Environmental Microbiology (2014) **16**(1), 72–83 doi:10.1111/1462-2920.12190

Macroecology of methane-oxidizing bacteria: the β-diversity of *pmoA* genotypes in tropical and subtropical rice paddies

Claudia Lüke,[1,2*] Peter Frenzel,[1] Adrian Ho,[1,3] Dian Fiantis,[4] Peter Schad,[5] Bellinda Schneider,[1] Lorenz Schwark[6,7] and Sri Rahayu Utami[8]

[1]*Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str.10, Marburg D-35043, Germany.*
[2]*Radboud University Nijmegen, Heyendaalsweg 135, Nijmegen 6525 AJ, The Netherlands.*
[3]*Laboratory for Microbial Ecology and Technology (LabMET), Faculty of Bioscience Engineering, Coupure Links 653, Ghent 9000, Belgium.*
[4]*Department of Soil Science, Faculty of Agriculture, Andalas University, Kampus Unand Limau Manis, Padang 25163, Indonesia.*
[5]*Department Ecology and Ecosystem Management, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Lehrstuhl für Bodenkunde, Freising-Weihenstephan D-85350, Germany.*
[6]*Institute of Geosciences, University Kiel, Ludewig-Meyn-Straße 10, Kiel 24118, Germany.*
[7]*WA-OIGC, Curtin University, Perth, WA 6845, Australia.*
[8]*Department of Soil Science, Faculty of Agriculture, Brawijaya University, Jalan Veteran, Malang 65145, East Java, Indonesia.*

Summary

Studies addressing microbial biogeography have increased during the past decade, but research on microbial distribution patterns is still in its infancies, and many aspects are only poorly understood. Here, we compared the methanotroph community in paddy soils sampled in Indonesia, Vietnam, China and Italy, focusing on the distance–decay relationship. We used the *pmoA* gene as marker for methanotroph diversity in terminal restriction fragment length polymorphism, microarray and pyrosequencing approaches. We could observe a significant increase of β-diversity with geographical distance across continents (12 000 km). Measured environmental parameters explained only a small amount of data variation, and we found no evidence for dispersal limitation. Thus, we propose historical contingencies being responsible for the observed patterns. Furthermore, we performed an in-depth analysis of type II methanotroph *pmoA* distribution at the sequence level. We used ordination analysis to project sequence dissimilarities into a three-dimensional space (multidimensional scaling). The ordination suggests that type II methanotrophs in paddy fields can be divided into five major groups. However, these groups were found to be distributed in all soils independent of the geographic origin. By including tropical field sites (Indonesia and Vietnam) into the analysis, we further observed the first paddy fields harbouring a methanotroph community depleted in type II methanotrophs.

Introduction

Understanding the mechanisms creating and sustaining biodiversity is important to predict ecosystem responses to future environmental changes. For macroorganisms, biogeographical patterns have been studied extensively, showing the decrease of community similarity with geographical distance as a general phenomenon (Nekola and White, 1999). Such relationship could be explained by different environmental conditions across space, thus niche-based processes. However, it can be also the result of dispersal limitations (neutrality, Hubbell, 2001), or a combination of both processes (Soininen, 2012). Microbes are different from plants and animals in many ways. They are unicellular and small in size, thus their dispersal might be not restricted by geographical barriers. Hence, 'everything [might be] everywhere – the environment selects' (Baas Becking, cited after De Wit and Bouvier, 2006). However, microbial biogeography has gone through advances during the past years indicating that, in contrast to first hypotheses, microbes do display biogeographical patterns (e.g. Whitaker *et al.*, 2003; Martiny *et al.*, 2006). Nevertheless, research is still in its infancies.

In contrast to the entire microbial community, a well-defined subpopulation provides less complexity, allowing

Received 31 January, 2013; revised 21 May, 2013; accepted 10 June, 2013. *For correspondence. E-mail clueke@science.ru.nl; frenzel@mpi-marburg.mpg.de; Tel. (+49) 6421 178 803/820; Fax (+49) 6421 178809.

ecological patterns to be more easily extracted. Aerobic methane-oxidizing bacteria (methanotrophs) represent a suitable model community for studying microbial biogeography. They can be linked to a defined ecosystem function, and their diversity has been studied extensively. In wetland rice fields, they have been tracked by different fingerprinting techniques as well as massively parallel pyrosequencing (Horz *et al*., 2001; Lüke and Frenzel, 2011; Ho *et al*., 2011b). Aerobic methanotrophs can be found within the bacterial phyla of *Proteobacteria* and *Verrucomicrobia*. However, the verrucomicrobial methanotrophs described so far seem to be restricted to extreme environments (Op den Camp *et al*., 2009). The proteobacterial methanotrophs can be divided into two major groups belonging to the families of *Methylococcaceae* (type I), *Methylocystaceae* and *Beijerinckiaceae* (both type II) (Semrau *et al*., 2010).

Methanotroph populations were investigated in paddy fields worldwide (e.g. Horz *et al*., 2001; Ferrando and Tarlera, 2009; Ho *et al*., 2011a); nevertheless, the geographic distribution of respective study sites is biased. So far, most knowledge of community composition and ecology was derived from work on Italian rice field soils. Here, a diverse methanotroph population dominated by the *Methylocystis/Methylosinus* group (type II) could be detected using various molecular techniques. This group was found to be well distributed through time and space in paddy field, whereas type I methanotrophs were shown to be more abundant in the rhizosphere (e.g. Horz *et al*., 2001). A similar high proportion of type II methanotrophs could be identified in Californian paddy fields (Macalady *et al*., 2002) and rice fields located in Uruguay (Ferrando and Tarlera, 2009).

Italy is the main rice producer in Europe, but at the global scale, it only plays a minor role. About 90% of the world's rice is cultivated in Asia where China and India are the top producing countries (FAO statistics, 2010; http://faostat.fao.org/). Here, field studies on methanotroph diversity increased during the past years, revealing again a high abundance of type II methanotrophs (Zheng *et al*., 2008; Vishwakarma *et al*., 2009; Ma and Lu, 2011; Ho *et al*., 2011b). In Japanese rice fields, type II were found again to represent the dominant fraction (Jia *et al*., 2007). Thus, a high proportion of the *Methylocystis/Methylosinus* group might be a common feature of wetland rice methanotrophs.

However, nearly all paddy fields studied intensively are located within subtropical climates and hardly anything is known on methanotroph diversity in tropical paddy fields, even though they contribute a substantial part to global rice production.

In this work, we compared the methanotroph population in tropical paddy fields distributed through the two main islands of Indonesia (Java and Sumatra), and from Southern Vietnam (Tien Giang province) to wetland rice fields located in the subtropical climate: the coastal area south of the Yangtze River Delta (Cixi, Zhejiang province, China) and the lowlands of the rivers Po and Sesia (Vercelli province, Italy). We used the *pmoA* gene as marker for methanotroph diversity in terminal restriction fragment length polymorphism (T-RFLP), microarray and pyrosequencing approaches. The *pmoA* gene encoding a subunit of the particulate methane monooxygenase has been shown to be an excellent marker for methanotroph diversity studies (McDonald *et al*., 2008). This key enzyme of the methane oxidation pathway is present in all known methanotrophs except the genera *Methylocella* and *Methyloferula* (Dedysh *et al*., 2000; Vorobev *et al*., 2011). However, these methanotrophs seem to play a minor role – if any – in paddy soils (Reim *et al*., 2012).

We focused on the distance–decay relationship of methanotrophs and combined ordination analysis of sequence distances with phylogenetic approaches to unravel the diversity of the dominant, but highly similar *pmoA* genotypes of type II methanotrophs.

