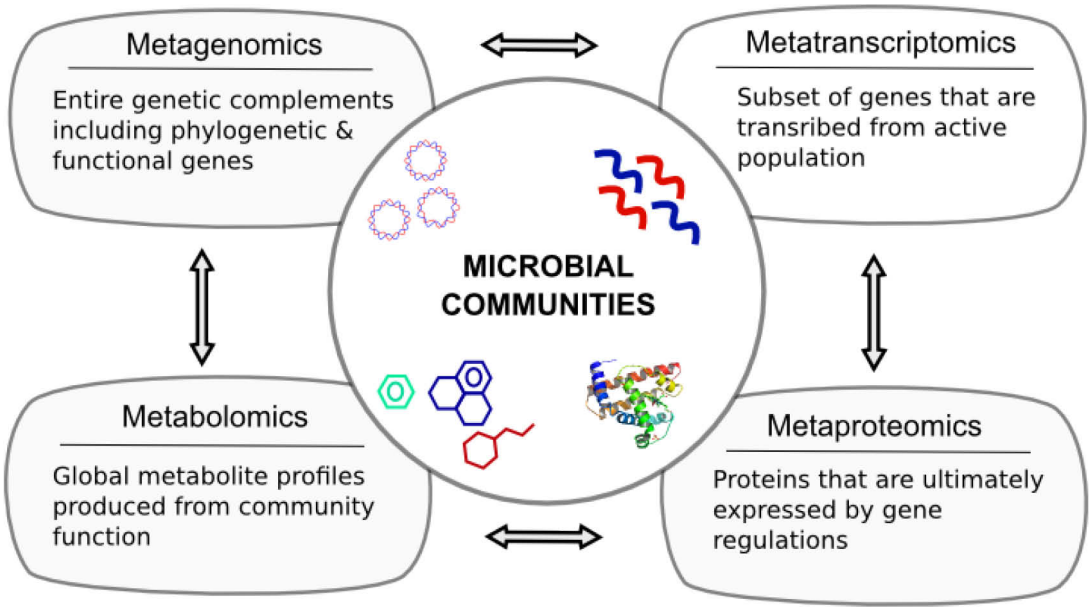

Figure 1.1. Meta-omics approaches to microbial community analysis.

1.2. Functional analysis of microbial community

It is reported that a considerable amount of the bacterial genome is dedicated to shaping the organisms' habitats and maintaining their ecosystems, representing a major commitment to community and niche (Phelan *et al*, 2011). Cultivation-independent genomic approaches have greatly advanced our understanding of the genetic diversity and potential of microbial communities. Metagenomics

enables us to study microorganisms by deciphering their genetic information directly from the environment without the need to cultivate. 16S rRNA gene surveys offer insights into the phylogenetic composition of microbial communities. However, environmental genomics and 16S rRNA gene surveys only provide limited clues about the functional significance of the observed genes, as metagenomic approaches may detect not only viable but also non-viable cells including dead populations (Fukui *et al*, 1996; Schmid *et al*, 2001).

Hence, *in situ* activities (and changes therein) in response to environmental factors or stress have to be monitored through functional units such as mRNA molecules or proteins. So far, metabolic activities of microbial communities were investigated by reverse transcription quantitative PCR (RT-qPCR) of marker genes specific for particular functional guilds such as nitrogen fixation (*nifH*), nitrite reduction (*nirS, nirK*), ammonia oxidation (*amoA*), or methane oxidation (*pmoA*) (Bürgmann *et al*, 2003; Chen *et al*, 2007; Ebie *et al*, 2004; Knauth *et al*, 2005; Kolb *et al*, 2005; Sharma *et al*, 2005). Basically, RT-qPCR requires sequence information of target genes to design primers and probes. As a consequence, the range of detection is limited by their target specificity. Zhou and his colleagues developed a functional gene array, which contains several thousands of oligonucleotide probes in hundreds of gene categories (He *et al*, 2007). Microarray-based techniques allow to overcome the constraints related to the number of genes whose expression levels can be analyzed and quantified simultaneously. However, functional genes often contain highly conserved regions and sequence motifs. Thus, potential cross-hybridization of highly related sequences and the need of complex quantification algorithms are major challenges in microarray applications. Massively parallel sequencing technologies enable high-throughput monitoring of community-wide expression profiles at reasonable costs (Frias-Lopez *et al*, 2008; Vila-Costa *et al*, 2010; Stewart *et al*, 2011). Community-wide patterns of mRNA expression and their variations in response to environmental change can be monitored via RNA-seq without prior knowledge. Therefore, metatranscriptomics lately became the state-of-the-art technology to investigate community-wide functional activity in marine environments.

Since proteins mediate diverse functions, such as catalytic activities and cellular building blocks (Graham *et al*, 2007), the identification of expressed proteins on a large scale would provide a profound picture of the functional and structural status of a microbial community. However, metaproteomics is not suitable to investigate community-wide changes in functional activity with high throughput (Zhang *et al*, 2010). Metabolomics has advantages in deciphering microbial metabolism. Metabolites are the currency of metabolic reactions occurring in cells. The levels of their concentrations change in response to genetic or physiological status of microbial communities (Raamsdonk *et al*, 2001). However, microbial metabolomics needs improvements in resolution and detection sensitivity (Cascante & Marin, 2008). Metabolomics may have greater potential to perform functional analysis of a microbial community when combined with metatranscriptomics or metaproteomics (Dunn, 2008).

1.3. Next-generation sequencing technology

Over the past 6 years, there has been a great progress in sequencing technology from conventional Sanger sequencing to next-generation sequencing (NGS), which is defined by massively parallel sequencing in a single run and, as a consequence, a price per base that is several orders of magnitude lower than that of previous technologies. The performance of NGS platforms which are currently available or will be available soon on the market is compared in **Table 1.1**. The Illumina sequencing platform generates millions of short reads (single reads: 75-120 bp; paired-end reads: 200 bp) per run with the lowest cost-per-base and is being used primarily for *de novo* and re-sequencing of genomes, including human, viral and bacterial genomes. Meanwhile, 454 GS Titanium/FLX+ offers the longest reads with average read lengths of 400 bp (GS FLX) and 750 bp (GS FLX+) among currently available sequencing technologies combined, however, with relatively low throughput of 1,000,000 reads per run. Increased read length has great advantages to determine the community composition and genetic structure of complex microbial communities, and

Platform	Library/template preparation	NGS chemistry	Mean Read length (bases)	Reds per run	output per run	Consensus accuracy	Run time	Pros	Cons
Sequencing by synthesis									
Roche/454 GS FLX +	Frag[a], MP[b]/ emPCR[c]	PS[d]	750	~1 milllion	700 Mb	99.997 %	23 h	Longer reads, fasta run time; good choice for de novo assembly	High reagent cost; high error rates in homopolymeric region
Illumina HiSeq 2000	Frag[a], MP[b]/ solid-phase	RTs[e]	2 x100	1 billion	600 Gb	> 80% of bases higher than Q30	8.5 d	Currently the most widely used platform	Shorter read length, less feasible for de novo assembly
IonTorrent PGM	Frag[a]/ emPCR[c]	Natural nucleotide	200	5 million	1 Gb	99.99 %	2 h	Very fast run time, cost effective, open source	Not as much throughput
Sequencing by ligation									
Life technology SOLiD 5500	Frag[a], MP[b]/ emPCR[c]	Cleavable probe SBL[f]	2 x 60	1.4 billion	90 Gb	99.99 %	7 d	Two-base encoding provides inherent error correction	Long run time
Single-molecule sequencing									
Helicos BioScience HeliScope	Frag[a], MP[b]/ Single molecule	RTs[e]	35	~ 1 billion	35 Gb	99.995 %	8 d	High multiplexing ability, no template amplification needed	Short read lenth, high error rates compared with RT-based platform
Pacific Bioscience PacBio RS	Frag[a] only/ Single molecule	Real-time	1,300	~45,000	50 Mb	99.999 %	~1 h	The longest read length at present; fast run time to result	Less high-throughput compared to other NGS platforms
Sequencing by hybridization									
Oxford Nanopore technology GridION	Frag[a] only/ Single molecules	Real-time	Ultra long (Mirror fragment size)	N/A	Tens of Gb	> 99.99%	within 24 h	Low cost, ultra long read length, high accuracy	Limited application to RNA-seq

Table 1.1. Comparison of next-generation sequencing platforms. Frag[a]: fragment run, MP[b]: mate-pair run, emPCR[c]: emulsion PCR, PS[d]: pyrosequencing, RTs[e]: reversible terminator, SBL[f]: sequencing by ligation

putative recombination events among their members (Konstantinidis *et al*, 2009; Mackelprang *et al*, 2011). The average read length of the current 454 GS Titanium/FLX+ chemistry covers 40% (GS FLX) or 80% (GS FLX+) of the average sequence length of bacterial genes (~950 bp) (Casjens, 1998). In particular, metatranscriptome data sets from soil microbial communities for which mostly a specific reference metagenome will not be available for mapping require sufficient read lengths for functional annotation and taxonomic binning with high significance (**Figure 1.2**). In this PhD project, 454 sequencing was, therefore, the method of choice to investigate bacterial mRNA expression profiles in soil.

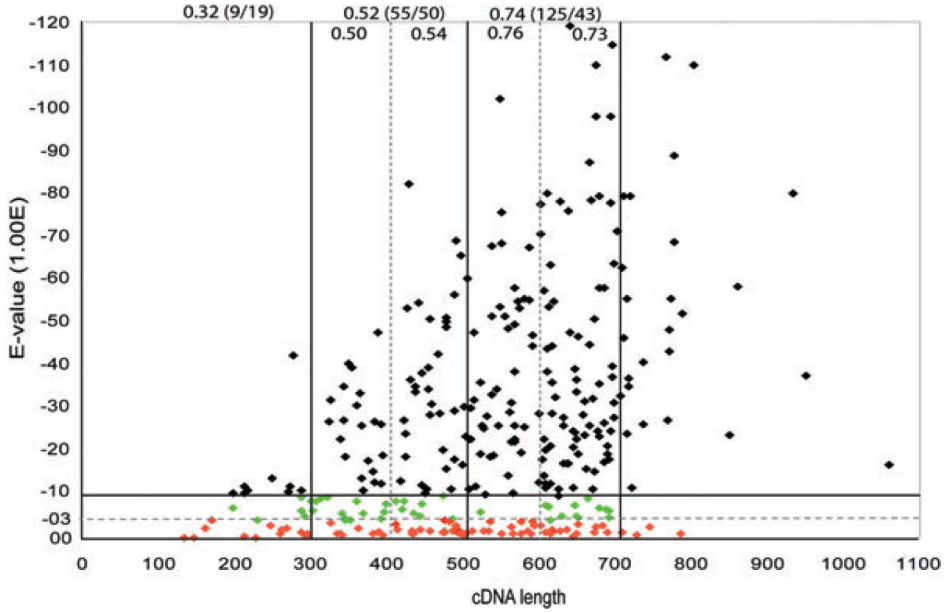

Figure 1.2. Plot of expressd sequence tag (EST) cDNA length versus BLASTX *E*-value. The proportion of assignable ESTs within the different length categories is indicated at the top. The number of assignable versus non-assignable ESTs is given in parenthesis (Shrestha *et al*, 2009)

1.4. Application of next-generation sequencing to environmental transcriptomics

Potential metabolic activity of microbial communities has been investigated by PCR-based detection and analysis of phylogenetic and functional markers,

combined with cellular incorporation of C^{13} or N^{15} offered in labeled substrates that are utilized only by particular functional guilds, such as methane oxidizers or methanogens (Lu & Conrad, 2005; Hori *et al*, 2007; Dumont *et al*, 2011; Pratscher *et al*, 2011). In 2005, the first attempt to monitor functional activity of a bacterial community *in situ* was made by Poretsky *et al*. (2005). cDNA libraries were constructed from mRNA that had been freshly extracted from marine environment. Their analysis revealed various transcripts involved in environmentally important processes, such as sulfur oxidation, C1 carbon assimilation and nitrogen fixation. Metatranscriptomics potentially reveals functional gene expression under natural environmental conditions without PCR-based bias towards known genes. Various metatranscriptome analyses in marine environments have followed this initial study, using massively parallel 454-pyrosequencing. Novel groups of small RNAs were found to be present in high numbers in metatranscriptome data sets (Shi *et al*, 2009). More recently, community-wide responses of marine microbial assemblages to substrate amendments (e.g., dimethylsulfoniopropionate, polyamines putrescine, and spermidine) were investigated to characterize the functional activity of those members able to utilize the substrates offered (Vila-Costa *et al*, 2010; McCarren *et al*, 2010; Mou *et al*, 2011). Differential expression of transporter genes triggered by dissolved organic carbon (DOC) was also examined by metatranscriptomics, revealing possible bioreactive components of the coastal DOC pool and ecological roles of certain bacterial taxa in carbon turnover (Poretsky *et al*, 2010).

As shown in Table 2, most of the metatranscriptome studies that have been done so far were focused on open ocean microbial assemblages. Only very few studies have analyzed the metatranscriptome of soil microbial communities. This is because (i) the extraction of highly-purified soil mRNA is very challenging, (ii) soil microbial communities are more diverse than marine assemblages and vary greatly in their composition between sites and, as a consequence, (iii) available reference genomes or metagenomes for functional annotation and taxonomic binning are poorly available as compared to marine microbial assemblages.

References	Radax *et al.* (2012)	Shi *et al.*(2011)	Hollibaugh *et al.* (2011)	McCarren *et al.* (2010)	Vila-Costa *et al.* (2010)	Shi *et al.* (2009)	Poretsky *et al.*(2009)	Shrestha *et al.* (2009)	Gilbert *et al.* (2008)	Urich *et al.* (2008)	Frias-Lopez *et al.* (2008)	Bailly *et al.* (2007)	Poretsky *et al.* (2005)
Study site	Marine sponge	Ocean	Ocean	Ocean	Ocean	Ocean	Ocean	Soil	Ocean	Soil	Ocean	Soil	Ocean
Template molecules	total RNA	enriched mRNA	enriched mRNA	total RNA	enriched mRNA	enriched mRNA	enriched mRNA	enriched mRNA	enriched mRNA	total RNA	enrichd mRNA	Eukaryotic mRNA	enriched mRNA
mRNA enrichment	not applied	SH[a]	Exo[b] followed by SH[a]	not applied	Exo[b] followed by SH[a]	SH[a]	Exo[b] followed by SH[a]	SH[a]	SH[a]	Not applied	SH[a]	Oligo-dT based SH[a]	SH[a]
% of non-rRNA	8.1	18-40	49.0	3.6	51	53.2	47.1	95	99.9	8.2	47.1	100	80
Template amplification	not applied	IVT[c]	IVT[c]	IVT[c]	IVT[c]	IVT[c]	IVT[c]	rtPCR[d]	MDA[e]	Not applied	IVT[c]	rtPCR[d]	rtPCR[d]
cDNA synthesis	random primer	random primer	random primer	random primer	random primer	random primer	random primer	SD14 primer	random primer	random primer	random primer	oligo-dT rimer	semi-random primers and SD14
Sequencing platform	GS FLX	GS FLX	GS FLX	GS FLX	GS FLX	GS 20	GS 20	Sanger sequencing	GS 20	GS 20	GS 20	Sanger sequencing	Sanger sequencing
Average read length (bp)	223	186	211	~200	209	103	99	534	98	107	114	not indicated	not indicated
Total no. of reads	262,292	3,149768	2,181,899	2,945,424	606,286	388,738	240,422	805	506,353	258,411	128,324	119	282

Table 1.2. Comparison of methods applied for analyzing microbial metatranscriptome. SH[a]: subtractive hybridization of rRNA, Exo[b]: 5'-monophosphate dependent exonuclease treatment, IVT[c]: *in vitro* transcription, rtPCR[d]: reverse transcriptase PCR, MDA[e]: multiple displacement amplification.

The first metatranscriptome study in soil focused on the functional diversity of eukaryotic (mainly fungal) microorganisms by selective sequencing of poly-adenylated transcripts (Bailly *et al*, 2007). Urich and colleagues made an attempt to relate taxonomic groups to their ecological function by assessing community structure and function simultaneously (Urich *et al*, 2008). cDNA of total community RNA was analyzed without prior mRNA enrichment. As a consequence, a very limited number of putative mRNA transcripts were detected. More recently, the functional status of paddy soil bacterial communities was surveyed by RT-PCR of enriched mRNA, using a primer (SD14 oligonucleotide) that specifically targets the Shine-Dalgarno sequence of bacterial transcripts (Shrestha *et al*, 2009). This study was effective in analyzing a high proportion of protein-coding transcripts, but had the use of PCR and primer bias as a major shortcoming. Therefore, we developed an improved method for extraction of total RNA from soil and assessed different strategies to enrich intact mRNA suitable for soil metatranscriptomics (Mettel *et al*, 2010).

1.5. Rice paddy as a model system

CH_4 is the second most important greenhouse gas with a global warming potential greater than that of CO_2. Flooded rice fields are one of the major biogenic sources of atmospheric methane. The emission of CH_4 from flooded rice field is the net balance between CH_4 production by methanogenic archaea in anoxic bulk soil and CH_4 oxidation by methanotrophic bacteria in oxic surface soil. Due to the importance of microbial action to produce and consume the greenhouse gas CH_4, flooded rice paddies have long been used as model systems to investigate community structure and functional groups such as, for example, methanogens, and methanotrophs. Besides methane production, flooding creates oxygen-limited conditions. Flooded rice paddy soil can be differentiated into three compartments having different physico-chemical conditions: oxic surface soil, anoxic bulk soil, and rhizosphere (**Figure 1.3**).The latter compartment may be further separated into rhizoplane and rhizosphere soil.

Between these compartments, microscale oxygen and chemical gradients were defined and provided information about the activity and spatial distribution of functional groups of microorganism. Against this background, flooded rice-planted microcosms were studied for temporal changes in community composition and active metabolic pathways under different conditions of oxygen availability.

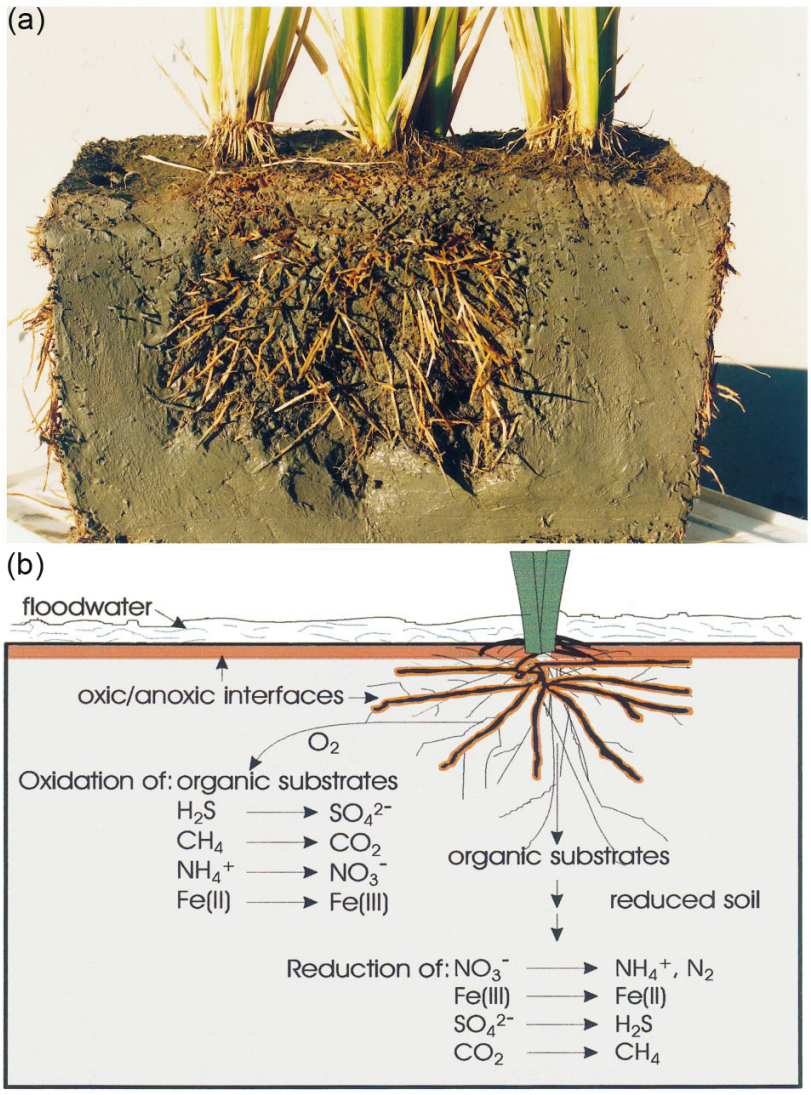

Figure 1.3. Cross-section through a drained rice microcosm (a) and schematic cross-section through the compartmentalized rice paddy soil (b). The rice was cultivated for 90 days under flooded condition in the greenhouse. Redox reactions characteristic of oxic and anoxic zones are shown (Liesack *et al*, 2000).

