Environmental Microbiology (2003) **5**(7), 566–582

Development and validation of a diagnostic microbial microarray for methanotrophs

Levente Bodrossy,[1]* Nancy Stralis-Pavese,[1] J. Colin Murrell,[2] Stefan Radajewski,[2] Alexandra Weilharter[1] and Angela Sessitsch[1]
[1]*Department of Biotechnology, Division of Life and Environmental Sciences, ARC Seibersdorf research GmbH, A-2444 Seibersdorf, Austria.*
[2]*Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, UK.*

Summary

The potential of DNA microarray technology in high-throughput detection of bacteria and quantitative assessment of their community structures is widely acknowledged but has not been fully realised yet. A generally applicable set of techniques, based on readily available technologies and materials, was developed for the design, production and application of diagnostic microbial microarrays. A microarray targeting the particulate methane monooxygenase (*pmoA*) gene was developed for the detection and quantification of methanotrophs and functionally related bacteria. A microarray consisting of a set of 59 probes that covers the whole known diversity of these bacteria was validated with a representative set of extant strains and environmental clones. The potential of the *pmoA* microarray was tested with environmental samples. The results were in good agreement with those of clone library sequence analyses. The approach can currently detect less dominant bacteria down to 5% of the total community targeted. Initial tests assessing the quantification potential of this system with artificial PCR mixtures showed very good correlation with the expected results with standard deviations in the range of 0.4–17.2%. Quantification of environmental samples with this method requires the design of a reference mixture consisting of very close relatives of the strains within the sample and is currently limited by biases inherent in environmental DNA extraction and universal PCR amplification.

Received 21 January, 2003; accepted 27 February, 2003. *For correspondence. E-mail levente.bodrossy@arcs.ac; Tel. (+43) (0) 50550 3548; Fax (+43) (0) 50550 3444.

Introduction

Methanotrophs are bacteria capable of utilizing methane as their sole source of carbon and energy. They are ubiquitous in nature and represent the largest biogenic sink for the greenhouse gas methane. They oxidize methane via methanol and formaldehyde to carbon dioxide or incorporate carbon from methane into cell biomass at the oxidation level of formaldehyde. The first step in the pathway is catalysed by one of the two types of the enzyme methane monooxygenase (MMO). The soluble MMO is found in only some of these bacteria whereas the particulate MMO (pMMO) is present in all known methanotrophs (Hanson and Hanson, 1996) except for one, *Methylocella palustris* (Dedysh *et al*., 2000). The sequence of the *pmoA* gene encoding the 27 kDa subunit of pMMO has been shown to reflect evolutionary relationships amongst the carrying organisms. The ammonia monooxygenase (AMO) of autotrophic ammonia oxidizing bacteria (AOB) is evolutionarily related to pMMO and *pmoA*, the gene encoding for the corresponding subunit of the AMO has a high degree of identity with *amoA* genes (McDonald and Murrell, 1997). Both *pmo*A and *amo*A genes can be present in one to three, generally highly, similar copies in the genomes of methane and ammonia oxidisers (McTavish *et al*., 1993; Semrau *et al*., 1995; Purkhold *et al*., 2000; Bourne *et al*., 2001). There are *pmoA*/*amoA* related genes cloned from environmental samples, where the nature of the encoded enzyme is not clear (Holmes *et al*., 1999; Henckel *et al*., 2000a; Bourne *et al*., 2001). There is a rapidly growing database (with over 700 entries) of *pmoA*, *amoA* and related gene sequences from cultivated strains and 'environmental' clones retrieved directly from the environment by PCR. Methanotrophs play an essential role in mitigating the greenhouse effect by metabolizing most of the biogenically produced methane. Understanding the factors influencing their diversity is thus of crucial importance. Denaturating gradient gel electrophoresis (DGGE) and clone library analysis have been used to monitor seasonal changes in methanotroph diversity or changes induced by environmental impacts, such as drainage or aeration (Henckel *et al*., 2000b; Henckel *et al*., 2001; Reay *et al*., 2001). The most comprehensive oligonucleotide probe set for methanotrophs so far was designed by Gulledge *et al*. (2001) and targeted the 16S rRNA gene.

DNA microarrays are a powerful tool for the parallel, high-throughput detection and quantification of many

genes. Originally developed for whole genome gene expression analyses (Schena *et al*., 1996; Tao *et al*., 1999), DNA microarrays have very strong application potential in many areas of microbiology. Upon availability of corresponding probe sets, they enable the detection of up to several thousand microbial strains, species, genera or higher clades (depending on the design of the probe) in a single assay. In clinical, veterinary and plant microbiology, food and water quality control, this means that a single test can be developed to detect all pathogenic/beneficial/contaminating bacteria which might be present in the sample being investigated. The potential for environmental microbiology is even stronger. By applying nested sets of oligonucleotide probes (Behr *et al*., 2000) which target genes reflecting the phylogeny of the target organism, it becomes possible to assess the whole prokaryotic diversity of an environment. The most obvious target for such studies are the 16S and 23S rRNA genes. Other (often referred to as 'functional', i.e. encoding for related enzymes carrying out a defined function) possible target genes can also be used, at least for a physiologically restricted group of microorganisms. The application of functional genes narrows down the analysis to a functionally (sometimes also phylogenetically) defined group of microbes. The main advantage of this approach is that it also enables the detection and analysis of uncultivated members of the microbial groups being investigated (Raskin *et al*., 1994; McDonald and Murrell, 1997; Purkhold *et al*., 2000; Radajewski *et al*., 2000; Lovell *et al*., 2001; Lueders *et al*., 2001). In contrast to 16S rRNA, environmental sequences belonging to a novel, but functionally related phylum can easily be recognised, and included in the analysis.

