Potential of *pmoA* Amplicon Pyrosequencing for Methanotroph Diversity Studies^{∇†}

Claudia Lüke* and Peter Frenzel

Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str. 10, D-35043 Marburg, Germany

Received 3 May 2011/Accepted 30 June 2011

We analyzed the potential of *pmoA* amplicon pyrosequencing compared to that of Sanger sequencing with paddy soils as a model environment. We defined operational taxonomic unit (OTU) cutoff values of 7% and 18%, reflecting methanotrophic species and major phylogenetic *pmoA* lineages, respectively. Major lineages were already well covered by clone libraries; nevertheless, pyrosequencing provided a higher level of diversity at the species level.

Next-generation sequencing technologies such as pyrosequencing have revolutionized the field of microbial ecology by enabling a hitherto unprecedented sampling depth. This technology was successfully applied for examining archaeal and bacterial 16S rRNA gene diversity in various environments (e.g., references 2 and 12). Nevertheless, the pyrosequencing technology also has its limitations. Problems occur in particular during sequencing of homopolymeric stretches. The length of homopolymers is inferred from light signal intensities, a procedure which is prone to errors. As a consequence, insertions or deletions of nucleotides (indels) represent the most frequent error modality (25), and sequence diversity is artificially enlarged in raw sequences. Different algorithms were developed to correct for this bias (14, 21). However, uncertainties remain for distinguishing between true and artificial indels.

In this work, we sequenced a protein-encoding gene—the *pmoA* gene—using pyrosequencing (GS FLX, titanium chemistry; Roche). Sequence analysis of protein-encoding genes has the great advantage that indels can be easily detected. They result in reading frameshifts that are directly evident in the amino acid alignment.

The *pmoA* gene, encoding a subunit of the particulate methane monooxygenase (pMMO), was shown to be an excellent phylogenetic marker for methanotrophs (9). This functional guild of bacteria provides a key function in the global carbon cycle by attenuating potential methane emissions to the atmosphere or even acting as a sink for atmospheric methane (24). Based on their phylogeny, aerobic methanotrophs can be divided into types Ia and Ib (both gammaproteobacteria) and type II (alphaproteobacteria) (24). Verrucomicrobial methanotrophs discovered recently seem to be restricted to extreme environments (20).

In rice paddies, aerobic methanotrophs have been studied for decades, resulting in substantial knowledge of their biodiversity, spatial heterogeneity, and biogeography (e.g., refer-

ences 16 and 19). Paddy soil methanotrophs were shown to be phylogenetically diverse, but so far, this diversity seemed to be tractable. Thus, they provide an excellent model community to evaluate the performance and potential of pyrosequencing.

In this study, amplicons were generated using the two primer sets providing the widest coverage of the known *pmoA* diversity (A189f/mb661r, A189f/A682r) (6, 13). We chose the soil of an Italian (Vercelli) and two Chinese (Cixi, Zhejiang Province) rice fields. The Chinese soils developed after diking and cultivation with rice for different time periods (50 years [young soil] and 2,000 years [old soil]) (5). Total DNA was extracted from 0.3 g freeze-dried soil using the Fast DNA spin kit for soil (MP Biomedicals) followed by PCR amplification as described before (19). Only sequences with read lengths above 400 bp were considered for further analysis: 3,511 reads obtained from the Italian soil, 5,234 reads from the young Chinese soil, and 3,591 reads from the old Chinese soil. Approximately 40% of these sequences contained obvious reading frameshifts (see Fig. S1 in the supplemental material) and were subsequently excluded. This demonstrates the high rate of indels that occur during sequencing and would have not been identified when comparing 16S rRNA gene sequences. Furthermore, sequencing errors were distributed asymmetrically, as some reads were not affected whereas other reads contained multiple indels. Hence, individual reads may be prone to error rates much higher than the 5% assumed so far (3). Such errors may strongly bias subsequent operational taxonomic unit (OTU) clustering.

