

ORIGINAL ARTICLE

Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice

Claudia Knief¹, Nathanaël Delmotte¹, Samuel Chaffron², Manuel Stark², Gerd Innerebner¹, Reiner Wassmann^{3,4}, Christian von Mering² and Julia A Vorholt¹

¹Institute of Microbiology, ETH Zurich, Zurich, Switzerland; ²Institute of Molecular Biology and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland; ³International Rice Research Institute, Metro Manila, Philippines and ⁴Institute for Meteorology and Climate Research, Karlsruhe Research Center, Karlsruhe, Germany

The above- and below-ground parts of rice plants create specific habitats for various microorganisms. In this study, we characterized the phyllosphere and rhizosphere microbiota of rice cultivars using a metaproteogenomic approach to get insight into the physiology of the bacteria and archaea that live in association with rice. The metaproteomic datasets gave rise to a total of about 4600 identified proteins and indicated the presence of one-carbon conversion processes in the rhizosphere as well as in the phyllosphere. Proteins involved in methanogenesis and methanotrophy were found in the rhizosphere, whereas methanol-based methylotrophy linked to the genus *Methylobacterium* dominated within the protein repertoire of the phyllosphere microbiota. Further, physiological traits of differential importance in phyllosphere versus rhizosphere bacteria included transport processes and stress responses, which were more conspicuous in the phyllosphere samples. In contrast, dinitrogenase reductase was exclusively identified in the rhizosphere, despite the presence of *nifH* genes also in diverse phyllosphere bacteria.

The ISME Journal advance online publication, 22 December 2011; doi:10.1038/ismej.2011.192

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: metaproteogenomics; phyllosphere; rhizosphere; *Oryza sativa*; microbial community; rice

Introduction

Rice is one of the most important food crops, nourishing approximately 50% of the world's population. The rice plants represent a habitat for diverse microorganisms, which colonize the aerial parts, referred to as phyllosphere, as well as the root surface (rhizoplane). A specific microbial community is also found in the zone around the root that is influenced by the plant, the rhizosphere (Kowalchuk *et al.*, 2010). Substantial research has been performed to elucidate the activities and functions of rice root-associated microbiota. On the one hand, beneficial effects were explored with respect to nutrient supply, in particular nitrogen, protection against pathogens and plant growth stimulation (Ladha and Reddy, 2000; Prasanna *et al.*, 2010). On the other hand, biogeochemical conversion processes of carbon, nitrogen, sulfur and iron were extensively studied, in particular for rice grown

under flooded conditions (Brune *et al.*, 2000; Liesack *et al.*, 2000). As oxygen is rapidly consumed in rice paddies upon flooding, the major carbon-cycling process is the fermentative degradation of organic matter. The resulting products are further oxidized coupled to the reduction of nitrate, iron (III) or sulfate, as long as these are available as electron acceptors, before the final degradation steps are taken over by syntrophic bacteria and methanogenic archaea. Thus, flooded rice paddies represent a major biogenic source of atmospheric methane, despite the fact that a substantial amount (10–40%) of this greenhouse gas is oxidized by aerobic methanotrophic bacteria before it reaches the atmosphere (Frenzel, 2000; Krüger *et al.*, 2001). These methanotrophs thrive in the oxic zones around the rice roots and the shallow soil-surface layer (Conrad, 2007).

The bacteria inhabiting the phyllosphere of rice and their physiological adaptations to the habitat have been less intensively studied. So far, a number of bacterial isolates from the rice phyllosphere have been characterized (Elbeltagy *et al.*, 2000; Madhaiyan *et al.*, 2007, 2009; Mano *et al.*, 2007) and potential beneficial interactions of phyllosphere bacteria with rice plants, such as plant

Correspondence: JA Vorholt or C Knief, Institute of Microbiology, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland.

E-mail: vorholt@micro.biol.ethz.ch or knief@micro.biol.ethz.ch
Received 29 August 2011; revised 31 October 2011; accepted 15 November 2011

growth promotion, for example, by bacterial nitrogen fixation or plant hormone production, and protection against pathogens have been studied (Madhaiyan *et al.*, 2004; Maliti *et al.*, 2005; De Costa *et al.*, 2008; Yang *et al.*, 2008; Chinnadurai *et al.*, 2009; Pedraza *et al.*, 2009).

Metagenome- and proteome-based analyses are global approaches that allow to identify members of a microbial community, give insight into the physiological potential of the community and enable the identification of the metabolic pathways in an ecosystem under given conditions. However, the application of metaproteomic methods to highly complex microbial plant-associated communities will remain challenging (Knief *et al.*, 2011). The first metaproteogenomic study of plant-associated microorganisms was concerned with the bacterial communities inhabiting the phyllosphere of *Arabidopsis*, soybean and clover plants, and revealed a remarkable consistency with respect to the dominant bacterial taxa and the proteins identified in populations from the different plants (Delmotte *et al.*, 2009). In that study, microorganisms were sampled from leaf material, and parallel metagenomics and metaproteomics analyses were performed using one-dimensional protein separation followed by liquid chromatography high-accuracy mass spectrometry. Metaproteomic studies of rhizosphere samples are still at its beginning and first reports were only published very recently (Wang *et al.*, 2011; Wu *et al.*, 2011). In those studies, proteins were directly extracted from rhizosphere samples (without a preceding physical enrichment step of microbial cells; Bastida *et al.*, 2009), separated by two-dimensional gel electrophoresis and identified by MALDI-TOF/TOF. The studies revealed that a direct protein extraction method restricted microbial protein

identification in the rhizosphere samples by the high recovery of plant proteins (>75% from 120 different identified proteins).

In the present study, a metaproteogenomic approach was applied to analyze the microbial community inhabiting the phyllosphere and rhizosphere of rice. We aimed at identifying the major physiological traits of the dominant rice-colonizing microorganisms and addressed the following research questions specifically: (1) how similar or different is the overall phyllosphere microbiota of this monocotyledonous tropically grown plant compared with the biota of the previously analyzed plant species; (2) is there evidence for microbial proteins specific for life in the phyllosphere and rhizosphere; (3) which are the potentially dominating catabolic processes of microorganisms living in association with rice, in particular in relation to one-carbon compound conversion; (4) to what extent can nitrogen fixation potential be demonstrated, considering the fact that phyllosphere bacteria and, even more so, stem endophytes have been reported to be able of nitrogen fixation, and rice is known to be colonized by diazotrophs (Elbeltagy *et al.*, 2001; Elbeltagy and Ando, 2008; Fürnkranz *et al.*, 2008; Pedraza *et al.*, 2009).

Materials and methods

Sample collection

Samples were collected from rice fields at the International Rice Research Institute, Los Baños, Philippines. Sampling took place 59 to 76 days after seedling transplantation in March 2009 (for details, see Figure 1 and Supplementary Table S1). For the analysis of phyllosphere microorganisms, the aerial

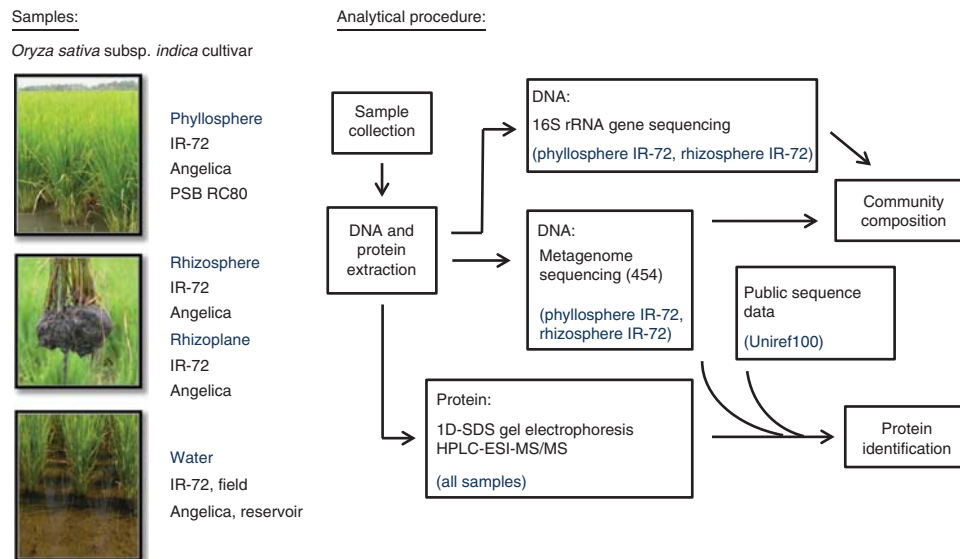


Figure 1 Overview of samples analyzed in the present study and the applied methods to get insight into the identity and physiology of the rice-associated microbiota. Further details about the samples are given in Supplementary Table S1.

parts of the plants of three rice varieties (*Oryza sativa* subsp. *indica* cv. Angelica, IR-72 and PSB RC80) were cut approximately 10 cm above the water level and immediately transferred to the laboratory at the International Rice Research Institute, where they were further processed within half a day. To wash off the microorganisms from the plant material, a bunch of plants was taken, dead leaf material and panicles were removed, and the material (200–250 g fresh weight) bagged into a polypropylene bag (excluding the lower part of the stem). TE buffer (250 ml) supplemented with 0.1% Silwet L-77 was added and the microorganisms were dislodged by alternate sonication and shaking for 150 s. Further enrichment of bacterial and archaeal cells by centrifugation of the suspension on top of a 2-ml Percoll layer in 50-ml tubes and washing was performed as described (Delmotte *et al.*, 2009). Between 6 and 10 kg of plant material was processed for each of the three phyllosphere samples.

