Wen, X., Yang, S., Liebner, S. (2016): Evaluation and update of cutoff values for methanotrophic pmoA gene sequences. - *Archives of Microbiology*, *198*, 7, pp. 629–636.

DOI: http://doi.org/10.1007/s00203-016-1222-8

Evaluation and update of cutoff values for methanotrophic *pmoA* gene sequences

Xi Wen[1,2], Sizhong Yang[1,3], Susanne Liebner[1]

[1] Helmholtz Center Potsdam, GFZ German Research Centre for Geosciences, Section 5.3 Geomicrobiology, Telegrafenberg, D-14473 Potsdam, Germany

[2] College of Electrical Engineering, Northwest University for Nationalities, Lanzhou, 730030, China

[3] State Key Laboratory of Frozen Soils Engineering, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China

Abstract

The functional *pmoA* gene is frequently used to probe the diversity and phylogeny of methane oxidizing bacteria (methanotrophs) in various environments. Here we compared the similarities between the *pmoA* gene and the corresponding 16S rRNA gene sequences of 77 described species covering gamma- and alphaproteobacterial methanotrophs (type I and type II MOB, respectively) as well as methanotrophs from the phylum *Verrucomicrobia*. We updated and established weighted mean *pmoA* gene cutoff values on the nucleotide level at 86%, 82%, and 71% corresponding to the 97%, 95%, and 90% similarity of the 16S rRNA gene. Based on these cutoffs, the functional gene fragments can be entirely processed at the nucleotide level throughout software platforms such as Mothur or QIIME which provide a user-friendly and command based alternative to amino-acid based pipelines. Type II methanotrophs are less divergent than type I both with regard to ribosomal and functional gene sequence similarity and GC content. We suggest that this agrees with the theory of different life strategies proposed for type I and type II MOB.

Keywords *pmoA* gene, cutoff, methanotrophs, Mothur, GC content

Introduction

Aerobic methane oxidizing bacteria (MOB), also known as methanotrophs, are commonly grouped into type I and type II MOB belonging to *Gammaproteobacteria* and *Alphaproteobacteria*, respectively. Newer studies have uncovered further aerobic and intra-aerobic methanotrophs among the *Verrucomicrobia* and the NC10 phylum (Dunfield et al. 2007; Nazaries et al. 2013). Aerobic methanotrophs catalyze the oxidation of methane through the enzyme methane monooxygenase (MMO) which exists in a soluble (sMMO) and a membrane-bound, particulate form (pMMO) (Nazaries et al. 2013). The type I methanotrophs use the ribulose monophosphate (RuMP) pathway to assimilate carbon while the type II utilizes the Serine pathway of carbon assimilation (Trotsenko and Murrell 2008). The alpha subunit of the pMMO, encoded by the *pmoA* gene, is highly conserved (Hakemian and Rosenzweig 2007) and commonly used as a functional gene marker to probe methanotrophs in the environment (Luesken et al. 2011). This functional gene marker can simultaneously provide functional and taxonomic records of environmental methanotrophs (Luesken et al. 2011) and phylogenetic information as congruent as the 16S rRNA gene (Holmes et al. 1995). In practice, the *pmoA* gene fragments were mostly amplified using the forward primer A189f and the reverse primers of A682r (Holmes et al. 1995) or mb661 (Costello and Lidstrom 1999).

In the recent years, extensive sequencing efforts on MOB communities have yielded a massive amount of *pmoA* gene sequences from variable habitats. These gene fragments were commonly analyzed after translating them into amino acid sequences. One of the major advantages with amino acid over nucleotide sequences is to guarantee all sequences pass the translation check from the nucleotide to the amino acid sequence. Given all sequences are firstly checked for correct open reading frames (ORF) using tools like FrameBot (http://fungene.cme.msu.edu/FunGenePipeline/framebot/form.spr), the functional high throughput data could thereafter be further processed at the nucleotide level. In other words, the functional gene fragments could be entirely processed at the nucleotide level throughout software platforms such as Mothur or QIIME which provide a user-friendly and command-based alternative to an amino-acid based pipeline (Fig. 1). This would also allow calculating

distance matrices using Mothur and QIIME which do not yet support amino acids. In this context, precise cutoff values at the nucleotide level are very important for reasonable OTU picking and further community analysis. Moreover, a dedicated reference database of *pmoA* gene sequences can enable consensus taxonomic assignation for custom data. Dumont and his colleagues have recently proposed a *pmoA* gene reference database (Dumont et al. 2014). From them, the taxonomy is given in a special format which is not ranked as commonly in classification schemes from phylum to species.