Results

Methanotroph diversity revealed by microarray and T-RFLP analysis

pmoA microarray analysis revealed a diverse, yet similar core inventory of methanotrophs in all paddy soils (Figs. 1 and S1). Next to *Methylocystis* and *Methylosinus*-like type II methanotrophs, diverse uncultured type Ib methanotrophs were found in all soils, in particular, two lineages covered by the microarray probes 501-375, fw1-641, fw1-286, LW21-391 (Fig. S1). These lineages comprise sequences retrieved from freshwater environments such as lakes and paddy fields. Accordingly, they were referred to as freshwater lineage-1 and lineage-2 (Lüke and Frenzel, 2011). In addition, the Japanese Rice Cluster 3 (JRC-3; probes JHTY2-562, JHTY2-578) was found to be highly abundant in several Indonesian soils. Within type Ia, *Methylosarcina*-like methanotrophs were most prominent (probes Mmb562, Mmb304). Furthermore, the *Crenothrix*-related group (probes Nit_rel351, Nit_rel470) was detected in high abundance in the Chinese and Italian soils. This group showed no hybridization signal for the Indonesian and Vietnamese soils. Sequences forming this lineage were retrieved from various habitats including high methane environments such as paddy fields (Jia *et al*., 2007; Lüke and Frenzel, 2011) and freshwater lakes (Kim *et al*., 2008), but also from soils consuming atmospheric methane (e.g. Knief *et al*., 2006). As no phylogenetically related isolates are available, the substrate of the respective monooxygenase is unknown. Here, we refer to all lineages with unknown substrate specificity as *pxmA*. This usage is an extension of the original definition by

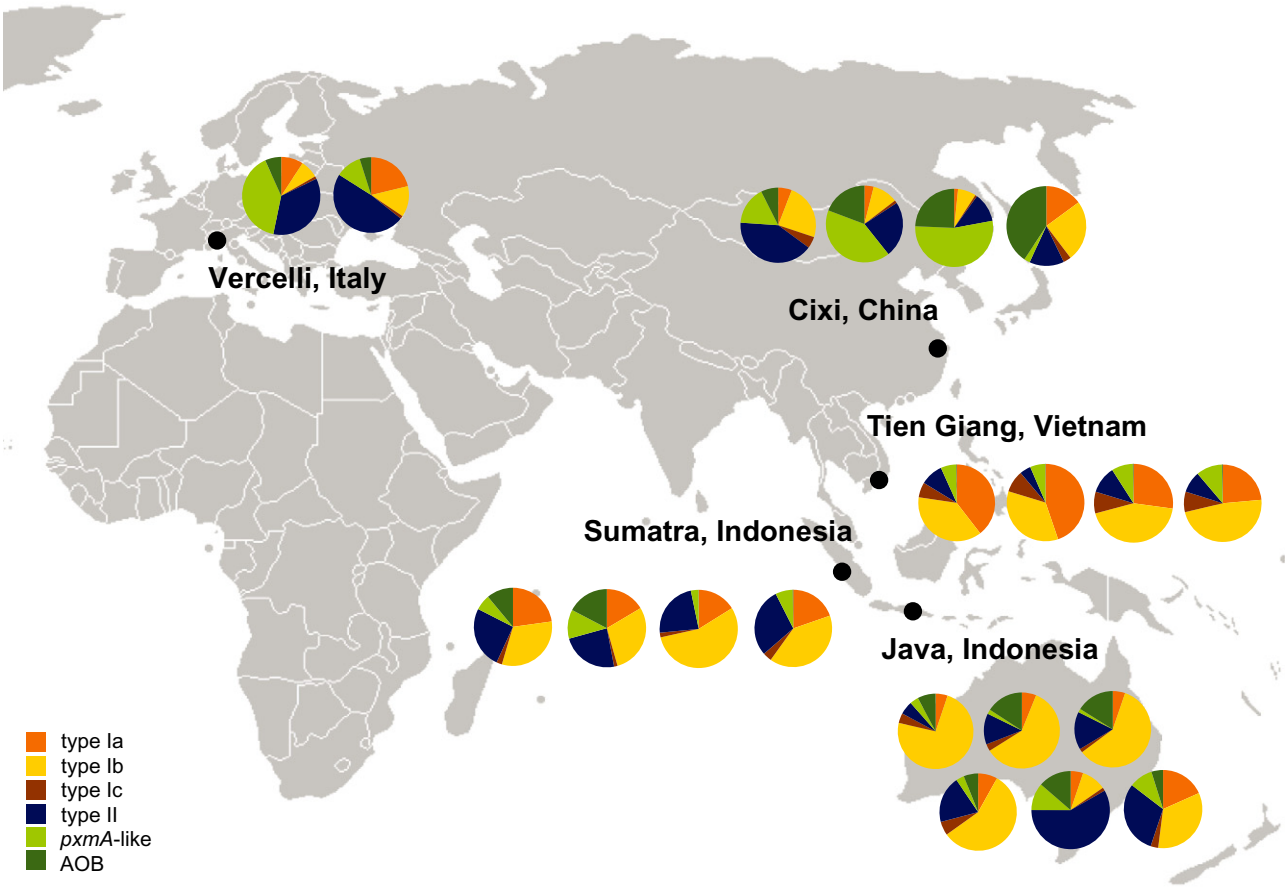

Fig. 1. α-Diversity of *pmoA* genotypes in paddy soils distributed over Europe and Asia. The pie charts are based on standardized microarray data outlined in Fig. S1. The individual sample sites from left to right are Italy: Vercelli-1, Vercelli-2; Indonesia (Java): Sumbermujur-1, Sumbermujur-2, Simo-1, Simo-2, Simo-3, Padas; Indonesia (Sumatra): Suntiang, Lukok, Piladang, Payobasuang; China: paddy 50a, paddy 100a, paddy 500a, paddy 1000a, paddy 2000a; Vietnam: My Tho-1, My Tho-2, My Tho-3, My Tho-4. Sample sites are described in Table S1.

Tavormina and colleagues (2011). They introduced the term *pxmA* for the distinct group encoding a second isoenzyme of type I methanotrophs. All these sequences are clustering phylogenetically in an intermediate position between *pmoA* and the homologous *amoA* gene encoding a subunit of the ammonia monooxygenase.

Remarkably, no upland soil clusters [e.g. upland soil cluster (USC)-α or USC-γ] assumed to be involved in oxidation of methane at atmospheric concentrations were detected in any of the paddy soils. However, a high fraction of USC-α was found in a bamboo forest located adjacent to the Sumatra paddy fields (Fig. S1).

Even though we could detect similar *pmoA* lineages in all of the studied paddy soils, each soil possesses its individual population structure, caused mainly by the relative contribution of the different lineages. The Italian and the Chinese soils are dominated by type II methanotrophs, ammonia oxidizers or *pxmA* lineages. Many of the soils located in tropical climate harbour a large fraction of type Ib methanotrophs (Fig. 1). Some of the tropical soils, in particular soils from Vietnam and Java, possessed a highly reduced fraction of type II methanotrophs. For the Indonesian samples, we additionally performed T-RFLP analysis of *pmoA* polymerase chain reaction (PCR) products amplified using the mb661 (Costello and Lidstrom, 1999) and A682r (Holmes *et al.*, 1995) reverse primers. Cluster analysis clearly separated the soils into two groups: one group characterized by type Ib T-RFs and a very low fraction of type II methanotrophs, and the other group dominated by type II T-RFs (Fig. S2).