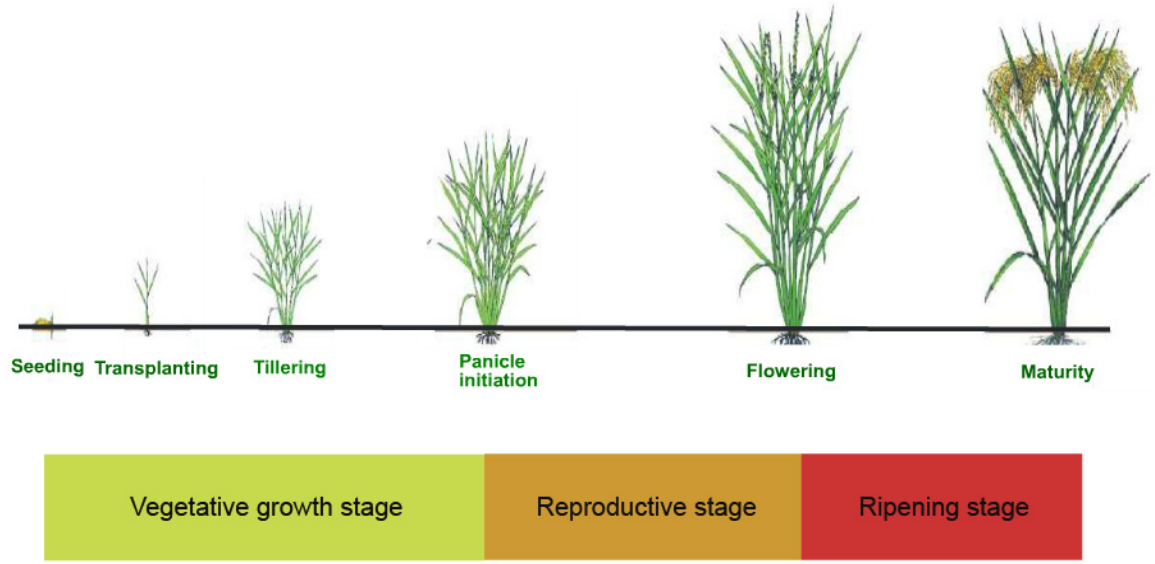

Figure 1.4. Growth stage of rice plant from germinating seed to a mature plant. (International Rice Research Institute: http://www.knowledgebank.irri.org/)

1.5.1. Growth stage of rice plant

The growth season of the rice plant is generally divided into the vegetative, reproductive, and ripening stage (**Figure 1.4**). The vegetative growth stage is characterized by active tillering and an increase in height. During this period, the seminal root develops and starts growing both vertically and laterally. The reproductive stage begins with panicle formation at the end of the last internode protected by leaf sheaths and, depending on varieties, lasts 30 to 45 days to flowering. Rice plant development in reproductive stage is most greatly affected by environmental conditions that determine crop yield, such as weather and nutritional status. During the ripening stage, the starch and sugars accumulated in the culms and leaf sheath are transferred to grains, increasing grain in size and weight. The carbohydrates are photosynthesized and fill grains. Therefore, the ripening stage is very sensitive to light intensity and temperature.

The amount and composition of exudates released from rice roots change over growth phase, resulting in considerable variation in CH_4 emission in relation to plant age (Aulakh *et al*, 2001; Wang & Adachi, 2000). Since root exudates

provide substrates for methanogenesis in anoxic bulk soil, diversity and community structure may change over time. Therefore, this study aimed to elucidate the taxonomic composition of microbial communities in the oxic surface layer and the anoxic bulk soil of rice-planted microcosms by use of a PCR-independent ribotag analysis.

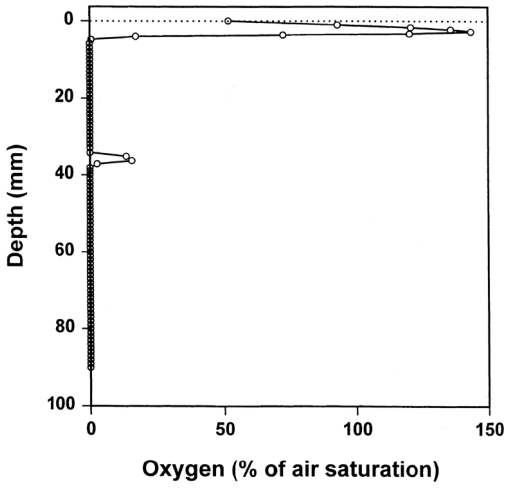

Figure 1.5. *In situ* profile of oxygen down to a depth of 90 mm in the soil. The vertical profile was measured 5 mm from the base of a rice plant. The upper O_2 maximum was due to microphytobenthic photosynthesis, whereas the lower maximum was associated with rice roots (Revsbech *et al*, 1999).

1.5.2. Various microbial processes in flooded rice paddy

After flooding of rice fields, oxygen is consumed by aerobic bacteria and chemical oxidation reactions. Dissolved O_2 is depleted within top 5 mm of surface layer in flooded rice-planted paddy soil microcosms (**Figure 1.5**) (Revsbech *et al*, 1999). In the oxic surface layer, apart from degradable organic matter, the diffusion of methane produced in the anoxic zone causes an increase in the biological oxygen demand (Gilbert & Frenzel, 1995). A significant amount of oxygen is consumed for the oxidation of reduced compounds such as sulfide, ammonium, and ferrous iron (Canfield *et al*, 1993). The gradients of terminal electron acceptors and reduced compounds govern the activities and spatial distribution of diverse functional groups of microorganisms in flooded rice paddy soil.

The degradation of organic matter to methane and carbon dioxide is well characterized in anoxic rice field soil (Chin *et al*, 1998). Organic matter that remains after harvest is hydrolyzed by a variety of cellulases, hemicellulases, and pectinases secreted from cellulolytic microorganisms. Groups of fermenting bacteria convert the resulting sugar monomers into acetate, alcohols, fatty acids, CO_2 and H_2. Among these fermentation products, more reduced ones, such as alcohols and fatty acids, are further oxidized by acetogenic bacteria in syntrophic association, producing mainly acetic acids as well as carbon dioxide and hydrogen. Methanogens utilize the intermediate or end products of hydrolysis, fermentation, and acetogenesis and convert them into CH_4. The concerted action of all functional groups of microorganisms involved in these processes drives the biogeochemical cycles in flooded rice paddies.

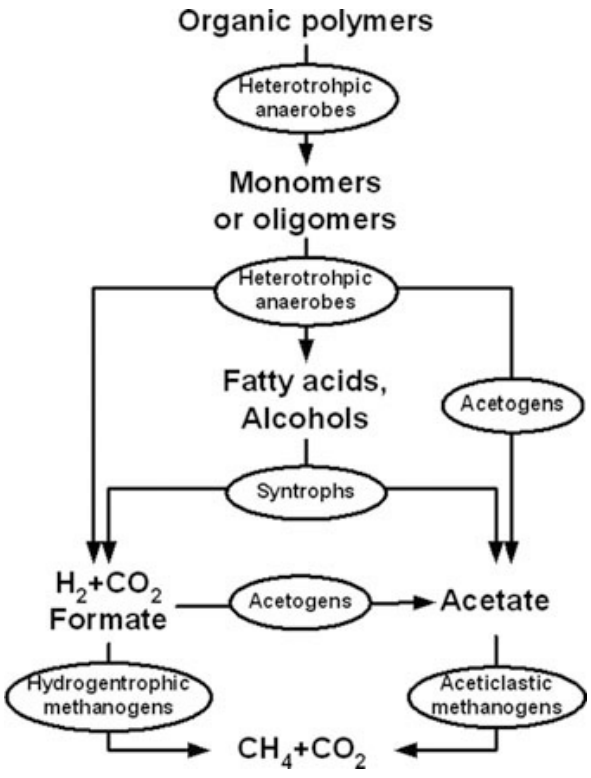

Figure 1.6. Anaerobic degradation pathway of organic matter to methane. Major functional groups of microbial organisms catalyzing the reactions are shown in ellipses (Liu & Whitman, 2008).

1.6. Aim of the project

The main objective of my PhD project was to establish the complete method procedure for soil metatranscriptomics. This had to include the extraction of total RNA, enrichment of high-quality mRNA, construction of cDNA libraries for 454-pyrosequencing, and setting up a bioinformatic analysis pipeline for large sequence data sets generated by NGS technology. Using this method procedure, the metatranscriptome of microbial communities inhabiting the oxic surface layer and anoxic bulk soil of flooded rice paddy soil microcosms should be analyzed. The microcosms were incubated in the greenhouse under normal day/night cycles. Samples for metatranscriptome analysis were taken from three different time points, corresponding to the different growth stages of the rice plant: tillering, flowering, and ripening. We hypothesized that the simultaneous assessment of community composition (total RNA) and functional activity (enriched mRNA) should provide information on which bacterial species are involved in biogeochemical cycling through which metabolic pathways. We expected to detect major differences in the taxonomic composition (ribotags) and in the expression of particular functional genes (mRNA-tags) when comparing metratranscriptome data sets obtained from communities adapted to different physico-chemical conditions, such as oxic versus anoxic paddy soil. Overall, we anticipated that the combined analysis of taxonomic composition and functional gene expression by this PCR-independent approach represents a new avenue to link microbial community structure to active functioning in flooded rice paddy soil and other soil environments.

For easy understanding, small subunit ribosomal RNA of microbial community was described with SSU rRNA as a collective term and small subunit rRNA of a bacterial organism was expressed by 16S rRNA throughout the text.

2. Methodology

2.1. Materials

2.1.1. Soil sample

Soil was taken from drained paddy field of the Italian Rice Research Institute in Vercelli, Italy. The soil characteristics have been described previously (Holzapfel-Pschorn *et al*, 1985).

2.1.2. Instruments

Items	Manufacturer
FastPrep®-24 bead beater	MP Biomedicals, USA
NanoDrop® ND-1000 UV-Vis spectrophotometry	NanoDrop Tech. Inc., USA
Qubit® 2.0 Fluorometer	Invitrogen, USA
Experion automated electrophoresis system	Bio-Rad, USA
Mastercycler Gradient	Eppendorf, Germany
Magnetic stand	Invitrogen, Germany
FlashGel™ Camera	Lonza, USA
FlashGel™ Dock system	Lonza, USA
QuantiFluor™	Promega, USA
Swing-bucket rotor A-2-DWP	Eppendorf, Germany
Centrifuge 5804R	Eppendorf, Germany
Ultra-Turrax tube drive Multipette® plus	IKA, Germany Eppendorf, Germany
454 GS Junior sequencer	Roche/454, USA

2.1.3. Chemicals and reagents

Items	Manufacturer
Absolve™	PerkinElmer, Boston, USA
Urea	Merck, Darmstadt, Germany
KH_2PO_4	Sigma, Steinheim, Germany
KCl	Fluka, Buchs, Switzerland
DEPC-treated water	Ambion, Austin, USA
Tris-HCl	Sigma, Steinheim, Germany
Polyvinylpyrrolidone K25	Fluka, Buchs, Switzerland
Water-saturated phenol (pH4.5)	Carl Roth, Karlsruhe, Germany
Phenol-chloroform-isoamyl alcohol (pH5.5)	Carl Roth, Karlsruhe, Germany
Chloroform-isoamyl alcohol [24:1 (v/v)]	Carl Roth, Karlsruhe, Germany
$MgCl_2$	Carl Roth, Karlsruhe, Germany
Na_2EDTA	Sigma, Steinheim, Germany
SDS	Merck, Darmstadt, Germany
Sodium acetate	Sigma, Steinheim, Germany
Isopropanol	Carl Roth, Karlsruhe, Germany
Ethanol (Nuclease-free)	Applichem, Darmstadt, Germany
RNase-free TE buffer	Applichem, Darmstadt, Germany
Q-sepharose® Fast Flow	Sigma-Aldrich, Sweden
NaCl	Carl Roth, Karlsruhe, Germany
Humic acids	Sigma, Steinheim, Germany
Glass beads (0.17-0.18 mm)	Satorius, Goettingen, Germany
Sodium hydroxide solution (10M)	Sigma, Buchs, Switzerland

2.1.4. Enzymes and kits

Items	Manufacturer
Non-stick RNase-free 1.5 ml microfuge tube	Ambion, Austin, USA
Turbo™ DNase	Ambion, Austin, USA
RNasin® Ribonuclease Inhibitor	Promega, Madison, USA
AutoSeq™ G-50	GE Health Care, Germany
RNA Clean & Concentrator™-5	Zymo Research, USA
Qubit® RNA assay kit	Invitrogen, Oregon, USA
Ribo-Zero™ rRNA removal Kit (Meta-Bacteria)	Epicentre, Madison, USA
Microb*Express*™ Bacterial mRNA Purification kit	Ambion, Austin, USA
Experion™ RNA HighSens Analysis Kit	Bio-Rad, Hercules, USA
SuperScript® II reverse transcriptase	Invitrogen, Carlsbad, USA
NEBNext® mRNA Library Prep Master Mix	NEB, Ipswich, USA
GS Rapid Library Prep Kit	Roche, Branford, USA
E.coli rRN A standard	Roche, Indianapolis, USA
Wizard® SV Gel and PCR Clean-up Kit	Promega, Madison, USA
Agencourt® Ampure® XP	Beckman Coulter, USA
FlashGel® DNA Cassette (1.2%)	Lonza, Rockland, USA
GS Junior Titanium emPCR Kit (Lib-L)	Roche, Branford, USA
GS Junior Titanium Sequencing Kit	Roche, Branford, USA
GS Junior Titanium PicoTiterPlate Kit	Roche, Branford, USA
DNA Smart Ladder	Eurogentec, Seraing, Belium

2.2. Methods

2.2.1. Seeding

Dry rice seeds (*Oryza sativa variety* KORAL type japonica) were spread and germinated on moist filter paper in a plastic petri dish at room temperature for 14 days.

2.2.2. Preparation of microcosms

Drained paddy soil collected from the Italian Rice Research Institute in Vercelli (Italy), was air-dried at room temperature. The soil was sieved (pore size: 2 mm) prior to its use. Soil (1.8 kg) was mixed with 940 ml of demineralized water and 60 ml of fertilizer solution containing 9.89 g of urea, 7.605 g of KH_2PO_4 and 7.07 g of KCl per liter. The soil suspension was then filled into a 2.5-liter pot. In the center of the pot, a nylon mash bag (25 µm mesh; 7 cm in depth and 9 cm radius) was placed to separate roots and rhizosphere from the bulk soil, but to allow free movement of water and root exudates between the two compartments. The prepared microcosms were flooded with demineralized water to a level 5 cm above the soil surface. The flooded microcosms were covered with aluminum foil and incubated for 5 days in the greenhouse.

2.2.3. Transplantation and incubation

After incubating the flooded microcosms for five days, one 14-day-old rice seeding was transplanted into the center of the nylon bag in each pot. Fertilizer solution (20 ml) was added to the flooding water. The day of plantation was taken as day zero of the rice paddy soil incubation. Flooded rice microcosms were incubated in the greenhouse with a relative humidity of 70%, 12 hours light period and 28/22 °C day/night temperature (Shrestha *et al*, 2008). The flooding water was kept at a level of 5 cm above soil surface by adding demineralized water

every two days over the complete incubation period. After 50 days of incubation, 20 ml of fertilizer solution was again added to the microcosms.

2.2.4. Sampling

Sampling points were chosen according to plant growth stage. After 25 days, 45 days and 90 days of incubation, flooded water was discarded. Oxic soil of the upper 3-mm surface layer was carefully sampled using a small spatula. Wet soil was collected in 1-gram aliquots in 2-ml screw cap tubes and shock-frozen in liquid nitrogen. In order to sample the anoxic bulk soil, the nylon bag containing the rice plant was removed from the microcosms. After rice plant removal, dark bulk soil was immediately collected in 2-ml screw tubes and shock-frozen in liquid nitrogen. For each sampling time point except 25 days, two pots were incubated and sampled to produce data sets from two independent replicate microcosms. Several soil aliquots were sampled from each replicate microcosm to ensure sufficient raw material for RNA extraction and processing. The shock-frozen samples were stored at -80 °C until RNA extraction.

2.2.5. Extraction of total RNA

Prior to extraction, glassware was made RNase-free by treatment with diethyl pyrocarbonate (DEPC) treated water (Blumberg, 1987). The working bench was treated with 2% Absolve™ (Perkin, USA). The frozen soil aliquots were mixed with an equal volume of glass beads (0.17 to 0.18 mm in diameter) and resuspended in 700 µl of TPM buffer (50 mM Tris-HCl [pH5.0], 1.7% [wt/vol] polyvinylpyrrolidone, 20 mM $MgCl_2$). Subsequently, the mixture was shaken in a bead beater at 6.0 ms^{-1} for 35 seconds (FastPrep®-2). Soil, glass beads and cell debris were pelleted by centrifugation at 20,000 × *g* for 1 minute at 4℃, and the supernatant was transferred to a fresh 2-ml microcentrifuge tube. 700 µl of PBL buffer (5 mM Tris-HCl [pH5.0], 5 mM Na_2EDTA, 0.1 % [wt/vol] sodium dodecyl

sulfate, and 6% [vol/vol] water-saturated phenol) was added to the soil pellet, and the lysis procedure was repeated as described above. The supernatant of the second lysis was mixed with that of the first lysis. The pooled supernatants were treated with 500 µl of water-saturated phenol (pH 4.5), and then homogenized by shaking vigorously. After centrifugation at 20,000 × *g* for 3 minutes at 4 ℃, the supernatant was extracted with phenol-chloroform-isoamyl alcohol (25:24:1 [vol/vol/vol], pH 5.5) and finally with chloroform-isoamyl alcohol (24:1 [vol/vol]). The resulting aqueous phase was mixed with 0.1 volume of 3 M sodium acetate (pH 5.7) and 0.7 volume of isopropanol, incubated at room temperature for 5 minutes, and centrifuged for 30 minutes at 20,000 × *g* and 4 ℃. The nucleic acid pellet was washed with 500 µl of 70 % ethanol, air dried for 2 minutes, and resuspended in 25 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH8.0]) (Ambion). Ten samples equal to 10 gram of soil were processed simultaneously. The collective amount of total RNA was pooled into one 1.5-ml non-stick RNase-free microfuge tube (Ambion). Subsequently, 5 U Turbo™ DNase (Ambion) and 10 U RNasin® Plus RNase inhibitor (Promega) were added to the pooled RNA sample, which was adjusted to 300 µl with 1× DNase buffer. After incubation for 60 minutes at 37 ℃, DNase-treated total RNA was immediately purified using Q-Sepharose chromatography.

2.2.6. RNA purification

An aliquot (750 ul) of Q-Sepharose Fast Flow (Sigma-Aldrich) was transferred to an empty illustra™ Autoseq® column, and packed by centrifugation with 650 × *g* for 10 seconds. A packed Q-Sepharose column was washed three times with 200 µl of DEPC-treated water (Ambion) by centrifugation at 650 × *g* for 10 seconds. The DEPC water was discarded after each washing step. The total RNA obtained from the DNase treatment was then loaded onto the resin, equilibrated for 1 minute, and then centrifuged at 400 × *g* for 7 seconds. The RNA was selectively eluted by loading 80 µl of 1.5 M NaCl in DEPC-treated water (pH 5.5) onto the resin and centrifugation at 400 × *g* for 5 seconds. This elution step was repeated

five times and the eluates were collected in a single 2-ml tube. The RNA eluate was precipitated as described above. The precipitate was dissolved in 100 µl RNase-free TE buffer and subjected to RNA Clean & Concentrator™-5 (Zymo Research) according to the small RNA elimination protocol, resulting in the removal of fragments < 200 nt. Total RNA was eluted twice with 20 µl of RNase-free TE buffer.

2.2.7. Determination of humic acid content and purity of RNA extracts

Serial dilutions of commercially available humic acids were prepared in distilled water. Using a NanoDrop ND-1000 spectrophotometer, a linear relationship between UV absorbance at 400 nm (A_{400}) and the concentration of humic acids was determined. The amount of co-extracted humic acids in purified RNA was measured by comparing the absorbance of each sample at 400 nm against a standard curve generated from commercial humic acids (Mettel *et al*, 2010). The purity of RNA was evaluated by measurement of absorbance at 230, 260, and 280 nm. The ratio of A_{260}/A_{280} and A_{230}/A_{260} indicates protein or phenol contamination and salt contamination, respectively. For highly purified RNA, A_{260}/A_{280} and A_{230}/A_{260} should be in the range of 1.8-2.0 and 0.3-0.9, respectively.

2.2.8. Quantitation of total RNA

The concentration of purified total RNA was determined applying a Qubit® fluorometer (Invitrogen), using Qubit® RNA assay kit. This assay is sufficiently sensitive to detect RNA in the range from 250 pg/µl to 100 ng/µl. Aliquots of total RNA (0.5 µl) were diluted 20-fold with RNase-free TE buffer and 2 µl of the dilution was used for quantification. Total RNA (350 ng) was aliquoted into fresh non-stick RNase-free tubes. Total RNA extracts from 25-day-old and 45-day-old microcoms were immediately converted into cDNA. One aliquot of total RNA from the 90-day-old microcosms was stored on ice during another aliquot was

subjected to mRNA enrichment. Both total RNA and enriched mRNA were then converted into cDNA libraries in a single procedure.

2.2.9. mRNA enrichment

In order to remove ribosomal RNA (rRNA), the Ribo-Zero™ rRNA Removal Kit (Meta-Bacteria) (Epicentre) was applied according to manufacturer's instructions. About 5 µg of total RNA was used as the starting material. Accordingly, 10 µl of Ribo-Zero RNA Removal Solution was added to the total RNA sample. As recommended in the manufacturer's manual, rRNA-depleted samples were purified by a column-based method using the RNA Clean & Concentrator™-5, following the manufacturer's small RNA elimination protocol. Enriched mRNA was eluted with 20 µl of TE buffer.