A previous study used 22 methanotrophic isolates and proposed 13% and 7% cutoffs at the nucleotide and protein level, respectively, corresponding to a 3% dissimilarity of the 16S rRNA gene (Degelmann et al. 2010). In their study, Degelmann and colleagues provided cutoff values for the species- but not for the genus-level. Another study proposed 10% and 17% dissimilarity of *pmoA* gene nucleotide cutoffs corresponding to 3% and 5% thresholds of the 16S rRNA gene (Lüke et al. 2010) by assuming a 3.5 times higher substitution rate (Heyer et al. 2002). However, the *pmoA* gene database has been increasingly expanded with new methanotrophic isolates. Numerous new sequences are not covered by already published *pmoA* gene cutoff values which, therefore, need to be updated. In this study we aimed to evaluate common *pmoA* gene cutoff values at the nucleotide level and to establish such values for the genus and family level taking into account recently isolated methanotrophs as well. The focus is on proteobacterial methanotrophs but *Verrucomicrobia* related species were also included. We also want to disclose the meaning of individual cutoff values for type I and type II MOB.

Accumulating evidences concerning the ecological characteristics of type I and type II MOB, and community level molecular analyses under different conditions suggest that the different MOB subgroups possess distinct traits (Ho et al. 2013). For example, stable isotope labeling experiments demonstrated that type I MOB exhibit significantly higher *pmoA* gene expression level and growth rates than the type II, and are predominantly active in many

important habitats with high methane emissions (Chen et al. 2007; Qiu et al. 2008; Zheng et al. 2008; Kip et al. 2010; Zheng et al. 2010; Dumont et al. 2011; Graef et al. 2011; Zheng et al. 2012; Ho et al. 2013). On the other hand, the type II MOB population is relatively stable and serves as microbial seed bank in the soil (Eller et al. 2005; Krause et al. 2012). These different traits are explained by different r/k strategies of type I and II MOB (Bodelier et al. 2012; Ho et al. 2013). The sum of these studies renders it convincing that these two strategies exist among methanotrophs. Based on the sequence information necessary for the primary goal of this study we want to compact this theory. If type I and type II MOB possessed different life strategies, this should be reflected in the *pmoA* gene's GC (guanine-cytosine) content of type I and type II methanotrophs because the GC content is hypothesized as one of the genomic traits that relates with the variation in selection and mutational bias (Birdsell 2002). An additional, a minor objective of this work is therefore to compare the *pmoA* gene sequences of type I and type II MOB with regard to their GC content.

Material and methods

A total of 516 *pmoA* nucleotide sequences of pure cultures were originally retrieved from the FunGene database (http://fungene.cme.msu.edu). Then the corresponding 16S rRNA sequences were sequentially searched from the NCBI database by using perl scripts. After removing short and low quality sequences both for the *pmoA* and the 16S rRNA genes, we recovered a total of 77 species (for details, please refer to Table S1 and Fig. S1). These species could be assigned to 19 genera within all known classes of methanotrophs. Two filamentous methanotrophs (fMOB) *Clonothrix* and *Crenothrix* were also included together with three other *pmo*-like (*pxmA*) gene sequences. The 16S rRNA and *pmoA* gene sequences were aligned against the silva reference file (v119) by using Mothur platform (Schloss et al. 2009) and the *pmoA* sequences were aligned with pre-aligned *pmoA* sequences provided by the FunGene pipeline database. Afterwards, the distance matrices

were calculated by the *R* package of ape (v3.3) (Paradis et al. 2004) for both the 16S rRNA and the functional *pmoA* gene sequences. The distances were plotted pairwise against each other using the *R* package of ggplot2 v. 0.9.3.1 (Wickham 2009). The linear and quadratic regressions, analysis of variance (ANOVA) were performed by functions provided by the basic package in *R* (R Core Team 2014). In addition, the regression cutoff values were further evaluated by counting the agreement of both 16S rRNA genes and *pmoA* genes (Fig 2. B-D). Briefly, this evaluation process is as follows, 1) set a line at the 97%, 95% or 90% similarity of the 16S rRNA gene and move a horizontal line upwards (i.e. increase the similarity of the *pmoA* gene), the intersection in the top right represents the consistent classifications based on both genes, 2) count the intersection points and calculate their fractions according to the total points which fell within the 16S rRNA and *pmoA* genes, respectively, 3) plot these fraction values as a line chart, the ideal thresholds in agreement with both genes should be close to the intersection point of the two lines. Comparing to these ideal values helps to estimate how confident the regression derived values are. Finally, all nucleotide sequences of 16S rRNA and *pmoA* genes in this study were subject to calculating the GC content by using Biopython scripts. A pairwise plot of the GC content between both gene sequences was generated by using basic packages in *R* (R Core Team 2014). Besides, we created a more comprehensive taxonomy database for the *pmoA* gene sequences which could be probed by the primer set combination of A189f and A682r. Sequences in this database were firstly retrieved from the NCBI database and progressively screened by Biopython or *R* scripts. The corresponding taxonomy was generally referred to the NCBI taxonomy if the taxonomic ranks from phylum to species are available. For those with ambiguous taxonomies given by the NCBI database, efforts have been made to blast against the Dumont's database (Dumont et al. 2014) to improve the taxonomic classification as appropriate.