Statistical analysis [non-metric multidimensional scaling (NMDS)] of the *pmoA* microarray data revealed a clear grouping of sample sites according to the geographical origin of the soils (Fig. 2A). A constrained correspondence analysis (CCA) confirmed the significance of the clustering ($P = 0.005$, the 'country' explained 53.1% of total inertia; data not shown). To characterize the relationship between community similarity and geographical distance, distance–decay curves were constructed (Fig. 2B). The

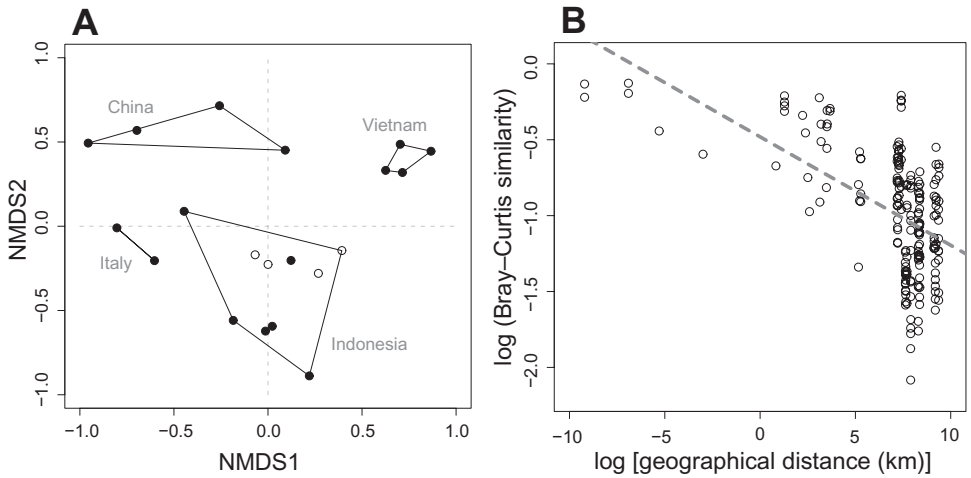

Fig. 2. β-Diversity of *pmoA* genotypes in paddy soils distributed over Europe and Asia. Filled circles within the Indonesian sites correspond to field sites in Java, unfilled circles represent Sumatra.
A. Non-metric multidimensional scaling (NMDS) analysis calculated from standardized microarray data (Stress = 0.14). An overview of the standardized microarray data for the individual soils is shown in Fig. S1.
B. Distance–decay plot of standardized microarray data. The grey line denotes the least squares linear regression across the different paddy soils. The slope is significantly less than zero ($P < 0.001$).

graph revealed a significant ($P < 0.001$) negative correlation indicating the decay of community similarity with increasing geographical distance and a high β-diversity across the sites. Indonesia and Southern Vietnam are located within the tropics whereas Italy and China are influenced by a subtropical climate. To evaluate the climatic influence on the observed pattern, we used the geographical latitude as a proxy for the different climate zones in a CCA (Fig. 3A). The latitude as constraint explained about 18% of the total inertia ($P = 0.005$). One factor that might have direct influence on the bacterial communities is the number of rice harvests per year, as each growing period implies an alternating cycle of soil flooding and drying and thereby, changing the redox states and substrate availability. As the number of rice harvests per year is dependent on the mean annual temperature, it partially correlates with the local climate. To focus on the effect of the rice cycles, we removed the background caused by the latitude (climate) from the ordination before constraining (Fig. 3B). The numbers of rice cycles showed a significant effect ($P = 0.005$) explaining 13.8% of the total inertia. In addition to the rice cycles, we could furthermore observe an influence ($P = 0.005$) of the soil pH on the methanotroph communities (Fig. 3B).

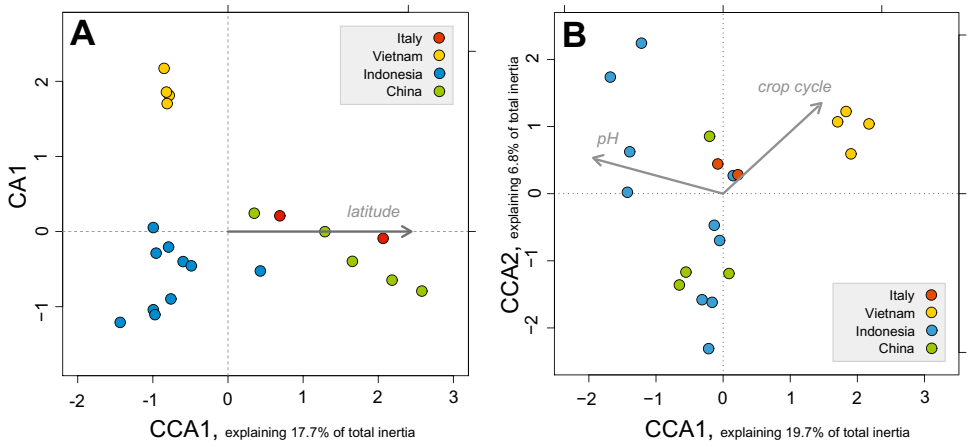

Fig. 3. Constraint correspondence analysis (CCA) calculated from standardized microarray data.
A. Geographical latitude (Table S1), as proxy for the different climate zones, was used as constraint.
B. The pH value and number of rice harvest per year (crop cycle, Table S1) were used as constraints. Here, the effect of latitude was removed by conditioning (Oksanen, 2012).