2.2.10. Quality assessment of total RNA and mRNA

Quantitative removal of rRNA and size distribution of enriched mRNA were assessed by automated capillary electrophoresis using an Experion™ system (Bio-Rad). Total RNA and enriched mRNA were denatured and loaded onto the Experion™ HighSense microfluidic chip. The capillary electrophoresis was carried out according to the manufacturer's protocol.

2.2.11. cDNA synthesis

Aliquots (350 ng) of total RNA or mRNA enriched by subtractive hybridization were converted into single-stranded cDNA using SuperScript® II reverse transcriptase (Invitrogen) with random hexamers. Template RNA molecules were denatured at 65 °C for 5 minutes and quickly chilled on ice for 3 minutes. Random primers were annealed at 25 °C for 10 minutes, and first-strand cDNA

synthesis was performed at 42 °C for 50 minutes. The reverse transcriptase was then inactivated by heating at 70 °C for 15 minutes. Double-stranded cDNA was synthesized using the NEBNext® second-strand synthesis system (New England Biolabs) employing RNase H and DNA polymerase I. The reaction was carried out at 16 °C for 3 hours. Double-stranded cDNA was purified using the Wizard® SV Gel and PCR Clean-Up kit (Promega) following the manufacturer's instructions and finally eluted with 18 µl of nuclease-free water.

2.2.12. 454 sequencing library construction

Purified cDNA was end-repaired and dA-tailed applying the NEBNext® End Prep Enzyme Mix and End Repair Reaction Buffer (New England Biolabs). The reaction mixture was incubated at 25 °C for 20 minutes and inactivated at 72 °C for 20 minutes. GS FLX Titanium Rapid Library MID adaptors were added to the mixture and ligated to dA-tailed cDNA using Quick T4 DNA ligase. The adaptor-ligated cDNA library was purified using Agencourt® AMPure® XP beads (Beckman Coulter Genomics) in combination with NEBNext® sizing buffer according to the manufacturer's protocol. At the elution step, the DNA-carrying beads were resuspended in 53 µl of TE buffer and then pelleted using a magnetic stand (Ambion). The supernatant was carefully transferred to a fresh microcentrifuge tube.

2.2.13. Quality assessment of 454 sequencing library

In order to assess the size distribution of the prepared 454 sequencing library, 4 µl of the library were mixed with 1 µl of 5 × FlashGel® Loading Dye. The mixture was loaded onto a FlashGel® DNA cassette of 1.2 % agarose and run at 250 volts for 6 minutes (Lonza).

2.2.14. Quantitation of 454 sequencing library

The adaptor-ligated 454 sequencing library was quantified using a QuantiFluor™ fluorometer and single-use cuvettes (Promega). A standard curve was generated from eight serially diluted standards with 2/3rd dilution factor, starting from the 2.5 × 10^9 molecules/µl solution of the RL Standard in GS Rapid Library Prep kit (Roche). The relative fluorescence unit (RFU) of each standard dilution was read and recorded using the blue channel, which was set to a standard value of 250. A linear trend line of standard dilutions was created using the Rapid Library Quantitation Calculator provided by Roche (http://my454.com/my454/tools-downloads/rapid-library-calc.asp). RFUs of sample libraries were recorded under the same parameter setting used for standard dilution and transferred back to its tube after measurement. Given that the correlation coefficient of the standard curve offered a minimum value of 0.9 and sample RFUs were within the range of the standard curve, samples were diluted to a working stock of 1 × 10^7 molecules/µl in TE buffer as indicated by Rapid Library Quantitation Calculator.

2.2.15. Pyrosequencing

Different 454 sequencing libraries were mixed in equal molar ratio. The libraries were constructed from total RNA of oxic or anoxic soil sampled from different plant growth stages. The mixture of libraries was mixed with the capture beads in a ratio of 100:1 and clonally amplified by emulsion PCR according to the emPCR amplification method manual (Roche). DNA templates on enriched beads were sequenced on a 454 GS Junior sequencer (Roche), following the protocol of the manufacturer. Sequencing libraries produced from enriched mRNA of 90-day-old rice microcosms were sequenced by GS FLX + at the Max Planck Genome Centre Cologne.

2.3. Bioinformatic analysis

2.3.1. Quality filtration of raw 454-pyrosequencing data

In order to remove chimeric reads, reads with undetermined bases (called “N”) or those showing low sequence complexity, PRINSEQ, a stand-alone tool (Schmieder & Edwards, 2011), was installed on our linux server. Reads which are shorter than 200 bp or those having mean quality scores lower than 20, were filtered out. If reads contained characters other than A, C, G, T and N or more than 1% of ambiguous signals (N), they were eliminated from the datasets. A complexity score was applied to remove low-complexity reads that contain homopolymers or dinucleotide repeats. The complexity of a sequence was calculated using the following equation:

where,

k = the alphabet size (43)
w = the window size (3)
n_i = the number of words *i* in a window
l = the number of possible words in a window of size 64
s = the scaling factor (100/31)

Complexity scores above 7 were used as low-complexity cutoff as suggested by Schmieder *et al.* (2011). Datasets obtained by metagenomics or metatranscriptomics were found to contain artificial duplicates, which may comprise 10-35 % of total reads (Gomez-Alvarez *et al*, 2009). In this study, exact duplicates and reverse complimentary reads sharing 100 % sequence identity and length were defined to be artificial replicates and therefore discarded. 3' end sequences of each read were trimmed off when quality scores were below 10. Filter and trim options applied are summarized in **Table 2.1.**

Filter or Trim options	Parameters defined
Minimum sequence length in bp	200
Minimum mean quality score	20
Maximum allowed rate of Ns in %	1
Removal of sequences with characters other than A, C, G T or N	Yes
Trim ends by quality scores from	3'-end only
Trim ends by quality scores while	Minimum score is less than 10
Low-complexity threshold	7 (using DUST)

Table 2.1. Sets of options applied for the filtering and trimming of raw 454 reads.

2.3.2. Differentiation of rRNA and non-rRNA sequences

The quality-filtered sequence libraries constructed from total RNA were screened to identify rRNA by BLASTN against SILVA small and large subunit ribosomal RNA reference databases (SSURef and LSURef, respectively), using an e-value cutoff of 1e-10. Some query sequences had matches to both SSURef and LSURef databases. In that case, the best hit alignments from both databases were taken and their bit score values were automatically compared. The query sequences were assigned to either SSU or LSU based on the higher bit score value, using a custom-coded python script. Sequence reads which had no significant hit to the rRNA databases were classified as non-rRNA sequences, and then subjected to mRNA-tag analysis.

2.3.3. SSU-ribotag analysis

SSU rRNA sequences obtained from total RNA libraries were used to determine the taxonomic composition of soil microbial communities. SSU rRNA sequences from the same or different cDNA libraries were grouped into operational taxonomic units (OTUs). The OTU cluster analysis was done using a non-

redundant version of the SILVA SSU Ref database as a template. The SILVA sequence to which preprocessed 454 reads were mapped with ≥ 95% sequence identity was used as the representative sequence for taxonomic assignment and phylogenetic analysis of the OTU. The number of preprocessed 454 reads mapped or assigned to each reference sequence was recorded along with the library identifier and used to calculate alpha- and beta-diversity via the QIIME pipeline, an open source software package (Caporaso *et al*, 2010).

2.3.4. mRNA-tag analysis

2.3.4.1. Removal of small, non-coding RNA

Small, non-coding RNA was detected by comparing non-ribosomal RNA sequences against a database of non-coding RNA families, Rfam 10.1(http://rfam.sanger.ac.uk), using BLASTN with the maximum e-value of 1e-10 (Gardner *et al*, 2011). Sequences identified as small RNA were excluded from furthur analysis.

2.3.4.2. Functional and taxonomic annotation of mRNA-tags

While MG-RAST (http://metagenomics.anl.gov/) was originally developed for analysis and comparison of metagenomic data sets, its ORF annotation pipeline can also be used for analysis of mRNA-tag data. Taxonomic origin of preprocessed putative mRNA was predicted based on the best hit in searches against the M5NR database with a maximum e-value of 1e-05. M5NR is a non-redundant, integrated database sharing similarity results between several databases for rapid annotation (**Table 2.2**). Functions of each putative mRNA sequence were hierarchically classified according to SEED subsystems with e-value cutoff of 1e-5. The statistical significance of differential taxonomic and functional distribution was determined by Fisher's exact test using STAMP (Statistical Analysis of Metagenomic Profiles) software package (Parks & Beiko,

2010). Active metabolic pathways were found by mapping annotated sequences to KEEG pathway map, and then compared using KEGG Mapper (Kanehisa *et al*, 2011)

Database	Source	Type	Total IDs	% IDs (74,560,976)	Sequences	% Sequences (16,255,122)	Functions	% Functions (5,793,086)	Organisms	% Organisms (398,255)
Greengenes	Greengenes	rna	407,309	0.546	358,750	2.207	1	0.000	11,948	3.000
LSU	SILVA	rna	207,909	0.279	175,146	1.077	1	0.000	37,237	9.350
RDP	RDP	rna	1,418,273	1.902	1,268,768	7.805	184	0.003	15,908	3.994
SSU	SILVA	rna	1,471,257	1.973	1,334,174	8.208	1	0.000	38,705	9.719
GenBank	NCBI	protein	14,687,492	19.699	10,232,124	62.947	3,646,400	62.944	327,873	82.327
IMG	JGI	protein	8,922,652	11.967	6,659,363	40.968	483,685	8.349	5,987	1.503
KEGG	KEGG	protein	6,063,465	8.132	5,413,730	33.305	912,278	15.748	1,413	0.355
PATRIC	VBI	protein	8,795,560	11.796	6,165,838	37.932	94,082	1.624	2,444	0.614
RefSeq	NCBI	protein	12,054,056	16.167	9,147,942	56.277	876,847	15.136	8,393	2.107
SEED	SEED	protein	3,918,073	5.255	3,327,192	20.469	222,853	3.847	4,440	1.115
SwissProt	UniProt	protein	525,207	0.704	441,541	2.716	114,900	1.983	12,398	3.113
TrEMBL	UniProt	protein	13,499,622	18.105	10,814,950	66.533	1,310,604	22.624	329,829	82.819
eggNOG	eggNOG	protein	2,483,276	3.331	2,249,085	13.836	42,507	0.734	630	0.158
COG	eggNOG	ontology	4,873	0.007	1,301,389	8.006	3,278	0.057	0	0
GO	GO	ontology	17,681	0.024	6,939,561	42.692	899,591	15.529	0	0
KO	KEGG	ontology	13,584	0.018	1,912,249	11.764	12,506	0.216	0	0
NOG	eggNOG	ontology	58,198	0.078	395,760	2.435	7,718	0.133	0	0
Subsystems	SEED	ontology	14,818	0.020	1,242,670	7.645	10,869	0.188	0	0

Table 2.2. Complete list of databases comprising M5NR (Last update: 2011-02-22)

3. Results

3.1. Sampling

Samples from the flooded rice paddy soil microcosms were taken at different time points. Temporal sampling was made according to the different plant growth stages (incubation time after transplantation of rice seedlings): tillering (25 days); flowering (45 days); and ripening (90 days). The oxic surface soil (light brown) could easily be distinguished from the anoxic bulk soil (dark grey) (**Figure 3.1**).

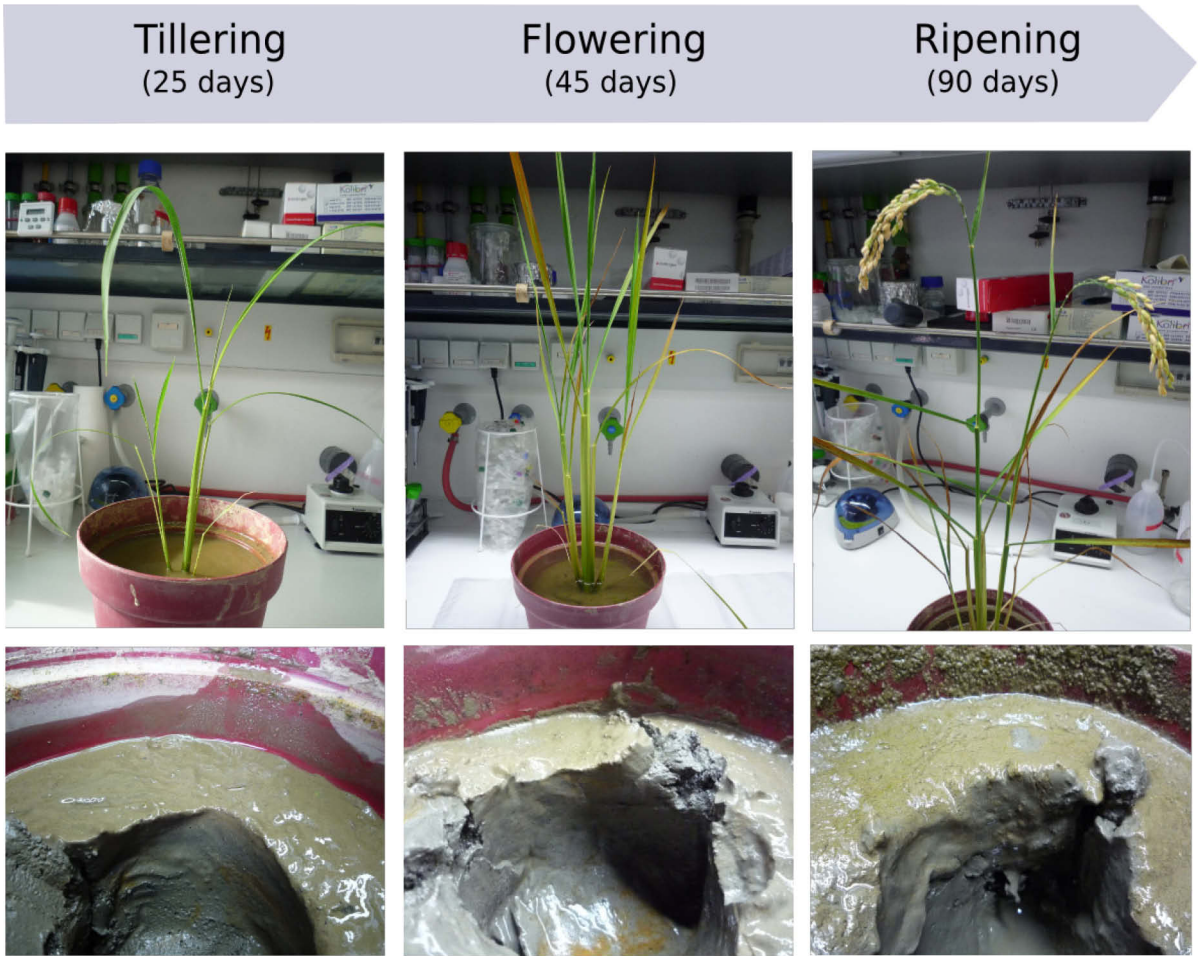

Figure 3.1. Plant growth stages (upper panel) and oxic surface soil versus anoxic bulk soil (lower panel) at each sampling time point.

3.2. RNA extraction and purification

Low-pH extraction (pH 4.5 – 5.0) of total RNA resulted in an effective removal of

humic acids in raw extracts. A crucial step in the extraction of total RNA is sequential Q-Sepharose column chromatography that eliminated 95% of the humic acid contaminants present in the raw extract, as judged by UV absorbance at 400 nm. An important criterion for RNA purity is the absorbance ratio between 260 nm and 280 nm (A260/A280). The value of A260/A280 was always greater than 1.7, which allowed further processing such as cDNA synthesis and mRNA enrichment. The RNA integrity was assessed by capillary electrophoresis with fluorescence detection. Commercial *E.coli* RNA was used as reference standard. It contained three dominant bands corresponding to 5S, 16S, and 23S rRNA. Total RNA extracted from the paddy soil microcosms showed two major bands at exactly the positions of the *E. coli* 16S and 23S rRNA (**Figure 3.2**). Low intensity of the 5S RNA peak indicated its efficient removal by RNA clean & concentrator™ column purification. The ratio of 16S and 23S rRNA band intensity indicated sufficient integrity of the paddy soil RNA extracts, according to the manufacturer's instructions (Bio-Rad). Using 0.7 gram of soil (wet weight), a total of about 500 ng and 350 ng of purified total RNA was obtained from oxic surface soil and anoxic bulk soil, respectively.

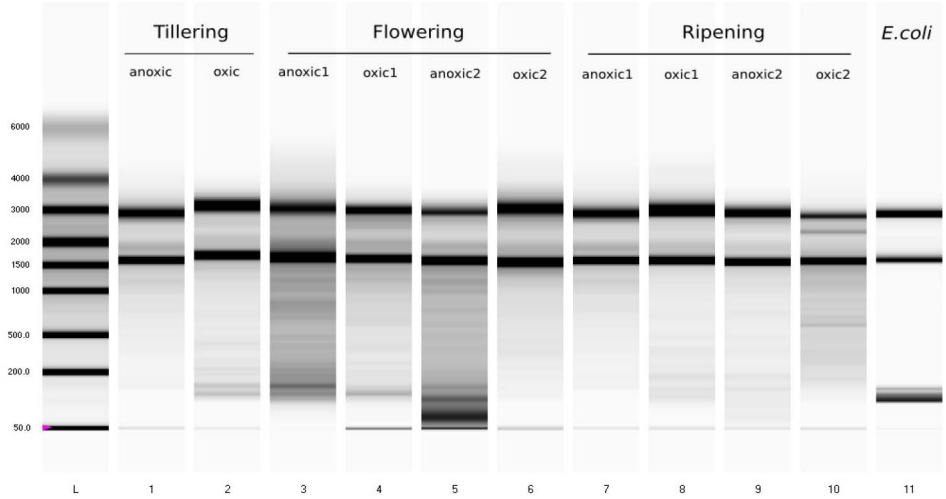

Figure 3.2. Total RNA extracted from the oxic and anoxic zones of flooded rice paddy soil microcosms at different plant growth stages. Experion™ RNA ladder was used as a size marker (lane L). Commercial *E. coli* RNA was used as a reference standard for identifying the positions of the 16S and 23S rRNA, using an Experion™ RNA HighSens Chip for analysis (lane 11).

3.3. mRNA enrichment

Two commercial kits that are based on subtractive hybridization of rRNA were tested for their rRNA removal efficiency: (i) MICROB*Express*™ bacterial mRNA enrichment kit (Ambion) and (ii) Ribo-Zero™ rRNA removal kit (Meta-Bacteria) (Epicentre). A considerable amount of 16S and 23S rRNA was eliminated by both kits, as judged by the reduction of the rRNA peaks in the Experion Bioanalyzer electropherograms (**Figure 3.3**). While still small peaks related to 16S and 23S rRNA were observed in the electropherogram after MICROB*Express*™ treatment, the electropherogram of the Ribo-Zero™-treated aliquot was close to the baseline, even at retention times where the 16S and 23S rRNA peaks would be positioned. Based on this finding, we anticipated that rRNA was more efficiently depleted by Ribo-Zero™ than by MICROB*Express*™. This assumption was later confirmed by 454-pyrosequening of cDNA libraries generated from mRNA enriched by either MICROB*Express*™ or Ribo-Zero™. Sequence data analysis revealed that about 10-20 % and 50-70% of total 454 reads were derived from putative mRNA after treatment of total RNA with MICROB*Express*™ and Ribo-Zero™, respectively.

3.4. cDNA synthesis and 454 library preparation

Double-stranded cDNA of total RNA and enriched mRNA was produced by random priming, followed by adaptor ligation and small fragment removal. Free adaptors and cDNA fragments smaller than 400 bp were eliminated from the cDNA libraries by a sizing step using AMP pure bead treatment (**Figure 3.4**). After removal of small fragments, the cDNA ranged from 750 bp to 1250 bp, regardless of whether total RNA or enriched mRNA was used as template. This size distribution allowed the use of the most recent versions of Roche 454 pyrosequencing chemistry (GS FLX Titanium and GS FLX+).

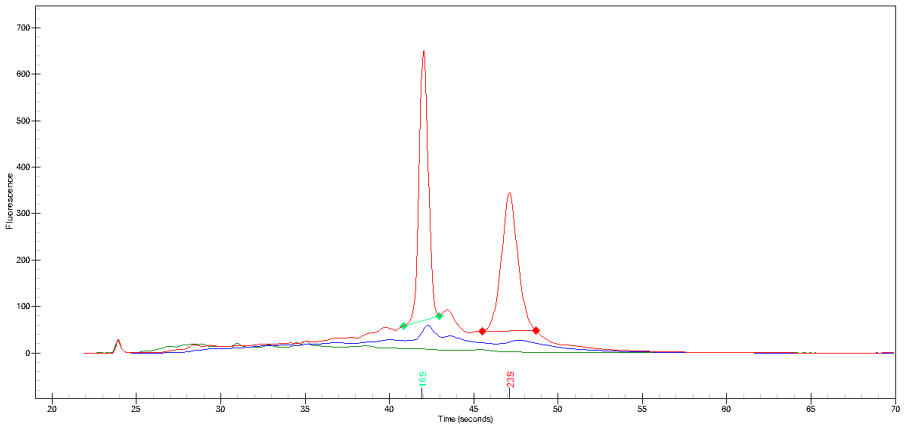

Figure 3.3. Overlay electropherogram of total RNA (red) and mRNA enriched by either MICROB*Express*™ bacterial mRNA enrichment kit (blue) or Ribo-Zero™ rRNA removal kit (green). The *x* axis shows the retention time of size-separated fragments in seconds, and the *y* axis shows the relative fluorescence intensity of each fragment.