Results

Briefly, the taxonomic database consists of 7809 unaligned *pmoA* gene nucleotide sequences of methanotrophs in fasta format and corresponding taxonomy files special for Mothur and QIIME. This database included unique functional sequences of *pmoA* gene and related (*pxmA*) genes in the methanotrophs within the phyla of *Proteobateria* and *Verrucomicrobia*. This library also contains some related *amoA* sequences which are frequently co-amplified by the above mentioned primer set. The taxonomic file is a two column tabular file, with the first column containing the sequence accession number and the second showing the taxonomic information, where the taxonomic levels are separated by semicolons. The database is published as supplementary data to this article and can be accessed via http://dx.doi.org/10.5880/GFZ.5.3.2016.001 (Yang et al. 2016). These files are suitable for assigning sequences to the taxonomy using open-source software such as Mothur and QIIME.

To determine the family, genus and species cutoff values for the *pmoA* gene we referred to the common thresholds of 90%, 95% and 97% sequence identities of 16S rRNA genes, respectively. Based on a linear correlation of the pairwise distances (R^2 = 0.7603), the thresholds of 74.44%, 82.06% and 85.10% were derived for *pmoA* gene. However, the quadratic fitting analysis which resulted in cutoff values of 70.95%, 81.20% and 86.18%, appears more appropriate (R^2 = 0.8047) (Fig. 2A). Since all cutoff values contain a certain degree of arbitrariness, we systematically investigated how well the classification based on the 16S rRNA gene and the *pmoA* gene agrees for different threshold calculations, shown in Figures 2B, C and D. The intersections from them suggest *pmoA* gene thresholds of 69.92%, 83.10% and 87.22% according to 16S rRNA gene cutoff values of 90%, 95%, and 97%. Ideally, the *pmoA* gene cutoff values should coincide with the maximum fraction of pairs classified based on both genes. Therefore, the statistical cutoff values by quadratic rather than the linear fitting (Fig. 2A) are again more preferable. In this regard, it is plausible to set the *pmoA* gene nucleotide cutoff values at 86% for the species, 82% for the genus, and 71% for the family level (Table 1).

Because type I and type II methanotrophs exhibit different variability in sequence similarity (insert Fig. 2), the cutoff values of *pmoA* gene were tentatively computed for the type I and type II, respectively. For the type I methanotrophs, if including the 92 outliers scattering in the bottom right of the plot, the 78% and 83.95% identity of *pmoA* gene fragment roughly correspond to 95% and 97% similarity of 16S rRNA gene. However, the R^2 of 0.2224 indicate a relatively low representative of these values. Excluding the minor part of outliers from the total caused by *Crenothrix* and *Methylococcaceae bacterium* M200 derived sequences substantially improves the fitting goodness (R^2 = 0.5359), leading to the corresponding cutoff values of 82.04% and 86.30% (Fig. S1). For the type II MOB, the thresholds are 82.13% and 87.19% equivalent to 95% and 97% of similarity of the 16S rRNA gene (Fig. S2). Thus, the overall cutoff is very robust and can also individually be applied for type I and type II MOB with a slight underestimation of type II species richness using the general cutoff.

The GC content of the *pmoA* gene fragments used in this study showed high variations among type I MOB while they only displayed a narrow range among type II MOB which also had a generally higher GC content (Fig. 3). ANOVA test suggests statistical significance (p < 0.001) in GC content between the 16S rRNA and *pmoA* gene fragments (the inset boxplot in Fig. 3), but not between the alpha- and gamma- 16S rRNA genes (p = 0.37). The narrow range of GC values may partly be biased by the uneven distribution of species per genotype. There are, for example, 22 species belonging to *Methylocystis* (Fig. S1), and they had to be expected to have similar GC values. However, species of genotypes *Methylobacter* and *Methylomonas* (type I MOB) vary greatly with regard to their *pmoA* gene's GC values despite their affiliation to the same genus.

Discussion

According to our results, the *pmoA* gene cutoff value (14% dissimilarity) for species is 4.7 times higher than the 3% cutoff for 16S rRNA gene, and modifies the Degelmann cutoff (13%) by 1%. Our substitution rate is higher than the previously published rate of 3.5 for partial *pmoA* genes of type II MOB (Heyer et al. 2002). The regression of the pairwise plot could cover most of the species (R^2 = 0.8047). The residuals are mainly due to *Crenothrix* and *Methylococcaceae bacterium* M200, which scattered as outliers in Figure 2. *Crenothrix* belongs to the gammaproteobacterial methanotrophs in terms of the 16S rRNA gene, but it shows a very divergent *pmoA* gene (Stoecker et al. 2006). The strain M200 shows highest homology on the 16S rRNA gene with various uncultured bacteria from different ecosystems, but it shares only 71-72% homology on the functional gene level with its closest relatives *Methylobacter marinus* and *Methylomicrobium album* (Kip et al. 2011; Tavormina et al. 2011). Moreover, the *pxm* operon has a different origin and the order of genes is uniquely organized in the non-canonical form different to that in all reported *amo* and *pmo* operons (Tavormina et al. 2011). These two species also contributed the major outliers in the pairwise plot of type I methanotrophs (Fig. S1). Among the type II methanotrophs, the outliers are mainly caused by *Methylocapsa acidiphila*, which forms a deep divergent clade to the other alphaproteobacteria methanotrophs (Degelmann et al. 2010).