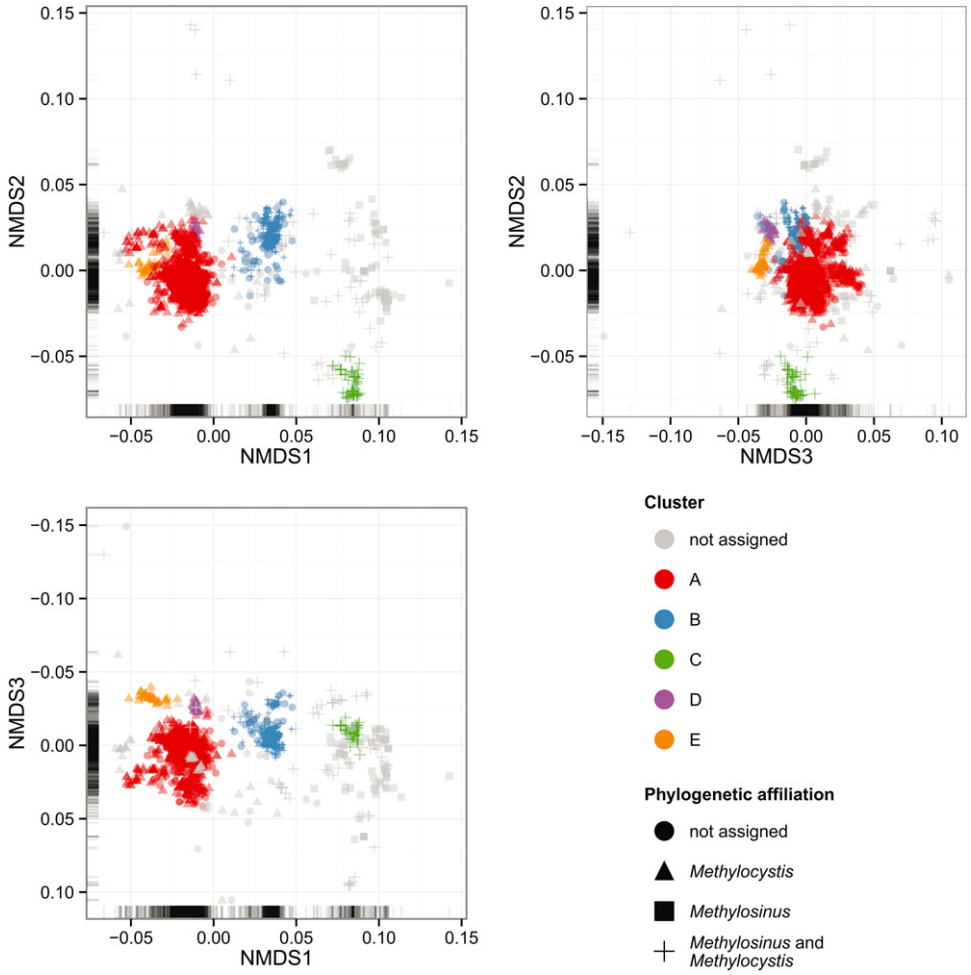

Fig. 4. Non-metric multidimensional scaling (NMDS) with three axes based on Jukes-Cantor corrected *pmoA* nucleotide sequence distances. A random subset of 600 sequences per sample was used. The unidimensional density distribution along the respective axes is indicated by rugs. Sequences were obtained using amplicon pyrosequencing from two Chinese soils (50a and 2000a), one Italian soil (Vercelli) and one Indonesian soil (Piladang, Sumatra) (Table S1). Only high-quality sequences clustering within the *Methylosinus*/*Methylocystis* group were selected. Chinese and Italian soil sequences have previously been assigned to phylogeny based on maximum likelihood/neighbour joining trees constructed in ARB (Lüke and Frenzel, 2011). A priori assignment to taxonomic units is indicated by different symbols. Clusters are defined by a density-based algorithm (DBSCAN, Ester *et al.*, 1996) and marked by different colours. Sequences not assigned to a specific cluster ('border' points in DBSCAN terminology) are printed in grey. To reduce overplotting, a transparency of 0.5 was used.

Methanotroph diversity revealed by pyrosequencing

The *pmoA* microarray provides a high phylogenetic resolution corresponding to the genus or even to the species level of methanotrophs (Bodrossy *et al.*, 2003; Lüke *et al.*, 2011). However, methanotroph speciation events in the different soils might be only visible at the sequence level. Thus, in addition to microarray analysis, we performed *pmoA* amplicon pyrosequencing. We focused on the distribution of *Methylosinus/Methylocystis*-like sequences as they constitute the largest fraction in most paddy soils. We used sequence data obtained from the Italian soil and two of the Chinese paddy fields (described in Lüke and Frenzel, 2011), and furthermore amplified the *pmoA* gene from an Indonesian paddy soil (Piladang, Sumatra). Pyrosequencing resulted in a data set of approximately 10 000 high-quality sequences clustering phylogenetically within the *Methylosinus/Methylocystis* group. Comparative sequence analysis using phylogenetic treeing methods is difficult to illustrate for such a large data set. Thus, we used an alternative approach. A subset of 600 sequences from each site was randomly selected, and a distance matrix was calculated based on Jukes-Cantor corrected nucleotide sequence distances. Duplicated sequences were removed prior analysis. This matrix was used for ordination analysis (NMDS; Fig. 4). Such an approach reduces the data complexity into a three-dimensional solution and thus, provides a convenient illustration of the underlying structure. We used a density-based clustering to identify distinct groups. The analysis resulted in five

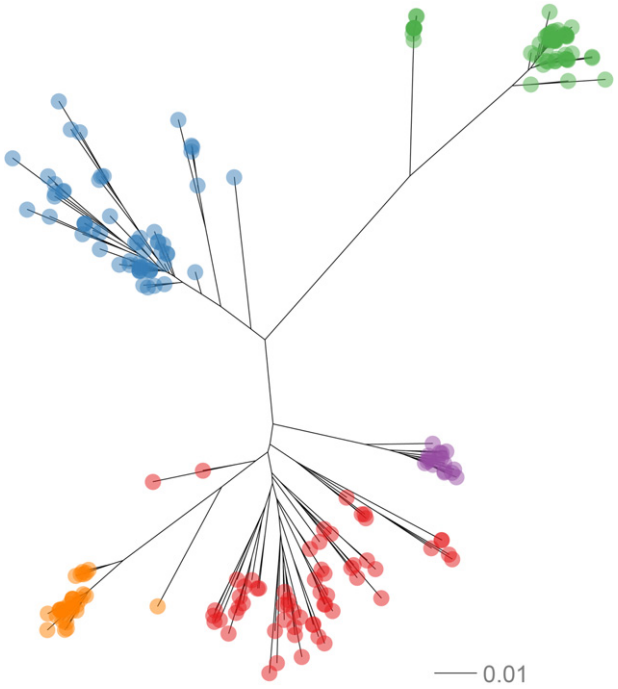

Fig. 5. Unrooted neighbour joining tree showing the phylogentic relationship of clusters defined by DBSCAN in Fig. 4. For clarity, all sequences not assigned to a cluster are omitted, and only a subset of up to 60 sequences per site has been used. Colour coding is identical to Fig. 4.

clusters of which the largest cluster (cluster A in Fig. 4) could be affiliated to the *pmoA* lineage of *Methylocystis*. In addition, cluster D and E contain sequences that were previously assigned to *Methylocystis*, based on phylogenetic trees including isolates (Lüke and Frenzel, 2011). Clustering within the ordination space was compared with neighbour joining trees (Figs. 5 and S3), confirming the grouping. Calculating the tree from a randomly selected subset of sequences, clusters E and D were located close to the main group of *Methylocystis*-like sequences (red in Fig. 5), but clearly separated. Including all sequences in the tree construction, these clusters become part of the *Methylocystis* group, but still form distinct units (Fig. S3). Clusters B and C could not be clearly affiliated to *Methylocystis* or *Methylosinus*. They seem to form two separate groups. The clustering algorithm based on density did not identify the *Methylosinus*-like sequences as separate cluster (squares in Fig. 4). However, only a limited amount of these sequences are present in this data set, and the number was probably too low and sequences clustered not dense enough to be recognized by the algorithm as a distinct group.

Sequences from the different soils were widely distributed throughout all clusters in the ordination, independent from the soil origin (Figs. S4 and S5). Nevertheless, a geographical tendency could be observed for cluster E, which contains sequences nearly exclusively retrieved from Italian paddy soil. To get further insights into the spatial distribution of type II methanotrophs, the distance–decay relationship was analysed. Sequences were grouped into classes based on the geographical distance between sampling sites. The sequence identities were plotted against the distance classes to depict a distance decay plot (Fig. 6). Overall, the sequence identity decreased with increasing distance, and because of the large number of distances considered, the decrease was highly significant ($P < 0.001$). However, with a slope of -0.0005 and r squared $= 0.0022$, this decrease can be regarded as ecologically meaningless.

Discussion

In this work, we showed that methanotroph communities in wetland rice fields exhibit the classical distance–decay patterns known from macroorganisms. To our knowledge, this is the first study addressing global bacterial distribution patterns within a man-made and thus, highly controlled environment. Based on *pmoA* microarray analysis, we found a significant increase of β-diversity with geographical distance. We furthermore investigated the distance–decay relationship for the most abundant group, type II methanotrophs, at the sequence level. Characterized by a vast amount of closely related sequences, this group is phylogenetically very difficult to resolve. We performed ordination analysis of sequence distances unravelling patterns within this group.