3.5. 454-pyrosequencing

We performed two independent runs on a 454 Junior sequencer using the GS FLX Titanium chemistry. On average, we obtained 206,148 key pass wells (96.1%) out of 214,490 raw wells for each run. A raw well is defined as a well showing any signal on a picotiter plate and a key pass well is the one that has four key bases next to the adaptor sequence. Among key pass wells, 24,436 wells (11.9%) were discarded due to poor nucleotide incorporation or interruptions. A certain proportion of wells (12.6%) were discarded due to multiple nucleotide incorporations, possibly occurring from a bead carrying two or more DNA fragments. Additionally, 65,999 wells (32.0%) failed to pass the data processing for shotgun sequencing due to poor quality score. Finally, a total of 80,000 reads of sufficient quality (38.8%) were obtained and subjected to further preprocessing. The 454 reads that passed all the automated quality checks featured lengths between 40 bp and 1,196 bp. The average read length and median read length was 377 bp and 437 bp, respectively.

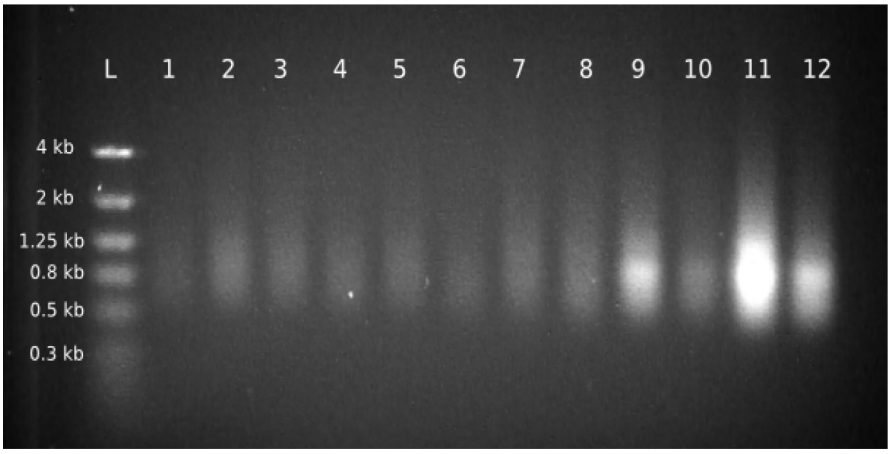

Figure 3.4. Size distribution of cDNA libraries analyzed by high-performance gel electrophoresis (Flashgel DNA cassette [1.2%]). Lane L: FlashGel™ DNA marker (100 bp -4 kb). Lane 1 to 10: libraries produced from total RNA; lane 1, 25D-anoxic bulk soil; lane 2, 25D-oxic surface soil; lane 3, 45D-anoxic bulk soil (1); lane 4, 45D-anoxic bulk soil (2); lane 5, 45D-oxic surface soil (1); lane 6, 45D-oxic surface soil (2); lane 7, 90D-anoxic bulk soil (1); lane 8, 90D-anoxic bulk soil (2); lane 9, 90D-oxic surface soil (1); lane 10, 90D-oxic surface soil (2). Lanes 11 and 12: libraries produced from enriched mRNA (lane 11, 90D-anoxic bulk soil (1); lane 12, 90D-oxic suface soil (1).

3.6. Preprocessing of 454 sequence data

The 454 reads were clustered into sample-specific data sets based on multiplex identifier. Each data set was processed individually. Basic statistics of each 454 cDNA library are summarized in **Table 3.1**. It was reported that the 454 platform systematically generates artificial sequence replications at different steps of whole-genome pyrosequencing. Those replicates accounted for 3.5 % - 18.1% of total reads in GS FLX sequencing data sets under experimental condition where every DNA template is completely unique (Dong *et al*, 2011). In this study, each cDNA library was found to contain 0.3 % to 1.6 % of artificial duplicates, which shared identical sequences of the same length. Using PrinSeq (Schmieder & Edwards, 2011), about 6 % of total reads of a single GS Junior run were detected to be artificial replicates. These replicates were randomly distributed. Finally, about 75 % of total reads passed the preprocessing, which included 200-bp size cutoff, quality filtration and trimming the 3'-end sequences. The mean size of

	25D oxic1	25D anoxic1	45D oxic1	45D oxic2	45D anoxic1	45D anoxic2	90D oxic1	90D oxic2	90D anoxic1	90D anoxic2
Minimum read length	22	19	20	22	19	20	18	19	19	21
Maximum read length	602	646	594	728	608	575	693	780	605	627
Average read length (¥bp)	403.25	389.19	381.19	391.29	313.56	338.69	406.8	355.29	397.03	365.79
*No. of raw reads	17249	16567	48987	14006	10430	15902	20441	7528	18169	16707
[1]No. of exact duplicates	125 (0.72%)	79 (0.48%)	786 (1.60%)	150 (1.07%)	80 (0.77%)	170 (1.07%)	330 (1.61%)	180 (2.39%)	81 (0.45%)	200 (1.18%)
[2]Preprocessed reads	13906 (80.62%)	12201 (73.65%)	33211 (67.80%)	10365 (74.00%)	8405 (80.58%)	12492 (78.56%)	16532 (80.88%)	6209 (82.48%)	13892 (76.46%)	13222 (79.14%)
[3]Non-passing reads	3343 (19.38%)	4366 (26.35%)	15776 (32.20%)	3641 (26.00%)	2025 (19.42%)	3410 (21.44%)	3909 (19.12%)	1319 (17.52%)	4277 (23.54%)	3485 (20.86%)
Average read length of preprocessed reads (¥bp)	440.9	428.0	437.5	444.1	423.2	430.8	469.2	424.4	435.0	451.2
[4]LSU rRNA	8166 (58.72%)	6814 (55.85%)	18596 (55.99%)	5091 (49.12%)	4059 (48.29%)	5962 (47.73%)	7977 (48.25%)	2990 (48.16%)	7298 (52.53%)	6694 (50.63%)
[5]SSU rRNA	5237 (37.66%)	5160 (42.29%)	14186 (42.71%)	4950 (47.76%)	4089 (48.65%)	6189 (49.54%)	8334 (50.41%)	3144 (50.64%)	6042 (43.49%)	5824 (44.05%)
[6]non-rRNA	503 (3.62%)	227 (1.86%)	429 (1.29%)	324 (3.13%)	257 3.06%)	341 (2.73%)	221 (1.37%)	75 (1.21%)	552 (3.97%)	704 (5.32%)

Table 3.1. Summary of sequence statistics for cDNA libraries of total RNA. The RNA was extracted from either the oxic surface soil (oxic) or the anoxic bulk soil (anoxic) of one (1) or two (2) independent replicate microcosms. [1]The percentage value means the relative abundance of exact duplicates in raw reads. [2]The number of raw reads which passed the preprocessing. [3]The number of raw reads which failed the preprocessing. [4,5,6]The number of LSU rRNA, SSU rRNA and non-rRNA sequences among the preprocessed reads. *No. is the abbreviation of number. ¥bp stands for base pair.

each individual cDNA library shifted up from between 314 bp and 406 bp to between 423 bp and 469 bp. Reads that passed preprocessing had mean quality scores greater than 30 (Q30). Plotting the quality score over read length showed that all preprocessed reads had high sequence quality (**Figure 3.5**).

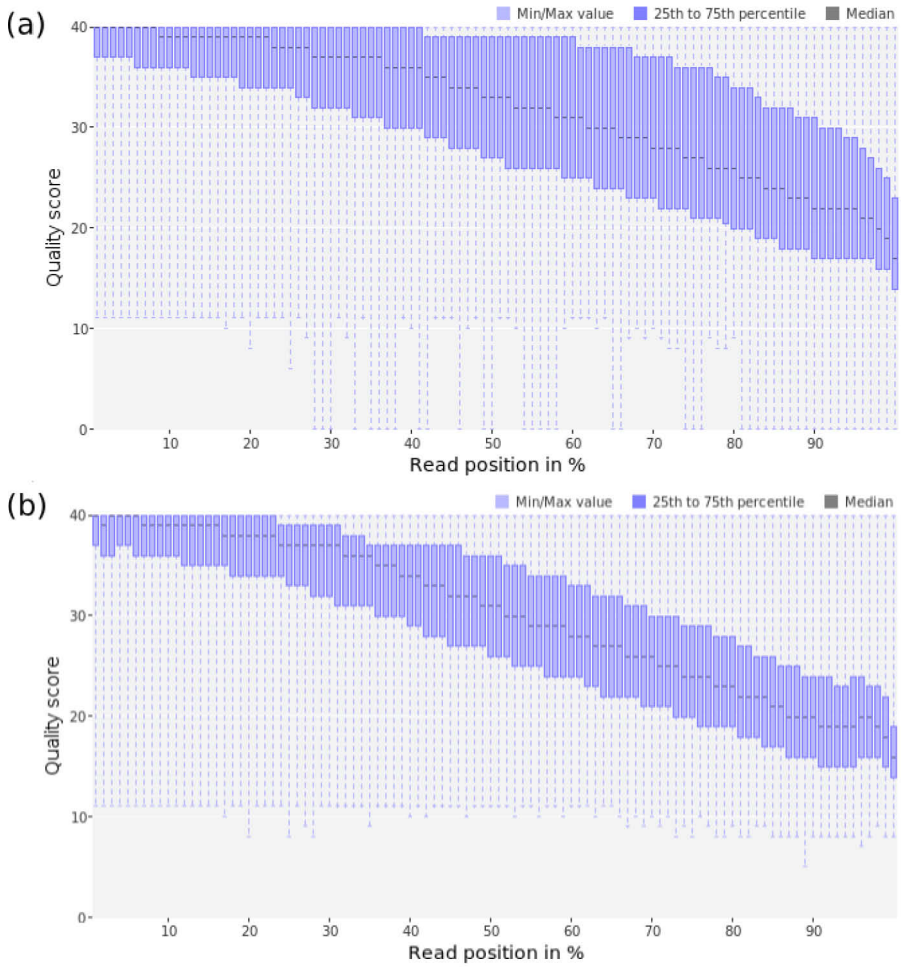

Figure 3.5. Comparison of the base quality distribution for a given 454 data set prior to and after preprocessing. Dotted lines indicate the range of the quality score. Blue boxes show the quality score of 25th to 75th percentile of all the reads.

3.7. Community structure analysis using ribosomal RNA tags

3.7.1. Extraction of small subunit rRNA sequences

The vast majority of preprocessed reads in total RNA data sets (95-98%) had significant BLASTN hits against the SILVA rRNA database (Pruesse *et al*, 2011). These reads were classified as rRNA-derived sequences (ribotags). On average, 45% of ribotags of a given sample-specific data set was assigned to small subunit rRNA (SSU-ribotags), which were used for analysis of community composition. The minimum number of SSU-ribotags analyzed for a given sample was 3,144 (sample 90D-oxic2) and the maximum number was 14,186 (sample 45D-oxic1). Overall, comparable numbers of SSU-ribotags (~5,000) were obtained for each sample (**Table 3.1**).

3.7.2. Clustering of SSU-ribotags into operational taxonomic units (OTUs)

In order to determine rarefaction curves, richness, and evenness for each sample-specific SSU-ribotag data set, the preprocessed reads were clustered for each data set into operational taxonomic units (OTUs). Since cDNA was generated by random priming, the SSU-ribotags are derived from different regions of SSU rRNA. SSU-ribotags were clustered into OTUs by mapping them to full-length rRNA sequences with a given sequence identity cutoff. In SILVA (http://www.arb-silva.de), the non-redundant version of the SSU Ref database (release 108) was used as a reference data set, which was built by dereplication of the full SSU Ref data set using a cutoff of 99 % sequence identity. The full-length rRNA sequence to which SSU-ribotags were mapped was taken as the reference standard for a given OTU. At the species level, defined by 97 % sequence identity, about 64 % of all the SSU-ribotags could be mapped to full-length rRNA reference sequences, and 36 % of SSU-ribotags were identified as representing novel groups or species. By lowering the sequence identity cutoff for OTU clustering, the more SSU-ribotags grouped together. However, 2 to 5 % of SSU-ribotags could not be mapped to any known 16S rRNA sequence, using a cutoff of 85 % sequence identity. These sequences may represent novel

microbial lineages at the phylum and subphylum levels.

3.7.3. Bacterial diversity and richness

In order to compute estimates of genus-level diversity within samples, a threshold of 95 % sequence identity was applied (Schloss & Handelsman, 2005). Using this cutoff, the numbers of unique OTUs in samples from the anoxic bulk soil were almost 1.5-fold greater than those in samples from the oxic surface layer, for both 45 day- and 90 day-old microcosms. In general, the rarefraction curves were composed of 1,500 to 2,000 OTUs (oxic surface layer) or 2,500 to 3,000 OTUs (anoxic bulk soil) and did not show any saturation, even after random sampling of up to 5,000 SSU-ribotags. However, the rarefaction curve of SSU-ribotags obtained from the anoxic bulk soil of the 25 day-old microcosm began to be saturated already after 1,800 OTUs when 2,900 SSU-ribotags had been randomly sampled (**Figure 3.6a**). Obviously, the microbial communities in the anoxic bulk soil were more diverse at the genus level than those in the oxic surface layer during flowering (45 days) and ripening (90 days). Comparing the number of OTUs observed in rarefraction analysis with the estimated number of OTUs determined by Chao1 richness estimator revealed that, on average, 30 % and 34.1 % of the estimated taxonomic richness were covered by SSU-ribotags from samples of the oxic surface layer and anoxic bulk soil (45D/90D), respectively. The SSU-ribotag data set from the anoxic bulk soil of the 25-day-old microcosm showed the highest coverage of 60.2 %, in good correspondence to the results of the rarefaction analysis.

The evenness of OTUs was estimated by Shannon diversity index. In order to avoid biases derived from different sampling efforts, Shannon diversity indices were plotted for each sample against the expected number of OTUs as estimated by Chao1 (**Figure 3.6b**). The Chao1 estimates were computed by randomly selecting 2,500 and 3,300 SSU-ribotags from each sample. The Shannon diversity index ranged from 8.0 to 9.1 in the oxic surface layer and from 9.7 to 10.8 in the anoxic bulk soil. Despite the lowest OTU richness, the sample

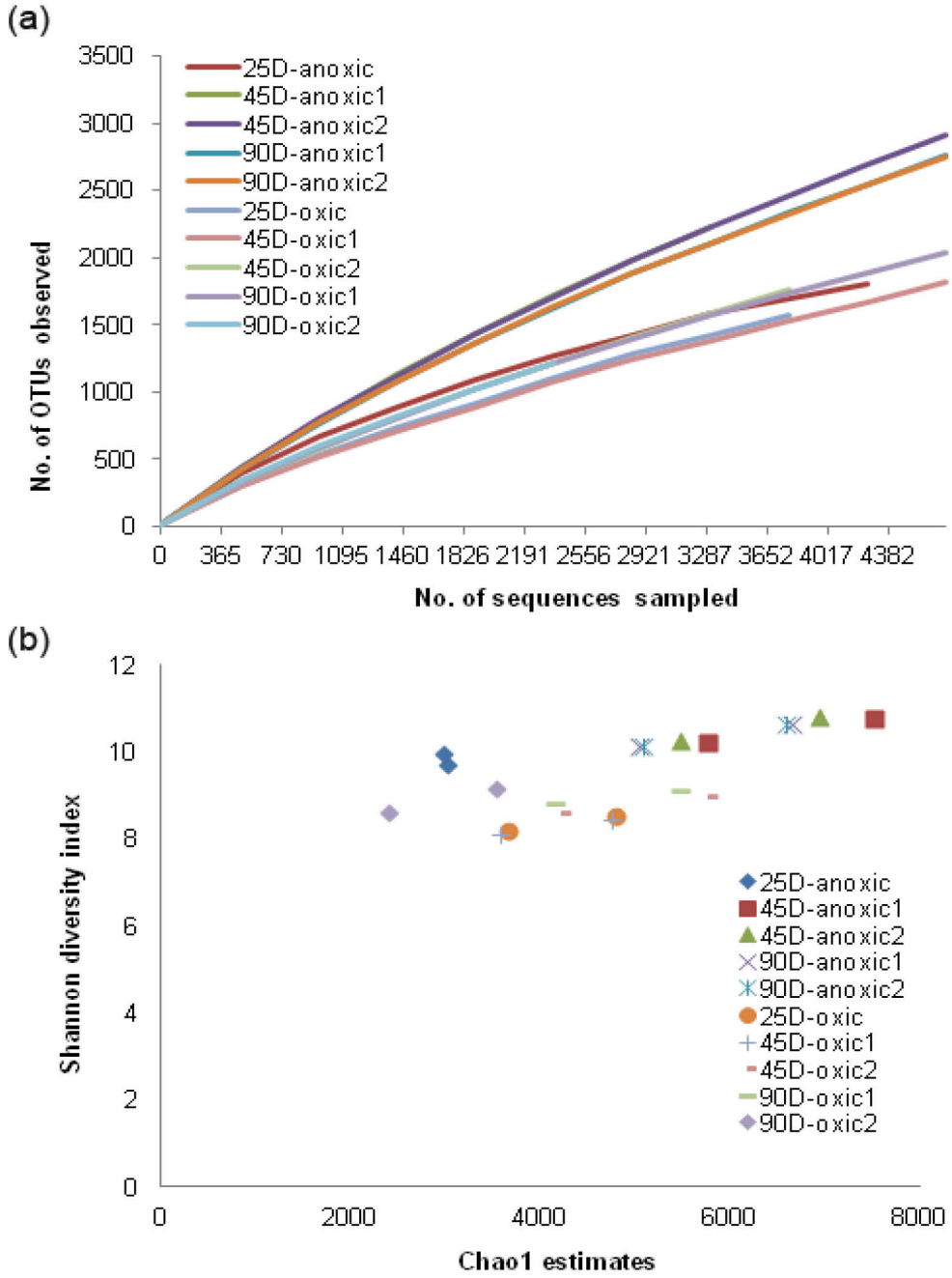

Figure 3.6. OTU-based microbial diversity analysis in samples from the oxic surface layer and the anoxic bulk soil of flooded rice paddy soil microcosms. Samples for analysis were taken at different time points (25, 45, and 90 days after transplantation of rice seedlings). (a) Rarefaction curves displaying the average number of OTUs observed by random sampling within each SSU-ribotag data set. (b) Plot of Shannon diversity index versus Chao1 estimates for total genus-level diversity.

from the 25-day-old microcosm showed a greater Shannon diversity index than the samples from the oxic surface layer. The comparison of the mean Shannon index values between the microbial communities from the oxic and anoxic zones indicate that the anaerobic communities are more evenly composed than the aerobic communities.

3.7.4. Comparative taxonomic analysis

The standard reference sequence of each OTU was taxonomically classified using the RDP Bayesian classifier with a confidence threshold of 80 %. The taxonomic composition of microbial communities in the oxic surface layer greatly differed from that in the anoxic bulk soil. Bacteria were the most abundant domain in both soil zones. The relative proportion of eukaryotes was significantly higher in the oxic surface layer than in the anoxic bulk soil, but archaea were present in a higher proportion in the anoxic soil. At phylum level, the microbial communities in the oxic surface layer were dominated by cyanobacteria (40-55 % of total SSU-ribotags), corresponding to a relatively low community richness and evenness. Other abundant bacterial phyla were *Proteobacteria* (13-30%), *Chloroflexi* (2.4%), *Actinobacteria* (2.3%), *Planctomycetes* (2.0%), and *Verrucomicrobia* (1.4 %). By contrast, the predominant bacterial phyla in the anoxic bulk soil were *Proteobacteria* (30%), *Firmicutes* (11.6–18.6%), *Actinobacteria* (11.4-17.5%), *Chloroflexi* (7.5%), and *Acidobacteria* (6.3%) (**Figure 3.7**). Principal Coordinate Analysis (PCoA) confirmed that the microbial communities in the oxic surface layer greatly differed from those in the anoxic bulk soil (**Figure 3.8**).