For the remaining sequence reads, distance matrices were calculated based on the amino acid alignment using the ARB software package (18). The sequences were assigned to OTUs using the average linkage algorithm implemented in the program mothur (ver. 1.13.0) (23), based on two different distance levels. The PmoA distance of 7% was shown to correspond to the 3% 16S rRNA distance level (8), thus representing an indication for a methanotroph species. In addition, we searched for an OTU cutoff value that would reflect the major *pmoA* lineages formed by phylogenetic tree construction (Fig. 1). These lineages were determined based on a database containing more than 5,000 *pmoA* and related sequences. Five percent of the database consists of homologous *amoA* sequences encoding a subunit of the ammonia monooxygenase. Neighbor-joining trees were calculated, including various sub-

* Corresponding author. Mailing address: Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str. 10, D-35043 Marburg, Germany. Phone: 49-6421-178803. Fax: 49-6421-178809. E-mail: lueke@mpi-marburg.mpg.de.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

[∇] Published ahead of print on 15 July 2011.

FIG. 1. Neighbor-joining tree depicting major *pmoA* lineages based on 136 deduced amino acid positions of approximately 2,000 sequences. The lineages were determined by comparing neighbor-joining and maximum-likelihood trees built from different data subsets covering the entire known *pmoA* diversity (more than 12,000 sequences in total). Lineages that were monophyletic in all tree constructions are marked with a black circle. Lineages comprising cultivated methanotrophs are shown in bold and referred to as the respective genera or species. The *Methylobacter*-like

sets of sequences covering the entire diversity of the database combined with the new sequences obtained in this study (see Fig. S2 in the supplemental material). Tree topologies were compared and analyzed for lineages that revealed a stable clustering in all trees. The neighbor-joining trees were furthermore compared with the topology of maximum-likelihood trees. Most major *pmoA* lineages were monophyletic in all tree constructions. However, some lineages within type I methanotrophs divided into nonmonophyletic subclusters (Fig. S2), suggesting that larger subsets within the type I phylogeny might form a “bush”-like continuum after further extensive sampling. Nevertheless, most of the *pmoA* lineages shown in Fig. 1 were stable, providing a solid taxonomic basis for studying methanotrophs, at least in environments such as paddy fields where *Methylocella* (7) and *Methyloferula* (29) are not dominant. These genera possess only a soluble methane monooxygenase (sMMO) and not the pMMO.

Besides reflecting methanotroph evolution, many of the *pmoA* lineages describe environmental distribution patterns. For instance, several lineages consist entirely of sequences recovered from marine or saline environments (*Methylobacterium japonense*-like, lineages deep sea-1 to -5) (e.g., 17, 27). Furthermore, sequences clustering within the upland soil cluster (USC)- α , USC- γ , JR-2, and JR-3 were nearly exclusively retrieved from upland soils consuming methane at atmospheric concentration. In addition, the main lineages within type Ib methanotrophs (lineages freshwater sediment-1 and -2) are highly dominated by rice paddy and freshwater lake sequences.

To find an OTU cutoff value that reflects the phylogenetic clustering in Fig. 1, sequences retrieved by pyrosequencing were first assigned to *pmoA* lineages based on phylogenetic tree construction in ARB. Then, these sequences were clustered into OTUs based on different distance levels using mothur. The results were compared using the Rand index (see Table S1 in the supplemental material). This index, ranging from 0 to 1, provides a measure of similarity between two data clustering methods, in which 0 means no similarity and 1 means identical clustering of the two methods (22). The OTU cutoff value of 18% amino acid sequence distance resulted in a nearly identical number of clusters and a Rand index of 0.99. Thus, these results are highly comparable to the *pmoA* lineages obtained by phylogenetic tree construction.

Comparison of the overall methanotroph communities in the three analyzed paddy soils revealed similar patterns (Fig. 1). All soils were characterized by a high diversity of members of type Ib, such as lineage freshwater sediment 2 and the rice paddy cluster 1 (RPC-1) (19) within lineage freshwater sediment 1. Besides type Ib, a high abundance of type II methanotrophs (*Methylocystis* and *Methylosinus*-like) could be detected. In contrast, the diversity within type Ia was primarily restricted

TABLE 1. Comparison of *pmoA* diversity within each paddy soil (α diversity) and diversity shared by two of the soils (β diversity)^a

Soil origin	OTU cutoff (%)	α diversity (Shannon index)	β diversity (Bray-Curtis index)	
			Italy	China (young soil)
Vercelli, Italy	7	1.41		
Cixi, China (young soil)	7	2.47	0.34	
Cixi, China (old soil)	7	2.09	0.55	0.54
Vercelli, Italy	18	1.02		
Cixi, China (young soil)	18	1.66	0.38	
Cixi, China (old soil)	18	1.67	0.57	0.64