The cutoff values based on the regression model are a weighted mean for the two groups. Since the type II MOBs have higher similarities and smaller variability than the type I MOB, the tradeoff thresholds should be a bit lower than the actual ones for the type II MOB while they should be slightly higher for the major type I MOB. The individual cutoff values for type II and type I reflect this difference (Fig. S1, 2), although the regression fitness is slightly poor for both groups. As a consequence, assigning OTUs based on the generalized cutoff values lead to slightly underestimated species richness for type II and slightly inflated diversity for type I MOB. Despite these minor limitations, the proposed cutoff values cover the majority of methanotrophic species and represent the most reliable values to date. They allow for a more precise estimation of methanotrophic diversity in the environment.

We have illustrated that the *pmoA* gene's GC content of the type II MOB is generally higher than that of type I MOB and show a rather narrower variability (Fig. 3). As mentioned earlier, the GC content is suggested to be a genomic trait that relates for example with the variation in selection (Birdsell 2002). Although the hypothesis that the GC content plays a vital role in temperature adaptation has been refuted (Hurst and Merchant 2001), a recent gene-centric association analysis demonstrated that correlation exists at least for certain genomic regions (Zheng and Wu 2010). Experimental evidence shows that the dormant type II MOB can become metabolically active with higher methane uptake rates in response to an exposure to elevated temperatures (Whittenbury et al. 1970; Ho and Frenzel 2012), suggesting the high-GC type II can positively respond to higher temperatures. Type I MOB are active in various environments with high methane emissions, while type II MOB are relatively stable and are assumed to be present often in dormant states (Ho et al. 2013). Type II MOB also exhibited slower growth rates reflected in lower mRNA transcripts per cell (Steenbergh et al. 2010). However, some type II MOBs are less dependent on the availability of other nutrients besides methane (Steenbergh et al. 2010). Some facultative type II MOBs are able to utilize more versatile substrates than the type I (Ho et al. 2013). Therefore, in some case, the type II MOB, although largely dormant, could ultimately dominate the total MOB population following disturbance as type I adversely response to the disturbance (Ho et al. 2011). These different features and life strategy give them a survival advantage. In the long process of evolution, type II MOB, being slow-growing as well as capable of dormancy under unfavorable conditions, tend to propagate to a much lower extent compared to type I populations, and may simultaneously have accumulated fewer mutations in the genomic sequences. This possibly also had an effect on the differences in GC content we described.

Acknowledgement

The financial support to Xi Wen provided by the China Scholarship Council (CSC) (Grant No. 201408620031) is gratefully acknowledged.

Reference

Birdsell JA (2002) Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Molecular Biology and Evolution 19:1181-1197

Bodelier PLE, Bar-Gilissen MJ, Meima-Franke M, Hordijk K (2012) Structural and functional response of methane-consuming microbial communities to different flooding regimes in riparian soils. Ecology and Evolution 2:106-127 doi: 10.1002/ece3.34

Chen Y, Dumont MG, Cébron A, Murrell JC (2007) Identification of active methanotrophs in a landfill cover soil through detection of expression of 16S rRNA and functional genes. Environmental Microbiology 9:2855-2869 doi: 10.1111/j.1462-2920.2007.01401.x

Costello AM, Lidstrom ME (1999) Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. Applied and Environmental Microbiology 65:5066-5074

Degelmann DM, Borken W, Drake HL, Kolb S (2010) Different atmospheric methane-oxidizing communities in European beech and Norway spruce soils. Applied and Environmental Microbiology 76:3228-3235 doi: 10.1128/Aem.02730-09

Dumont MG, Luke C, Deng YC, Frenzel P (2014) Classification of *pmoA* amplicon pyrosequences using BLAST and the lowest common ancestor method in MEGAN. Frontiers in Microbiology 5:34. doi: 10.3389/fmicb.2014.00034 doi: 10.3389/fmicb.2014.00034

Dumont MG, Pommerenke B, Casper P, Conrad R (2011) DNA-, rRNA- and mRNA-based stable isotope probing of aerobic methanotrophs in lake sediment. Environmental microbiology 13:1153-1167 doi: 10.1111/j.1462-2920.2010.02415.x

Dunfield PF, Yuryev A, Senin P, Smirnova AV, Stott MB, Hou SB et al. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. Nature 450:879-U818 doi: 10.1038/nature06411