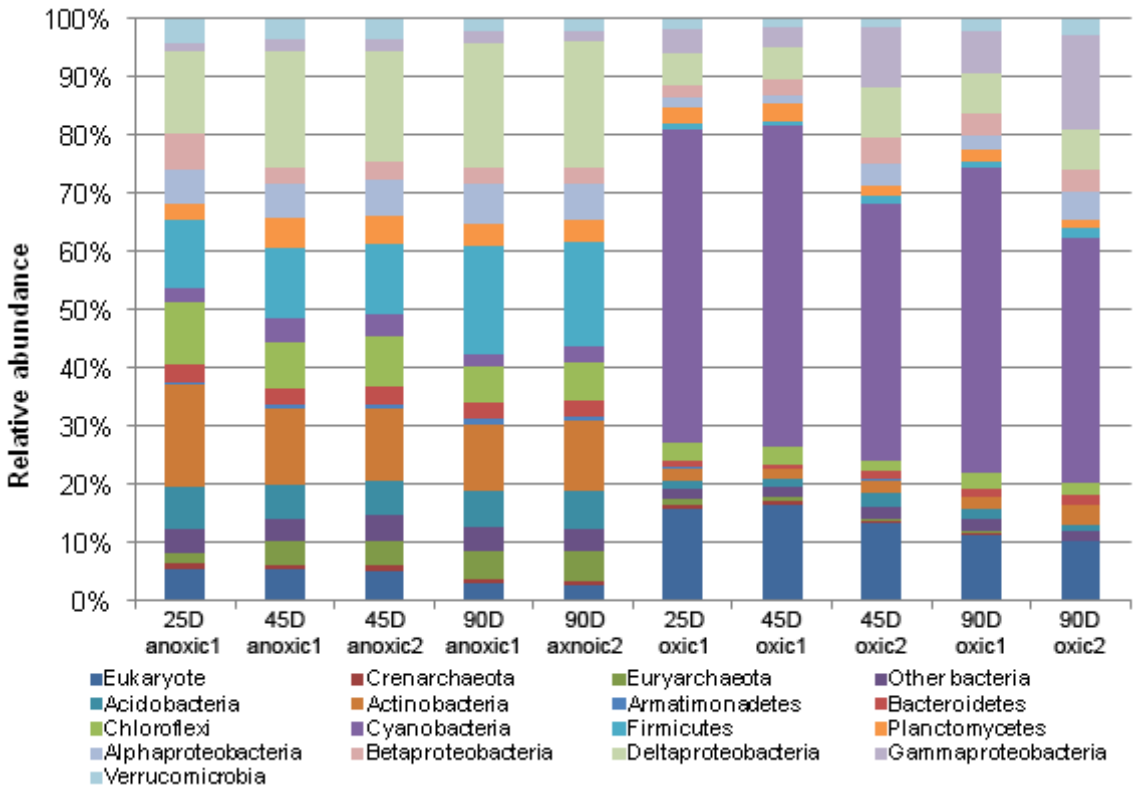

Figure 3.7. Relative abundance of phyla, including major subclasses of *Proteobacteria*. Abundance measurement was made based on the number of SSU-ribotags assigned to the respective group.

Due to the dominance of the *Cyanobacteria,* members of other major groups were generally less abundant in the oxic surface layer than in the anoxic bulk soil, except for the *Gammaproteobacteria*. These accounted for about 9 % of total SSU-ribotags in samples from the oxic surface layer, which is almost four times greater than their relative abundance in the anoxic bulk soil (2.3%). *Xanthomonadales* was the characteristic group of *Gammaproteobacteria* in the oxic surface layer. Four OTUs were affiliated with the *Xanthomonadales*. One of the OTUs whose standard reference sequence is GenBank ID EU131032.1.1502 contains 586 SSU-ribotags. Among the non-cyanobacterial OTUs, it was the OTU with the greatest number of SSU-ribotags assigned to it (after three cyanobacteria-affiliated OTUs). Among the methane-oxidizing bacteria, *Crenothrix*-like bacteria were prominent in the oxic surface layer, while type II methanotrophs (*Methylocystis*) were detected primarily in the anoxic bulk soil.

Type I methanotrophs were detected similarly in both the oxic conditions. Over all samples, the methane-oxidizing bacteria accounted, on average, for 4 % and 2.8 % of total SSU-ribotags in the oxic surface layer and the anoxic bulk soil, respectively. Non-cyanobacterial groups at the order level, which showed greatest relative abundance, were highly representative of the anoxic bulk soil. Among those, the groups that were most differential for the comparison between oxic surface layer and anoxic bulk soil were *Actinobacteria*, *Clostridia, Geobacter*, *Anaeromyxobacter, Bacillus,* and methanogens.

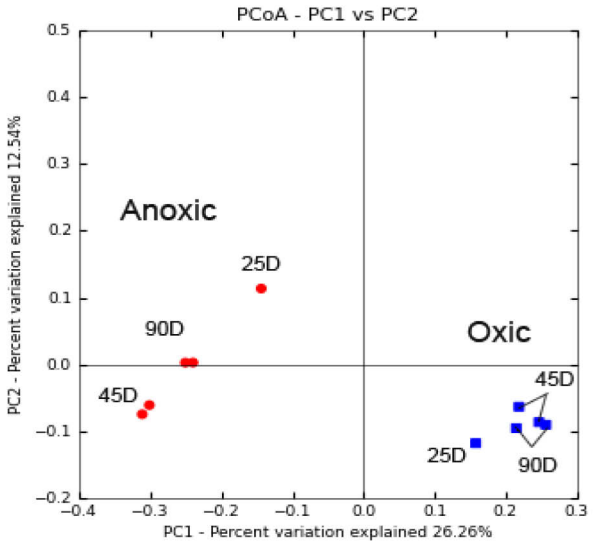

Figure 3.8. Principal coordinate analysis based on Unifrac distance matrix. Blue dots represent SSU-ribotag data sets from samples of the oxic surface layer, while red dots are SSU-ribotag data sets from samples of the anoxic bulk soil.

3.7.5. Microbial community succession over time

The taxonomic assignments of SSU-ribotags showed highly similar community compositions at the different plant growth stages. No major temporal changes in the community composition were observed, except for a few groups. In the oxic surface layer, the relative abundance of eukaryotic organisms decreased from 16% to 11% of total SSU-ribotags. In addition, there was a slight increase in the

relative abundance of the *Burkholderiales* and *Myxococcales*, and a doubling in the relative abundance of the *Xanthomonadales.* Changes in relative abundance primarily occurred between flowering(45D) and ripening (90D) (**Figure 3.9a**).

In the anoxic bulk anoxic soil, the microbial communities exhibited more gradual changes over time than those in the oxic surface layer. The most significant decline in the relative abundances was observed for the *Actinobacteria, Burkholderiales,* and *Verrucomicrobia* (**Figure 3.9b)**. Microorganisms assumed to be characteristic of anoxic soil, such as *Geobacter, Clostridia,* and methanogens, were enriched during plant growth. The percentage of SSU-ribotags assigned to these three groups increased from 1.5 % to 5.5 %, 8.5 % to 13.3 % and 1.4% to 4.6 %, respectively. The most abundant methanogens were *Methanosaeta* and members of the *Methanosarcinaceae*.

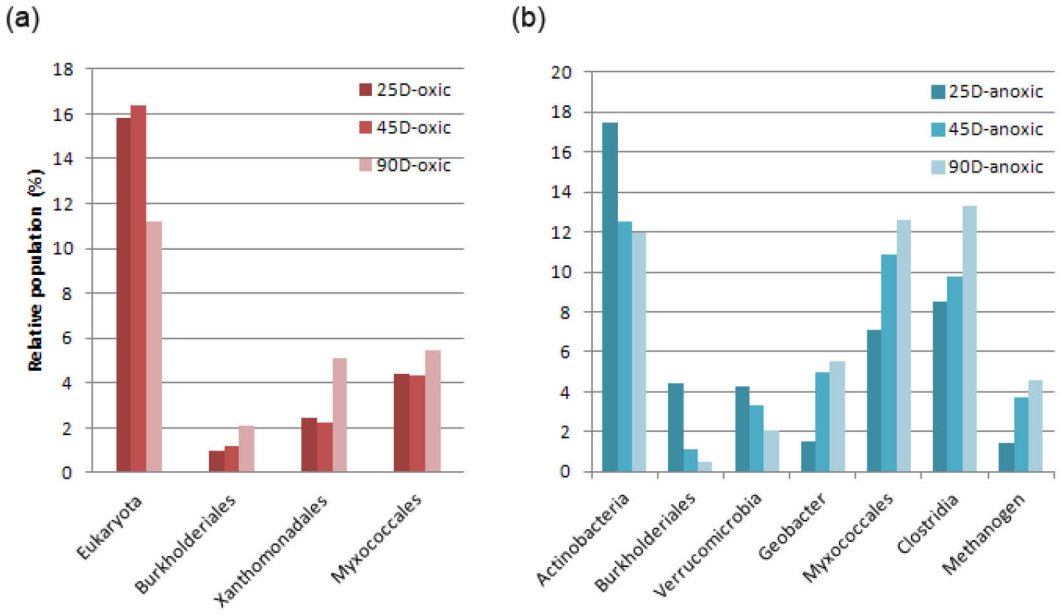

Figure 3.9. Temporal changes in the relative abundance of particular microbial groups over time, as shown for the oxic surface soil (a) and the anoxic bulk soil (b).

3.8. Functional analysis of paddy soil microbial communities

PCR-independent SSU-ribotag analysis was applied to assess the taxonomic composition of aerobic versus anaerobic microbial communities at different plant

growth stages. Overall, there occurred only little changes in the relative abundance of particular microbial groups; slightly more pronounced in the anoxic bulk soil than in the oxic surface layer. Therefore, we chose 90-day-old flooded rice paddy soil microcosms (ripening stage) to explore functional activities of aerobic versus anaerobic microbial communities. In SSU-ribotag analysis, these microbial communities had shown greatest differences in their taxonomic composition, in particular with regard to the presence and relative abundance of *Geobacter*, *Clostridia,* and methanogens. The mRNA samples from the oxic surface layer and the anoxic bulk soil were enriched by subtractive hybridization of rRNA. The cDNA libraries produced from the enriched mRNA were analyzed using the 454 GS FLX+ chemistry. In total, we obtained 29,859 and 80,027 raw reads from the anoxic and oxic samples, respectively. The average read length within the two cDNA libraries was 349.5 and 455.5 bp. The short average read length of the data set from the anoxic bulk soil was due to a relatively high proportion of short fragments (< 200 nucleotides). These short sequence reads failed to pass the quality filtration (**Table 3.2**). As a consequence, the average read length of preprocessed reads increased up to 529.3 bp for the data set from the anoxic bulk soil, but the number of reads used for further bioinformatic analysis decreased to 18,274 (61.2%). The average read length of the data set from the oxic surface layer was 525.2 bp.

	Anoxic bulk soil	Oxic surface layer
No. of raw reads	29,859	80,027
No. of preprocessed reads	18,274 (61.2%)	65,622 (82.0%)
No. of rRNA	6,115 (33.5%)	16,042 (24.4%)
No. of small RNA	122 (0.67%)	626 (0.95%)
No. of putative mRNA	12,037 (85.8%)	48,954 (74.6%)

Table 3.2. Statistics of cDNA libraries constructed from enriched mRNA. Percentage values are given in parenthesis. The relative proportion of preprocessed reads was calculated in relation to the total number of raw reads (= 100%), and the relative proportion of reads derived from rRNA, small RNA, and putative mRNA was calculated in relation to the total number of preprocessed reads (= 100%).

3.8.1. Removal of rRNA sequences

Within the two preprocessed data sets, about 33.5% and 24.4% of preprocessed reads had significant matches to known rRNA sequences (**Table 3.2**). When compared to the sequence composition of the data sets that were obtained from total RNA, almost all rRNA (98-99%) had been removed by the mRNA enrichment method applied. The remaining rRNA sequences were mainly composed of LSU rRNA. The capture probe specificity and sensitivity presumably introduces some bias in the taxonomic composition of the non-captured rRNA, in particular with regard to a strong relative increase of eukaryotic rRNA. All the 454 reads identified as being derived from rRNA were excluded from further bioinformatic analysis.

3.8.2. Detection of small RNA

A total of only 122 reads out of 12,159 non-rRNA sequences in the data set from the anoxic bulk soil and 626 reads out of 49,580 non-rRNA sequences in the data set from the oxic surface layer were identified as small RNA, thereby accounting for less than 1% of the non-rRNA sequence data sets. Most of the small RNA reads were assigned to eukaryotic organisms (**Figure 3.10**). The majority of the bacteria-derived small RNA reads was annotated as bacterial RNase P class A, a ubiquitous endoribonuclease. The RNA component of bacterial RNase P varies in length from 338 to 444 nucleotides (Brown & Pace, 1992). Physical elimination of small cDNA fragments (< 400 nucleotides) during 454 library preparation removes most but not all small RNA-derived sequences, thereby resulting in a biased and limited diversity of small RNA in cDNA libraries produced to analyze putative mRNA.

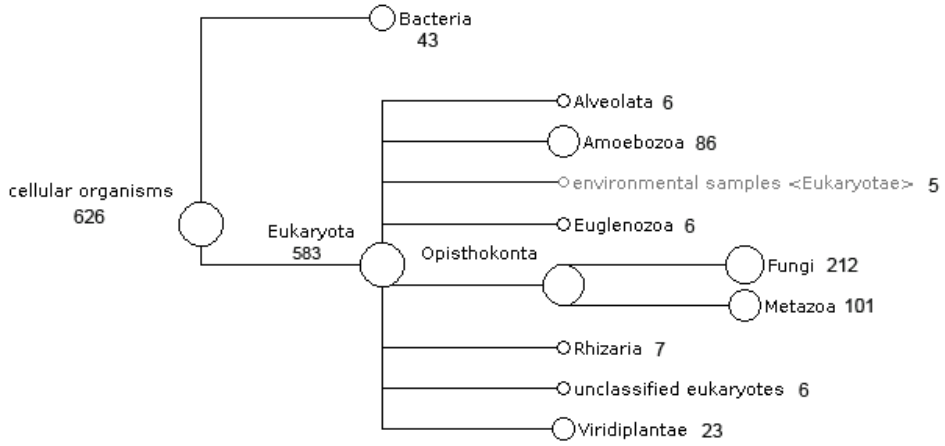

Figure 3.10. Taxonomic affiliation of small RNA detected in the preprocessed data set from the oxic surface layer, including the number of small RNA reads assigned to each taxonomic group.

3.8.3. Functional annotation of putative mRNA-tags

Data sets of 12,037 and 48,954 putative mRNA-tags were finally obtained from the anoxic bulk soil and oxic surface layer, respectively. The putative mRNA-tags were functionally annotated by searching against the integrated protein database. The proportion of putative mRNA-tags that could be functionally annotated was higher in the data set from the oxic surface layer (60%) than in that from the anoxic bulk soil (40%), suggesting that the anaerobic microbial community is less well represented by public genomic databases. The mRNA-tags that had homologs in public databases were classified into hierarchical functional categories using the SEED subsystem. Differences in the frequency with which particular transcripts (oxic versus anoxic zone) were assigned to the different subcategories were statistically analyzed using STAMP (Statistical Analysis of Metagenomic Profile) (Parks & Beiko, 2010). Overall, the functional repertoire in both data sets were highly similar, as concluded from the relative distribution of mRNA-tags at the level 1 functional subsystem (**Figure 3. 11**). Similar to metatranscriptome studies of marine assemblages, almost half of the functionally annotated mRNA-tags were associated with carbohydrate, amino acids or protein metabolism (Gilbert *et al*, 2008). This finding reflects the high transcript level of

house-keeping genes in microorganisms, regardless of whether in soil or marine systems.

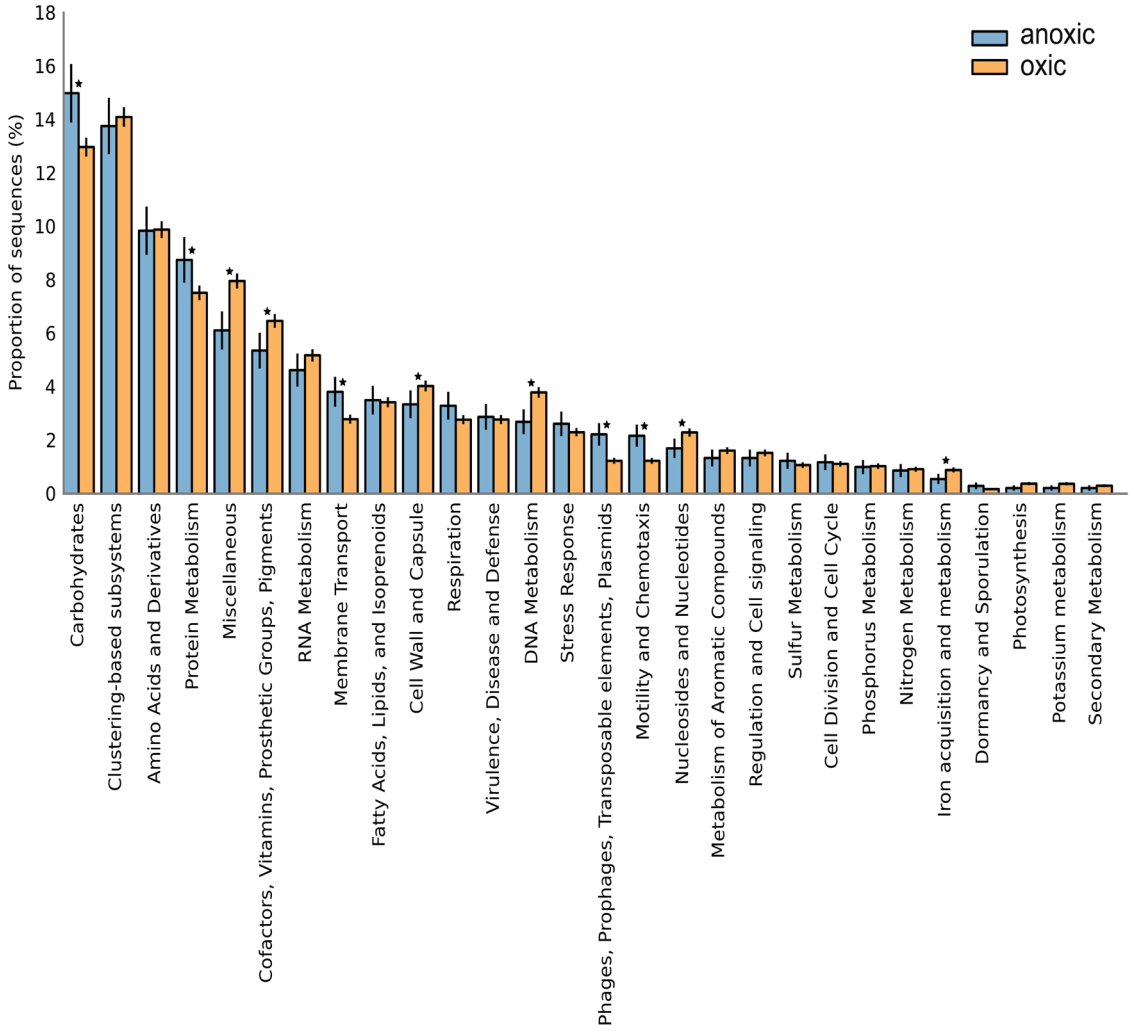

Figure 3.11. Functional classification of mRNA-tags obtained from aerobic (oxic surface layer) and anaerobic (anoxic bulk soil) microbial communities. The transcriptome data sets were subjected to the automated annotation process into the SEED subsystem using MG-RAST. The proportion of sequences (y-axis) assigned to a particular level 1 category was calculated by dividing the number of mRNA-tags assigned to this category by the total number of mRNA-tags assigned to the SEED subsystem. Subsystems that showed significant differences in the number of mRNA-tags ($P<0.05$; Fisher's exact test) between the oxic and anoxic zone are marked by an asterisk (*).

Subsystems involved in carbohydrate and protein metabolism were significantly overrepresented in the data set from the anoxic zone. mRNA-tags related to DNA metabolism and non-protein components (cofactors, vitamins, and prosthetic groups) as well as cell wall metabolism were present in higher proportions in the data set from the oxic surface layer than in that from the anoxic bulk soil (**Figure 3.11**). Among the carbohydrate utilization profiles, the central carbohydrate metabolism showed the most differential distribution of mRNA-tags within the subcategories (**Figure 3.12**). mRNA-tags assigned to pyruvate metabolism to form acetate via acetyl-coA were significantly overrepresented in the data set from the anoxic bulk soil, while those related to anaplerotic reactions to supplement intermediates of pyruvate metabolism were significantly more present in the data set from the oxic surface layer. Ethylmalonyl-CoA pathway of C2 assimilation was also one of the subcategories in central carbohydrate metabolism, which showed significantly higher number of mRNA-tags in the data set from the anoxic bulk soil than in that from the oxic surface layer (**Figure 3.13**). Multiple copies of particulate methane monooxygenase (pMMO) transcripts were detected only in the data set from the oxic zone, suggesting methane oxidation activity in the oxic surface layer around the time of sampling.

Differential functional activities between the aerobic and anaerobic microbial communities were also concluded for other subcategories. The 31 subcategories, listed in **Figure 3.13**, showed with high significance different frequencies with which mRNA-tags from the two data sets (oxic versus anoxic zone) were assigned to them (P-value < 0.01). Among these subcategories, transcripts involved in methanogenesis was most indicative of the anoxic bulk soil (p-value of 1.20e-9). These transcripts accounted for approximately 0.5% of total mRNA-tags from the anaerobic microbial community. The relatively high level of transcripts encoding proteins involved in methanogenesis, including methyl-coenzyme M reductase, agrees well with the fact that methanogenesis is one of the key metabolic processes in the anoxic bulk soil of flooded rice fields. mRNA-tags affiliated with bacterial tRNA and rRNA modification, purine conversion, ribosome biogenesis and RNA methylation were more frequently detected in the data set from the oxic surface layer than in that from the anoxic bulk soil (**Figure**

3.13).