Biogeography of methanotrophs

Distance–decay relationships for bacteria were discovered in several natural environments and at various distance scales, ranging from few centimetres to thousands of kilometres. Studied environments include terrestrial (e.g. Horner-Devine *et al.*, 2004; Martiny *et al.*, 2011; Monroy *et al.*, 2012) as well as marine habitats (e.g. Pommier *et al.*, 2007; McAllister *et al.*, 2011). In contrast to all previous work, we focused here on microbial communities within an 'artificial' environment, the rice agroecosystem. Such a system is controlled and managed by standardized agricultural practices. Paddy fields are plowed regularly, and they are dominated by one plant species during rice-growing season. Thus, they provide a comparably homogenous habitat, and accordingly, bacterial communities display no horizontal structure within one paddy field (Krause *et al.*, 2009). Because of this low complexity, wetland rice fields represent a suitable model habitat for studying biogeography. We furthermore focused on a defined subset of microbial diversity by targeting methanotrophic bacteria, facilitating the data interpretation. Our study showed that even in this standardized

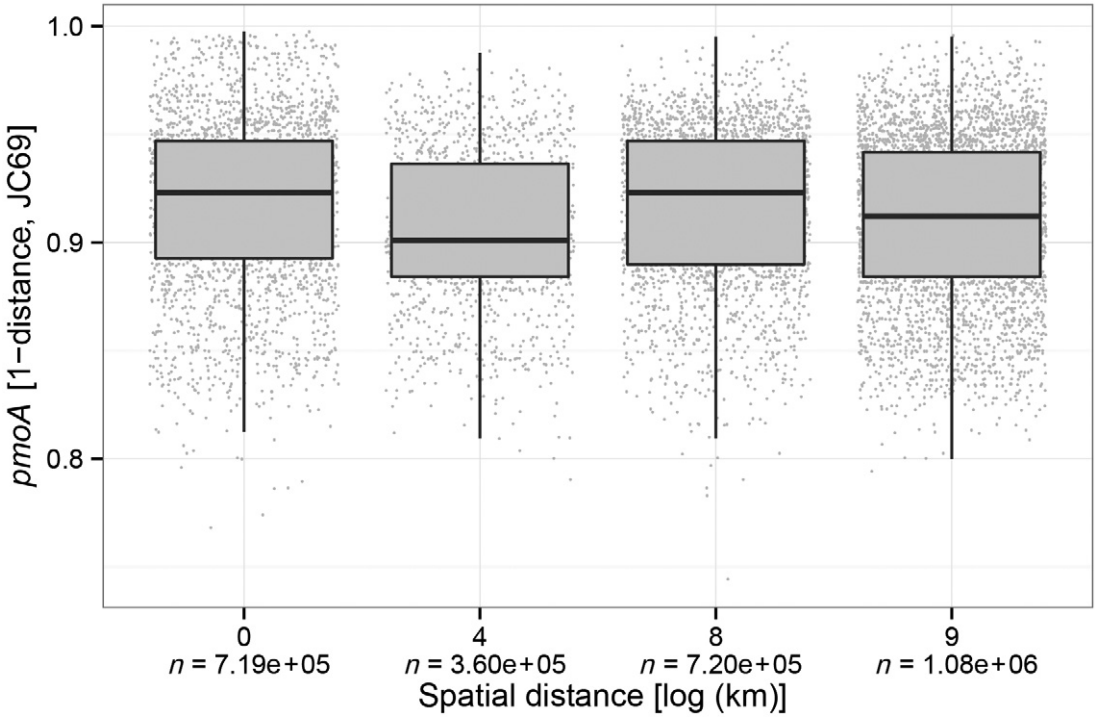

Fig. 6. Distance–decay plot of pyrosequencing data. Geographical distances between the sampling sites were grouped into distance classes. Jukes-Cantor corrected distances were calculated and plotted against the distance classes. The graph is based on a random subset of 104 data points. The slope of the regression of the full data set is −0.0005 (r squared = 0.0022, $n = 2.88 \cdot 106$).

environment, significant distance–decay patterns are displayed for methanotroph communities in paddy fields worldwide. Such patterns can be explained by two underlying processes: environmental selection and dispersal limitations (Soininen, 2012). As environmental variables tend to be autocorrelated, observed patterns could be solely generated by environmental differences. Despite highly comparable field conditions, the field sites investigated here are located within different climate zones. Indeed, the latitude had a small but significant influence on the methanotroph community. However, the largest part of the data variance could not be affiliated to any known environmental parameter. Thus, dispersal limitations are potentially involved in generating the observed pattern. The different field sites were largely characterized by the same methanotrophs, but in different relative abundances. Nevertheless, a certain restriction of dispersal is needed to even create these patterns of different abundances. Methanotrophs are able to form exospores and dormant cells surviving adverse conditions for decades (Whittenbury *et al*., 1970; Rothfuss *et al*., 1997). Thus, they provide a seed bank within the paddy soil from which a subset can be selected by suitable environmental conditions (Eller *et al*., 2005; Ho and Frenzel, 2012). Conditions change over time, and different subsets are stimulated. Here, the sum of environmental conditions acting over time is referred to as historical contingencies (Martiny *et al*., 2006), and we propose that these contingencies are responsible for shaping the individual seed bank. As environmental conditions tend to be spatially autocorrelated, they support the formation of distance–decay patterns.

In general, detection of distance–decay relationships seems to very much depend on the heterogeneity of the investigated habitat and the strength of underlying environmental gradients. Accordingly, for magnetotactic bacteria in aquatic environments including a strong salinity gradient within study sites, only a correlation to salinity and not to geography could be found (Lin *et al*., 2012). In addition, soil bacteria on the continental scale were shown to scatter largely independent of geographic distance, but correlated with soil pH (Fierer and Jackson, 2006).

Spatial distribution of type II methanotrophs

In this study, the geographic pattern was solely created by different species abundances and not different species. Nevertheless, we might have missed possible speciation events by focusing on microarray data, which might provide insufficient phylogenetic resolution. Thus, we investigated the spatial distribution of type II methanotrophs at the sequence level. However, compara-

tive analysis of *pmoA* genotypes revealed no distribution pattern and no evidence for evolutionary diversification (e.g. Figs 6 and S4). *pmoA* affiliated to *Methylocystis* was detected in highest abundance in all soils. This seems to be a general observation in environmental studies and could be biased and the result of different cell lysis characteristics. Comparing different methanotrophic pure cultures harvested at logarithmic growth phase, *Methylocystis* species were found to be most resistant to cell lysis (Meyer *et al*., 1986). However, in the environment, microbial cells undergo various physiological stages, and they are often present in their resting forms. In contrast to *Methylocystis*, *Methylosinus* species form very robust exospores (Whittenbury *et al*., 1970) that can even survive passages through a French press (Titus *et al*., 1982). Therefore, they might be more resistant to disruption in standard DNA extraction procedures from soil, which could explain the low detection by molecular tools. In isolation approaches, both, *Methylocystis* and *Methylosinus* species were frequently retrieved from paddy fields (e.g. Heyer *et al*., 2002).

In this work, we plotted sequence distances into the ordination space to unravel patterns in diversity. Especially for large data sets of highly similar sequences, it might be a suitable complementation to treeing methods. Clusters can be easily identified and verified by phylogenetic approaches. For type II methanotrophs, the ordination exposed five groups of *pmoA* genotypes in paddy fields. Two of these clusters (blue and green in Fig. 5) are large and separated from *Methylocystis* and *Methylosinus* in the neighbour joining tree. As they cluster in between the described genera, they might even form new genera of methanotrophs within the family of *Methylocystaceae*. Our analysis showed that this very dense group of closely related type II methanotroph sequences is more heterogeneous than expected. This analysis only included rice field methanotrophs, but when extending the data set with sequences from various habitats, it might be possible to not only resolve patterns, but also to assign ecological meaning to individual clusters.

Methanotrophs in tropical paddy fields

The dominance of *Methylocystis* and *Methylosinus*-like type II methanotrophs appeared to be a general feature of wetland rice fields. Type I methanotrophs were repeatedly reported to predominate on the roots of rice plants and within the rhizosphere (Horz *et al*., 2001; Wu *et al*., 2009; Lüke *et al*., 2010), but type II were found to dominate in the bulk soil covering the main volume within a paddy field. In some of the Indonesian as well as the Vietnamese soils, we found for the first time a dominance of type I methanotrophs in the bulk soil. In particular, probes for two *pmoA* lineages belonging to type Ib methanotrophs gave the strongest hybridization signal in the microarray analysis (freshwater sediment-2 and JRC-3). As there are no cultured representatives available to date, nothing about the resting stages of these methanotrophs is known. However, they were found frequently in paddy fields, and in particular, the lineage freshwater sediment-2 might play a key function as it was often detected in high abundances (Vishwakarma and Dubey, 2010; Lüke and Frenzel, 2011; Ho *et al*., 2011b).