Eller G, Kruger M, Frenzel P (2005) Comparing field and microcosm experiments: a case study on methano- and methylo-trophic bacteria in paddy soil. FEMS Microbiology Ecology 51:279-291 doi: 10.1016/j.FEMSec.2004.09.007

Graef C, Hestnes AG, Svenning MM, Frenzel P (2011) The active methanotrophic community in a wetland from the High Arctic. Environmental Microbiology Reports 3:466-472 doi: 10.1111/j.1758-2229.2010.00237.x

Hakemian AS, Rosenzweig AC (2007) The biochemistry of methane oxidation. Annual Review of Biochemistry 76:223-241 doi: 10.1146/annurev.biochem.76.061505.175355

Heyer J, Galchenko VF, Dunfield PF (2002) Molecular phylogeny of type II methane-oxidizing bacteria isolated from various environments. Microbiology 148:2831-2846

Ho A, Frenzel P (2012) Heat stress and methane-oxidizing bacteria: Effects on activity and population dynamics. Soil Biology and Biochemistry 50:22-25 doi: http://dx.doi.org/10.1016/j.soilbio.2012.02.023

Ho A, Kerckhof F-M, Luke C, Reim A, Krause S, Boon N et al. (2013) Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies. Environmental Microbiology Reports 5:335-345 doi: 10.1111/j.1758-2229.2012.00370.x

Ho A, Luke C, Frenzel P (2011) Recovery of methanotrophs from disturbance: population dynamics, evenness and functioning. ISME J 5:750-758 doi: http://www.nature.com/ismej/journal/v5/n4/suppinfo/ismej2010163s1.html

Holmes AJ, Costello A, Lidstrom ME, Murrell JC (1995) Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. FEMS Microbiology Letters 132:203-208 doi: 10.1111/j.1574-6968.1995.tb07834.x

Hurst LD, Merchant AR (2001) High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. Proceedings of the Royal Society B: Biological Sciences 268:493-497 doi: 10.1098/rspb.2000.1397

Kip N, Ouyang WJ, van Winden J, Raghoebarsing A, van Niftrik L, Pol A et al. (2011) Detection,isolation, and characterization of acidophilic methanotrophs from sphagnum mosses. Applied and Environmental Microbiology 77:5643-5654 doi: 10.1128/Aem.05017-11

Kip N, van Winden JF, Pan Y, Bodrossy L, Reichart GJ, Smolders AJP et al. (2010) Global prevalence of methane oxidation by symbiotic bacteria in peat-moss ecosystems. Nature Geoscience 3:617-621 doi: 10.1038/Ngeo939

Krause S, Luke C, Frenzel P (2012) Methane source strength and energy flow shape methanotrophic communities in oxygen-methane counter-gradients. Environmental Microbiology Reports 4:203-208 doi: 10.1111/j.1758-2229.2011.00322.x

Luesken FA, Zhu BL, van Alen TA, Butler MK, Diaz MR, Song B et al. (2011) *pmoA* primers for detection of anaerobic Methanotrophs. Applied and Environmental Microbiology 77:3877-3880 doi: 10.1128/Aem.02960-10

Lüke C, Krause S, Cavigiolo S, Greppi D, Lupotto E, Frenzel P (2010) Biogeography of wetland rice methanotrophs. Environmental Microbiology 12:862-872 doi: 10.1111/j.1462-2920.2009.02131.x

Nazaries L, Murrell JC, Millard P, Baggs L, Singh BK (2013) Methane, microbes and models: fundamental understanding of the soil methane cycle for future predictions. Environmental microbiology 15:2395-2417 doi: 10.1111/1462-2920.12149

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289-290 doi: 10.1093/bioinformatics/btg412

Qiu QF, Noll M, Abraham WR, Lu YH, Conrad R (2008) Applying stable isotope probing of phospholipid fatty acids and rRNA in a Chinese rice field to study activity and composition of the methanotrophic bacterial communities in situ. ISME Journal 2:602-614 doi: 10.1038/ismej.2008.34

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing. In, R version 3.0.3 edn, Vienna, Austria

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-7541 doi: 10.1128/Aem.01541-09

Steenbergh AK, Meima MM, Kamst M, Bodelier PLE (2010) Biphasic kinetics of a methanotrophic community is a combination of growth and increased activity per cell. FEMS Microbiology Ecology 71:12-22 doi: 10.1111/j.1574-6941.2009.00782.x

Stoecker K, Bendinger B, Schoning B, Nielsen PH, Nielsen JL, Baranyi C et al. (2006) Cohn's *Crenothrix* is a filamentous methane oxidizer with an unusual methane monooxygenase. Proceedings of the National Academy of Sciences of the United States of America 103:2363-2367 doi: 10.1073/pnas.0506361103

Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MSM, Klotz MG (2011) A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. Environmental Microbiology Reports 3:91-100 doi: 10.1111/j.1758-2229.2010.00192.x

Trotsenko YA, Murrell JC (2008) Metabolic aspects of aerobic obligate methanotrophy. In: Laskin AI, Sariaslani S, Gadd GM (eds) Advances in Applied Microbiology. Academic Press, London, pp 183-229