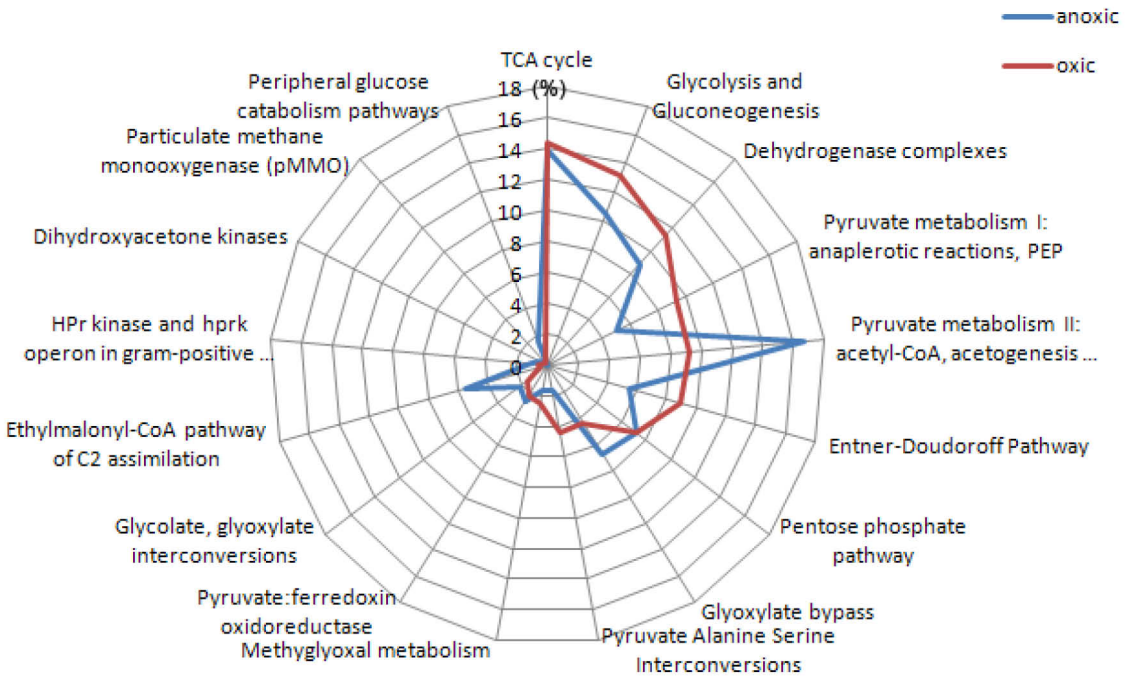

Figure 3.12. Relative distribution of mRNA-tags assigned to the different subcategories of central carbohydrate metabolism. The results for the data set from the anoxic bulk soil is shown in blue, while those for the data set from the oxic surface layer are indicated in red. The percentage scale indicates the proportion of mRNA-tags which are assigned to the respective subcategory in relation to the total number of mRNA-tags assigned to the central carbohydrate metabolism.

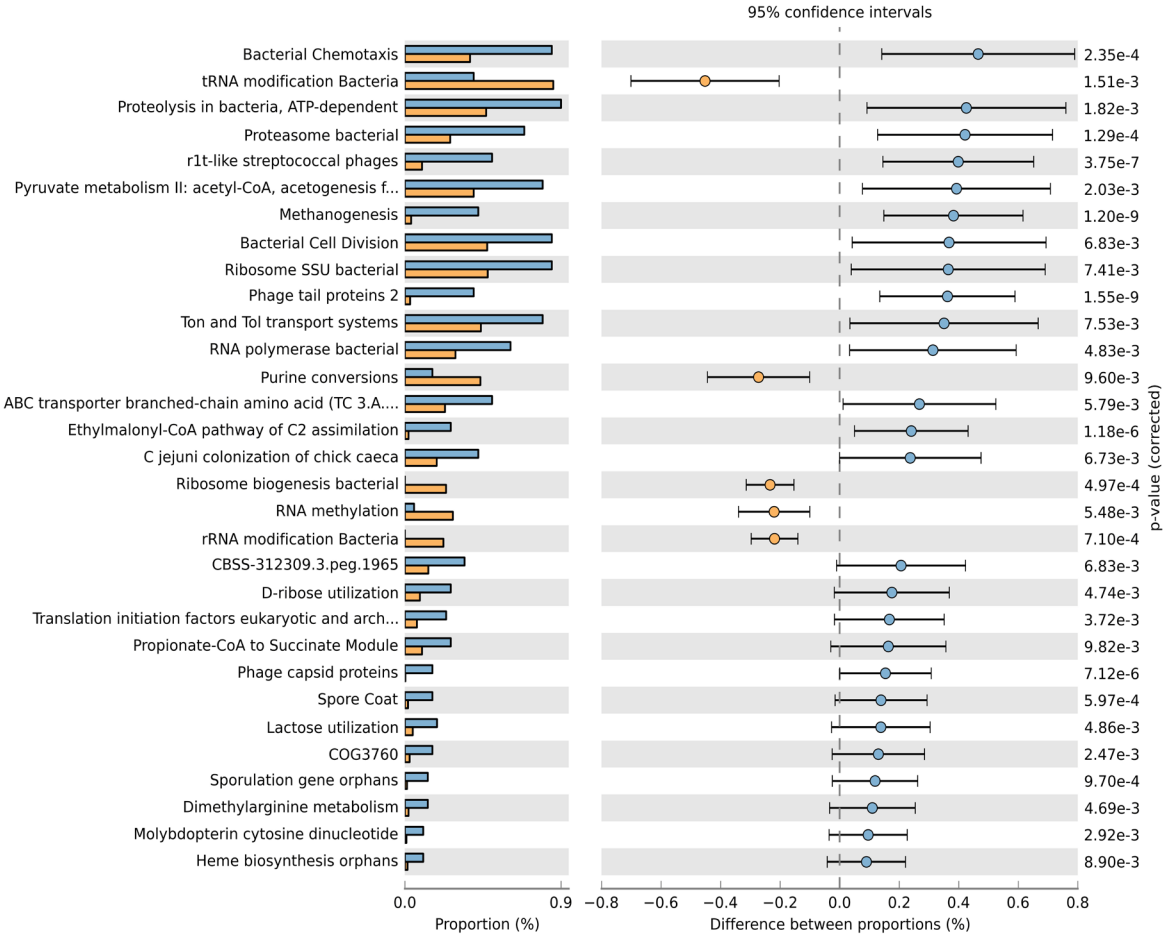

Figure 3.13. SEED level 3 subsystems to which mRNA-tags were assigned with significantly different frequencies (P-value < 0.01; Fisher's exact test), when comparing the data sets from the oxic surface layer (yellow) and anoxic bulk soil (blue).

3.8.4. Taxonomic binning of putative mRNA-tags

The mRNA-tags were taxonomically binned based on the best homologous hit in the M5NR protein database using MG-RAST server. The majority of mRNA-tags (94.9% in the data set from the oxic surface layer and 90.4% in that from the anoxic bulk soil) were derived from bacteria. Both archaea and eukaryotes were slightly overrepresented in the data set from the anoxic bulk soil as compared to that from the oxic surface layer (3.8% vs 1.3% and 5.0% vs. 3.5%, respectively). Overall, taxonomic binning of mRNA-tags was less powerful than SSU-ribtotag analysis in differentiating between the aerobic and anaerobic microbial

communities (**Figure 3.14**).

At the phylum and class levels, when comparing the relative abundance of each group in the two mRNA data sets (oxic versus anoxic zone), *Actinobacteria*, *Deltaproteobactera*, *Clostridia*, *Bacilli*, and *Methanomicrobia* were overrepresented in the data set from the anoxic bulk soil. *Cyanobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* and *Flavobacteria* were detected in the oxic surface layer with significantly higher proportion than in the anoxic bulk soil (*P<0.01*; Fisher's exact test) (**Figure 3.15**), in reasonable agreement with the SSU-ribotag analysis. However, SSU-ribotag analysis suggested that *Alphaproteobacteria* was overrepresented in the anoxic bulk soil, while taxonomic binning of mRNA-tags identified a greater proportion of *Alphaproteobacteria* in the data set from the oxic surface layer. In samples from the oxic surface layer, the proportion of mRNA-tags assigned to *Flavobacteria* were 20-fold greater than that assigned to *Flavobacteria* in the corresponding SSU-ribotag data set (2.77% vs 0.2 %).

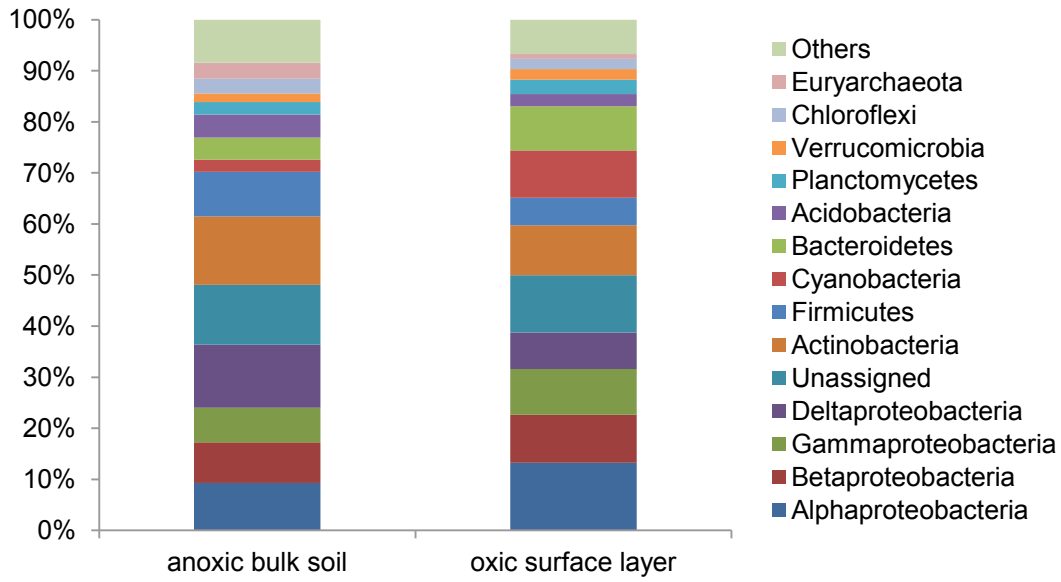

Figure 3.14. Taxonomic binning of mRNA-tags based on the best homologous protein. An e-value cutoff of 1e-05 was taken as threshold to consider best hits in BLASTX search significant. The relative abundance was calculated based on the number of mRNA-tags assigned to the respective group in relation to the total number of mRNA-tags that could be taxonomically binned.

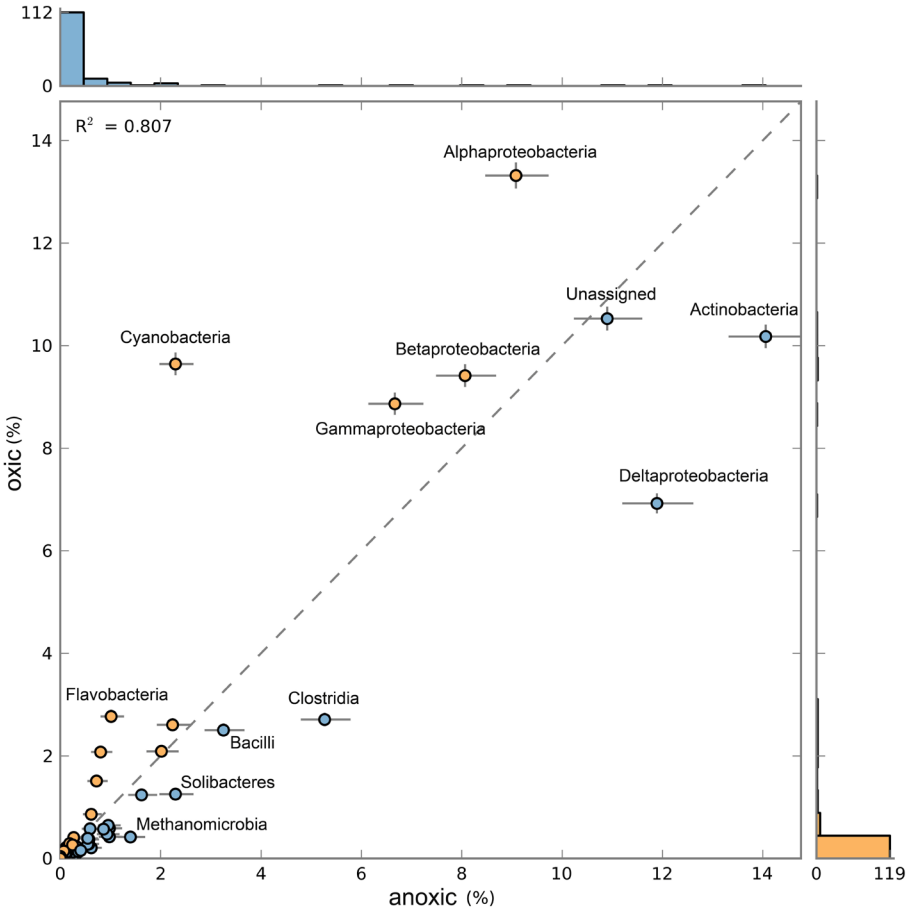

Figure 3.15. Profile scatter plot showing the percentage proportion of the different phyla and classes in the mRNA data sets from the oxic surface layer and the anoxic bulk soil.

4. Discussion

4.1. Sampling and RNA extraction

RNA has a short half-life in bacterial cells, so that cellular expression profiles may change rapidly in response to environmental cues. Therefore, the preservation of the RNA expression profiles at the time of sampling is critical in the analysis of functional activity by metatranscriptomics (Auer *et al*, 2003). RNA*later*®, a commercial RNA stabilization reagent, is being widely used to extract RNA from tissues, marine samples, and soil (Mutter *et al*, 2004; Meyer *et al*, 2006; Frias-Lopez *et al*, 2004). According to my own experience, RNA*later*® treatment of soil has a negative effect on the removal efficiency of organic compounds, including humic acids, during the extraction and purification of total RNA. Therefore, we decided to avoid the RNA*later*® treatment which, however, increased the procedural challenge to achieve the extraction of total RNA with sufficiently high integrity. After the assessment of various procedural steps, the combination of shock-freezing and low-pH conditions proved to be the optimal procedure to extract RNA of high integrity and quantity. Shock-freezing of soil samples with liquid nitrogen replaced the RNA*later*® treatment in order to maintain the metatranscriptome expression profile at the time of sampling. The shock-frozen samples were immediately subjected to the extraction of total RNA. The use of low-pH buffers had shown to stabilize the RNA during the extraction procedure. Co-extracted humic acids and other organic compounds, which inhibit chemical and enzymatic downstream processing, were successively removed by Q-Sepharose column chromatography.

4.2. mRNA enrichment

The predominance of rRNA in total cellular RNA is a major technical challenge in metatranscriptome analysis using RNA-seq. While eukaryotic mRNAs can be selectively converted into cDNA by targeting their poly(A) tails, prokaryotic mRNAs need to be enriched prior to cDNA synthesis in order to increase the proportion of mRNA sequences in RNA-seq data sets. Commercially available

kits for mRNA enrichment are based on the removal of rRNA by either subtractive hybridization or 5'-monophosphate-dependent exonuclease digestion. We evaluated both methods with respect to the removal efficiency of rRNA and the preservation of mRNA diversity. The exonuclease treatment was found to degrade not only rRNA but also a considerable amount of soil mRNA (Mettel *et al*, 2010). mRNAs carrying a monophosphate group at the 5' end were reported to generally occur during mRNA decay in prokaryotes (Celesnik *et al*, 2007). Thus, the exonuclease treatment would direct metatranscriptomic studies toward the analysis of unprocessed mRNAs that are 5'-triphosphorylated. We therefore decided to use subtractive hybridization of rRNA in order to cover full diversity of mRNA transcripts. Two mRNA enrichment kits that are based on subtractive hybridization were tested: MICROB*Express*™ Bacterial mRNA enrichment kit and Ribo-Zero™ rRNA removal kit (Meta-Bacteria). Their performance was assessed by capillary electrophoresis and small-scale sequencing. The MICROB*Express*™ Bacterial mRNA enrichment kit depleted rRNA in soil extracts with a removal efficiency of about 95%, as determined by comparing the fluorescence signal intensities of rRNA in total RNA with those in the enriched mRNA. One limitation of MICROB*Express*™ is that the oligonucleotide probes used for capturing of 16S and 23S rRNA do not target archaeal and eukaryal rRNA. In addition, rRNA of some particular bacterial taxa is not captured such as, for example, that from members of the *Planctomycetes* (a complete list is available on the manufacturer's website). The rRNA removal efficiency of MICROB*Express*™ declined with increasing partial fragmentation of the total RNA used for mRNA enrichment. Fragmentation of rRNA, particularly of 23S rRNA, occurs in many bacteria depending on growth phase (Selenska-Pobell & Evguenieva-Hackenberg, 1995; Klein *et al*, 2002). In addition, some physical fragmentation during the extraction of total RNA from soil cannot be completely avoided. The Ribo-Zero™ rRNA removal kit (Meta-Bacteria) was assessed because its manufacturer claimed that the kit is able to capture both intact and partially degraded rRNA. In fact, we observed a dramatic increase in the proportion of mRNA-derived reads after treatment with Ribo-Zero™. The exact procedure on how this high rRNA removal efficiency is achieved is not yet released by the

manufacturer. Taken together, the rRNA removal efficiency achieved by use of Ribo-Zero™ was tremendously greater than that observed by use of MICROB*Express*™ (**Figure 4.1**). Based on these findings, we used the Ribo-Zero™ rRNA removal kit for mRNA enrichment in this study.

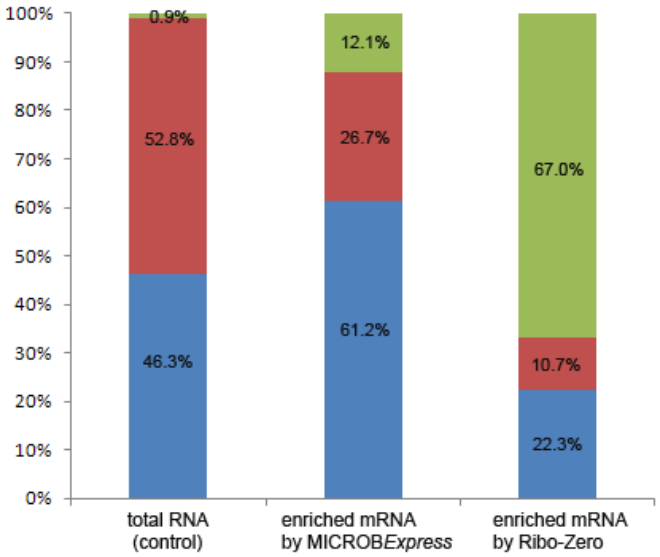

Figure 4.1. Comparison of metatranscriptome libraries after mRNA enrichment using Microb*Express* or Ribo-Zero, shown in relation to total RNA (non-enriched control). Blue, red and green indicate the proportion of LSU rRNA, SSU rRNA, and non-rRNA sequences, respectively.

4.3. cDNA synthesis and 454 library preparation

Metranscriptome studies performed in marine systems, even the most recently published study (Shi *et al*, 2011), amplified the enriched mRNA by *in vitro* transcription (IVT) prior to cDNA synthesis to have sufficient starting amount of cDNA for 454 library preparation. The cDNA rapid library preparation method for use of the GS FLX Titanium chemistry requires a minimum of 200 ng of enriched mRNA as the starting material, but avoids IVT amplification. We were able to produce sufficient amount of enriched mRNA for the rapid library preparation method by pooling ten aliquots of total RNA extracted from the same microcosm.

We assume that the direct conversion of enriched mRNA into cDNA avoids potential bias which may be introduced during RNA amplification via IVT. Therefore, RNA-seq data sets obtained after the use of the rapid library preparation method may reflect the mRNA composition in the metatranscriptome more accurately than those that involved the use of IVT amplification. It should also be noted that in all published studies, metatranscriptome libraries produced by IVT amplification have been sequenced using 454 GS FLX chemistry with read lengths not greater than 200 bp (**Table 1.2**), rather than the Titanium chemistry with increased read lengths as applied by us.

4.4. Preprocessing of 454 sequence data

Raw data of next-generation-sequencing contain low-quality sequences, sequence artifacts, sequence contaminations, sequence replicates and ambiguous sequence motifs. In *de novo* whole-genome sequencing, high coverage may compensate for some of these uncertainties in sequence reads. However, quality control and preprocessing of sequence data is crucial for correct functional annotation and taxonomic binning. This is true in particular in soil metatranscriptomics due to the complexity of community structure and, compared to marine metatranscriptomics, the limited amount of genomic and metagenomic reference data to which the soil metatransciptome data can be compared or mapped. Therefore, raw image data of 454 GS FLX Titanium runs have to pass through image and signal processing. Chimeric reads, reads with undetermined bases (called “N”) or those showing low sequence complexity will need to be removed by supplementary read filtering strategies. In this study, various quality filters were applied to raw reads data to omit ambiguous reads (**Table 2.1**). Additionally, reads with poor sequence quality at their 3’-ends were trimmed off to prevent incorrect annotation caused by low-quality base calls. The accuracy of base-calling in pyrosequencing decreases towards the end of reads (Balzer *et al*, 2010). Consequently, sequencing errors occur more frequently at the 3’ end (Balzer *et al*, 2011). Quality trimming allows to keep more reads in the data set and to use a reliable portion of the sequences for downstream analysis.

4.5. Removal of non-coding RNA

Cellular RNA is primarily composed of ribosomal RNA (rRNA) and transfer RNA (tRNA). In ocean metatranscriptome data sets, a significant fraction of cDNA sequences that could not be assigned to rRNA or known protein-coding genes was found to comprise well-known small RNA as well as new groups of previously unrecognized putative small RNA (Shi *et al*, 2009). The non-coding RNA, such as rRNA and tRNA but also small RNA, has to be identified and removed from the metatranscriptome libraries prior to the functional annotation of the mRNA-tags. Otherwise, they could be misclassified as protein-coding transcripts and then be added to the protein database as putative or hypothetical proteins. For example, it is reported that one conserved region of 23S rRNA was consistently misclassified to create spurious Pfam protein family (PF10695) with the function of cell wall hydrolase (Tripp *et al*, 2011). In our study, all the preprocessed reads were screened against the most updated SILVA rRNA databases and subsequently against the Rfam database, using an e-value cutoff of 1e-10.