Conclusions

Distance–decay relationships exist for bacteria. The challenge of future research is to unravel the underlying processes causing this pattern. Here, we could show that the methanotroph community in paddy fields becomes more dissimilar with increasing geographical distance. We could not find any evidence for dispersal limitation and propose historical contingencies being responsible for this pattern. However, we only focused on the distribution of one gene. Methanotrophs within the different paddy fields might be different, and possible speciation events might have happened, becoming only apparent when comparing multiple genes or whole genomes.

Previous studies of wetland rice methanotrophs focused on paddy fields located in temperate climate zones. Tropical countries have so far been neglected even though they belong to the top producers of rice worldwide. Studying the methanotroph community in Indonesian and Vietnamese paddy soils, we observed the first paddy soils depleted in type II methanotrophs. Thus, expanding the focus beyond the well-studied 'model soils' provides new insights into methanotroph diversity. Tropical countries contribute a substantial part to global rice production, and their importance will even increase in the future. Thus, further work should be directed towards investigating the microbial diversity in these soils, including studies of resistance and resilience and the implications of a community depleted in type II methanotrophs as this could have a marked impact on methane emission and consequently, climate change.

Experimental procedures

Soil sampling

All soils were sampled before the rice-growing season started. Indonesian paddy soils were sampled in March 2010 from different sites distributed over the islands Sumatra and Java. About 250–300 g of soil was collected with a scoop in soil pits excavated the same day. Samples were stored in plastic containers at 4–8°C during the 10 days field campaign and frozen to −18°C when transferred to the lab. For analysis, samples were freeze-dried and homogenized. The Chinese soils were sampled in May 2008 (Ho *et al*., 2011b). The Italian paddy soil was sampled in March 2011 from a field located in

Vercelli. Two cores (inner diameter 6.5 cm) 5 m apart were sampled down to a depth of 10 cm. Soil was transferred into plastic bags, kneaded to homogenize and stored at 4°C until further processed. The Vietnamese samples were collected in spring 2012 from two fields and with 10 cm sampling distance within one field. Soil characteristic and geographical coordinates of sampling sites are listed in Table S1.

DNA extraction from soil, T-RFLP and microarray analysis

Total DNA was extracted from 0.25 g of dried soil using the FastDNA Spin Kit for soil (MP Biomedicals, Illkirch, France), following the manufacturer's instructions.

T-RFLP analysis was carried out as described before (Lüke *et al.*, 2010) using the two primer pairs A189f/A682r (Holmes *et al.*, 1995) and A189f/mb661 (Costello and Lidstrom, 1999). Briefly, the *pmoA* gene was amplified using a FAM-labelled forward primer followed by *MspI* digestion. The purified terminal restriction fragments were separated using the Applied Biosystems 3130 Genetic Analyser (Applied Biosystems, Darmstadt, Germany) and further evaluated using the GeneMapper Version 4.0 (Applied Biosystems).

pmoA microarray analysis was performed according to the procedure developed by (Bodrossy *et al.*, 2003) with few modifications: the primer set A189f/T7-A682r, containing a T7 promoter site attached to the reverse primer, was used to prepare the target for *in vitro* transcription. The generated RNA was fragmented and hybridized overnight in a hybridization oven (ThermoLifeScience, Engelsbach, Germany) with the shaking platform set at maximum speed.

454 pyrosequencing of the pmoA gene

pmoA amplicons were generated from total DNA extracted out of Indonesian soil (Piladang, Sumatra). A modified A189f/A682r primer set was used for amplification: the forward primer contained a barcode (four nucleotides) and adapter A. Adapter B was attached to the reverse primer. Adapter A and B were essential for subsequent binding to beads and emulsion PCR. In PCR amplification, five reactions of 50 µl were carried out. A 100 ng template DNA was mixed with 2.5 U of Taq polymerase (Invitrogen, Darmstadt, Germany), 33 pmol of each primer, 0.02 mg bovine serum albumin (Roche, Basel, Switzerland), 5% (v/v) dimethyl sulfoxide, 25 ml Masteramp 2x PCR Premix F (Epicentre Biotechnologies, Hessisch Oldendorf, Germany) and filled up with molecular grade water (Sigma, Hamburg, Germany). The touchdown PCR was carried out with an initial denaturing step at 94°C for 5 min, followed by 20 cycles of 1 min at 94°C, 30 s at 62°C (touchdown 0.5°C per cycle) and 45 s at 72°C. Further, 15 cycles were carried out for 1 min at 94°C, 30 s at 52°C and 45 s at 72°C. Final elongation was performed for 10 min at 72°C. PCR products were pooled and gel purified using the NucleoSpin Extract II Kit (Macherey-Nagel, Düren, Germany). Products were sent for 454 amplicon sequencing (GATC, Konstanz, Germany). Nucleotide sequences found in this study were deposited at the European Molecular Biology Laboratory Sequence Read Archive under the study accession number ERP002645.

Statistical analyses of T-RFLP and microarray data

Statistical analysis was performed using the **R** software environment for statistical computing and graphics version 2.10.0 (R Development Core Team, 2012). For analysis of T-RFLP patterns, T-RFs were binned to operational taxonomic units based on an *in silico* analysis of public database sequences and 454 sequences from this study. After binning, a data set was generated consisting of T-RF sizes in base pairs (bp) with peak heights in fluorescence units. T-RF profiles were standardized according to Dunbar and colleagues (2000). Cluster analysis was performed using the *heatmap.2* function implemented in **gplots** version 2.7.4 (Warnes *et al.*, 2009).

Microarray data were standardized against the mean of overall array intensity in a first step, and in a second step, standardized against an experimentally determined reference value of positive detection (Bodrossy *et al.*, 2003). The probes on the array target the *pmoA* phylogeny at different hierarchical levels; thus, they partially cover the same organisms (Bodrossy *et al.*, 2003). Including all probes in the statistical analysis would strongly bias the result. Thus, a subset of probes was selected (Lüke *et al.*, 2011). This subset is highlighted in Fig. S1.

NMDS and CCA were performed using *metaMDS* and *cca* provided by the **vegan** package version 1.15-4 (Oksanen, 2012). In NMDS analysis, the Bray–Curtis distance was chosen for creating the dissimilarity matrix. The effect of the latitude in CCA was removed by using a condition within the formula of *cca*. By conditioning a variable in **vegan**, the effect of this variable is removed from the analysis before constraining with the remaining variables (Oksanen *et al* 2012).

Statistical analysis of pyrosequencing data

Sequences were imported into the ARB software package (Ludwig *et al.*, 2004) for quality check, alignment and phylogenetic tree construction. Sequences containing errors resulting in reading frame shifts or shorter than 140 amino acids were manually excluded from further analysis. Sequences clustering phylogenetically within the *Methylosinus/Methylocystis* group were exported as FASTA files and imported into **R** version 2.15.2 (R Development Core Team, 2012). The analysis was based on 409 nucleotide positions starting from position 41 within the *pmoA* alignment. The start of the alignment was defined by the first base of the A189f forward primer sequence. Graphics were done using **ggplot2** version 0.9.2.1 (Wickham, 2009). Sequences were handled and analysed using functions from package **seqinr**, version 3.0-6 (Charif and Lobry, 2007) and **ape**, version 3.0-6 (Paradis, 2006; Popescu *et al.*, 2012). Ordination was done in **vegan**, version 2.0-5 (Oksanen *et al.* 2012), applying **metaMDS** to a distance matrix provided by function **dist.dna** from **ape** with model = 'JC69' (Jukes and Cantor, 1969). Sequences were clustered in ordination space using density-based clustering (Ester *et al.*, 1996) as implemented in function **dbscan** in package **fpc** version 2.1-4 (Hennig, 2012). For distance–decay curves, geographical distances were calculated using the haversine formula as implemented in package **geosphere**, version 1.2-28 (Hijmans *et al.*, 2012).