^a Shannon and Bray-Curtis indices were calculated for the pyrosequencing data set based on OTU definition of 7% and 18% Kimura-corrected amino acid sequence distances using mothur. Distance matrices were computed based on 136 deduced amino acid positions using ARB, and clustering was performed using the average neighbor algorithm implemented in mothur. If clustering on a specified distance level was not possible, the next lower distance level was chosen.

to two genera: *Methylobacter* and *Methylosarcina*. Furthermore, pyrosequencing revealed a high number of sequences grouping between the *pmoA* gene of methanotrophs and the *amoA* gene of ammonia oxidizers. One of these clusters was previously referred to as *pxmA*, as the physiological function of the encoded protein is still unknown (28). Here, we use *pxmA* for all clusters between *amoA* and *pmoA* with uncertain substrate specificity. Remarkably, despite the great sequencing effort, no sequences belonging to the upland soil clusters (e.g., USC- α and USC- γ) were recovered, indicating that these potentially high-affinity methanotrophs (10) indeed play no role in irrigated paddy fields. This observation might seem trivial; however, the paddy soils are not flooded permanently but are aerated during winter season. Furthermore, high- and low-affinity kinetics for methane oxidation were measured in Italian paddy soils (4). Rain-fed rice fields have even been shown to act as a net sink for atmospheric methane when not flooded (26). Thus, in seasonally flooded paddy fields, type II methanotrophs harboring a second pMMO isoenzyme with a high affinity to methane (1) might be responsible for this process. Indeed, we found a significant number of *pmoA2* sequences encoding a subunit of this isoenzyme (Fig. 1).

Shannon indices calculated for the pyrosequencing data set at 7% and 18% amino acid distance level revealed higher α diversity of methanotrophs within the Chinese soils (Table 1). However, β diversity between the Italian and the old Chinese soil—including shared OTU richness and abundance (Bray-Curtis indices)—revealed values nearly as high as the β -diversity values between the two Chinese soils (Table 1). The Italian paddy soil also has a long history, having been cultivated for

lineage comprises additionally the closely related genus *Methylosoma*. Lineages lacking isolates are named according to representative clones or to the environment in which they were predominantly or initially found (RPC, rice paddy cluster; JRC, Japanese rice cluster; TUSC, tropical upland soil cluster; USC, upland soil cluster; JR, Jasper Ridge, California; MOB, methane-oxidizing bacteria; AOB, ammonia-oxidizing bacteria). GenBank accession numbers of representative clones are given in parentheses. The lineages RPC-2 and deep sea-4 are clustering within either type Ia or type Ib, depending on subset or treeing method. Colored squares next to *pmoA* lineages represent the paddy soils from which sequences were retrieved using pyrosequencing: red, Italian paddy soil; blue, young Chinese paddy soil; yellow, old Chinese paddy soil. The scale bar represents 0.1 changes per amino acid position.

TABLE 2. OTUs within paddy soils^a

Soil origin	OTU cutoff (%)	No. of OTUs after Sanger sequencing	No. of OTUs after pyrosequencing	No. of shared OTUs	Chao1 Sanger sequencing	Chao1 pyrosequencing
Vercelli, Italy	7	12	25	9	27	30
	18	9	13	8	10.5	13.5
Cixi, China (young soil)	7	14	57	10	16	78
	18	10	17	9	10.3	17
Cixi, China (old soil)	7	14	40	14	16	62
	18	9	16	9	9	16.5

^a Observed number of OTUs within the paddy soils revealed by Sanger and by pyrosequencing. OTUs shared by both sequencing technologies and Chao1 richness estimators are also reported. Calculations were performed using the mothur software.

over a century (19). Thus, a specific paddy soil-adapted methanotroph community might develop over centuries.