To collect rhizosphere samples, three soil cores per rice variety (Angelica and IR-72) were taken by punching a stainless steel corer (inner diameter 5.4 cm) into the ground over a cropped rice tuft. The upper 6 cm of the core contained most of the root bale and was further processed. The root-attached soil was washed from the root material of all the three cores using 700 ml of TE-buffer plus 0.1% Silwet. Large clumps of non-rooted soil were removed before they disaggregated in the suspension. The obtained soil suspension was processed further in a similar way as the plant material, starting with the dislodgement of the bacteria from the soil particles by shaking and sonication in 50-ml tubes. The enrichment of bacterial and archaeal cells was done on top of a 3-ml Percoll layer and followed by washing steps as described for the phyllosphere samples.

Rhizoplane samples were collected by pulling out seven tufts of rice plants per variety and washing off the attached soil under tap water. Excess water was removed from the root material with paper towels. The root material was cut into pieces, transferred into 50-ml tubes and the bacteria were washed from the material and collected as described above.

Flooding water (4.5 l) from the field (IR-72) and from a basin (size approximately 1 m³; feeding the field on which rice cultivar Angelica was grown) was collected and filtrated through cellulose membrane filters with 0.22 µm pore size (GSWP04700, Millipore, Zug, Switzerland). Filters were frozen until further processing. To recover the microorganisms, the membrane filters were placed into 50-ml tubes and 35 ml of TE-buffer plus 0.1% Pefabloc and 0.1% Silwet L-77 were added. Microorganisms were dislodged by alternate sonication and shaking for 3 min. Enrichment of bacterial and archaeal cells was achieved by centrifugation of the suspension on top of a 6-ml Percoll layer and followed by washing steps.

Extraction of DNA and proteins

DNA and proteins were extracted from the collected material using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Hilden, Germany) as described (Delmotte *et al.*, 2009) with slight modifications. The bead-beating time in the tissue lyser was increased to 6 min and a second lysis step was introduced. To this end, the pellet obtained after the first lysis step was resuspended in 750 µl kit-supplied RLT buffer and shaken in a capsule-mixing unit (Cap Mix; 3M ESPE, Rüslikon, Switzerland) for 90 s. Upon centrifugation, the supernatants of both lysis steps were combined, the DNA in the suspension bound to two kit-supplied columns and washed. DNA was obtained from the column by two sequential elutions with 100 µl elution buffer.

Although DNA extraction using the AllPrep DNA/RNA/Protein Mini Kit resulted in high-quality DNA for the phyllosphere samples, it failed in case of the root samples. To obtain DNA for metagenome sequencing in this case, additional material of the IR-72 rhizosphere sample was taken for DNA extraction using the FastDNA SPIN Kit for Soil (Qbiogene, Heidelberg, Germany). In all, 1.5 g of material was extracted in three parallel assays. Cell lysis was performed using the kit-supplied materials and applying the procedure described above. Further extraction and purification with guanidine isothiocyanate (performed twice) was done as described (Knief *et al.*, 2005).

16S rRNA gene-based community analyses

Bacterial 16S rRNA gene-based clone libraries were constructed from the IR-72 phyllosphere and rhizosphere sample using the TOPO TA Cloning Kit (Life Technologies, Grand Island, NY, USA) as described (Delmotte *et al.*, 2009). Archaeal 16S rRNA genes were amplified in the root samples using primers Ar109f and Ar912rt (Lueders and Friedrich, 2002) in a PCR of 30 cycles (94 °C, 45 s; 52 °C, 45 s; 72 °C, 90 s). Sequences were deposited in the DDBJ/EMBL/GenBank databases under accession numbers HE589809 to HE589931.

Metagenome analysis

Metagenome sequence libraries were established as in a previous study (Delmotte *et al.*, 2009). Sequences were generated by shotgun sequencing on the Roche 454 Genome Sequencer FLX system (454 Life Sciences, Branford, CT, USA) at the Functional Genomics Center Zurich and are available under accession number SRA047327.1 (NCBI BioProject PRJNA75059). Assembly resulted in 184 273 contigs (average fragment length 605 bp) and 2 029 672 non-assembled reads totaling 831 769 586 bp with an average fragment length of 375 bp for the phyllosphere sample and 10 279 contigs (average fragment length 458 bp) and 1 016 703 non-assembled reads totaling 395 652 345 bp with an average fragment length of 385 bp for the rhizosphere sample. Metagenome open reading frame (ORF) prediction and annotation was performed as described in Delmotte *et al.* (2009). Similarity searches using

BLAST against the database UniRef90 were used to transfer protein and Pfam annotations. A total of 1 340 274 phyllosphere ORFs and 749 569 rhizosphere ORFs could be annotated with confidence (expected E-value cutoff of 0.0001 and minimum bitscore of 60). All non-annotated phyllosphere ORFs (23 169 756) and non-annotated rhizosphere ORFs (11 039 787) were kept in the metagenome database for MS identification. The taxonomic composition of the phyllosphere and the rhizosphere sample was analyzed based on metagenome data using MLTreeMap (Stark *et al.*, 2010).

Protein identification

The extracted protein fraction was processed as described before (Delmotte *et al.*, 2009). The number of spectra detected per sample is given in Supplementary Table S1. Spectra are deposited in the PRIDE database under accession number 1689. Data files were converted to peak lists and analyzed with Mascot 2.3 (Matrix Science, London, UK) and X! Tandem Tornado (2008.12.01.1; The Global Proteome Machine Organization). Database searches were performed against a database concatenated from Uniref100 (10 246 365 entries, release June 2010) and 36 299 386 sequences issued from the metagenomics ORF databases of this study; it had a total of 46 545 751 entries. Search parameters were as follows: taxonomy, all entries; fixed modification, cysteine carbamidomethylation; variable modifications, methionine oxidation; enzyme, trypsin; maximum number of missed cleavages, 1; peptide tolerance, 5 ppm; and MS/MS tolerance, 0.5 Da. Results were validated with Scaffold 3.0 (Proteome Software Inc., Portland, OR, USA). Protein identifications were accepted if they could be established at >99.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al.*, 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. False discovery rate at protein level was estimated to be below 0.5%. Note, however, that despite the high confidence in protein identification itself, the assignment of identified proteins to genera might mask assignment to closely related strains for which no protein sequence information is currently available. The lists of identified proteins reported by the two search engines were merged to a single list by creating the union of hits obtained by both algorithms for each spectrum.

Metaproteome data analyses

The complete list of identified proteins was filtered in a non-supervised fashion to remove eukaryotic proteins by parsing protein taxonomy annotations and excluding proteins from species not belonging to the NCBI bacteria (including archaea), viruses or environmental sample phylogenetic divisions of the

NCBI taxonomy. This filtered list was used for all downstream analyses. For taxonomic and functional analysis of the metaproteome, a version of MLTreeMap was designed that requires protein sequences as input instead of nucleotide sequences.

The similarity between sample proteomes was analyzed based on expressed Pfam protein domains. Fractional spectral counting was performed to semi-quantitatively estimate protein abundance. A given spectrum can be ambiguous and identified several proteins, and therefore this spectrum count is fractionally assigned to these ambiguous hits. Additionally, spectral counts were normalized by the total number of spectra identified in one sample and by the length of the matching protein. For each Pfam, these abundances were then aggregated based on protein/domain mappings (all domains of a given identified protein were considered). Samples were clustered using the Ward hierarchical clustering algorithm (using R, package *hclust*) on Euclidean distances computed using their Pfam fractional count profiles after log transformation. Cluster reliability was assessed using the R package *pvclust* (Suzuki and Shimodaira, 2006).