Oligonucleotide probe sets spotted onto nylon or nitrocellulose membranes ('macroarrays') have been used for the diagnosis of bacteraemia (Anthony *et al*., 2000), food-contamination (Rudi *et al*., 2002), detection of enterococci (Behr *et al*., 2000) or cyanobacteria (Rudi *et al*., 2000). A special microarray format consisting of individual polyacrylamide gel micropads with immobilized oligonucleotides (Liu *et al*., 2001) was used for the characterization of aromatic hydrocarbon degrading consortia (Koizumi *et al*., 2002) and identification of rifampicin resistant strains of *Mycobacterium tuberculosis*. (Mikhailovich *et al*., 2001) A 'traditional' glass microarray consisting of seven probes was developed for *Staphylococcus* diagnosis (Hamels *et al*., 2001). Different genotypes of rotaviruses were specifically detected by an oligonucleotide microarray (Chizhikov *et al*., 2002). Cho and Tiedje (2001) applied whole genome DNA–DNA hybridization for the detection and community analysis of microorganisms. A high-density Affymetrix GeneChip containing over 30 000 16S rRNA targeting oligo probes was used to identify culture collection species and subsequently to characterize populations of airborne bacteria at the level of higher phylogenetic taxa (Wilson *et al*., 2002). A 16S rRNA gene based oligonucleotide microarray targeting and covering the entire known diversity of sulphate reducers was developed and successfully validated with environmental samples (Loy *et al*., 2002).

In most cases the target consists of labelled large gene fragments (several hundred nucleotides long) increasing the potential for the accumulation of background signal arising from a low rate of non-specific hybridization. Alternative labelling approaches such as terminal transferase labelling of specific oligonucleotides (Rudi *et al*., 2002) and the application of end labelled specific stacking probes in conjunction with immobilized capture probes (Small *et al*., 2001) may improve the detection limit of diagnostic microbial microarrays by decreasing the above mentioned non-specific background hybridization.

Microbial community structures can be assessed by the quantitative analysis of micro or macroarray results. A simple and elegant method to quantify specific microbial genes from a sample has recently been published (Cho and Tiedje, 2002). In this method, gene fragments (500–900 bp in length) were applied as probes. Each spot consisted of a mixture of an individual probe and a common reference gene fragment. Hybridization was done with a Cy3-labelled environmental mixture and a Cy5-labelled reference DNA. Quantification was based on the Cy3/Cy5 ratios. Unfortunately, the same principle cannot be applied to the quantification of oligonucleotide chip results because of the inherent differences in hybridization efficiencies between oligonucleotide probes. In practice, a different reference oligo for each probe should be designed.

Even though high expectations exist for diagnostic microbial microarrays, applications utilizing their full potential (high throughput detection of large numbers of microbes and quantitative assessment of their community structures) are still very limited (Loy *et al*., 2002). Part of the reason for this is that a technical platform of reasonable cost, based on readily available consumables and equipment, with demonstrated potential in high-throughput quantitative microbial diagnostics, is missing.

Here we describe the development and validation of a diagnostic microbial microarray for the high-throughput detection and community structure analysis of methanotrophic and functionally related bacteria. A set of techniques based on well established and widely applied technologies and commercially available consumables and equipment is reported.

Results and discussion

Protocol optimization

One of the main problems when designing oligonucleotide microarrays is to achieve nearly identical melting temper-

atures for all the probes on the array. One potential approach to achieving this is the use of hybridization buffers containing tertiary amine salts. Tetramethyl ammonium chloride (TMACl) or tetraethyl ammonium chloride (TEACl) have been successfully applied to enable GC-content independent hybridization on nitrocellulose or nylon membranes (Wood *et al*., 1985; Spiro *et al*., 2000). The present weakness of this approach is the lack of detailed information on the thermodynamics of hybridizations in such solutions, making it impossible to predict the effects of further factors (as detailed later in this paper) on the behaviour of a given oligo probe. Further, the effect of tertiary amine salts on the different present and future surface chemistries is largely unknown. Thus we chose another approach by designing oligos which should have nearly identical melting temperatures in 'traditional' hybridization buffers.

Many independent studies have shown that steric effects (interference of the solid support on the hybridization properties of the immobilized oligos and steric hindrance resulting from the crowding of immobilized oligos) can seriously hinder the accessibility of immobilized oligonucleotide probes (Gou *et al*., 1994; Shchepinov *et al*., 1997; Brown and Anthony, 2000). These effects are successfully mitigated by the application of spacer molecules. In our hands, a C12 linker and an extra five thymidine residues at the 5′ end provided optimal spacing; addition of further thymidine residues had no significant effect on the hybridization capacity (accessibility) of the oligos tested (data not shown).

Dimethylsulphoxide (DMSO) (50% v/v in dH_2O) was selected as the printing buffer because it did not dry out during long spotting rounds and provided uniform spots on the slides. Standard deviations in signal intensities between replicate spots were 10–15% as opposed to 20–30% for arrays printed from 3× SSC (data not shown). Spotting was done with a single pin to avoid variations inherent in spotting with multiple pins.

Several alternative approaches (methods are available from the authors upon request) were tested for target preparation. These included direct incorporation of Cy-labelled dNTPs into dsDNA during PCR, application of a labelled PCR primer for the generation of dsDNA targets and application of a labelled and a biotinilated PCR primer for the generation of ssDNA targets via subsequent separation of the two strands using streptavidin coated magnetic beads. As the secondary structure of the target plays a very significant role in determining the maximal hybridization signal obtainable with a given probe ('hybridization capacity'), it was an absolute necessity to fragment the target before hybridization. RNA targets were generated because RNA can be fragmented in a random manner via chemical fragmentation (Hughes *et al*., 2001). A further advantage of using RNA targets is that the direct incorporation of the Cy-labelled nucleotides by the T7 RNA polymerase is very efficient. (Only a fourfold decrease was observed in RNA yield when 50% of UTP was replaced with Cy3-UTP during *in vitro* transcription.) Yields of target preparation were in the range of 50 ng μl^{-1} concentration (2500 ng) with every 10th to 12th nucleotide being labelled. Fragmentation of the target RNA to an average fragment size of 50 nucleotide (Hughes *et al*., 2001) resulted in a significant enhancement of hybridization. In spite of the sensitivity of the Cy dyes to nucleophilic attack applied for RNA fragmentation, an increase by over an order of magnitude in the Cy3, as well as in the Cy5, signals was observed (compared to signals obtained with unfragmented target RNA). Furthermore, in many cases fragmentation decreased the differences between the hybridization capacities of probes (data not shown). Hybridization conditions were chosen which were relaxed enough to enable the hybridization of both perfect match (PM) and single mismatch (1 MM) targets.