Whittenbury R, Davies SL, Davey JF (1970) Exospores and cysts formed by methane-utilizing bacteria. Microbiology 61:219-226 doi:10.1099/00221287-61-2-219

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer New York, 2009. In:

Yang S, Wen X, Liebner S (2016) pmoA gene reference database (fasta-formatted sequences and taxonomy). GFZ Data Services, Potsdam, http://dx.doi.org/10.5880/GFZ.5.3.2016.001

Zheng H, Wu H (2010) Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. BMC Bioinformatics 11:S7 doi: 10.1186/1471-2105-11-S11-S7

Zheng Y, Liu XZ, Zhang LM, Zhou ZF, He JZ (2010) Do land utilization patterns affect methanotrophic communities in a Chinese upland red soil? Journal of Environmental Sciences 22:1936-1943 doi: 10.1016/S1001-0742(09)60342-9

Zheng Y, Yang W, Sun X, Wang SP, Rui YC, Luo CY et al. (2012) Methanotrophic community structure and activity under warming and grazing of alpine meadow on the Tibetan Plateau. Applied Microbiology and Biotechnology 93:2193-2203 doi: 10.1007/s00253-011-3535-5

Zheng Y, Zhang LM, Zheng YM, Di HJ, He JZ (2008) Abundance and community composition of methanotrophs in a Chinese paddy soil under long-term fertilization practices. Journal of Soils and Sediments 8:406-414 doi: 10.1007/s11368-008-0047-8

Table 1. cutoff values for *pmoA* gene nucleotide sequences

	Species	Genus	Family	Reference
pmoA	87%	-	-	Degelmann et al. 2010
pmoA	90%	83%	-	Lüke et al. 2010
pmoA	86%	82%	71%	This study
Type I MOB	86%	82%	-	This study
Type II MOB	87%	82%	-	This study