4.6. Community structure analysis based on SSU-ribotags

Microbial community structure in complex environments has been widely studied using the 16S rRNA gene as phylogenetic marker. Next-generation sequencing technologies, in particular 454 pyrosequencing with average read lengths of up to 800 bp, enables the high-throughput analysis of microbial community composition. However, PCR-based amplicon sequencing may introduce bias due to primer selectivity and exponential amplification. DNA-based approaches are a good indicator of the genetic presence and potential but do not provide information on those community members that are active. Another concern is the possible presence of free DNA (Prosser, 2007). Some of these limitations can be solved by analysis of total RNA rather than DNA, because cDNA is synthesized by random priming. The RNA approach avoids the potential bias introduced by PCR amplification, identifies the microbial groups active or at least viable at the time of sampling and, when combined with the analysis of enriched mRNA, provides a

linkage between community composition and functional activities.

4.7. Functional activity in the oxic surface layer

Ribotag analysis combined with metatranscriptome data provides insights into the functional activities of microbial communities. For example, one of the biogeochemical characteristics of the oxic surface layer is methane oxidation, while the anoxic rice paddy bulk soil is defined by methane production and its emission to the atmosphere. However, a significant portion of the methane produced in the bulk soil is immediately oxidized by methanotrophic activity prior to emission. The amount of methane that is oxidized before it reaches the atmosphere is in the range from 45% to 60% (Khalil *et al*, 1998). Thus, the activity of methanotrophic bacteria in the oxic surface layer, but also in the partially oxygenated rhizosphere, represents a biofilter that reduces the emission of methane to the atmosphere. In full agreement with the methanotrophic pathway, transcripts of genes encoding particulate methane monooxygenase were detected only in the oxic surface layer but not in the anoxic bulk soil. In SSU-ribotag analysis, we observed a spatial distribution pattern of methanotrophs. Ribotags affiliated with *Crenotrichaceae* were primarily detected in the oxic surface layer, while those of type II methanotrophs were retrieved from both oxygen zones, but more frequently from the anoxic zone. Ribotags of type I methanotrophs were detected in both oxic surface layer and anoxic bulk soil. At the rRNA level, the detection of type I and type II methanotrophs in the anoxic zone is not unexpected. Previous PCR-based studies already revealed the presence of methanotrophs in the anoxic bulk soil, in particular type II methanotrophs (Horz *et al*, 2001). It has been suggested that some methanotrophs are capable of fermentation (Roslev & King, 1995). In fact, Vecherskaya *et al*. (2009) showed that a *Methylocystis parvus* strain is capable of fermentative metabolism of poly-3-hydroxybutyrate (PHB) under anaerobic conditions. Moreover, type II methanotrophs may persist as “resting stages”, such as cysts (*Methylocystis*) or spores (*Methylosinus*), under unfavorable conditions (Hanson & Hanson, 1996).

Under periodic light condition, a filamentous green matrix developed early on the surface layer of the flooded rice microcosms. Ribotag analysis revealed that this biofilm was primarily formed by *Cyanobacteria*. Nearly half of the SSU-ribotags from all the sampling time points were assigned to this phylum (**Figure 3.7**). The primary function of the *Cyanobacteria* is oxygenic photosynthesis, thereby explaining why the O_2 concentrations in the surface soil layer of flooded rice paddy microcosms is higher in the light than in the dark (Frenzel *et al*, 1992). In the mRNA-tag data set obtained from the oxic surface layer, *Cyanobacteria* ranked second after *Proteobacteria* in the number of mRNA-tags taxonomically assigned at the phylum level, accounting for about 10% of annotated mRNA-tags. About 5% of the *Cyanobacteria*-derived mRNA-tags were annotated as proteins involved in photosynthesis. Vijayan *et al.* (2011) reported that genes related to photosynthesis, ribosome and RNA polymerase are most highly transcribed in *Synechococcus elongates* PCC 7942. The distribution of *Cyanobacteria*-derived mRNA-tags to functional categories in our environmental data set differed from the distribution of mRNA-tags in the pure culture transcriptome of *S. elongates* PCC 7942. This discrepancy may have both biological and method-inherent reasons. Members of the *Cyanobacteria* were reported to exhibit relatively rapid mRNA decay rates of 2.4 minutes (Steglich *et al*, 2010). The rapid turnover rates may lead to rapid changes in the expression profiles in response to environmental cues such as stress that may have affected cyanobacterial cells during soil sampling. Another possibility is that among the cyanobacterial mRNA-tags, basic functional categories such as carbohydrate, protein and amino acids metabolism are overrepresented due to misclassification in taxonomic binning. I refer to section 4.7 for further details.

In SSU-ribotag analysis, *Gammaproteobacteria* ranked second after the *Cyanobacteria* in the number of tags assigned at the phylum level. The OTU with the greatest number of SSU-ribotags was affiliated with the order *Xanthomonadales*, accounting for 5% to 7% of the SSU-ribotags in the oxic surface layer over all samplings. Its 16S rRNA reference sequence (**see section 2.3.3**) had been obtained in an environmental study from the heavy DNA fraction of a stable-isotope probing (DNA-SIP) experiment using $^{13}CH_4$. It was indicative

of an uncultured bacterium assumed to be detected and rRNA-labeled due to cross-feeding of $^{13}CH_4$-derived compounds. Alternatively, this bacterium may represent a novel, uncultivated methanotroph present in the alkaline coal mine soil studied (Han *et al*, 2009). Thus, the high relative abundance of *Xanthomonadales*, including the prevalence of a few particular phylotypes, in the oxic surface layer may be related directly or indirectly to methane oxidation and calls for further research on the putative functional role of this *Xanthomonadales* population.

4.8. Functional activity in the anoxic bulk soil

Under anaerobic condition in the dark, microorganisms gain their energy by anaerobic respiration or fermentation. Different microbial guilds interact to break down and degrade complex organic matter in the anaerobic food chain. Diverse organisms exist and perform different types of fermentation or anaerobic respiration using various electron acceptors (McInerney *et al*, 2009). In our study, the taxonomic composition of the microbial communities showed the greater diversity in the anoxic bulk soil with higher richness and evenness than in the oxic surface layer. Over the different plant growth stages, there occurred no major changes in community composition at the phylum level, but shifts in their relative abundances. However, in the anoxic bulk soil, the richness of genus-level OTUs in the 25-day-old rice microcosm (tillering) was remarkably lower than that in the 45-day-old (flowering) and 90-day-old (ripening) rice microcosms (**Figure 3.6**). In order to experimentally define our sampling material as anoxic soil, roots and rhizosphere soil were separated from the bulk soil by a gaze bag. In contrast to the later plant growth stages, the roots of 25 day-old plants did not sufficiently grow to fill the bag. Thus, plants in the tillering stage may not have released sufficient amount of root exudates to promote the same OTU diversity as observed in the anoxic bulk soil of the 45-day-old and 90-day-old microcosm.

In the anoxic bulk soil, microorganisms typically involved in the anaerobic biodegradation of organic matter were detected with high frequency in both the SSU-ribotag and mRNA-tag data sets (**Figure 3.7 and Figure 3.15**). *Geobacter,*

Anaeromyxobacter, *Clostridia,* and methanogens represented the dominant groups of the anaerobic microbial community, but were not present or only present as minor population in the oxic zone. Previous studies have shown that *Geobacter* and *Anaeromyxobacter* have functional significance in flooded rice paddy soil as dissimilatory Fe(III) reducers (Jackel *et al*, 2005; Ratering & Schnell, 2000; Hori *et al*, 2007). Our findings are corroborated by the results of a previous study that detected *Geobacter*-related mRNA-tags specifically in the anoxic zone of flooded paddy soil by RT-PCR using a primer that targets bacterial Shine-Dalgarno sequences (SD14 primer), followed by amplicon cloning and Sanger sequencing of 800 random clones (Shrestha *et al*, 2009). Comparative analysis of our mRNA data sets suggests that different pyruvate metabolism pathways were expressed in the oxic and anoxic zones of the flooded rice paddy soil microcosms (**Figure 3.12**). Pyruvate metabolism II, defined by the formation of acetate via acetyl-CoA, was greatly overrepresented in the anoxic bulk soil, as judged by the frequency of mRNA-tags related to this pathway. The genome sequences of organisms having a syntrophic life style such as, for example, *Geobacter sulfurreducens, Pelobacter carbinolicus, Desulfovibrio desulfuricans* G20, *Syntrophobacter fumaroxidans, and Syntrophus aciditrophicus*, revealed that genes encoding pathways of acetate formation are present in multiple gene copies (Kosaka *et al*, 2008; Butler *et al*, 2009; McInerney *et al*, 2007).

Methane production is a energy-yielding metabolism unique to methanogens in anoxic environments (Thauer *et al*, 2008). In rice field soil, methane is mainly produced by the reduction of carbon dioxide (perfomed by hydrogenotrophic methanogens) or the fermentation of acetate (performed by acetoclastic methanogens) (Conrad & Claus, 2005). Approximately 65% of CH_4 produced in anoxic rice field soil is derived from acetate (Chidthaisong *et al*, 1999). *Methanosarcina* (generalist) and *Methanosaeta* (specialist) are the only methanogens that can utilize acetate to form methane. In good agreement with previous studies based on PCR, both hydrogenotrophic and acetoclastic methanogens were detected in the SSU-ribotag data sets with similar read frequency (Frenzel *et al*, 1999; Krüger *et al*, 2005). However, twice as much of the methanogenesis-related mRNA-tags were taxonomically assigned to

Methanosarcina and *Methanosaeta* than to hydrogenotrophic methanogens. Particularly, mRNA-tags of the methyl-coenzyme M reductase (MCR), which is involved in the terminal step of methane formation, were mostly derived from acetoclastic methanogens. The strong expression of acetate-producing pyruvate metabolism II pathways among members of the bulk soil microbial community along with the high transcriptional activity of acetoclastic methanogens reflects that acetate is an important metabolite in anoxic rice paddy soil and a major substrate for methane production (Chidthaisong *et al*, 1999). These findings are also a good example of the new insights into the functional activity that can be gained by soil metatranscriptomics, even if used as a stand-alone approach. However, in particular, it will be a powerful tool if combined with other approaches, such as ribotag analysis, metagenomics, and/or process measurements.

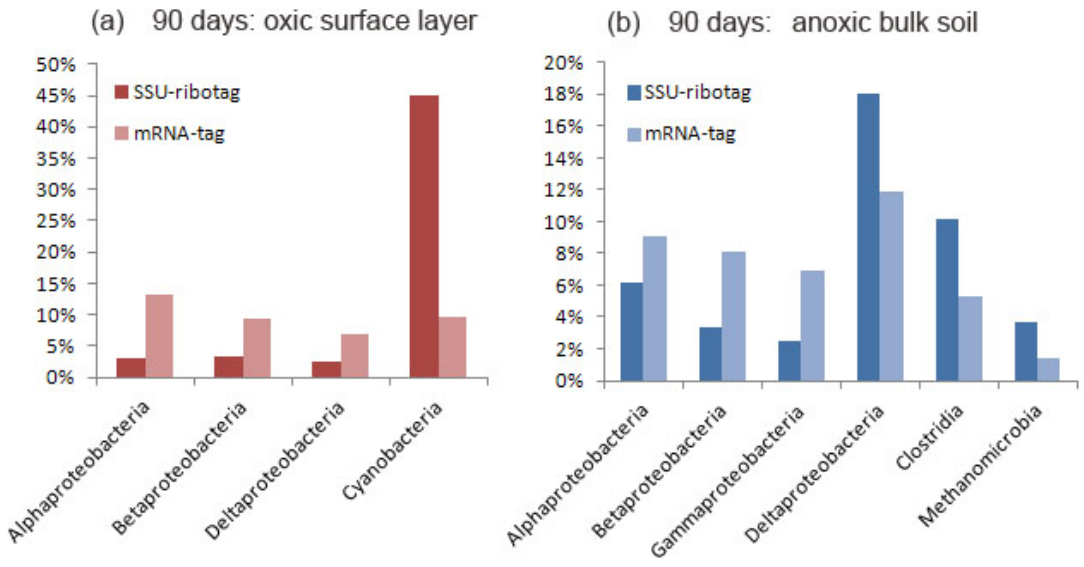

Figure 4.2. Comparison of the relative proportion of SSU-ribotags and mRNA-tags assigned to respective taxonomic groups.

4.9. Discrepancies between SSU-ribotag and mRNA-tag data analysis

In principle, we observed a similar trend between the taxonomic assignment of SSU-ribotags and the taxonomic binning of mRNA transcripts in the

corresponding data sets from either oxic surface layer or anoxic bulk soil. Taxonomic groups that were more frequently detected in the SSU-ribotag data set from the oxic surface layer than in that from the anoxic bulk soil (and vice versa) were also overrepresented in the corresponding mRNA-tag data set. However, when performing a more detailed inspection of the corresponding SSU-ribotag and mRNA-tag data sets, it became obvious that considerable discrepancies occur in the relative abundances of certain taxonomic groups (**Figure 4.2**). These discrepancies may be explained by a combination of several factors: (1) Differences in the SSU-ribotag and mRNA-tag frequencies may reflect the native situation because the number of SSU-ribotags represents the cellular biomass of the respective taxon, while mRNA-tags are indicative of the *in situ* cellular activity (Weller & Ward, 1989; Urich *et al*, 2008). Since even starving cells contain a certain amount of rRNA, high biomass of inactive cells may lead to a low frequency of mRNA-tags relative to SSU-ribotags. (2) The discrepancy could be due to incorrect taxonomic binning of mRNA-tags. These were functionally and taxonomically annotated by searching translated cDNA against non-redundant protein databases, because protein sequences are evolutionarily more conserved than nucleotide sequences (Mount, 2001). The amino acid sequences of various proteins involved in basic cellular function, stability, or reproduction are highly conserved among microorganisms, regardless of phylogenetic affiliation (Gaasterland *et al*, 2000). It has been reported that the accurate taxonomic binning of mRNA is negatively affected by potential lateral gene transfer (Poretsky *et al*, 2005; Shrestha *et al*, 2009). Therefore, as exemplified in **Figure 4.3**, it should be noted that taxonomic binning of genes which are highly conserved and distributed among many organisms has to be interpreted with care, in particular if assignments are made at the genus or even species level (Burke *et al*, 2011). (3) The relative proportion of mRNA-tags taxonomically assigned to *Proteobacteria* was greater than the relative proportion of SSU-ribotags assigned to this group (**Figure 4.2**). Among the 8448 bacterial genome projects including finished and on-going projects, almost half of them focus on members of the *Proteobacteria*, and 45% of the metagenome projects are surveying aquatic systems that are dominated by *Proteobacteria* (**Figure 4.4**). As

a consequence, the taxonomic binning of mRNA-tags may lead to a bias toward an underrepresentation of microbial groups that are not well represented by genome sequences (Urich *et al*, 2008).

Figure 4.3. BLASTX matches of an exemplary mRNA-tag (HH02YNVO2FO0UB) in the data set from the anoxic bulk soil. Top 29 hits out of 94 matches are shown. The taxonomic origin of the different BLASTX hits is highly diverse. They have nearly identical sequence similarity scores. The function of all hit sequences was annotated as potassium uptake protein.

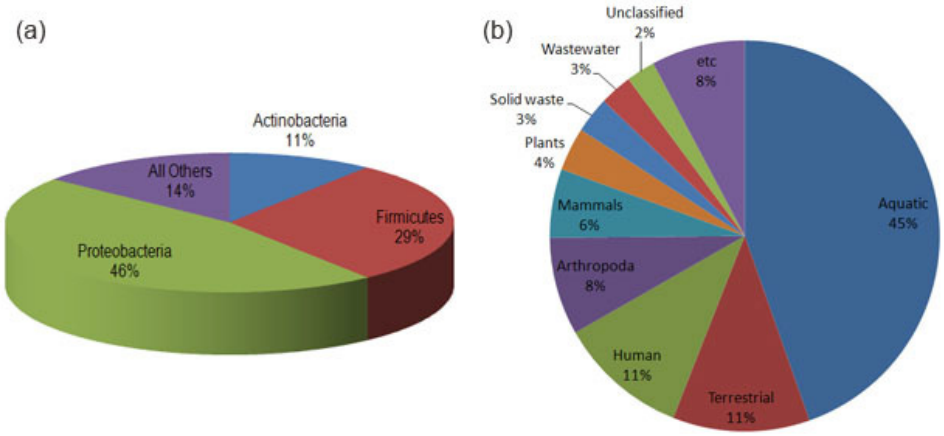

Figure 4.4. (a) Phylogenetic distribution of bacterial genome projects (October 2011, 8448 projects) and (b) metagenome classification depending on ecosystem categories (March 2012, 334 projects). The genome and metagenome data are available at http://www.genomesonline.org/cgi-bin/GOLD/index.cgi.

4.10. Current status and perspectives

In this study, we analyzed aerobic and anaerobic microbial communities in flooded rice paddy soil microcosms through soil metatranscriptomics, a novel molecular ecology approach to gain insights into the functional activities of soil microbial communities. We observed a similar functional distribution of mRNA-tags related to the maintenance of the basic cellular machinery (**Figure 3.11**). This finding may not be surprising as the basic cellular machinery is highly conserved among bacteria, regardless of whether they possess an aerobic or anaerobic metabolism. Common metatranscriptome patterns of basic cellular machinery have also been observed between different study sites in marine systems (Stewart *et al*, 2011)**.** On the other hand, gene expression in response to particular environmental conditions was observed. As one example, transcripts encoding particulate methane monooxygeanase (methane oxidation) were detected only in the oxic surface layer, while *those encoding* methyl-coenzyme *M* reductase (methanogenesis) were identified only in the anoxic bulk soil.

In addition to the SSU-ribotag datasets, about 48,000 and 12,000 mRNA-

tags were obtained from the oxic surface layer and the anoxic bulk soil, respectively. In order to further search for candidates of ecologically relevant genes and to have mRNA-tag datasets of comparable size, we intend to increase the number 454 pyrosequences from the anoxic bulk soil to about 40,000 to 50,000 mRNA-tags. We annotated and compared functional activities of microbial communities adapted to either oxic or anoxic conditions, using a hierarchical system of functional categories. Given a sufficient number of mRNA-tags, the metatranscriptome level of particular genes or gene families among members of a microbial community can be monitored to quantify their ecological importance.

The differential representation of *Cyanobacteria*-derived reads in the SSU-ribotag and mRNA-tag datasets from the surface layer shall be addressed in further studies. The cyanobacterial biofilm was formed soon after transplantation of the rice seedling and may already have been aged at the time of sampling to analyze enriched mRNA. Thus, a large portion of the biofilm may have been composed of inactive cells, thereby explaining the low relative abundance of mRNA-tags affiliated with cyanobacteria, in particular with oxygenic photosynthesis. This may also explain why a much greater number of functional subcategories had a siginificantly higher metatranscriptome level in the anoxic bulk soil than in the oxic surface layer.Therefore, we intend to perform an analysis of total RNA and enriched mRNA from the oxic zone of flooded paddy soil microcosms shortly after a cyanobacterial biofilm is visible. We expect that the mRNA-tag analysis in early plant growth stage, when the biofilm is freshly formed, could give some clue to explain the relationship between metabolic activity and biomass of *Cyanobacteria*.