As the hybridization between oligonucleotides and any other type of nucleic acid (oligonucleotide, gene fragment, RNA, etc.) is a reversible process, all hybridizations were carried out overnight (16 h) to ensure complete hybridization. In static microarray hybridizations diffusion is the only process providing mixing. Statistical errors in the final signal are inherent in such a diffusion limited system. Furthermore, such circumstances usually result in the edges of probe sets hybridizing more efficiently, thus yielding a relatively strong signal for the edge of the spot and a much weaker one for the centre. To overcome these limitations, a custom modified BellyDancer laboratory shaker and sticky hybridization chambers with significantly higher volumes than those which exist between a microarray and a traditional coverslip, were used. The tiny bubbles unavoidably formed within this chamber were slowly moved to the edge due to the motion of the BellyDancer, and this provided enough extra mixing to ensure a reasonably uniform hybridization across the whole microarray.

Standard deviation of results for (triplicate) spots of individual probes was 3% to 30% between parallel slides hybridized with targets prepared in parallel (slide-to-slide variation).

The extent to which the application of different *pmoA*/*amoA* specific PCR primers influenced the results was investigated. Two alternative reverse primers were used for target amplification. The primer *pmoA*682 can amplify *pmoA*, *amoA* and similar sequences and yields a product of 531 bp with the forward primer *pmoA*189. The primer mb661 is specific for *pmoA*-type sequences and yields a slightly shorter (508 bp) product (Bourne *et al*., 2001). As many of the environmental *pmoA* clones were obtained with the latter primer, both primer pairs were applied. Standard deviations between results generated with the different primer pairs were 10–30%, not exceeding that

between parallel experiments using the same primer pairs.

Probe set design and validation

A database of over 700 *pmoA*/*amoA* sequences was established from public database entries and in part from unpublished sequences (courtesy of A. Auman, S. Dedysh, P. Dunfield, W. Liesack, I. McDonald, S. Morris and S. Nold). Full-length and nearly full-length [i.e. covering positions 206–661 of the *Mc. capsulatus* (Bath) *pmoA* gene] as well as some of the (unique) shorter sequences (514 entries) were used to create a phylogenetic tree.

One hundred and eighteen *pmoA*/*amoA* based oligonucleotide probes were designed. A nested set of 68 probes (Table 1) was selected by omitting redundant probes. The

Table 1. Set of oligonucleotide probes synthesized and spotted for evaluation. Positions of mismatches with target sequences are indicated by boldfaced and underlined characters. Melting temperatures were calculated by the nearest neighbour method.