Results and discussion

The leaf- and root-associated microbiota of two rice cultivars grown at the International Rice Research Institute were analyzed in this study using a metaproteomic approach (Figure 1). Samples were taken from *O. sativa* subsp. *indica* cv. Angelica and IR-72 at the growth stage of flowering. An additional phyllosphere sample was taken from rice variety PSB RC80 (early flowering stage). The data of these samples were contrasted with two metaproteomic reference datasets of microbial communities not directly associated with plants: the microorganisms residing in the flooding water of the field on which cultivar IR-72 was grown and those of a water reservoir, which was used for flooding the field on which cultivar Angelica was grown. To increase the number of identified proteins, metagenomics shotgun sequencing was performed for one phyllosphere (IR-72) and one rhizosphere (IR-72) sample.

Bacterial community composition in the rice phyllosphere and rhizosphere according to DNA-based analyses

Information about the microbial community composition was gained from complementary approaches. The analysis of metagenome data by MLTreeMap uses the phylogenetic information contained in protein-coding marker genes that occur in single copy in all living organisms. This allows assessing the relative abundance of the members in the microbial community (Stark *et al.*, 2010). MLTreeMap analysis revealed the dominance of Alpha-proteobacteria (35%) and Actinobacteria (38%) in the phyllosphere of rice cultivar IR-72 (Figure 2).

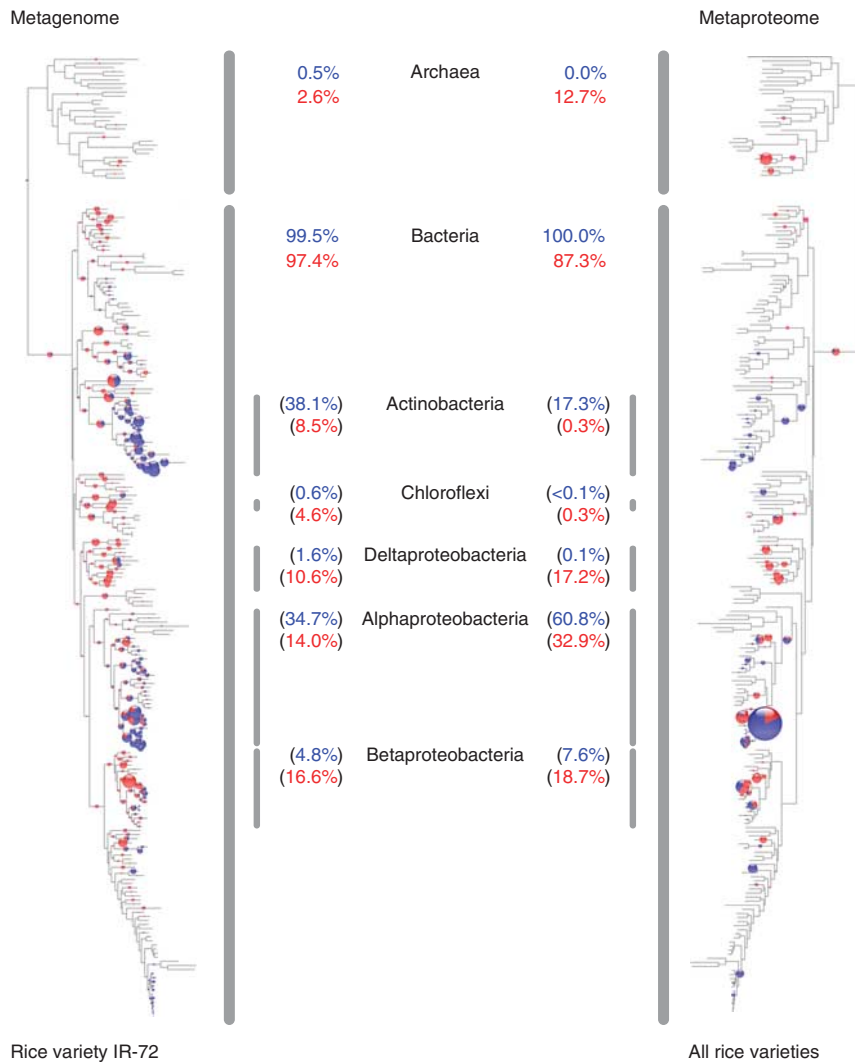


Figure 2 Bacterial and archaeal diversity in the metagenome and metaproteome datasets. An ML-TreeMap analysis was performed to assess the microbial community composition in the phyllosphere (blue) and rhizosphere (red) of rice variety IR-72 (left tree). The backbone tree was calculated based on aligned sequences of 40 phylogenetic marker genes from fully sequenced organisms. Dots indicate the placement of metagenomic sequence reads containing these marker genes, whereby the size of a dot corresponds to the frequency of recovery. In the metaproteome tree (right tree), blue dots indicate the phylogenetic placement of identified proteins of all the three phyllosphere samples, whereas red dots represent proteins identified in the rhizosphere and rhizoplane samples. The relative frequency with which genome reads and proteins of selected taxa were recovered is highlighted in the center. Detailed trees are available as supplementary material (Supplementary Figure S1).

Moreover, Bacteroidetes, Firmicutes, Beta- and Gammaproteobacteria contributed mainly to the bacterial community (Figure 2 and Supplementary Figure S1). 16S rRNA gene-based clone library data were in agreement with these results and allowed the identification of bacterial community members at higher taxonomic resolution. The Alphaproteobacteria appeared to be primarily represented by the genera *Rhizobium* and *Methylobacterium* (Supplementary Table S2 and Figure S2). Among the Actinobacteria, the genus *Microbacterium* was predominantly detected. In all, 13% of the bacterial clone sequences could not be assigned to known bacterial genera. A rarefaction analysis based on clone library data revealed that the complexity of

the bacterial phyllosphere community was comparable to that of previously analyzed plants (Supplementary Figure S3a).

The microbial community composition in the rhizosphere of rice cultivar IR-72 was clearly distinct from that in the phyllosphere, both in terms of composition and complexity (Figure 2, Supplementary Figure S1 and S3, Table S2). Our findings with regard to complexity and composition were consistent with those of earlier rhizosphere studies (Lu *et al.*, 2006; Kim *et al.*, 2008) and similar to those of paddy soil studies (Shrestha *et al.*, 2009). The higher complexity of the rhizosphere sample in comparison with the phyllosphere was also reflected by the lower degree of assembly of the

rhizosphere metagenome data (see Materials and methods). According to MLTreeMap analysis, Alpha-, Beta- and Deltaproteobacteria were most abundant (contributing to >10% of the bacterial community). Further abundant taxa (>5% each) included the Firmicutes, Actinobacteria, Gammaproteobacteria and the Deinococcus-Thermus phylum. Archaea were more abundant in the rhizosphere than in the phyllosphere. Archaeal rhizosphere inhabitants comprised in particular diverse methanogens (Methanobacteriales, Methanomicrobiales, Methanosarcinales and Methanocellales; Supplementary Figures S1 and S2c), which is in agreement with previous reports (Conrad, 2007). The percentage of unknown taxa was clearly higher in the rhizosphere compared with the phyllosphere; 40% of the clone sequences originated from unknown genera (Supplementary Table S2). This finding is also reflected by the higher number of rhizosphere metagenome reads that mapped to deep-branching positions in the MLTreeMap tree (Figure 2), reflecting the distant relatedness to sequences of genome-sequenced organisms.

Protein identification and assignment to bacterial and archaeal taxa

A total of 4628 different proteins were identified (Supplementary Table S3), of which 70% were annotated as bacterial and archaeal proteins. Protein identification was most successful in the phyllosphere samples (762–959 bacterial proteins per sample; Table 1). Metagenome data significantly improved protein identification: 60% of the proteins in the metagenome-sequenced sample IR-72 were identified based on metagenome sequence data and

slightly more than 50% in samples Angelica and PSB RC80 (Table 1). The higher complexity of the microbial community composition in the root samples together with the higher percentage of uncultivated genera largely affected protein identification. The number of identified proteins in the more complex rhizosphere and rhizoplane samples was much lower; only between 126 and 350 bacterial and archaeal proteins were identified. The metagenome data generated for the rhizosphere sample IR-72 increased protein identification not as strongly as observed for the phyllosphere sample, only 7–25% of the proteins in the rhizosphere and rhizoplane were identified based on the metagenome data. The phyllosphere metagenome did not substantially improve protein identification in the rhizosphere and vice versa, which reflects the distinct community composition in these compartments.