5. References

Auer H, Lyianarachchi S, Newsom D, Klisovic MI, Marcucci G & Kornacker K (2003) Chipping away at the chip bias: RNA degradation in microarray analysis. *Nature Genetics* **35:** 292-293

Aulakh MS, Wassmann R, Bueno C, Kreuzwieser J & Rennenberg H (2001) Characterization of Root Exudates at Different Growth Stages of Ten Rice (Oryza sativa L.) Cultivars. *Plant biology* **3:** 139-148

Bailly J, Fraissinet-Tachet L, Verner M-C, Debaud J-C, Lemaire M, Wésolowski-Louvel M & Marmeisse R (2007) Soil eukaryotic functional diversity, a metatranscriptomic approach. *The ISME journal* **1:** 632-642

Balzer S, Malde K & Jonassen I (2011) Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* **27:** i304-i309

Balzer S, Malde K, Lanzén A, Sharma A & Jonassen I (2010) Characteristics of 454 pyrosequencing data—enabling realistic simulation with flowsim. *Bioinformatics* **26:** i420-i425

Bloem J (1995) Fluorescent staining of microbes for total direct counts. *Molecular microbial ecology manual* **1:** 861-873

Blumberg DD (1987) Creating a ribonuclease-free environment. *Methods in Enzymology* **152:** 20-24

Brown JW & Pace NR (1992) Ribonuclease P RNA and protein subunits from bacteria. *Nucleic Acids Research* **20:** 1451-1456

Burke C, Steinberg P, Rusch D, Kjelleberg S & Thomas T (2011) Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 14288-14293

Butler JE, Young ND & Lovley DR (2009) Evolution from a respiratory ancestor to fill syntrophic and fermentative niches: comparative fenomics of six Geobacteraceae species. *BMC Genomics* **10:** 103

Bürgmann H, Widmer F, Sigler WV & Zeyer J (2003) mRNA Extraction and Reverse Transcription-PCR Protocol for Detection of nifH Gene Expression by Azotobacter vinelandii in Soil. *Applied and Environmental Microbiology* **69:** 1928-1935

Canfield DE, Thamdrup B & Hansen JW (1993) The anaerobic degradation of organic matter in Danish coastal sediments - Iron reduction, manganese reduction, and sulfate reduction. *Geochimica et Cosmochimica Acta* **57:** 3867-3883

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, Mcdonald D, Muegge BD, Pirrung M, Reeder J, *et al* (2010) correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nature Publishing Group* **7:** 335-336

Cascante M & Marin S (2008) Metabolomics and fluxomics approaches. *Essays in Biochemistry* **45:** 67-81

Casjens S (1998) The diverse and dynamic structure of bacterial genomes. *Annual Review of Genetics* **32:** 339-77

Celesnik H, Deana A & Belasco JG (2007) Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. *Molecular Cell* **27:** 79-90

Chen Y, Dumont MG, Cébron A & Murrell JC (2007) Identification of active methanotrophs in a landfill cover soil through detection of expression of 16S rRNA and functional genes. *Environmental Microbiology* **9:** 2855-2869

Chidthaisong A, Rosenstock B & Conrad R (1999) Measurement of Monosaccharides and Conversion of Glucose to Acetate in Anoxic Rice Field Soil. *Applied and Environmental Microbiology* **65:** 2350-2355

Chin K-J, Rainey FA, Janssen PH & Conrad R (1998) SYSTEM4TIC AND APPLIED MICROBIOLOGY Methanogenic Degradation of Polysaccharides and the Characterization of Polysaccharolytic Clostridia from Anoxic Rice Field Soil. *Systematic and Applied Microbiology* **21:** 185-200

Christensen H, Hansen M & Sørensen J (1999) Counting and Size Classification of Active Soil Bacteria by Fluorescence In Situ Hybridization with an rRNA Oligonucleotide Probe. *Applied and Environmental Microbiology* **65:** 1753-1761

Conrad R & Claus P (2005) Contribution of methanol to the production of methane and its C-13-isotopic signature in anoxic rice field soil. *Biogeochemistry* **73:** 381-393

Dong H, Chen Y, Shen Y, Wang S, Zhao G & Jin W (2011) Artificial duplicate reads in sequencing data of 454 Genome Sequencer FLX System. *Acta Biochimica et Biophysica Sinica* **43:** 496-500

Dumont MG, Pommerenke B, Casper P & Conrad R (2011) DNA-, rRNA- and mRNA-based stable isotope probing of aerobic methanotrophs in lake sediment. *Environmental Microbiology* **13:** 1153-1167

Dunn WB (2008) Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Physical Biology* **5:** 011001

Ebie Y, Noda N, Miura H, Matsumura M, Tsuneda S, Hirata A & Inamori Y (2004) Comparative analysis of genetic diversity and expression of amoA in wastewater treatment processes. *Applied Microbiology and Biotechnology* **64:** 740-744

van Elsas JD, Semenov AV, Costa R & Trevors JT (2011) Survival of Escherichia coli in the environment: fundamental and public health aspects. *The ISME journal* **5:** 173-83

Fenchel T (2002) Microbial behavior in a heterogeneous world. *Science* **296:** 1068-1071

Frenzel P, Bosse U & Janssen PH (1999) Rice roots and methanogenesis in a paddy soil: ferric iron as an alternative electron acceptor in the rooted soil. *Soil Biology and Biochemistry* **31:** 421-430

Frenzel P, Rothfuss F & Conrad R (1992) Oxygen Profiles and Methane Turnover in a Flooded Rice Microcosm. *Biology and Fertility of Soils* **14:** 84-89

Frias-Lopez J, Bonheyo GT & Fouke BW (2004) Identification of Differential Gene Expression in Bacteria Associated with Coral Black Band Disease by Using RNA-Arbitrarily Primed PCR. *Applied and Environmental Microbiology* **70:** 3687-3694

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW & Delong EF (2008) Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 3805-3810

Fukui M, Suwa Y & Urushigawa Y (1996) High survival efficiency and ribosomal RNA decaying pattern of Desulfobacter latus, a highly specific acetate-utilizing organism, during starvation. *Fems Micriobiology Ecology* **19:** 17-25

Gaasterland T, Sczyrba A, Thomas E, Aytekin-Kurban G, Gordon P & Sensen CW (2000) MAGPIE/EGRET annotation of the 2.9-Mb Drosophila melanogaster Adh region. *Genome Research* **10:** 502-510

Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR & Bateman A (2011) Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Research* **39:** D141-D145

Gich FB, Amer E, Figueras JB, Abella CA, Balaguer MD & Poch M (2000) Assessment of microbial community structure changes by amplified ribosomal DNA restriction analysis (ARDRA). *International microbiology the official journal of the Spanish Society for Microbiology* **3:** 103-106

Gilbert B & Frenzel P (1995) Methanotrophic bacteria in the rhizosphere of rice microcosms and their effect on porewater methane concentration and methane emission. *Biology and Fertility of Soils* **20:** 93-100

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P & Joint I (2008) Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PLoS ONE* **3:** 13

Gomez-Alvarez V, Teal TK & Schmidt TM (2009) Systematic artifacts in metagenomes from complex microbial communities. *The ISME journal* **3:** 1314-1317

Graham RL, Graham C & McMullan G (2007) Microbial proteomics: a mass spectrometry primer for biologists. *Microbial Cell Factories* **6:** 26

Han B, Chen Y, Abell G, Jiang H, Bodrossy L, Zhao J, Murrell JC & Xing X-H (2009) Diversity and activity of methanotrophs in alkaline soil from a Chinese coal mine. *Fems Micriobiology Ecology* **70:** 40-51

Hanson R & Hanson T (1996) Methanotrophic bacteria. *Microbiological Reviews* **60:** 439-471

He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, Huang Z, Wu W, Gu B, Jardine P, Criddle C & Zhou J (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *The ISME journal* **1:** 67-77

Heuer H, Weiland G, Schönfeld J, Schönwälder A, Newton CM & Smalla K (2001) Bacterial community profiling using DGGE or TGGE analysis. In *Environmental molecular Microbiology Protocols and Application*, Rochelle PA (ed) p 264 pp. Horizon Scientific Press

Hogan CM (2009) Encyclopedia of Earth Overgrazing Syndromes leading to overgrazing Phenomenology. *Whats New*

Holzapfel-Pschorn A, Conrad R & Seiler W (1985) Production, oxidation and emission of methane in rice paddies. *FEMS Microbiology Letters* **31:** 343-351

Hori T, Noll M, Igarashi Y, Friedrich MW & Conrad R (2007) Identification of acetate-assimilating microorganisms under methanogenic conditions in anoxic rice field soil by comparative stable isotope probing of RNA. *Applied and environmental microbiology* **73:** 101-9

Horz H-P, Yimga MT & Liesack W (2001) Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of pmoA, mmoX, mxaF, and 16S rRNA and ribosomal DNA, including pmoA-based terminal restriction fragment length polymorphism profiling. *Applied and environmental microbiology* **67:** 4177-4185

Horz HP, Rotthauwe JH, Lukow T & Liesack W (2000) Identification of major subgroups of ammonia-oxidizing bacteria in environmental samples by T-RFLP analysis of amoA PCR products. *Journal of Microbiological Methods* **39:** 197-204

Jackel U, Russo S & Schnell S (2005) Enhanced iron reduction by iron supplement: A strategy to reduce methane emission from paddies. *Soil Biology and Biochemistry* **37:** 2150-2154

Kanehisa M, Goto S, Sato Y, Furumichi M & Tanabe M (2011) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* **40:** 109-114

Khalil MAK, Rasmussen RA & Shearer MJ (1998) Effects of production and oxidation processes on methane emissions from rice fields. *Journal of Geophysical Research - Atmospheres* **103:** 25233-25239

Klein F, Samorski R, Klug G & Evguenieva-Hackenberg E (2002) Atypical Processing in Domain III of 23S rRNA of Rhizobium leguminosarum ATCC 10004T at a Position Homologous to an rRNA Fragmentation Site in Protozoa. *Journal Of Bacteriology* **184:** 3176-3185

Knauth S, Hurek T, Brar D & Reinhold-Hurek B (2005) Influence of different Oryza cultivars on expression of nifH gene pools in roots of rice. *Environmental Microbiology* **7:** 1725-1733

Kolb S, Knief C, Dunfield PF & Conrad R (2005) Abundance and activity of uncultured methanotrophic bacteria involved in the consumption of atmospheric methane in two forest soils. *Environmental Microbiology* **7:** 1150-1161

Konstantinidis KT, Braff J, Karl DM & DeLong EF (2009) Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Applied and Environmental Microbiology* **75:** 5345-5355

Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T & Watanabe K (2008) The genome of Pelotomaculum thermopropionicum reveals niche-associated evolution in anaerobic microbiota. *Genome Research* **18:** 442-8

Krüger M, Frenzel P, Kemnitz D & Conrad R (2005) Activity, structure and dynamics of the methanogenic archaeal community in a flooded Italian rice field. *Fems Micriobiology Ecology* **51:** 323-331

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

Liu Y & Whitman WB (2008) Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Annals Of The New York Academy Of Sciences* **1125:** 171-89

Lu Y & Conrad R (2005) In situ stable isotope probing of methanogenic archaea in the rice rhizosphere. *Science (New York, N.Y.)* **309:** 1088-90

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, Rubin EM & Jansson JK (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **05:** 368-371

McCarren J, Becker JW, Repeta DJ, Shi Y, Young CR, Malmstrom RR, Chisholm SW & DeLong EF (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 16420-16427

McInerney MJ, Rohlin L, Mouttaki H, Kim U, Krupp RS, Rios-Hernandez L, Sieber J, Struchtemeyer CG, Bhattacharyya A, Campbell JW & Gunsalus RP (2007) The genome of Syntrophus aciditrophicus: life at the thermodynamic limit of microbial growth. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 7600-7605

McInerney MJ, Sieber JR & Gunsalus RP (2009) Syntrophy in anaerobic global carbon cycles. *Current Opinion in Biotechnology* **20:** 623-632

Mettel C, Kim Y, Shrestha PM & Liesack W (2010) Extraction of mRNA from Soil. *Applied and Environmental Microbiology* **76:** 5995-6000

Meyer H, Kaiser C, Biasi C, Hämmerle R, Rusalimova O, Lashchinsky N, Baranyi C, Daims H, Barsukov P & Richter A (2006) Soil carbon and nitrogen dynamics along a latitudinal transect in Western Siberia, Russia. *Biogeochemistry* **81:** 239-252

Mou X, Vila-Costa M, Sun S, Zhao W, Sharma S & Moran MA (2011) Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environmental Microbiology Reports* **3:** 798-806

Mount BDW (2001) The first step toward a definitive bioinformatics textbook Bioinformatics: Sequence and Genome Analysis. *Bioinformatics* **28:** 2001-2001

Mutter GL, Zahrieh D, Liu C, Neuberg D, Finkelstein D, Baker HE & Warrington JA (2004) Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. *BMC Genomics* **5:** 88

Newman DK & Banfield JF (2002) Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. *Science* **296:** 1071-1077

Pace NR (1997) A Molecular View of Microbial Diversity and the Biosphere. *Science* **276:** 734-740

Parks DH & Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26:** 715-721

Phelan VV, Liu W-T, Pogliano K & Dorrestein PC (2011) Microbial metabolic exchange—the chemotype-to-phenotype link. *Nature Chemical Biology* **8:** 26-35

Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, Pickering M, Pate WM, Moran MA & Hollibaugh JT (2005) Analysis of Microbial Gene Transcripts in Environmental Samples. *Applied and Environmental Microbiology* **71:** 4121-4126

Poretsky RS, Sun S, Mou X & Moran MA (2010) Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology* **12:** 616-627

Pratscher J, Dumont MG & Conrad R (2011) Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 4170-5

Prosser JI (2007) Microorganisms Cycling Soil Nutrients and Their Diversity. In *Modern Soil Microbiology*, Van Elsas JD Jansson JK, & Trevors JT (eds) pp 237-261. CRC Press

Pruesse E, Quast C, Yilmaz P, Ludwig W, Peplies J & Glöckner FO (2011) SILVA: comprehensive databases for quality checked and aligned ribosomal RNA sequence data compatible with ARB. In *Handbook of Molecular Microbial Ecology I Metagenomics and Complementary Approaches*, De Bruijn FJ (ed) pp 393-398. John Wiley & Sons

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, Van Dam K & Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology* **19:** 45-50

Ratering S & Schnell S (2000) Localization of iron-reducing activity in paddy soil by profile studies. *Biogeochemistry* **48:** 341-365

Revsbech NP, Pedersen O, Reichardt W & Briones A (1999) Microsensor analysis of oxygen and pH in the rice rhizosphere under field and laboratory conditions. *Biology and Fertility of Soils* **29:** 379-385

Rondon MR, Goodman RM & Handelsman J (1999) The Earth's bounty: assessing and accessing soil microbial diversity. *Trends in biotechnology* **17:** 403-409

Roslev P & King GM (1995) Aerobic and Anaerobic Starvation Metabolism in Methanotrophic Bacteria. *Applied and Environmental Microbiology* **61:** 1563-1570

Schink B (2000) Syntrophism among prokaryotes. *The Prokaryotes* **1:** 276-299

Schloss PD & Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology* **71:** 1501-1506

Schmid M, Schmitz-Esser S, Jetten M & Wagner M (2001) 16S-23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and in situ detection. *Environmental Microbiology* **3:** 450-459

Schmieder R & Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27:** 863-864

Selenska-Pobell S & Evguenieva-Hackenberg E (1995) Fragmentations of the large-subunit rRNA in the family Rhizobiaceae. *Journal Of Bacteriology* **177:** 6993-6998

Sharma S, Aneja MK, Mayer J, Munch JC & Schloter M (2005) Diversity of Transcripts of Nitrite Reductase Genes (nirK and nirS) in Rhizospheres of Grain Legumes. *Applied and Environmental Microbiology* **71:** 2001-2007

Shi Y, McCarren J & DeLong EF (2011) Transcriptional responses of surface water marine microbial assemblages to deep-sea water amendment. *Environmental Microbiology* **14:** 191-206

Shi Y, Tyson GW & DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-269

Shrestha M, Abraham W-R, Shrestha PM, Noll M & Conrad R (2008) Activity and composition of methanotrophic bacterial communities in planted rice soil studied by flux measurements, analyses of pmoA gene and stable isotope probing of phospholipid fatty acids. *Environmental Microbiology* **10:** 400-412

Shrestha PM, Kube M, Reinhardt R & Liesack W (2009) Transcriptional activity of paddy soil bacterial communities. *Environmental Microbiology* **11:** 960-970

Steglich C, Lindell D, Futschik M, Rector T, Steen R & Chisholm SW (2010) Short RNA half-lives in the slow-growing marine cyanobacterium Prochlorococcus. *Genome biology* **11:** R54

Stewart FJ, Ulloa O & Delong EF (2011) Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **2:** 23-40

Surette MG, Miller MB & Bassler BL (1999) Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: A new family of genes responsible for autoinducer production. *Proceeding of the National Academy of Sciences of the United States of America* **96:** 1639-1644

Thauer RK, Kaster A-K, Seedorf H, Buckel W & Hedderich R (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature reviews. Microbiology* **6:** 579-91

Tripp HJ, Hewson I, Boyarsky S, Stuart JM & Zehr JP (2011) Misannotations of rRNA can now generate 90% false positive protein matches in metatranscriptomic studies. *Nucleic Acids Research* **39:** 1-11

Urich T, Lanzén A, Qi J, Huson DH, Schleper C & Schuster SC (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* **3:** 13

Vecherskaya M, Dijkema C, Saad HR & Stams AJM (2009) Microaerobic and anaerobic metabolism of a Methylocystis parvus strain isolated from a denitrifying bioreactor. *Environmental Microbiology Reports* **1:** 442-449

Vijayan V, Jain IH & O'Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biology* **12:** R47

Vila-Costa M, Rinta-Kanto JM, Sun S, Sharma S, Poretsky R & Moran MA (2010) Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. *The ISME journal* **4:** 1410-1420

Wang B & Adachi K (2000) Differences among rice cultivars in root exudation , methane oxidation , and populations of methanogenic and methanotrophic bacteria in relation to methane emission. *Nutrient Cycling in Agroecosystems* **58:** 349-356

Weinbauer MG, Beckmann C & Höfle MG (1998) Utility of Green Fluorescent Nucleic Acid Dyes and Aluminum Oxide Membrane Filters for Rapid Epifluorescence Enumeration of Soil and Sediment Bacteria. *Applied and Environmental Microbiology* **64:** 5000-5003

Weller R & Ward DM (1989) Selective Recovery of 16S rRNA Sequences from Natural Microbial Communities in the Form of cDNA. *Applied and Environmental Microbiology* **55:** 1818-1822

Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 6578-6583

Yu CEPS & Harch BGHBD (2001) Capacity of fatty acid profiles and substrate utilization patterns to describe differences in soil microbial communities associated with increased salinity or alkalinity at three locations in South Australia. *Biology and Fertility of Soils* **33:** 204-217

Zelles L & Bai QY (1994) Fatty-Acid Patterns of Phospholipids and Lipopolysaccharides in Environmental-Samples. *Chemosphere* **28:** 391-411

Zhang W, Li F & Nie L (2010) Integrating multiple “omics” analysis for microbial biology: application and methodologies. *Microbiology* **156:** 287-301

List of abbreviations

amoA	ammonium monooxygenase alpha subunit
bp	base pairs
C	carbon
cDNA	complementary DNA
DEPC	diethylpyrocarbonate
DOC	dissolved organic carbon
EDTA	Ethylenediaminetetraacetic acid
emPCR	emulsion PCR
ESTs	expressed sequence tags
LSU rRNA	large subunit ribosomal RNA
mcrA	methane Co-M reductase alpha subunit
mmo	methane monooxygenase
mRNA	messenger RNA
N	nitrogen
NGS	next-generation sequencing
ORF	open reading frame
OTUs	operational taxonomic units
P	phosphorus
Pg	Petagram
pmoA	particulate methane monooxygenase alpha subunit
RFU	relative fluorescence unit
rtPCR	reverse transcription polymerase chain reaction
RT-qPCR	reverse transcription quantitative PCR
SSU rRNA	small subunit ribosomal RNA

Index of tables

Index of figures

Curriculum Vitae

Yongkyu Kim

Education

2008 – present	Ph.D. study at the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology,Marburg, Germany. Supervisor: Dr. Werner Liesack
2005 – 2007	M.Sc. Department of Bioinformatics and Life Science, Soongsil University, Seoul, South Korea. Supervisor: Prof. Hang-Cheol Shin
2001-2005	B. Sc. Department of Bioinformatics and Life Science, Soongsil University,Seoul, South Korea.

Professional Experience

2007 – 2008	Researcher, Biotech Research Center of Vexxon, Inc., Seoul, South Korea Research project: Protein engineering of Insulin
2005 – 2006	Teaching Assistant, Department of Bioinformatics and Life Science, Soongsil University,Seoul, South Korea.
2004	Research Intern, Equispharm Co., Ltd., Seoul, South Korea. Research Project: Adenosine A3 receptorstructure analysis and 3D structure modeling

Reward and scholarships

2008 – 2010	Doctoral scholarship from International Max Planck Research School for Environmental, Cellular and Molecular Microbiology
2006	The Brain Korea 21 BT-IT integration program scholarship by Korean Research Foundation

2001- 2004	Four-years of special scholarship for distinguished matriculate, Soongsil University

Publication

Mettel C[¥], **Kim Y[¥]**, Shrestha PM & Liesack W (2010) Extraction of mRNA from Soil. *Applied and Environmental Microbiology* **76:** 5995-6000

[¥] These authors contributed equally to this work

[M.Sc.thesis] **Kim Y** (2006) The identification of a role of a linker region between b' domain and a' daomin in ratPDI. Soongsil University

Kim Y, Kim J, Amgalan B, and Kim S. (2006) The relationships of pathways in KEGG database (PATHtoPATH) *Bioinformatics and Biosystems* **3**: 171-173

Pledge

I certify that the present thesis entitled:

“**Metatranscriptome analysis of microbial communities in rice microcosms”**

was carried out without any unlawful devices. I did not use any other than the described literature sources or technical devices. This work has never been submitted before in this or similar form to any other university and has not been used before any examination.

Marburg, April 2012

Yongkyu Kim

Acknowledgements

First of all, I wish to express my sincere gratitude and appreciation to my supervisor, Dr. Werner Liesack, for introducing me a metatranscriptomics and for valuable suggestions, comments, patience, deep understanding during the entire period of my Ph.D.

Next, I would like to especially thank my thesis committee, Prof. Martin Thanbichler, Prof. Michael Bölker, and Prof. Uwe Maier, for their helpful guidance and valuable suggestions.

I would like to thank my lab mates and collegues. Specially to Carsten Mettel for the collaboration to establish RNA extraction from soil. I always wish him best in his future endeavors. Carl-Eric Wegner for his lots of helps and encouragements. Bomba Dam and Somasri Dam for the great working environments. Former group members, Pravin Malla Shrestha and Mohamed Baani, for their kindness.

I was lucky to conduct my Ph.D. as a member of International Max Planck Research School for Environmental, Cellular and Molecular Microbiology. Thanks for the funding of my fellowship. So much thanks to Susanne Rommel for her help in finding a solution both in official as well as personal matters.

Now my thanks go to my family. They always encouraged my study and their love kept me going through many difficulties.

And finally to my beloved wife, Hyun Suk Kim. She was always by my side with *a* supportive smile, a trusting hand and full of love.