Name[a]	Intended specificity	Sequence 5′ → 3′[b]	Length	T_m	Selected[c]	MM[d]
Mb460	*Methylobacter*	GACAGTTACAGCGGTAATCGGTGG	24	60.9	+	
Mb478	*Methylobacter*	TGGTATGGC**A**TGGGGTCTGT	20	59.7	+	T
Mb271	*Methylobacter*	TTGTGGTGGCGTTACCGT	18	58.0	+	
511–436	*Methylobacter*	GTTTTGATGCTGTCTGGCAG	20	55.5	+	
BB51-299	*Methylobacter*	GCGCGGTTGTTTGTGTCT	18	57.4	–	
Mb292	*Methylobacter*	**C**CGTTACCGTCTGCCTT**TC**G	20	59.1	+	T AT
Mm275	*Methylomonas*	GTGGTGGAGATACCGTTTGCC	21	59.2	+	
Mm391	*Methylomonas*	ATTTGCTTCCCATCTAACCTG	21	54.0	–	
PS80-291	clone PS-80	ACCAATAGGCGCAACACTTAGT	22	58.3	+	
peat1–3-287	*Metylomonas*-related peat clones	AACTGCCTTTAGGCGCTACC	20	58.6	+	
Mb_SL#1–421	soda lake *Methylobacter* isolates and clones	GCGATCGTATTAGACGTTATCCTG	24	56.4	–	
Mb_SL-299	soda lake *Methylobacter* isolates and clones	GGGGTGCAACT**C**TGTGTATCTTAGG	25	60.5	+	T
Jpn284	clone Jpn 07061	ACCGTATCGCATGGGGTG	18	58.0	+	
Mm_pel467	*Methylomicrobium pelagicum*	ACTGCGGTAATCGATGGTTTGGC	23	61.6	+	
Est514	*Methylomicrobium*-related clones	AATTGGCCTATGGTTGCGCC	20	59.9	+	
LP20-644	*Methylomicrobium*-related clones	GGTACACTGCGTACTTTCGGTAA	23	58.2	+	
Mmb303	*Methylomicrobium album*	CAATGCTGGCTGTTCTGGGC	20	60.3	+	
Ia193	Type I a (*M. bacter-M. monas-M. microbium*)	GACTGGAAAGATAGACGTCT	20	51.9	–	
Ia577	Type I a (*M. bacter-M. monas-M. microbium*	TGGCTGACT**T**GCAAGGTTACC	21	58.9	+	A
Nc_oce426	*Nitrosococcus oceani*	CTTGGATGCCATGCTTGCGA	20	59.8	+	
Mth413	*Methylothermus*	CACATGGCGATCTTTTTAGACGTTG	25	58.3	+	
Mc396	*Methylococcus*	CCC**T**GCCTCGCTGGT**G**C	17	61.9	+	C A
501–286	*Methylococcus*-related clones	GTCAGCCGTGGGGCG	15	59.0	–	
fw1–639	*Methylococcus-Methylocaldum* related clones	GAAGGGCACG**C**TGCGTAC**G**	19	62.0	+	T C
M90-201	*Methylocaldum*-related clones	CGGCTGCTGTACAGGCGTTC	20	61.8	+	
Mcl408	*Methylocaldum*	GGTTCCGGGTGCGATTTTG	19	57.8	+	
Ib453	*Methylococcus-Methyocaldum* and related	GGCAGCTACCTGTTCAC**C**GC	20	61.7	+	G
Ib559	*Methylothermus-Methylococcus-Methyocaldum* and related	GGCA**T**GCTGATGTC**G**AT**T**GCCG	22	60.5	+	C C C
Mcy262	*Methylocystis*	CAGGC**G**TTCTGGTGGGT**G**AA	20	61.0	+	T T
Mcy409	*Methylocystis* and peat clones	AT**C**GTTCCGGCG**A**TCTGGC	19	61.0	+	U C +hairpin
Peat264	peat clones	GGCGTTTTTCTGGGTCAACTTCC	23	60.3	+	
Msi520	*Methylosinus*	GCGATCGCGGCTCTGCA	17	61.6	+	
Msi_tri309	*Methylosinus*	CGCGG**T**TCTGGG**T**CTG**C**T**C**	19	61.4	+	C C A G
Msi270	*Methylosinus*	GTTCTTCTGGGAGAACTTCAAGC	23	57.1	–	
Msi232	*Methylosinus*	**C**CT**G**GGCGTGAC**C**TTC**GC**	18	61.0	+	T C G TG
II510	Type II methanotrophs	CGAACAACTGGCCGGCG	17	60.0	+	
II630	Type II methanotrophs	CATGGTCGAGCGCGGC	16	59.7	+	
RA14-598	RA14 related clones	**AA**CGT**T**CG**T**ACCTCGATGCC	20	58.3	+	TT C C
B2rel260	*Methylocapsa*-related clones	GCCCAGTATTATTTCTGGACCCCAT	25	60.4	+	Most of G/C at the ends
B2-400	*Methylocapsa*	ACCTCTTTGGTCCCGGCTG	19	60.5	+	
B2all343	*Methylocapsa* and related clones	AACCGCTACACCAA**T**TTCTGGCG	23	61.8	+	C
pmoAMO3-400	clone pmoA-MO3	CCCAGATGATCCCGTCGGC	19	60.8	+	
xb6–539	Methanotroph-related clones	AGGCCGCCGAGGTCGAC	17	63.0	+	
LP21-190	Methanotroph-related clones	ATCGACTTCAAGGATCGCCG	20	58.2	–	
LP21-232	Methanotroph-related clones	ATCGTCGCCATGTGCTTCGC	20	61.9	+	
mtrof173	Universal	GGbGACTGGGACTTCTGG	18	58.2	+	
mtrof362-I	Methanotrophs	TGGGGCTGGACCTACTTCC	19	59.5	–	
mtrof656	Methanotrophs	ACCTTCGGTAAGGACGT	17	53.2	+	
mtrof661	Methanotrophs	GGTAARGACGTTGCKCCGG	19	61.9	+	
mtrof662-I	Methanotrophs	GGTAAGGACGTTGCGCCGG	19	61.9	–	
NmNc533	*Nitrosomonas-Nitrosococcus*	**C**AACCCAT**T**TGCCAATCGTTGTAG	24	58.6	+	G C

Table 1. *cont.*

Name[a]	Intended specificity	Sequence 5′ → 3′[b]	Length	T_m	Selected[c]	MM[d]
Nsm_eut381	*Nitrosomonas eutropha*	CCACTCAATTTTGTAACCCCAGGTAT	26	59.0	+	
Pl6–306	*Nitrosomonas-Nitrosococcus* related clones	GGCACTCTGTATCGTATGCCTGTTAG	26	60.5	+	
PS5-226	*Nitrosomonas-Nitrosococcus* related clones	ACCCCGATTGTTGGGATGATGTA	23	59.9	+	
NsNv207	*Nitrosospira-Nitrosovibrio*	TCAATGGTGGCCGGT**G**G	17	58.5	+	G
NsNv363	*Nitrosospira-Nitrosovibrio*	TACTGGTGGTC**G**CA**C**TA**C**CC	20	59.6	+	A T T
Nit_rel223	AOB related clones	GTCACACCGATCGTAGAGGT	20	56.9	+	
Nit_rel351	AOB related clones	GTTTGCCTGGTACTGGTGGG	20	59.2	+	
Nit_rel470	AOB related clones	CGATA**T**TCGGGGTATGGGCG	20	58.4	+	A
Nit_rel304	AOB related clones	CGCTCTGCATTCTGGCGCT	19	61.8	+	
M84P105-451	environmental clones of uncertain identity	AACAGCCTGACTGTCACCAG	20	58.1	+	
WC306–54–385	environmental clones of uncertain identity	AACGAAGTACTGCCGGCAAC	20	59.2	+	
M84P22-514	environmental clones of uncertain identity	AACTGGGCCTGGCTGGG	17	61.0	+	
gp23–454	environmental clones of uncertain identity	AACGCGCTGCTCACTGCG	18	62.3	+	
MR1-348	environmental clones of uncertain identity	AATCTTCGGTTGGCACGGCT	20	61.1	+	
gp391	environmental clones of uncertain identity	ATCTGGCCGGCGACCATG	18	61.1	+	
gp2–581	environmental clones of uncertain identity	ACATGATCGGCTACGTGTATCCG	23	60.0	+	
RA21-466	clone RA21 – environmental clone of uncertain identity	CGGCGTTCTTGGCGGCAT	18	62.4	+	

a. Numbers at the end of the probe names refer to their relative positions on the *Mc. capsulatus* (Bath) *pmoA* gene.
b. Sequences are of the sense strand.
c. Oligonucleotide probes of the final probe set are indicated by +0. Probes not selected are indicated by –.
d. Nucleotide residue(s) at mismatch position(s). Other factors included in the calculation of weighed mismatches are also indicated.