Acknowledgements

This work, as part of the European Science Foundation EUROCORES Programme EuroEEFG, was supported from funds to PF by the Deutsche Forschungsgemeinschaft (DFG). We furthermore thank Stefan Hotes for providing soil samples.

References

Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5:** 566–582.

Charif, D., and Lobry, J.R. (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*. Bastolla, U., Porto, M., Roman, H.E., and Vendruscolo, M. (eds). New York, USA: Springer Verlag, pp. 207–232.

Costello, A.M., and Lidstrom, M.E. (1999) Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. *Appl Environ Microbiol* **65:** 5066–5074.

De Wit, R., and Bouvier, T. (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say. *Environ Microbiol* **8:** 755–758.

Dedysh, S.N., Liesack, W., Khmelenina, V.N., Suzina, N.E., Trotsenko, Y.A., Semrau, J.D., *et al*. (2000) *Methylocella palustris* gen. nov., sp nov., a new methane-oxidizing acidophilic bacterium from peat bags, representing a novel subtype of serine-pathway methanotrophs. *Int J Syst Evol Microbiol* **50:** 955–969.

Dunbar, J., Ticknor, L.O., and Kuske, C.R. (2000) Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. *Appl Environ Microbiol* **66:** 2943–2950.

Eller, G., Krüger, M., and Frenzel, P. (2005) Comparing field and microcosm experiments: a case study on methano- and methylo-trophic bacteria in paddy soil. *FEMS Microbiol Ecol* **51:** 279–291.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996) *A density-based algorithm for discovering clusters in large spatial databases with noise*, *KDD-96 Proceedings*, 226–231.

Ferrando, L., and Tarlera, S. (2009) Activity and diversity of methanotrophs in the soil-water interface and rhizospheric soil from a flooded temperate rice field. *J Appl Microbiol* **106:** 306–316.

Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103:** 626–631.

Hennig, C. (2012) *fpc: flexible procedures for clustering*. [WWW document]. URL http://CRAN.R-project.org/package=fpc.

Heyer, J., Galchenko, V.F., and Dunfield, P.F. (2002) Molecular phylogeny of type II methane-oxidizing bacteria isolated from various environments. *Microbiology* **148:** 2831–2846.

Hijmans, R.J., Williams, E., and Vennes, C. (2012) *geosphere: spherical trigonometry*, 1.2–28. [WWW document]. URL http://CRAN.R-project.org/package=geosphere.

Ho, A., and Frenzel, P. (2012) Heat stress and methane-oxidizing bacteria: effects on activity and population dynamics. *Soil Biol Biochem* **50:** 22–25.

Ho, A., Lüke, C., Cao, Z., and Frenzel, P. (2011a) Ageing well: methane oxidation and methane oxidizing bacteria along a chronosequence of 2000 years. *Environ Microbiol Rep* **3:** 738–743.

Ho, A., Lüke, C., and Frenzel, P. (2011b) Recovery of methanotrophs from disturbance: population dynamics, evenness and functioning. *ISME J* **5:** 750–758.

Holmes, A.J., Costello, A., Lidstrom, M.E., and Murrell, J.C. (1995) Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol Lett* **132:** 203–208.

Horner-Devine, M.C., Lage, M., Hughes, J.B., and Bohannan, B.J.M. (2004) A taxa-area relationship for bacteria. *Nature* **432:** 750–753.

Horz, H.P., Yimga, M.T., and 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. *Appl Environ Microbiol* **67:** 4177–4185.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ, USA: Princeton University Press.

Jia, Z.J., Sugano, A., Asari, N., Tun, C.C., Ishihara, R., Asakawa, S., and Kimura, M. (2007) Molecular characterization of methane-oxidizing bacteria associated with rice straw decomposition in a rice field. *Soil Sci Plant Nutr* **53:** 729–739.

Jukes, T.H., and Cantor, C. (1969) Evolution of protein molecules. In *Applied and Environmental Microbiology*. Munro, M.N. (ed.). New York, USA: Academic Press, pp. 21–132.

Kim, O.-S., Junier, P., Imhoff, J.F., and Witzel, K.-P. (2008) Comparative analysis of ammonia monooxygenase (*amoA*) genes in the water column and sediment–water interface of two lakes and the Baltic Sea. *FEMS Microbiol Ecol* **66:** 367–378.

Knief, C., Kolb, S., Bodelier, P.L.E., Lipski, A., and Dunfield, P.F. (2006) The active methanotrophic community in hydromorphic soils changes in response to changing methane concentration. *Environ Microbiol* **8:** 321–333.

Krause, S., Lüke, C., and Frenzel, P. (2009) Spatial heterogeneity of methanotrophs: a geostatistical analysis of *pmoA*-based T-RFLP patterns in a paddy soil. *Environ Microbiol Rep* **1:** 393–397.

Lin, W., Wang, Y., Li, B., and Pan, Y. (2012) A biogeographic distribution of magnetotactic bacteria influenced by salinity. *ISME J* **6:** 475–479.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.

Lüke, C., and Frenzel, P. (2011) Potential of *pmoA* amplicon pyrosequencing for methanotroph diversity studies. *Appl Environ Microbiol* **77:** 6305–6309.

Lüke, C., Krause, S., Cavigiolo, S., Greppi, D., Lupotto, E., and Frenzel, P. (2010) Biogeography of wetland rice methanotrophs. *Environ Microbiol* **12:** 862–872.

Lüke, C., Bodrossy, L., Lupotto, E., and Frenzel, P. (2011) Methanotrophic bacteria associated to rice roots: the cultivar effect assessed by T-RFLP and microarray analysis. *Environ Microbiol Rep* **3:** 518–525.

Ma, K., and Lu, Y.H. (2011) Regulation of microbial methane production and oxidation by intermittent drainage in rice field soil. *FEMS Microbiol Ecol* **75:** 446–456.

Macalady, J.L., McMillan, A.M.S., Dickens, A.F., Tyler, S.C., and Scow, K.M. (2002) Population dynamics of type I and II methanotrophic bacteria in rice soils. *Environ Microbiol* **4:** 148–157.

McAllister, S., Davis, R.E., McBeth, J.M., Tebo, B.M., Emerson, D., and Moyer, C.L. (2011) Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing zetaproteobacteria. *Appl Environ Microbiol* **77:** 5445–5457.

McDonald, I.R., Bodrossy, L., Chen, Y., and Murrell, J.C. (2008) Molecular ecology techniques for the study of aerobic methanotrophs. *Appl Environ Microbiol* **74:** 1305–1315.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D., and Horner-Devine, M.C. (2011) Drivers of bacterial beta-diversity depend on spatial scale. *Proc Natl Acad Sci USA* **108:** 7850–7854.

Meyer, J., Haubold, R., Heyer, J., and Böckel, W. (1986) Contribution to the taxonomy of methanotrophic bacteria: correlation between membrane type and GC-value. *J Basic Microbiol* **26:** 155–160.

Monroy, F., Van der Putten, W.H., Yergeau, E., Mortimer, S.R., Duyts, H., and Bezemer, T.M. (2012) Community patterns of soil bacteria and nematodes in relation to geographic distance. *Soil Biol Biochem* **45:** 1–7.

Nekola, J.C., and White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *J Biogeogr* **26:** 867–878.