Figures

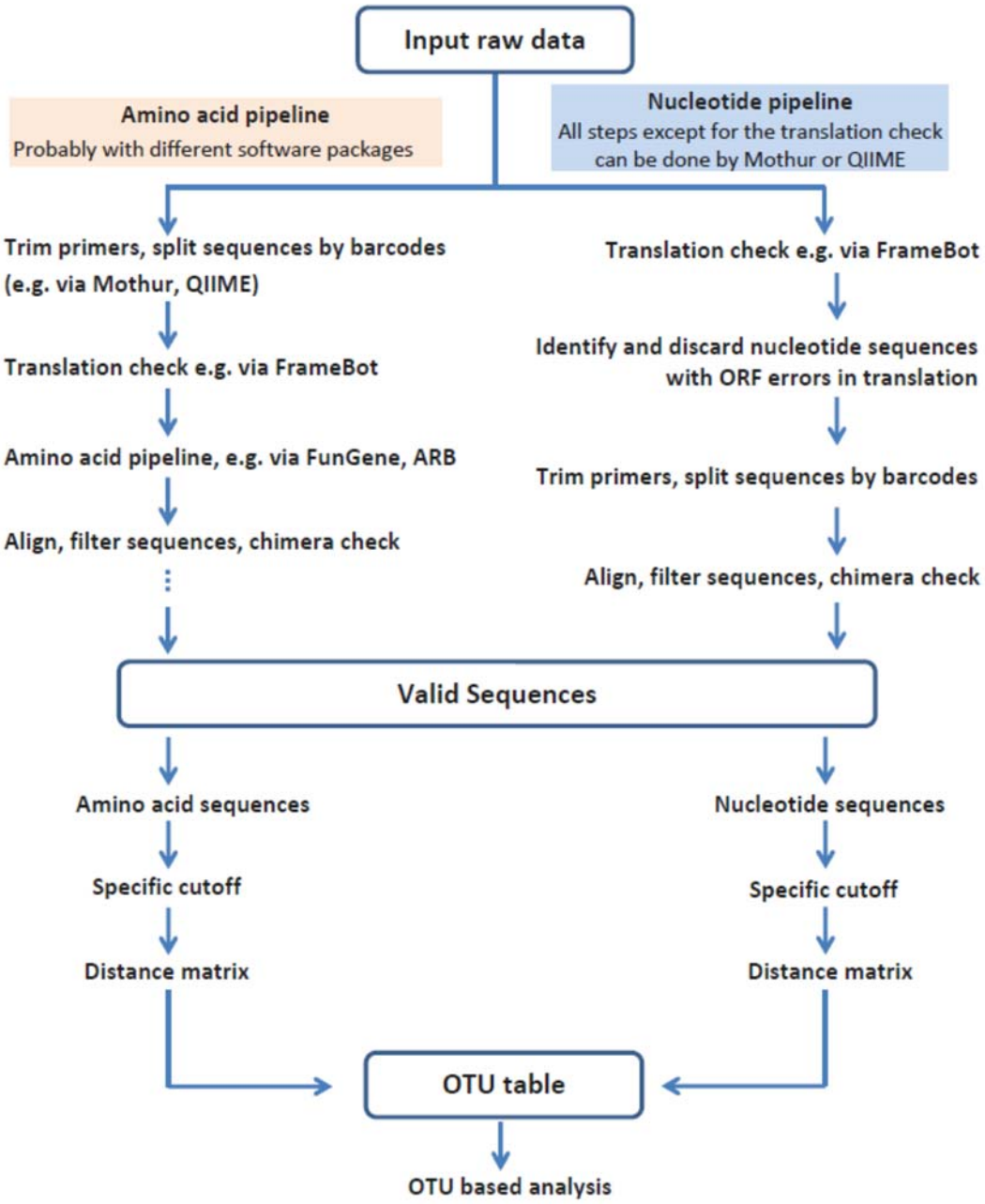

Fig 1. Schematic flowchart illustrating pipelines at nucleotide and amino acid levels for processing *pmoA* gene NGS data. Note that except for the very initial step the nucleotide pipeline can be performed throughout the Mothur or QIIME platforms.

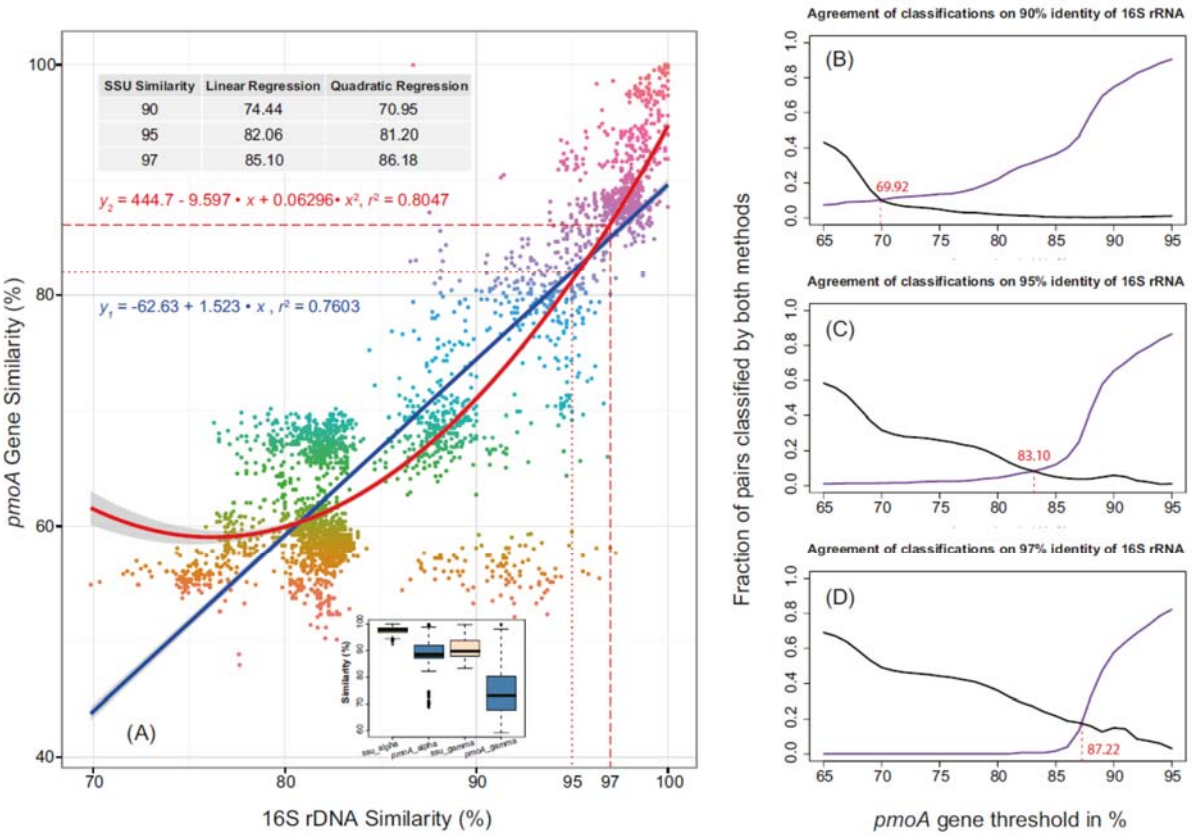

Fig. 2. Pairwise correlation of similarities based on the 16S rRNA gene versus the *pmoA* gene. The color of the dots in Fig. 2A indicates the numerical range of *pmoA* gene similarity. The polynomial and linear fitting lines in 2A are in red and blue, respectively. The inset boxplot describes the quantile statistic of similarities of 16S rRNA and *pmoA* genes of type I (gammaproteobacterial) and type II (alphaproteobacterial) methanotrophs. ANOVA test displayed statistical significance in sequence similarities between the two genes ($p < 0.001$). In 2B, C, D fixed cutoff-values for the 16S rRNA gene are assumed at 90%, 95%, and 97% sequence identity while the functional *pmoA* gene threshold is variable. The purple lines show the fraction of pairs classified on the *pmoA* gene level. Evidently, for a low functional threshold this ratio is 1, while for large values it drops to zero. Conversely, the black lines show the fraction of pairs classified on the 16S rRNA gene level. This value increases to 1 for large functional thresholds.