We compared the diversity coverage of the pyrosequencing data set with the sequence diversity of traditional clone libraries obtained in previous studies (19; A. Ho, C. Lüke, Z. Cao, and P. Frenzel, submitted for publication) (see Fig. S3 in the supplemental material). At a 7% cutoff distance, rarefaction curves show a steep slope and the observed number of OTUs in the pyrosequencing data set is larger than the OTU number obtained by cloning. Nevertheless, at an 18% cutoff distance, the number of OTUs retrieved by pyrosequencing is similar to the OTU numbers retrieved by cloning and rarefaction curves show only a minor increase. Sanger and pyrosequencing data sets share most OTUs at this distance level (Table 2). Thus, with increasing sampling efforts from approximately 120 to over 2,000 sequences, only a few additional *pmoA* lineages were detected. The Chao1 richness estimator confirms this trend (Table 2). Only one sequence formed a novel lineage clustering between *pmoA* and *amoA* (Fig. 1). The other additionally detected lineages all comprised known diversity, almost entirely type Ib and *pxmA*. Together with the *Methylosinus-Methylocystis* lineage (type II), type Ib represents the main fraction of methanotrophs in paddy fields. The latter group itself is highly diverse, consisting of eight major *pmoA* lineages (Fig. 1). With the exception of *Methylothermus*, all of these type Ib lineages are present in paddy soils. Thus, the increased sampling effort did not unravel new lineages but, instead, resulted in an increased diversity within the dominant group.

Within the *pxmA* sequences clustering between *pmoA* and *amoA*, several lineages (type Ia *pxmA*-like, RA21, putative ethane monooxygenases, *Verrucomicrobia*) contain sequences retrieved with primer sets different from those applied in our study (28; C. L. Chen, Z. Chen, X. Wang, H. Z. Yuan, H. L. Qin, M. N. Wu, and W. X. Wei, unpublished data: accession numbers FJ210305 and FJ210327), or represent complete coding sequences retrieved from isolates (11; T. Nakamura, T. Hoaki, S. Hanada, A. Maruyama, Y. Kamagata, and H. Fuse, unpublished data: accession numbers AB453960 and AB453961). They provide longer sequence reads than the remaining *pmoA* amplicons. In fact, these sequences contain up to four mismatches (RA21), six mismatches (type I MOB *pxmA*-like cluster), or even up to eight mismatches (putative ethane monooxygenases, *Verrucomicrobia*) to the A682r reverse primer used in our study. Thus, many of these *pxmA* sequences might have been missed before due to primer bias but could be recovered by deep sequencing.

We found that more than 40% of all reads included at least one insertion or deletion resulting in reading frameshifts in the amino acid alignment. These errors would not have been detected in sequence analysis of 16S rRNA gene amplicons. Thus, for microbial guilds catalyzing well-defined functions, such as methanotrophs, protein-encoding genes should be the target of choice for pyrosequencing analyses.

Excluding inaccurate reads, the detected diversity was virtually covered by the already existing *pmoA* database. Rice paddies are a well-studied habitat for methanotrophs, similarly to freshwater and marine environments, and the majority of *pmoA* sequences available in public databases were retrieved from rice paddies. The fact that deep sequencing does not result in a significant number of new phylotypes confirms that methanotrophic key players in these paddy fields may have already been identified and strengthens the value of methanotroph ecology studies performed in the past. Nevertheless, pyrosequencing of *pmoA* amplicons has great potential for exploring new habitats. Furthermore, recovered diversity is always restricted by the primers used in PCR amplification. The only new *pmoA* lineage found in this study (I14INRXW; Fig. 1) clusters phylogenetically within *pxmA* lineages where traditional primers possess mismatches to existing sequences. Thus, for *pmoA* diversity studies, a great potential of deep sequencing might be the recovery of new diversity beyond the borders of the traditional primer sets.

Nucleotide sequence accession numbers. Nucleotide sequences found in this study were deposited at the EMBL Sequence Read Archive (SAR) under the study accession number ERP000779 (<http://www.ebi.ac.uk/ena/data/view/ERP000779>).

This work, as part of the framework “Biogeochemistry of Paddy Soil Evolution” and of the European Science Foundation EUROCORES Programme EuroEEFG, was supported from funds by the Deutsche Forschungsgemeinschaft (DFG).

We thank Bellinda Schneider for excellent technical assistance.

REFERENCES

- Baani, M., and W. Liesack. 2008. Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in *Methylocystis* sp. strain SC2. *Proc. Natl. Acad. Sci. U. S. A.* **105**:10203–10208.
- Bates, S. T., et al. 18 November 2010, posting date. Examining the global distribution of dominant archaeal populations in soil. *ISME J.* [Epub ahead of print.] doi:10.1038/ismej.2010.171.
- Behnke, A., et al. 2011. Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ. Microbiol.* **13**:340–349.
- Bender, M., and R. Conrad. 1992. Kinetics of CH₄ oxidation in oxic soils exposed to ambient air or high CH₄ mixing ratios. *FEMS Microbiol. Lett.* **101**:261–270.