Oksanen, J. (2012) *Multivariate analysis of ecological communities in R: vegan tutorial*. [WWW document]. URL http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*. (2012) *vegan: community ecology package*, 2.1–13/r2105 edn. [WWW document]. URL http://CRAN.R-project.org/package=vegan.

Op den Camp, H.J.M., Islam, T., Stott, M.B., Harhangi, H.R., Hynes, A., Schouten, S., *et al*. (2009) Environmental, genomic and taxonomic perspectives on methanotrophic verrucomicrobia. *Environ Microbiol Rep* **1:** 293–306.

Paradis, E. (2006) *Analysis of Phylogenetics and Evolution with R*. New York, USA: Springer.

Pommier, T., Canbäck, L., Riemann, L., Boström, K.H., Simu, K., Lundberg, P., *et al*. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16:** 867–880.

Popescu, A.-A., Huber, K.T., and Paradis, E. (2012) ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28:** 1536–1537.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*, 2.15.1 Edn. Vienna, Austria: R Foundation for Statistical Computing.

Reim, A., Lüke, C., Krause, S., Pratscher, J., and Frenzel, P. (2012) One millimetre makes the difference: high-resolution analysis of methane-oxidizing bacteria and their specific activity at the oxic–anoxic interface in a flooded paddy soil. *ISME J* **6:** 2128–2139.

Rothfuss, F., Bender, M., and Conrad, R. (1997) Survival and activity of bacteria in a deep, aged lake sediment (Lake Constance). *Microb Ecol* **33:** 69–77.

Semrau, J.D., DiSpirito, A.A., and Yoon, S. (2010) Methanotrophs and copper. *FEMS Microbiol Rev* **34:** 496–531.

Soininen, J. (2012) Macroecology of unicellular organisms – patterns and processes. *Environ Microbiol Rep* **4:** 10–22.

Tavormina, P.L., Orphan, V.J., Kalyuzhnaya, M.G., Jetten, M.S.M., and Klotz, M.G. (2011) A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ Microbiol Rep* **3:** 91–100.

Titus, J.A., Reed, W.M., Pfister, R.M., and Dugan, P.R. (1982) Exospore formation in *Methylosinus trichosporium*. *J Bacteriol* **149:** 354–360.

Vishwakarma, P., and Dubey, S.K. (2010) DNA microarray analysis targeting *pmoA* gene reveals diverse community of methanotrophs in the rhizosphere of tropical rice soils. *Curr Sci India* **99:** 1090–1095.

Vishwakarma, P., Dumont, M.G., Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., and Dubey, S.K. (2009) Ecological and molecular analyses of the rhizospheric methanotroph community in tropical rice soil: effect of crop phenology and land-use history. *Curr Sci India* **96:** 1082–1089.

Vorobev, A.V., Baani, M., Doronina, N.V., Brady, A.L., Liesack, W., Dunfield, P.F., and Dedysh, S.N. (2011) *Methyloferula stellata* gen. nov., sp nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. *J Syst Evol Microbiol* **61:** 2456–2463.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., *et al*. (2009) *gplots: various R programming tools for plotting data*. [WWW document]. URL http://CRAN.R-project.org/package=gplots.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301:** 976–978.

Whittenbury, R., Davies, S.L., and Davey, J.F. (1970) Exospores and cysts formed by methane-utilizing bacteria. *J Gen Microbiol* **61:** 219–226.

Wickham, H. (2009) *Ggplot2. Elegant Graphics for Data Analysis, Vol. Use R*. Dordrecht, Netherlands: Springer.

Wu, L.Q., Ma, K., and Lu, Y.H. (2009) Rice roots select for type I methanotrophs in rice field soil. *Syst Appl Microbiol* **32:** 421–428.

Zheng, Y., Zhang, L.M., Zheng, Y.M., Di, H.J., and He, J.Z. (2008) Abundance and community composition of methanotrophs in a Chinese paddy soil under long-term fertilization practices. *J Soils Sediments* **8:** 406–414.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Heatmap showing microarray results of *pmoA* amplicons retrieved from various soils using the A682r reverse primer. Microarray signals were normalized against the mean of overall array intensity and against reference values determined for each probe individually (Bodrossy *et al*., 2003; Lüke *et al*., 2011). Relative fractions were calculated. Probes used for statistical analyses are highlighted in blue. A phylogenetic tree depicting an overview of the probe phylogeny as well as the rational for the probe selection is described in Lüke and colleagues (2011). Sample sites are described in Table S1.

Fig. S2. Cluster analysis of standardized T-RFLP data obtained from Indonesian paddy soils. The heatmap shows the relative abundance of individual T-RFs obtained using the *pmoA* reverse primers mb661 (A) and A682r (B). Affiliation of T-RFs: 33 bp – AOB, MO3 cluster, type Ib, *pxmA* – 36 bp – type Ib – 46 bp – AOB – 58 bp – *pxmA* – 79 bp – type Ib, RPC-2, *pxmA* – 114 bp – AOB, *pmoA-2*, *pxmA* – 203 bp – unknown – 208 bp – type Ia – 226 bp – type Ib – 241 bp – type Ia – 244 bp – *Methylosinus*/*Methylocystic* type II – 349 bp – type Ia, *pmoA-2*–377 bp – type Ia – 437 bp – type Ia, *pmoA-2*, *pxmA* – 505 bp – type Ia.

Fig. S3. Unrooted neighbour joining tree showing the phylogentic relationship of clusters defined by DBSCAN in Fig. 4. In contrast to Fig. 5, all sequences have been used for tree construction. Colour coding for individual clusters is identical to Fig. 4.

Fig. S4. Non-metric multidimensional scaling (NMDS) with three axes based Jukes-Cantor corrected *pmoA* nucleotide sequence distances. The data set and ordination are identical to Fig. 4, but here, sequences are colour-coded based on their soil origin. To reduce overplotting, a transparency of 0.5 was used.

Fig. S5. Mosaic plot showing the relative importance of clusters defined in Fig. 4. Data set and colour-coding are identical to Fig. 4.

Table S1. Description of sampling sites. The Indonesian site selected for additional *pmoA* amplicon pyrosequencing is highlighted in red. Fully developed paddy soils are called Anthrosols according to the international soil classification system (IUSS Working Group WRB, 2007). The parent soils that have developed naturally on the Indonesian and Chinese sites before having been altered by paddy use are Fluvisols, Vertisols, Ferralsols/Acrisols and Andosols/Umbrisols. The Fluvisols in Cixi (China) consist of silt-rich marine sediments and were brought to terrestrial soil formation by human land reclamation. Vertisols (Java) are clay-rich soils, and the clay fraction is dominated by strongly swelling and shrinking clays leading to the formation of strong and large aggregates. In Ferralsols and Acrisols (Sumatra), the clay fraction consists mainly of kaolinites allowing only small nutrient reserves and a weak macroaggregation. Andosols (Java and Sumatra) develop from volcanic ashes, are very loose and rich in organic matter and Fe oxides. Some soils from volcanic ashes do not entirely meet the Andosol criteria and have to be classified as Umbrisols.

Supplementary file 1 (*pyro_typeII_dna*). High-quality subset of the *pmoA* pyrosequencing data obtained from various paddy soils. These sequences were used for the analysis of type II methanotrophs depicted in Figs 4–6 and Figs S3–S5. The data set was created by importing all raw sequences into ARB, removing all sequences shorter than 400 bp and manually checking sequences for frame shifts. Only good quality sequences longer than 399 bp and without frame shifts were maintained. These sequences were classified by calculation of neighbour joining trees and only sequences related to the *Methylosinus*/*Methylocystis* group were selected.

Supplementary file 2 (*pyro_typeII_dna.nds*). Sequence associated information (metadata) referring to the data set in supplementary file 1.