Differences between phyllosphere and rhizosphere communities were also visible at the level of assigned proteins to bacterial and archaeal taxa (Figure 2). In the phyllosphere samples, the majority of proteins (60%) matched to members within the class Alphaproteobacteria, in particular to the genera *Methylobacterium* (559 proteins) and *Rhizobium/Agrobacterium* (89 proteins). The Actinobacteria, which were present at roughly equal abundance as the Alphaproteobacteria according to MLTreeMap analysis of the metagenome data, were underrepresented in the proteome fraction of the phyllosphere. This most likely resulted from insufficient genomic information of the Actinobacteria for protein identification. Genome-sequenced strains closely related to those present in the root samples are not yet available in public databases (Supplementary Figure S2a) and thus did not contribute to protein identification; on the other hand, metagenomic sequencing was not deep enough to cover the genomic diversity in the sample to fully compensate the lack of publicly available data compared with other taxa (see also Supplementary Table S3).

Likewise as in the phyllosphere, the majority of proteins in the rhizosphere and rhizoplane samples were identified within the Alphaproteobacteria (33%), however, in these samples proteins were assigned to diverse genera, in particular to *Bradyrhizobium*, *Rhodopseudomonas*, *Azospirillum*, *Methylobacterium*, *Magnetosprillum* and *Methylosinus*. Furthermore, a substantial part of the identified proteins was assigned to genera within the Betaproteobacteria (*Dechloromonas*, *Acidovorax* and *Herbaspirillum*) and Deltaproteobacteria (*Anaeromyxobacter*, *Geobacter* and *Desulfovibrio*), consistent with the already mentioned observation that these represent dominant taxa in the rhizosphere of sample IR-72 based on metagenome reads (Figure 2) and clone library analysis (Supplementary Table S2). Proteins of proteobacteria were also identified as the most prominent group in a recent metaproteomic rice rhizosphere study (Wang et al., 2011).

Table 1 Protein identifications in the different samples

Sample	Identified proteins	Bacterial and archaeal proteins	Unknown proteins ^a	% Bacterial and archaeal proteins identified exclusively based on metagenome data
<i>Phyllosphere</i>				
IR-72	1664	959	158	61.1
Angelica	1509	938	93	51.1
PSB RC 80	1135	762	23	52.6
<i>Rhizosphere</i>				
IR-72	302	208	8	20.7
Angelica	406	350	3	21.1
<i>Rhizoplane</i>				
IR-72	440	274	9	24.8
Angelica	255	126	1	7.1
<i>Water</i>				
Field, IR-72	993	713	14	38.4
Reservoir, Angelica	1036	727	1	4.7

^aThese proteins were identified based on metagenome sequence reads with an ORF of unknown identity.

Comparison of the different metaproteome datasets and identification of phyllosphere- and rhizosphere-enriched protein families

The proteome composition of the different samples was compared using cluster analysis with multi-scale bootstrap analysis. In order to take into account functional redundancy, the proteins were analyzed according to their assignment to Pfam domains (Figure 3). A clear clustering of phyllosphere,

rhizosphere/rhizoplane and water samples was obvious based on this analysis, whereas a further differentiation of the four root samples was not evident. Therefore, these samples were combined in the following and considered as ‘root samples’.

Differences between the phyllosphere, root and water samples were analyzed further by identifying significantly enriched Pfam domains (Figure 4,

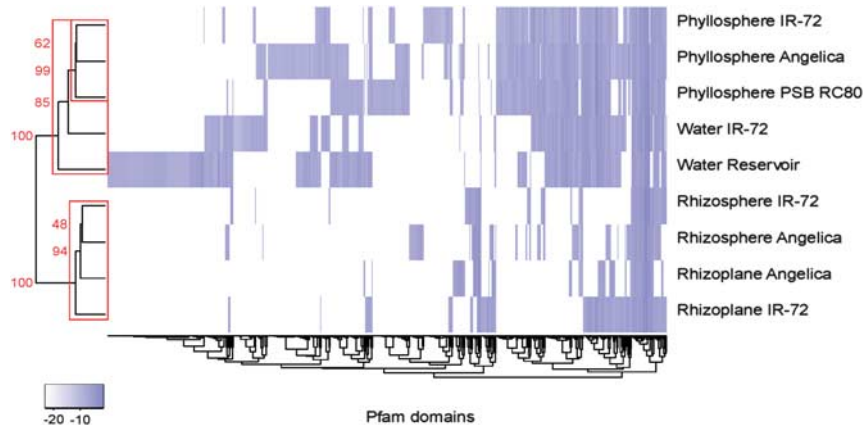


Figure 3 Comparison of the proteome composition in the different samples using hierarchical cluster analysis. The relative frequency of Pfam domains was compared across samples based on protein fractional spectral counting. The analysis was done using log-transformed data, differences between samples were calculated as Euclidean distances and samples grouped based on similarity using the Ward clustering algorithm. Significant grouping of samples is indicated based on multiscale bootstrap resampling, which results in approximately unbiased *P*-values (expressed in %). Clusters with *P*-values > 95% are highlighted by red rectangles.

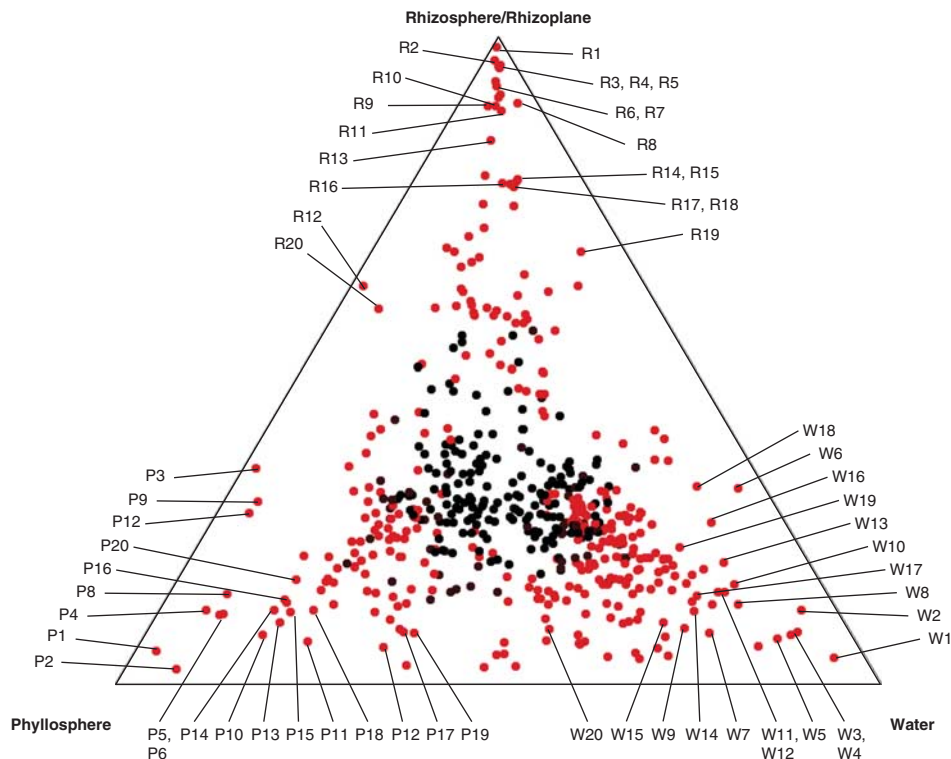


Figure 4 Proteome functions specifically enriched in the above- and below-ground parts of rice in comparison with flooding water. The position of each Pfam domain in the triangle was calculated based on fractional spectral counting. Significantly enriched Pfam domains are highlighted in red (*P*-value < 0.001) and the 20 most enriched Pfam domains per sample source are labeled. The identity of these is listed in Table 2.