probes targeted different species, groups of species, genera as well as higher taxonomic groups of methanotrophs and related bacteria. Several broad specificity probes targeting AOBs were also designed and included in order to improve the potential of the array for analysing various environments, including those potentially dominated by AOBs. Two probes (mtrof173 and mtrof661) were designed to target the PCR primers *pmoA*189 and mb661 respectively. A third probe 'universal' to methanotrophs (mtrof362-I) was designed for a region in the middle of the *pmoA* sequence, which is reasonably conserved amongst methanotrophs.

The most critical step of the probe design process is to fine tune the probe set in a way that all probes in the set display hybridization behaviour as identical as possible. In the first stage, an attempt was made to design oligos with predicted melting temperatures [according to the nearest neighbour model (Breslauer et al., 1986)] of $60 \pm 2^{\circ}$C. In some cases this was not possible because of the limited length of differentiative sequence regions. When no alternative probe sites were found, the probes with suboptimal melting temperatures were accepted and synthesized. As present models can only predict melting temperatures of free oligos but not of those bound to solid surfaces (see detailed discussion below), probes with suboptimal predicted melting temperature do not necessarily perform suboptimally.

The hybridization behaviour of oligonucleotide probes immobilized onto a solid surface depends on several factors. Length, GC content and exact sequence of the probe: these together are considered when predicting T_m for the oligos by using the nearest neighbour method (Breslauer *et al.*, 1986). Position of GC and AT pairs: the middle of the probe is more important in stabilizing hybridization, thus a probe with most of its GC content in the middle binds to its target more strongly than another one with homogenous GC distribution (but with identical length and GC content) (Guo *et al.*, 1994; Shchepinov *et al.*, 1997; Hughes *et al.*, 2001). Secondary structures of the probe and of the corresponding target: when any of these two are of significant strength, compared to the strength of hybridization between the probe and the target, a significant drop in hybridization efficiency is expected. The exact nature of the overhanging nucleotides on the target (nucleotides immediately next to the area targeted by the probe): this comes from the nearest method model, but isn't normally accounted for because the overhangs of the target sequence are not considered. Number and type of mismatches: some mismatches have little while others have very strong destabilizing effect (Sugimoto *et al.*, 2000). Position of mismatches: mismatches in the middle are more destabilizing than mismatches at end positions (Fotin *et al.*, 1998). Factors arising from the immobilized nature of the probes: steric effects can hinder the formation of hybrids between the target and the bound probe. This effect is much stronger for the immobilized end of the probe. Thus, the bound end of the probe plays a lesser role in the hybridization than the free end (Guo *et al.*, 1994; Shchepinov *et al.*, 1997; Hughes *et al.*, 2001). This applies for the position of GC and AT pairs as well as for the position of mismatches. Hybridization between DNA oligos and RNA fragments, as in our case, has slightly different thermodynamics to that of DNA–DNA hybridization (Hung *et al.*, 1994; Sugimoto *et al.*, 2000).

Based on the above criteria, on initial results from testing the hybridization behavior of our probes and on published data (Shchepinov *et al*., 1997; Fotin *et al*., 1998; Meroueh and Chow, 1999; Sugimoto *et al*., 2000) a set of simple findings was compiled which significantly improved the prediction of the hybridization behaviour of the probes. These findings are:

(i) U→C changes in the target sequence leads to an rG-dT bond which is almost as strong as the original rA-dT bond (for perfect match cases). These are not considered as mismatches.
(ii) G→A changes in the target sequence leads to an rU-dG bond which is almost as strong as the original rC-dG bond (for perfect match cases). These mismatches are considered only if they are present together with other types of mismatches.
(iii) Mismatches in the end positions are not considered. Mismatches adjacent to the end positions are considered only if they are present together with other types of mismatches.
(iv) If most of the GC content of a probe is close to the 5′ (immobilized) end, the probe will display a significantly lower melting temperature than originally predicted. The same is true for a G or C residue in the 3′ end position.
(v) Probes with strong hairpin structures ($\Delta G \geq 2.0$) were considered as having an extra mismatch to the target.
(vi) High GC content probes shorter than 17 nucleotides in length display unreliable hybridization behavior under the experimental conditions applied.

Considering the above points, another nine probes were excluded and mismatch values were updated to weighed mismatch values. The resulting set of 59 oligonucleotide probes was validated with a reference set of 59 pure strains and environmental clones covering almost the entire known diversity of methanotrophsand bacteria carrying *pmoA*/*amoA* homologues (listed in Fig. 2). Several AOB clones were also included in the validation. Hybridizations were carried out at 55°C with the aim of allowing for perfect match and single (weighed) mismatch probe-target pairs to hybridize. Figure 1 shows some typical hybridization results. Hybridization between a probe and a target was considered positive if the signal was at least 5% of the signal obtained for mtrof173 on the same array. There were, unfortunately, groups of clones for which no representative was available. Figure 2 shows the predicted hybridization behaviour of the probe set and the results obtained.