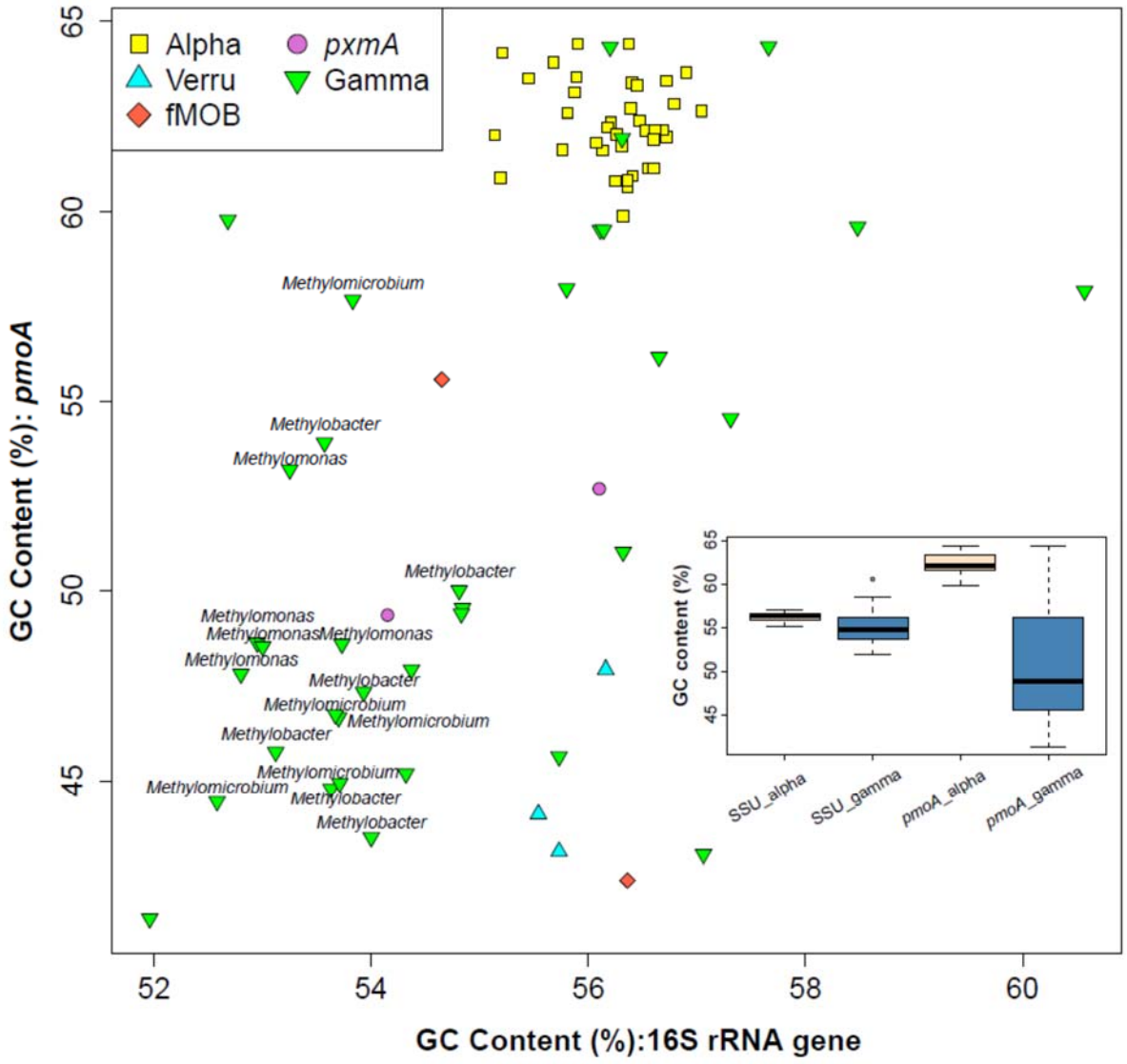

Fig. 3 Pairwise scatter plot of GC contents between the functional and 16S rRNA gene fragments across MOB used in this study. Specified symbols are used to discriminate different subgroups of methanotrophs (Alpha: type II MOB; Gamma: type I MOB; *pxmA*: methanotrophs with *pxmA* gene fragment; fMOB: filamentous methanotrophic *Crenothrix polyspora* and *Clonothrix fusca*; Verru: *Verrucomicrobia*). The inset boxplot illustrates the statistical quantiles of GC content of the 16S rRNA gene and the *pmoA* gene fragments.