Table 2). Pfams that were specific for the phyllosphere communities comprised proteins involved in substrate uptake such as porins (Pf02530, Pf00267 and Pf05736) and components of ABC transport systems (Pf00496, Pf00497, Pf02608, Pf04069, Pf00532, Pf01547 and Pf01297). Several proteins that have been reported to occur abundantly in phyllosphere bacteria before (Delmotte *et al.*, 2009) were in this study shown to be indeed enriched in bacteria residing in this particular plant compartment. These include proteins involved in stress response (Pf02566), especially proteins dealing with reactive oxygen species (Pf05443, Pf00210, Pf00141, Pf05532, Pf06628, Pf00199 and Pf00081), the fasciclin domain (Pf02469), PQQ-dependent methanol dehydrogenase as well as methanol dehydrogenase-like protein (Pf01011, Pf10527 and Pf10535) or the invasion-associated locus B-family protein (Pf06776). This latter protein was assigned to different Alphaproteobacteria, in particular to members of the genus *Methylobacterium*, where it was amongst the most abundant proteins. In addition to the above-mentioned proteins, several proteins of unknown function, both cytosolic- and membrane-associated (Pf07244, Pf01103, Pf06823, Pf04338, Pf03780 and Pf09917), were detected at higher frequency in the phyllosphere communities.

In the root samples, protein domains of enzymes involved in methanogenesis (Pf00296, Pf02240, Pf02241, Pf02783, Pf00374, Pf01993, Pf02745, Pf02249 and Pf04208) as well as methane oxidation (Pf04744, Pf02332, Pf02964 and Pf08714) and nitrogenase (Pf00142), and proteins involved in chemotaxis and motility (Pf00015, Pf00700, Pf00669 and Pf07196) were specifically enriched.

Most prominent in the water samples were Pfam domains of proteins involved in photosynthesis, CO₂ fixation (RubisCO) and the Calvin cycle (Pf00427, Pf02788, Pf00016, Pf00936, Pf00502, Pf00485, Pf00162, Pf01116, Pf02531, Pf01383, Pf03320, Pf00101, Pf00223 and Pf00421), which were assigned to Cyanobacteria, in particular *Synechococcus* and *Cyanobium*.

Overall, the observed significant differences were mostly due to protein families that were specifically assigned to distinct functional guilds, in particular in the root and water samples. In contrast to these, specific Pfam domains in the phyllosphere were more often seen in several different taxa, thus, reflecting general adaptations of the phyllosphere microbiota rather than specific metabolic capacities of distinct taxa.

One-carbon metabolism of rice-associated bacteria and archaea

The proteome analysis suggests the occurrence of diverse microbial one-carbon conversion processes in association with rice plants. In all three phyllosphere samples, enzymes involved in aerobic methylotrophy prevailed, whereby the large subunit of

methanol dehydrogenase (MxaF) (Anthony and Williams, 2003), methanol dehydrogenase-like protein XoxF (Schmidt *et al.*, 2010) and formaldehyde-activating enzyme Fae (Vorholt *et al.*, 2000) were among the most frequently detected proteins in the metaproteomes. They were assigned to the genus *Methylobacterium*, a dominant member of the bacterial phyllosphere community. This finding is in agreement with observations on other plant species and underlines the importance of one-carbon metabolism for this genus upon phyllosphere colonization (Sy *et al.*, 2005; Delmotte *et al.*, 2009; Schmidt *et al.*, 2010). Furthermore, a methanol:NDMA oxidoreductase was identified and assigned to *Amycolatopsis* (von Ophem *et al.*, 1993), suggesting that other bacteria may also benefit from plant-released methanol in the rice phyllosphere, even though they are apparently less numerous.

In the water and root samples, proteins known to be involved in methylotrophy were less prominent, but also detectable. Such enzymes assigned to *Methylobacterium* were for instance detected in the water, but, in contrast to the phyllosphere, they were not among the top hits within this genus, possibly suggesting that the methylotrophic lifestyle might be of less importance for *Methylobacterium* when residing in the flooding water. Moreover, a PQQ-dependent methanol dehydrogenase assigned to *Methylotenera* and a XoxF-like protein assigned to *Leptothrix* (Lcho_3106) were detected in the water samples. A formaldehyde-activating enzyme (Fae) assigned to *Variovorax* was seen in the root samples. Methylotrophy in these genera or closely related strains has been reported earlier (Anesti *et al.*, 2005; Nakatsu *et al.*, 2006; Kalyuzhnaya *et al.*, 2008).

Based on proteome data, a dominating one-carbon conversion process in the root samples was methanogenesis, as already evident from the specific enrichment of corresponding Pfam domains (Figure 4, Table 2). Although the diverse methanogens contributed only about 3% to the total microbial community (Figure 2), numerous proteins of these strictly anaerobic archaea were identified and dominant in the metaproteomes of the root samples. This overrepresentation is probably due to the fact that the relatively high number of genome-sequenced strains within this group of organisms enhanced protein identification. Moreover, enzymes involved in methanogenesis and in particular methyl-CoM reductase are known to be present at high abundance in methanogenic archaea (Thauer, 1998; Zhu *et al.*, 2004). The most abundant proteins were subunits of the methyl coenzyme M reductase, methylenetetrahydromethanopterin reductase, F₄₂₀-dependent methylenetetrahydromethanopterin dehydrogenase, CoM-CoB heterodisulfide reductase, tetrahydromethanopterin S-methyltransferase and coenzyme F₄₂₀ hydrogenase. The detection of both acetotrophic and hydrogenotrophic methanogens might indicate that different substrates were

Table 2 The 20 most specific Pfam domains significantly enriched in the phyllosphere, the rhizosphere/rhizoplane and the water samples

Label number	Pfam	Name of the Pfam category	Detection frequency ^a		
			Phyllosphere	Rhizosphere/ Rhizoplane	Water
<i>Phyllosphere</i>					
P1	Pf02530	Porin subfamily	+++	+	0
P2	Pf02469	Fasciclin domain	+++	0	+
P3	Pf07244	Surface-antigen variable number repeat	++++	++	0
P4	Pf01011	Pyrrroquinoline quinone enzyme repeat	++++	+++	+++
P5	Pf10535	Pyrrroquinoline quinone coenzyme, C-terminus	++++	++	++
P6	Pf10527	Pyrrroquinoline quinone coenzyme, N-terminus	++++	++	++
P7	Pf01103	Surface antigen	+++	+	0
P8	Pf06776	Invasion-associated locus B IalB protein	+++	+	+
P9	Pf00267	Gram-negative porin	++++	++	+
P10	Pf05443	ROSMUCR transcriptional regulator protein	+++	0	+
P11	Pf01389	OmpA-like transmembrane domain	+++	+	++
P12	Pf05736	OprF membrane protein	+++	0	++
P13	Pf00496	Bacterial extracellular solute binding proteins, family 5	+++	0	+
P14	Pf06823	Protein of unknown function DUF1236	++	0	0
P15	Pf04338	Protein of unknown function DUF481	++	0	+
P16	Pf10604	Polyketide cyclase dehydrase and lipid transport	++	0	0
P17	Pf00497	Bacterial extracellular solute-binding proteins, family 3	+++	+	++
P18	Pf00210	Ferritin-like domain	+++	+	++
P19	Pf00044	Glyceraldehyde 3-phosphate dehydrogenase, NAD-binding domain	+++	+	++
P20	Pf03328	Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain	+++	+	++
<i>Rhizosphere/Rhizoplane</i>					
R1	Pf00296	Luciferase-like monooxygenase	+	+++	0
R2	Pf00465	Iron-containing alcohol dehydrogenase	+	+++	0
R3	Pf02241	Methylcoenzyme M reductase, beta subunit, C-terminal domain	0	+++	0
R4	Pf02240	Methylcoenzyme M reductase, gamma subunit	0	+++	0
R5	Pf02783	Methylcoenzyme M reductase, beta subunit, N-terminal domain	0	+++	0
R6	Pf00374	Nickel-dependent hydrogenase	0	++	0
R7	Pf02745	Methylcoenzyme M reductase, alpha subunit, N-terminal domain	0	++	0
R8	Pf04744	Monooxygenase subunit B protein	0	+++	+
R9	Pf02249	Methylcoenzyme M reductase, alpha subunit, C-terminal domain	0	++	0
R10	Pf01993	Methylene-5,6,7,8-tetrahydromethanopterin dehydrogenase	0	++	0
R11	Pf00015	Methyl accepting chemotaxis protein (MCP) signaling domain	0	++	0
R12	Pf09361	Phasin protein	+++	+++	+
R13	Pf00142	4Fe4S iron sulfur cluster-binding proteins, NifH/frxC family	0	++	0
R14	Pf00215	Orotidine 5-phosphate decarboxylase HUMPS family	0	++	0
R15	Pf02310	B12-binding domain	0	++	0
R16	Pf02332	Methane/phenole/toluene hydroxylase	0	++	0
R17	Pf01314	Aldehyde ferredoxin oxidoreductase	0	++	+
R18	Pf02607	B12-binding domain	0	++	0
R19	Pf00890	FAD-binding domain	0	++	++
R20	Pf08714	Formaldehyde-activating enzyme Fae	+++	+++	+
<i>Water</i>					
W1	Pf00427	Phycobilisome linker polypeptide	0	0	+++
W2	Pf00936	BMC domain	0	+	++
W3	Pf00016	Ribulose biphosphate carboxylase, large chain, catalytic domain	0	0	++
W4	Pf02788	Ribulose biphosphate carboxylase, large chain, N-terminal domain	0	0	++
W5	Pf04966	Carbohydrate-selective porin, OprB family	+	+	++
W6	Pf00502	Phycobilisome protein	+++	+++	++++
W7	Pf07991	Acetohydroxy acid isomeroreductase, catalytic domain	+	0	++
W8	Pf01383	CpcD/allophycocyanin linker domain	0	0	++
W9	Pf01450	Acetohydroxy acid isomeroreductase, catalytic domain	++	0	++
W10	Pf00162	Phosphoglycerate kinase	+	0	++
W11	Pf00485	Phosphoribulokinase, uridine kinase family	0	0	++
W12	Pf02531	PsaD	0	0	++
W13	Pf00578	AhpC-TSA family	+	+	++
W14	Pf01116	Fructosebiphosphate aldolase, class II	+	+	++
W15	Pf02151	UvrB/uvrC motif	++	+	++
W16	Pf00395	S-layer homology domain	+	+	++
W17	Pf00582	Universal stress protein family	+	0	++
W18	Pf00463	Isocitrate lyase family	0	+	++
W19	Pf03320	Bacterial fructose-1,6-biphosphate glpX encoded	0	0	++
W20	Pf00180	Isocitrate-isopropyl-malate dehydrogenase	++	0	++