Out of 59 probes in the probe set, we were unable to obtain reference targets for seven which were thus not possible to validate. Most (42) probes displayed hybridization behavior as predicted. Two probes, II510 and II630 were expected to have unreliable hybridization behaviour because of their shortness (in combination with a strong secondary structure in case of II510). These two probes were left in the probe set as we were unable to identify a better region for a probe specific to the Type II methanotrophs. Eight probes displayed some unpredicted results. The unexpected positive result of Mcy409 came from a combination of a strong secondary structure and a mismatch. The unexpected negative results of Msi232 was obtained with targets displaying two adjacent mismatches right at the 3′ end plus a third one at different internal locations of the probe. Probe Ia577 displayed unexpected negative results to three targets all having the a single central mismatch U(r)A change, resulting in an rU-dT pair replacing the perfect match rA-dT pair.

Fifty out of the 59 probes were succesfully validated (seven probes with no reference targets available and two suboptimal probes, II510 and II630 were not). Of the 2950 individual hybridization reactions (50 validated probes × 59 reference strains/sequences), 2931 (99.3%) yielded the expected result by either showing detectable signal where expected or by no hybridization where a negative result was predicted. Only 19 of the hybridization reactions (0.7%) resulted in false negative or positive hybridization. Forty-two out of the 50 probes considered behaved 100% as predicted (in all of the hybridization reactions). This success rate is acceptable when redundant probe sets (three or more probes for each species or higher taxonomic group targeted) are and can be designed. There is however, a need for an improved method to predict hybridization behaviour of oligonucleotide probes. About half of the unpredicted results were associated with complicated cases where additional parameters influencing hybridization behaviour had to be considered together with mismatches. Mismatches, especially when their relative positions are also considered, can reliably be accounted for only by a nearest-neighbour method based algorithm. Software applying such an algorithm, which considers further effects arising from the immobilized nature of the probes, as well as the secondary structure of the probe and the target via user defined parameters, is badly needed. A computer program under development called CALCOLIGO is aiming exactly at filling this gap (J. Csontos, Bay Zoltán Institute for Biotechnology, Szeged, Hungary, pers. comm.).

Despite its apparent shortcomings, the probe set can diagnose almost the entire known diversity of methanotrophs and bacteria carrying *pmoA*/*amoA* homologues. The gaps in the validation of probe set represent probes against unique clones or small groups of clones which seem to be very poorly represented in the environments investigated so far. Results from environmental samples need to be referred to the validation results with the reference set, rather than to the predicted ones, thus minimizing the chances of false interpretation of results.

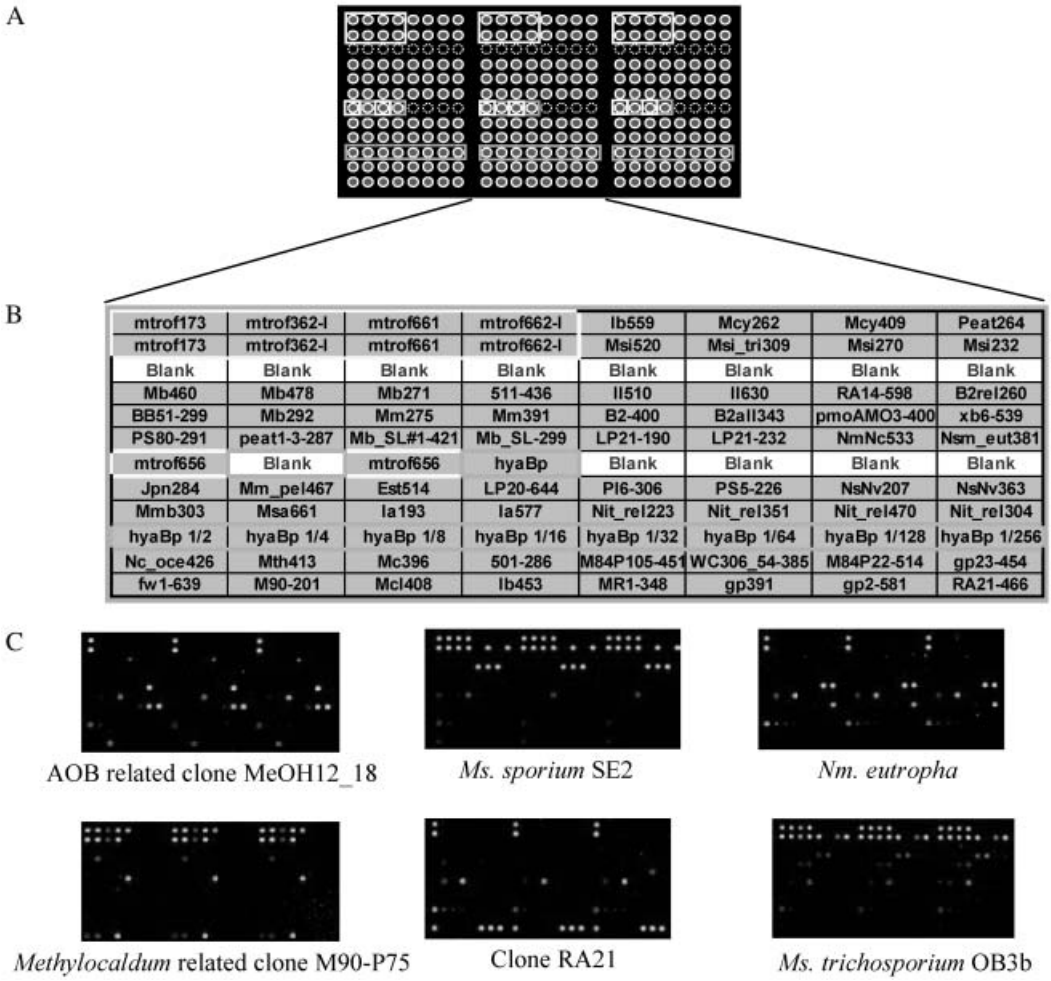


Fig. 1. A. Schematic diagram of the microarray design. Arrays were spotted in triplicate. Frames indicate universal ('mtrof') probes spotted in multiple copies and spots with an external positive control probe ('hyaBp'; results of this were not considered or used in the present study).
B. Detailed design of a single array with exact positions for each probe.
C. Representative hybridizations with reference strains or environmental clones. Probe hyaBp targets an independent gene (*hyaB* of *E. coli*). It can be applied as an alternative control spot for normalization after spiking of the *in vitro* transcription reaction with *hyaB* PCR product. Note that results of probe hyaBp were not considered throughout the work presented here. Microarray images were adjusted for best viewing (quantitative conclusions drawn from the image may be misleading).