5. Cheng, Y. Q., L. Z. Yang, Z. H. Cao, E. Ci, and S. X. Yin. 2009. Chronosequential changes of selected pedogenic properties in paddy soils as compared with non-paddy soils. *Geoderma* **151**:31–41.
6. Costello, A. M., and M. E. Lidstrom. 1999. Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. *Appl. Environ. Microbiol.* **65**:5066–5074.
7. Dedys, S. N., et al. 2000. *Methylocella palustris* gen. nov., sp. nov., a new methane-oxidizing acidophilic bacterium from peat bogs, representing a novel subtype of serine-pathway methanotrophs. *Int. J. Syst. Evol. Microbiol.* **50**:955–969.
8. Degelmann, D. M., W. Borken, H. L. Drake, and S. Kolb. 2010. Different atmospheric methane-oxidizing communities in European beech and Norway spruce soils. *Appl. Environ. Microbiol.* **76**:3228–3235.
9. Dumont, M. G., and J. C. Murrell. 2005. Community-level analysis: key genes of aerobic methane oxidation. *Environ. Microbiol.* **39**:413–427.
10. Dunfield, P. F. 2007. The soil methane sink, p. 152–170. *In* D. S. Reay, N. Hewitt, K. A. Smith, and J. Grace (ed.), *Greenhouse gas sinks*. CABI Publishing, Wallingford, United Kingdom.
11. Dunfield, P. F., et al. 2007. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**:879–882.
12. Hollister, E. B., et al. 2010. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *ISME J.* **4**:829–838.
13. Holmes, A. J., A. Costello, M. E. Lidstrom, and J. C. Murrell. 1995. Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* **132**:203–208.
14. Huse, S. M., D. M. Welch, H. G. Morrison, and M. L. Sogin. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* **12**:1889–1898.
15. Reference deleted.
16. Krause, S., C. Lüke, and P. Frenzel. 2009. Spatial heterogeneity of methanotrophs: a geostatistical analysis of *pmoA*-based T-RFLP patterns in a paddy soil. *Environ. Microbiol. Rep.* **1**:393–397.
17. Lin, J. L., et al. 2004. Molecular diversity of methanotrophs in Transbaikalian soda lake sediments and identification of potentially active populations by stable isotope probing. *Environ. Microbiol.* **6**:1049–1060.
18. Ludwig, W., et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363–1371.
19. Lüke, C., et al. 2010. Biogeography of wetland rice methanotrophs. *Environ. Microbiol.* **12**:862–872.
20. Op den Camp, H. J. M., et al. 2009. Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. *Environ. Microbiol. Rep.* **1**:293–306.
21. Quince, C., et al. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* **6**:639–641.
22. Rand, W. M. 1971. Objective criteria for evaluation of clustering methods. *J. Am. Stat. Assoc.* **66**:846–850.
23. Schloss, P. D., et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**:7537–7541.
24. Semrau, J. D., A. A. DiSpirito, and S. Yoon. 2010. Methanotrophs and copper. *FEMS Microbiol. Rev.* **34**:496–531.
25. Shendure, J., and H. L. Ji. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**:1135–1145.
26. Singh, J. S., A. S. Raghubanshi, V. S. Reddy, S. Singh, and A. K. Kashyap. 1998. Methane flux from irrigated paddy and dryland rice fields, and from seasonally dry tropical forest and savanna soils of India. *Soil Biol. Biochem.* **30**:135–139.
27. Tavormina, P. L., W. Ussler, S. B. Joye, B. K. Harrison, and V. J. Orphan. 2010. Distributions of putative aerobic methanotrophs in diverse pelagic marine environments. *ISME J.* **4**:700–710.
28. Tavormina, P. L., V. J. Orphan, M. G. Kalyuzhnaya, M. S. M. Jetten, and M. G. Klotz. 2011. A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ. Microbiol. Rep.* **3**:91–100.
29. Vorobev, A. V., et al. 19 November 2010, posting date. *Methyloferula stellata* gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium possessing only a soluble methane monooxygenase. *Int. J. Syst. Evol. Microbiol.* [Epub ahead of print.] doi:10.1099/ijls.0.028118-0.