The corresponding graphic representation of all data is shown in Figure 4. Pfams involved in DNA and RNA processing (replication, transcription and translation) were excluded from the list.

^aDetection frequency of Pfam domains was determined based on fractional spectral counts as explained in the methods and is indicated by symbols here (0, not detectable; +, >0 and ≤25; ++, >25 and ≤125; +++, >125 and ≤625; and +++, >625).

reduced to methane. In addition, members of the genus *Methanosarcina* can use methanol as carbon and energy source and its conversion is initiated by methanol:corrinoid methyltransferase (Hagemeyer *et al.*, 2006), which was also detected in this study.

Moreover, enzymes involved in aerobic methane oxidation, that is, methanotrophy, were detected exclusively in the root samples (Figure 4, Table 2). These were assigned to alpha- as well as gamma-proteobacterial methanotrophs, consistent with the 16S rRNA gene clone library analysis described above. The presence of both groups in the rice rhizosphere and rhizoplane is known from several previous studies (Bodelier *et al.*, 2000; Eller and Frenzel, 2001; Horz *et al.*, 2001; Shrestha *et al.*, 2008; Wu *et al.*, 2009). Both, the soluble and particulate methane monooxygenase were detected with a roughly equal number of spectra. The soluble methane monooxygenase is found in many type II methanotrophs and some type I methanotrophs; it is known to be expressed only under low copper conditions ($<0.8 \mu\text{M}$) and has a lower affinity for methane compared with the membrane-bound enzyme (Hanson and Hanson, 1996; Hakemian and Rosenzweig, 2007).

Since the discovery of methane production by plant leaves and a few reports about the isolation of methanotrophs from plant material, it has been speculated that methanotrophs, and thus methane-oxidizing activity, might be of relevance in the plant phyllosphere (Keppler *et al.*, 2009). However, no study in which the microbial community composition in the phyllosphere has been analyzed based on cultivation-independent methods has reported the presence of known methanotrophs. As rice plants release methane that is formed by methanogenic archaea via the aerenchyma and the aerial plant parts to the atmosphere (Frenzel, 2000; Wassmann and Aulakh, 2000), the abundance of methanotrophic bacteria in the phyllosphere may be higher on these plants. Therefore, the question about the occurrence and putative role of methanotrophs in the phyllosphere was revisited in this study. There was no evidence for the presence of methane monooxygenase in the phyllosphere metaproteome, and the encoding genes were not detected in the metagenome of the rice phyllosphere microbiota generated here, while they were in the rhizosphere. Thus, known methanotrophic bacteria are not apparent as dominant players in the microbial community of the rice phyllosphere. However, methanotrophic Alpha- and Gammaproteobacteria were detectable in the phyllosphere samples using PCR targeting a subunit of the membrane-bound methane monooxygenase (*pmoA*) and by cultivation (data not shown). Nevertheless, their contribution to methane oxidation is most likely insignificant compared with the activity of the methanotrophs residing in association with the rice root as their cell number in the phyllosphere is much lower and they are exposed to methane mixing ratios <1000 ppmv

(Bosse and Frenzel, 1997), at which several methanotrophic genera become inactive (Knief and Dunfield, 2005). Only in association with the lower part of rice plant stems, where methane concentrations are higher, the presence and activity of methanotrophic bacteria has been reported (Bosse and Frenzel, 1997; Watanabe *et al.*, 1997). Taken these findings together, it appears unlikely that the well-known methanotrophs that can be isolated from the phyllosphere of plants have a major role in the oxidation of plant-released methane.

Potential for nitrogen fixation in rice-associated bacteria

The question of dinitrogen fixation by endophytic bacteria and possible beneficial effects for rice plants has already been addressed in several studies, as it is assumed to hold potential for the improvement of rice cultivation and grain yield. Transcripts of *nifH* genes of diverse diazotrophs were detected in roots and stems of cultivars (Elbeltagy and Ando, 2008; Prakamhang *et al.*, 2009), and $^{15}\text{N}_2$ fixation in *Herbaspirillum*-inoculated plant seedlings of old rice cultivars could be demonstrated (Elbeltagy *et al.*, 2001). However, the fixation rate under the tested conditions was rather low and it remained unclear whether labeled nitrogen was transferred to the plant material. According to the metagenomic analysis in the present study, the potential for nitrogen fixation in a field grown rice variety (IR-72) was substantial. Genes encoding dinitrogen reductase (*nifH*) and dinitrogenase (*nifD* and *nifK*) were detected in the rhizosphere as well as in the phyllosphere metagenome. Closer inspection of *nifH* diversity revealed that the gene was present in different taxa in the phyllosphere compared with the rhizosphere communities (Figure 5). In the phyllosphere, the most frequently detected *nifH* sequence types were those of *Azorhizobium* and *Rhodopseudomonas*. In the rhizosphere, *nifH* was found across diverse taxa, including, for instance, *Rhizobium*, *Methylococcus*, *Dechloromonas*, *Anaeromyxobacter*, *Syntrophobacter* and some methanogenic archaea.

Remarkably, in contrast to the large diversity of nitrogenase genes in the phyllosphere and rhizosphere at the metagenome level, the detection of the protein dinitrogen reductase was restricted to root-associated bacteria (Figure 5), whereby the identified NifH protein subunits were most closely related to the proteins from *Bradyrhizobium*, *Magnetospirillum* and *Azospirillum*, respectively (Supplementary Table S3). Few peptides of dinitrogenase reductase were identified in the rhizoplane; the oxygen-sensitive enzyme was predominantly seen in the rhizosphere, probably favored by the micro-oxic to anoxic conditions. To what extent the plant may profit from microbial nitrogen fixation under these circumstances remains to be shown. It has been stated that endophytic bacteria may transfer