Evaluation of the microarray with environmental samples

To assess the applicability of the developed methanotroph microarray in environmental studies, two different environmental samples were analysed.

The first experiment was carried out with a soil sample from a landfill site, collected at the end of the summer. Microarray results indicated that strains related to the genera of *Methylocaldum* and *Methylocystis* were most abundant in this sample. In addition to *Methylocaldum* and *Methylocystis pmoA* specific probes, general *pmoA* probes for the B subgroup of Type I methanotrophs and for the Type II methanotrophs also showed positive hybridization. Two further probes, Mb478 and Msi232 were positive. Mb478 was known from the validation set to hybridize strongly to *pmoA* of *Methylocaldum tepidum*, thus this signal was accounted for as arising from *Methylocaldum*-related bacteria. Msi232 is a probe with a single mismatch towards *pmoA* from most *Methylocystis* strains. During validation experiments it was shown to display strong hybridization signal with *pmoA* from one clade of *Methylocystis*. Thus (weak) positive results with Msi232 were accounted for as arising from *Methylocystis* related bacteria. To confirm these results, a *pmoA* clone library was constructed. Out of 100 clones sequenced, 91 contained inserts with high homology to *pmoA* sequences. Sixty-five per cent of these clones were related to *pmoA* from *Methylocaldum* and 31% to *pmoA* from *Methylocystis*, confirming the results of the microarray analysis. A further 3% of the clones showed highest similarity to