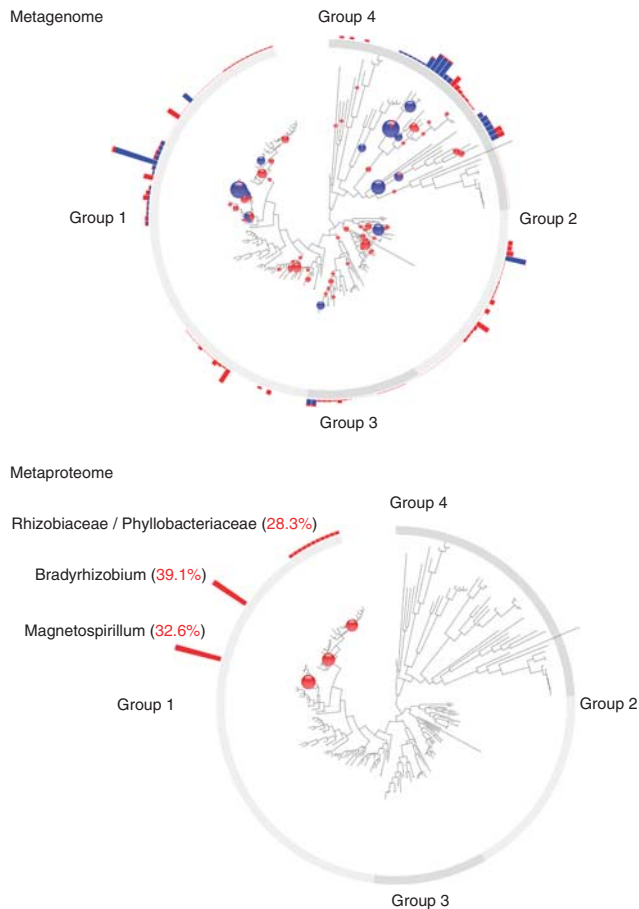


Figure 5 Diversity and frequency of *nifH* genes in the metagenome of sample IR-72 and of NifH in the metaproteome of all phyllosphere (blue) and root (red) samples. A high diversity of *nifH* was observed at metagenomic level, but only a couple of genes were shown to be expressed specifically in the root samples. The relative number of metagenome reads and protein spectra, respectively, that matched to a strain is indicated as colored stacked bar. Assignments to higher taxonomic ranks are equally fractionated across all strains representing the respective taxon. The figures were constructed with MLTreeMap. Higher-resolution images are provided as supplementary material (Supplementary Figure S4).

nitrogen more efficiently to the host plant compared with rhizosphere bacteria because of their closer association with the plant (Beattie, 2006). Moreover, despite the detection of Nif proteins, care has to be taken whether nitrogen fixation is an active process in plant-associated bacteria as they may not necessarily produce an active enzyme. These rice-associated bacteria may nevertheless exert a positive effect on plant growth, according to the hypothesis that phytohormone production rather than the nitrogen-fixing activity is responsible for the frequently described growth-stimulating effect of diazotrophic bacteria (Barea *et al.*, 2005).

Besides the utilization of dinitrogen by some microbial taxa, the rice-associated microorganisms were prepared to use diverse other sources of nitrogen. Proteins involved in the assimilation of ammonium and ABC transport systems for the

import of peptides were detected in the phyllosphere. Moreover, subunits of ABC-dependent transport systems for different amino acids were seen in the phyllosphere and rhizosphere. Amino acids are known to occur on leaf surfaces and in root exudates (Derridj, 1996; Hao *et al.*, 2010).

Conclusion

The present study extends knowledge about the physiology and adaptations of the plant-associated microbiota and provides for the first time a reference set of genomic and proteomic data of microbial communities associated with the above- and below-ground parts of rice. Compared with previously studied bacterial phyllosphere communities from dicotyledonous plants grown under different climatic and geographic conditions, the microbial proteome in the phyllosphere of rice was remarkably similar. The comparison of leaf- and root-associated communities allowed to strengthen previous hypotheses on compartment-enriched proteins of plant-associated bacteria, such as proteins involved in stress response, methanol utilization, the fasciclin protein or the invasion-associated locus B protein in the phyllosphere. Analysis of the metabolic make-up of rice-associated microbial communities underlines the importance of one-carbon compound cycling in the rhizosphere/rhizoplane as well as in the phyllosphere. In terms of nitrogen metabolism, the plant-associated microbiota was found to exhibit dinitrogen fixation potential; however, gene expression was exclusively found associated with the rhizosphere. This finding demonstrates the advantage of functional genomic approaches compared with metagenomic alone to infer the *in situ* physiology of microbial communities.

Acknowledgements

We thank Agnes Padre and Enrique Montserrat for support at the International Rice Research Institute during sample collection. We thank FGCZ for access to the proteomics facility and support. The study was supported by ETH Zurich.

References

- Anesti V, McDonald IR, Ramaswamy M, Wade WG, Kelly DP, Wood AP. (2005). Isolation and molecular detection of methylotrophic bacteria occurring in the human mouth. *Environ Microbiol* 7: 1227–1238.
- Anthony C, Williams P. (2003). The structure and mechanism of methanol dehydrogenase. *BBA - Proteins Proteom* 1647: 18–23.
- Barea JM, Pozo MJ, Azcon R, Azcon-Aguilar C. (2005). Microbial co-operation in the rhizosphere. *J Exp Bot* 56: 1761–1778.

- Bastida F, Moreno JL, Nicolas C, Hernandez T, Garcia C. (2009). Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *Eur J Soil Sci* **60**: 845–859.
- Beattie GA. (2006). Plant-associated bacteria: survey, molecular phylogeny, genomics and recent advances.. In: Gnanamanickam SS (ed). *Plant-associated bacteria*. Springer: Dordrecht, pp 1–56.
- Bodelier PL, Roslev P, Henckel T, Frenzel P. (2000). Stimulation by ammonium-based fertilizers of methane oxidation in soil around rice roots. *Nature* **403**: 421–424.
- Bosse U, Frenzel P. (1997). Activity and distribution of methane-oxidizing bacteria in flooded rice soil microcosms and in rice plants (*Oryza sativa*). *Appl Environ Microbiol* **63**: 1199–1207.
- Brune A, Frenzel P, Cypionka H. (2000). Life at the oxic-anoxic interface: microbial activities and adaptations. *FEMS Microbiol Rev* **24**: 691–710.
- Chinnadurai C, Balachandar D, Sundaram SP. (2009). Characterization of 1-aminocyclopropane-1-carboxylate deaminase producing methylobacteria from phyllosphere of rice and their role in ethylene regulation. *World J Microb Biot* **25**: 1403–1411.
- Conrad R. (2007). Microbial ecology of methanogens and methanotrophs. *Adv Agron* **96**: 1–63.
- De Costa DM, Samarasinghe SST, Dias HRD, Dissanayake DMN. (2008). Control of rice sheath blight by phyllosphere epiphytic microbial antagonists. *Phytoparasitica* **36**: 52–65.
- Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R et al. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc Natl Acad Sci USA* **106**: 16428–16433.
- Derridj S. (1996). Nutrients on the leaf surface. In: Morris CE, Nicot PC, Nguyen-The C (eds). *Aerial plant surface microbiology*. Plenum Press: New York, pp 25–42.
- Elbeltagy A, Ando Y. (2008). Expression of nitrogenase gene (NIFH) in roots and stems of rice, *Oryza sativa*, by endophytic nitrogen-fixing communities. *Afr J Biotechnol* **7**: 1950–1957.
- Elbeltagy A, Nishioka K, Sato T, Suzuki H, Ye B, Hamada T et al. (2001). Endophytic colonization and in planta nitrogen fixation by a *Herbaspirillum* sp. isolated from wild rice species. *Appl Environ Microbiol* **67**: 5285–5293.
- Elbeltagy A, Nishioka K, Suzuki H, Sato T, Sato YI, Morisaki H et al. (2000). Isolation and characterization of endophytic bacteria from wild and traditionally cultivated rice varieties. *Soil Sci Plant Nutr* **46**: 617–629.
- Eller G, Frenzel P. (2001). Changes in activity and community structure of methane-oxidizing bacteria over the growth period of rice. *Appl Environ Microbiol* **67**: 2395–2403.
- Frenzel P. (2000). Plant-associated methane oxidation in rice fields and wetlands. *Adv Microb Ecol* **16**: 85–114.
- Fürnkranz M, Wanek W, Richter A, Abell G, Rasche F, Sessitsch A. (2008). Nitrogen fixation by phyllosphere bacteria associated with higher plants and their colonizing epiphytes of a tropical lowland rainforest of Costa Rica. *ISME J* **2**: 561–570.
- Hagemeyer CH, Krüer M, Thauer RK, Warkentin E, Ermler U. (2006). Insight into the mechanism of biological methanol activation based on the crystal structure of methanol:cobalamin methyltransferase complex MtaBC from *Methanosarcina barkeri*. *Proc Natl Acad Sci USA* **103**: 18917–18922.
- Hakemian AS, Rosenzweig AC. (2007). The biochemistry of methane oxidation. *Annu Rev Biochem* **76**: 223–241.
- Hanson RS, Hanson TE. (1996). Methanotrophic bacteria. *Microbiol Rev* **60**: 439–471.
- Hao WY, Ren LX, Ran W, Shen QR. (2010). Allelopathic effects of root exudates from watermelon and rice plants on *Fusarium oxysporum* f.sp. *niveum*. *Plant Soil* **336**: 485–497.
- Horz HP, Yimga MT, Liesack W. (2001). Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of *pmoA*, *mmoX*, *mxoF*, and 16S rRNA and ribosomal DNA, including *pmoA*-based terminal restriction fragment length polymorphism profiling. *Appl Environ Microbiol* **67**: 4177–4185.
- Kalyuzhnaya MG, Hristova KR, Lidstrom ME, Chistoserdova L. (2008). Characterization of a novel methanol dehydrogenase in representatives of *Burkholderiales*: implications for environmental detection of methylo-trophy and evidence for convergent evolution. *J Bacteriol* **190**: 3817–3823.
- Keppler F, Boros M, Frankenberg C, Lelieveld J, McLeod A, Pirttilä AM et al. (2009). Methane formation in aerobic environments. *Environ Chem* **6**: 459–465.
- Kim MC, Ahn JH, Shin HC, Kim T, Ryu TH, Kim DH et al. (2008). Molecular analysis of bacterial community structures in paddy soils for environmental risk assessment with two varieties of genetically modified rice, Iksan 483 and Milyang 204. *J Microbiol Biotechnol* **18**: 207–218.
- Knief C, Dunfield PF. (2005). Response and adaptation of different methanotrophic bacteria to low methane mixing ratios. *Environ Microbiol* **7**: 1307–1317.
- Knief C, Delmotte N, Vorholt JA. (2011). Bacterial adaptation to life in association with plants—a proteomics perspective from culture to *in situ* conditions. *Proteomics* **11**: 3086–3105.
- Knief C, Vanitchung S, Harvey NW, Conrad R, Dunfield PF, Chidthaisong A. (2005). Diversity of methanotrophic bacteria in tropical upland soils under different land uses. *Appl Environ Microbiol* **71**: 3826–3831.
- Kowalchuk GA, Yergeau E, Leveau JHJ, Sessitsch A, Bailey M. (2010). Plant-associated microbial communities. In: Liu W-T, Jansson JK (eds). *Environmental molecular microbiology*. Caister Academic Press, pp 131–148.
- Krüger M, Frenzel P, Conrad R. (2001). Microbial processes influencing methane emission from rice fields. *Global Change Biol* **7**: 49–63.
- Ladha JK, Reddy PM. (2000). *The quest for nitrogen fixation in rice*. International Rice Research Institute: Los Baños, Philippines.
- Liesack W, Schnell S, Revsbech NP. (2000). Microbiology of flooded rice paddies. *FEMS Microbiol Rev* **24**: 625–645.
- Lu Y, Rosencrantz D, Liesack W, Conrad R. (2006). Structure and activity of bacterial community inhabiting rice roots and the rhizosphere. *Environ Microbiol* **8**: 1351–1360.
- Lueders T, Friedrich MW. (2002). Effects of amendment with ferrihydrite and gypsum on the structure and activity of methanogenic populations in rice field soil. *Appl Environ Microbiol* **68**: 2484–2494.

- Madhaiyan M, Poonguzhali S, Kwon SW, Sa TM. (2009). *Methylobacterium phyllosphaerae* sp. nov., a pink-pigmented, facultative methylotroph from the phyllosphere of rice. *Int J Syst Evol Microbiol* **59**: 22–27.
- Madhaiyan M, Poonguzhali S, Sa T. (2007). Influence of plant species and environmental conditions on epiphytic and endophytic pink-pigmented facultative methylotrophic bacterial populations associated with field-grown rice cultivars. *J Microbiol Biotechnol* **17**: 1645–1654.
- Madhaiyan M, Poonguzhali S, Senthilkumar M, Seshadri S, Chung HY, Yang JC et al. (2004). Growth promotion and induction of systemic resistance in rice cultivar Co-47 (*Oryza sativa* L.) by *Methylobacterium* spp. *Bot Bull Acad Sin* **45**: 315–324.
- Maliti CM, Basile DV, Corpe WA. (2005). Effects of *Methylobacterium* spp. strains on rice *Oryza sativa* L. callus induction, plantlet regeneration, and seedlings growth *in vitro*. *J Torrey Bot Soc* **132**: 355–367.
- Mano H, Tanaka F, Nakamura C, Kaga H, Morisaki H. (2007). Culturable endophytic bacterial flora of the maturing leaves and roots of rice plants (*Oryza sativa*) cultivated in a paddy field. *Microbes Environ* **22**: 175–185.
- Nakatsu CH, Hristova K, Hanada S, Meng XY, Hanson JR, Scow KM et al. (2006). *Methylibium petroleiphilum* gen. nov., sp. nov., a novel methyl tert-butyl ether-degrading methylotroph of the Betaproteobacteria. *Int J Syst Evol Microbiol* **56**: 983–989.
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**: 4646–4658.
- Pedraza RO, Bellone CH, de Bellone S, Sorte PMB, Teixeira KRD. (2009). *Azospirillum* inoculation and nitrogen fertilization effect on grain yield and on the diversity of endophytic bacteria in the phyllosphere of rice rainfed crop. *Eur J Soil Biol* **45**: 36–43.
- Prakamhang J, Minamisawa K, Teamtaisong K, Boonkerd N, Teaumroong N. (2009). The communities of endophytic diazotrophic bacteria in cultivated rice (*Oryza sativa* L.). *Appl Soil Ecol* **42**: 141–149.
- Prasanna R, Nain L, Pandey AK, Nayak S. (2010). Exploring the ecological significance of microbial diversity and networking in the rice ecosystem. In: Dion P (ed). *Soil biology and agriculture in the tropics*. Springer Berlin Heidelberg, pp 139–161.
- Schmidt S, Christen P, Kiefer P, Vorholt JA. (2010). Functional investigation of methanol dehydrogenase-like protein XoxF in *Methylobacterium extorquens* AM1. *Microbiology* **156**: 2575–2586.
- Shrestha M, Abraham WR, 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. *Environ Microbiol* **10**: 400–412.
- Shrestha PM, Kube M, Reinhardt R, Liesack W. (2009). Transcriptional activity of paddy soil bacterial communities. *Environ Microbiol* **11**: 960–970.
- Stark M, Berger SA, Stamatakis A, von Mering C. (2010). MLTreeMap - accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* **11**: 461.
- Suzuki R, Shimodaira H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540–1542.
- Sy A, Timmers AC, Knief C, Vorholt JA. (2005). Methylotrophic metabolism is advantageous for *Methylobacterium extorquens* during colonization of *Medicago truncatula* under competitive conditions. *Appl Environ Microbiol* **71**: 7245–7252.
- Thauer RK. (1998). Biochemistry of methanogenesis: a tribute to Marjory Stephenson. *Microbiology* **144**: 2377–2406.
- von Ophem PW, van Beeumen J, Duine JA. (1993). Nicotinoprotein [NAD(P)-containing] alcohol/aldehyde oxidoreductases. Purification and characterization of a novel type from *Amycolatopsis methanolica*. *Eur J Biochem* **212**: 819–826.
- Vorholt JA, Marx CJ, Lidstrom ME, Thauer RK. (2000). Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J Bacteriol* **182**: 6645–6650.
- Wang HB, Zhang ZX, Li H, He HB, Fang CX, Zhang AJ et al. (2011). Characterization of metaproteomics in crop rhizospheric soil. *J Proteome Res* **10**: 932–940.
- Wassmann R, Aulakh MS. (2000). The role of rice plants in regulating mechanisms of methane emissions. *Biol Fertil Soils* **31**: 20–29.
- Watanabe I, Hashimoto T, Shimoyama A. (1997). Methane-oxidizing activities and methanotrophic populations associated with wetland rice plants. *Biol Fertil Soils* **24**: 261–265.
- Wu LK, Wang HB, Zhang ZX, Lin R, Zhang ZY, Lin WX. (2011). Comparative metaproteomic analysis on consecutively *Rehmannia glutinosa*-monocultured rhizosphere soil. *Plos One* **6**: e20611.
- Wu LQ, Ma K, Lu YH. (2009). Rice roots select for type I methanotrophs in rice field soil. *Syst Appl Microbiol* **32**: 421–428.
- Yang JH, Liu HX, Zhu GM, Pan YL, Xu LP, Guo JH. (2008). Diversity analysis of antagonists from rice-associated bacteria and their application in biocontrol of rice diseases. *J Appl Microbiol* **104**: 91–104.
- Zhu W, Reich CI, Olsen GJ, Giometti CS, Yates 3rd JR. (2004). Shotgun proteomics of *Methanococcus jannaschii* and insights into methanogenesis. *J Proteome Res* **3**: 538–548.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)