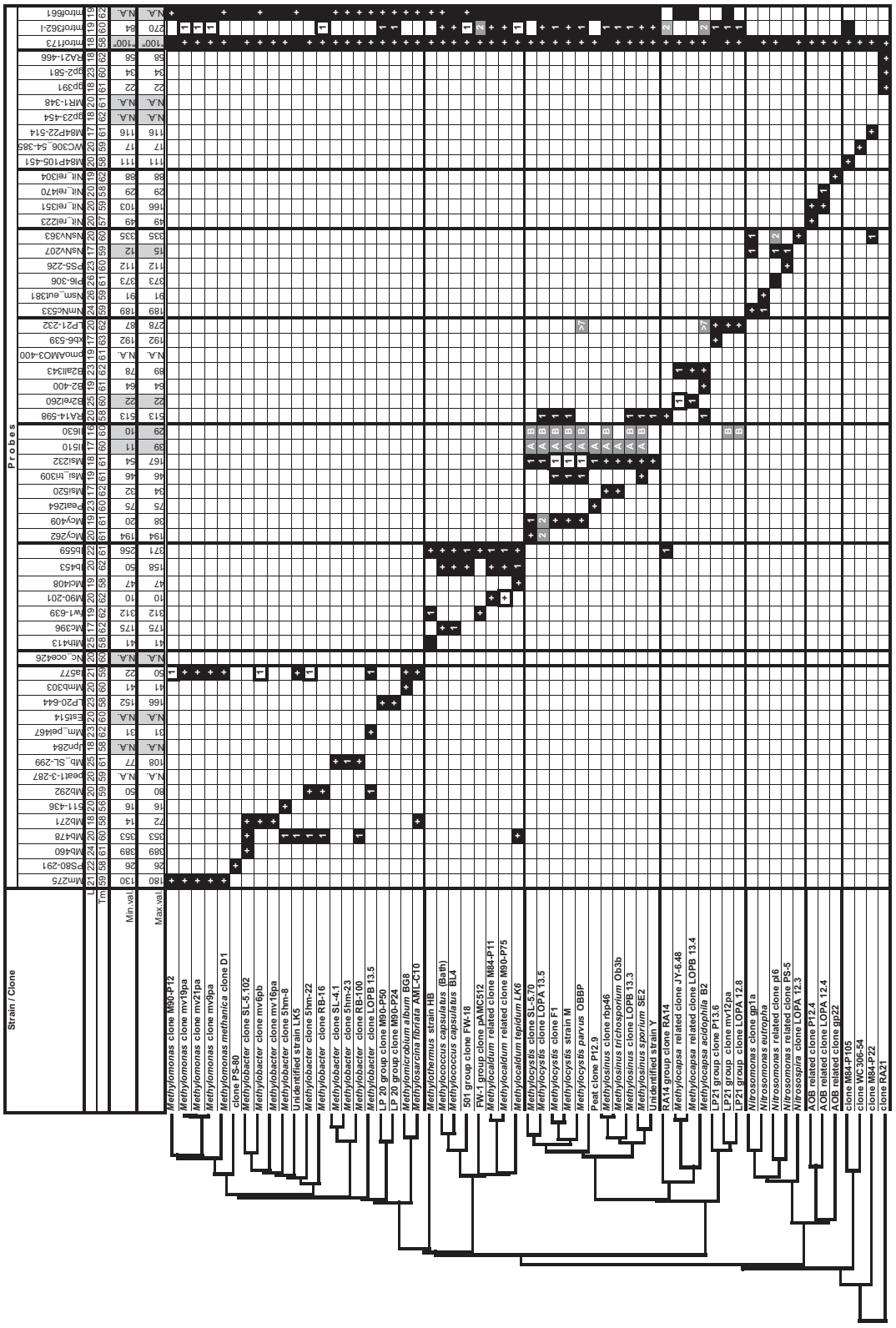


Fig. 2. Range of strain coverage for oligonucleotide probes targeting *pmoA*/*amoA* genes of methanotrophs, AOBs and bacteria carrying *pmoA*/*amoA* homologues. A similar table with all 514 sequences considered (without hybridization results) is available from the authors upon request. Under 'Probes' black fill indicates expected positive results, grey fill indicates positive results not predicted and thick black framing indicates negative results where hybridization was predicted. White numbers inside black and grey boxes indicate the number of 'mismatch equivalents' as described in the relevant section of *Results*. Letters in grey boxes indicate probes of unreliable hybridization behaviour: 'A', a short probe with significant secondary structure; 'B' a short probe. Names of environmental clones are preceded by an indication of their predicted immediate phylogenetic relationship. Highest and lowest signal values (% of that of mtrof 173) obtained with full match targets are indicated ('max.val.' and 'min.val.'). Grey boxes indicate data which are not representative (Jpn284, Est514, Nc_oce426, gp23–454 and MR1-348: no full match reference target available; Il510 and Il630: short probes of unreliable hybridization behaviour; B2rel260 and NsNV207: data from single mismatch targets; mtrof 173: reference probe; mtrof661: targeting the reverse primer applied in most amplifications). A *pmoA*/*amoA* DNA neighbour tree is added to the left providing further insight into the phylogenetic relationships of the reference set. The scale bar indicates the estimated number of base changes per nucleotide sequence position.

pmoA from *Methylomicrobium album* strains. The presence of the *Methylomicrobium album* related strains was not detected by the microarray (relevant specific probe: Mmb303). This is due to the current detection limit which is about 5% of cells in relation to the total bacterial community analysed (i.e. containing *pmoA*/*amoA* genes). There is also a statistical uncertainty in the percentage values derived from the analysis of only 91 clones. Microarray and clone library results are shown on Fig. 3.

The second experiment was done using a sample from a soil microcosm incubated with CH_4 as described by Radajewski *et al*. (2002). Microarray experiments indicated the presence of methanotrophs belonging to two clades. The first clade, the 'peat clones' group, belongs to Type II methanotrophs and is only known from acidic (peat) environments so far. There are no cultured representatives of this group yet. The second clade, the LP21 group, is related to the *Methylocapsa* and to the RA14 groups. This clade consists of environmental clones of various origins and of the 'unusual' second *pmoA* copies found in some Type II methanotrophs (Dunfield *et al*., 2002). This environmental sample has already been analysed by clone library construction and sequencing (Radajewski *et al*., 2002). Fifty *pmoA*/*amoA* clones were analysed (Fig. 4). Eighteen per cent of the clones belonged to the peat clones group and 70% to the LP21 group, confirming the microarray results detecting these two groups as the dominant ones. Four per cent of the clones analysed belonged to the *Nitrosomonas/Nitrosococcus* group and 8% to the gp2 group. The latter two groups were not detected by the microarray analysis (relevant probes were NmNc534 for the *Nitrosomonas/Nitrosococcus* group and gp391 and gp2–581 for the gp2 group). The abundance of these groups was very near to the detection level. Furthermore, the clone library and the microarray analyses were done from different PCR reactions, carried out from the same environmental DNA stock, but in different laboratories. Thus, different PCR biases may account for the discrepancy between the microarray and clone library analysis results for these low abundance sequences.

Quantification of methanotrophs and related bacteria based on *pmo*A/*amo*A sequences is potentially biased by the different number of *pmo*A/*amo*A gene copies per cell. This has to be taken into account when interpreting such results.

Quantification potential

Quantification potential of the developed microarray was tested with artificial mixtures of *pmoA* sequences. This way it was possible to avoid the introduction of biases inherent in DNA purification from environmental samples and PCR with degenerate primers (Reysenbach *et al*., 1992; Witzingerode *et al*., 1997; Polz and Cavanaugh, 1998; Ishii and Fukui, 2001) and the results show the potential of the array to reflect the composition of the PCR mixture (rather than that of the original environmental DNA or the original microbial community). By employing a reference mixture of known composition, it was possible to normalize variations in spot morphology and local differences in hybridization efficiency, as well as for the significant variation in hybridization capacities between different probes.

Our results, shown in Fig. 5, showed very good correlation between the true composition of the artificial mixtures and the results of quantitative analysis of the hybridization results. Standard deviation from expected ratios were in the range of 0.4–17.2%. These results show the potential of the microarray approach to reflect the ratios within the PCR product (used as template for target preparation). As the first steps of the procedure include environmental DNA purification and PCR with universal primers, the microarray approach is also prone to the bias inherent in these techniques.

Quantification potential was further tested with the landfill site and microcosm environmental samples (Fig. 6). Results from the first analyses (hybridization with Cy3-labelled target prepared from environmental DNA) were used to gain a rough estimate of the relative abundance of methanotrophs in these samples. Based on these results, a mixture of reference sequences covering the observed diversity was designed and labelled with Cy5. Competitive ('two-colour') hybridization with the Cy3-labelled environmental and the Cy5-labelled reference target was used to refine quantitative assessment of methanotroph community composition.

In this quantification scheme a one-colour hybridization is carried out first giving a rough estimation of the community structure by referring back the relative intensities on the array to the results from the reference strain/clone set. This information can also be used to select a subset of reference strains/clones to be used in the next stage where the same target is then hybridized against the selected reference set and quantitative data are drawn from the ratios of the two signals. The basic requirement of such a two-colour quantification approach is the ability to identify and create an appropriate reference set. This should consist of sequences as similar to those in the sample as possible. Failure to do so will result in skewed predicted ratios. Even though this limits the application potential of this approach, it can be very useful in studies where the same community is analysed over time or under different conditions.

By comparing the results of the two-colour microarray hybridizations to the composition of the corresponding clone libraries (landfill site: 14–30% predicted for *Methylocaldum* and 21–28% for *Methylocystis* versus 65% and

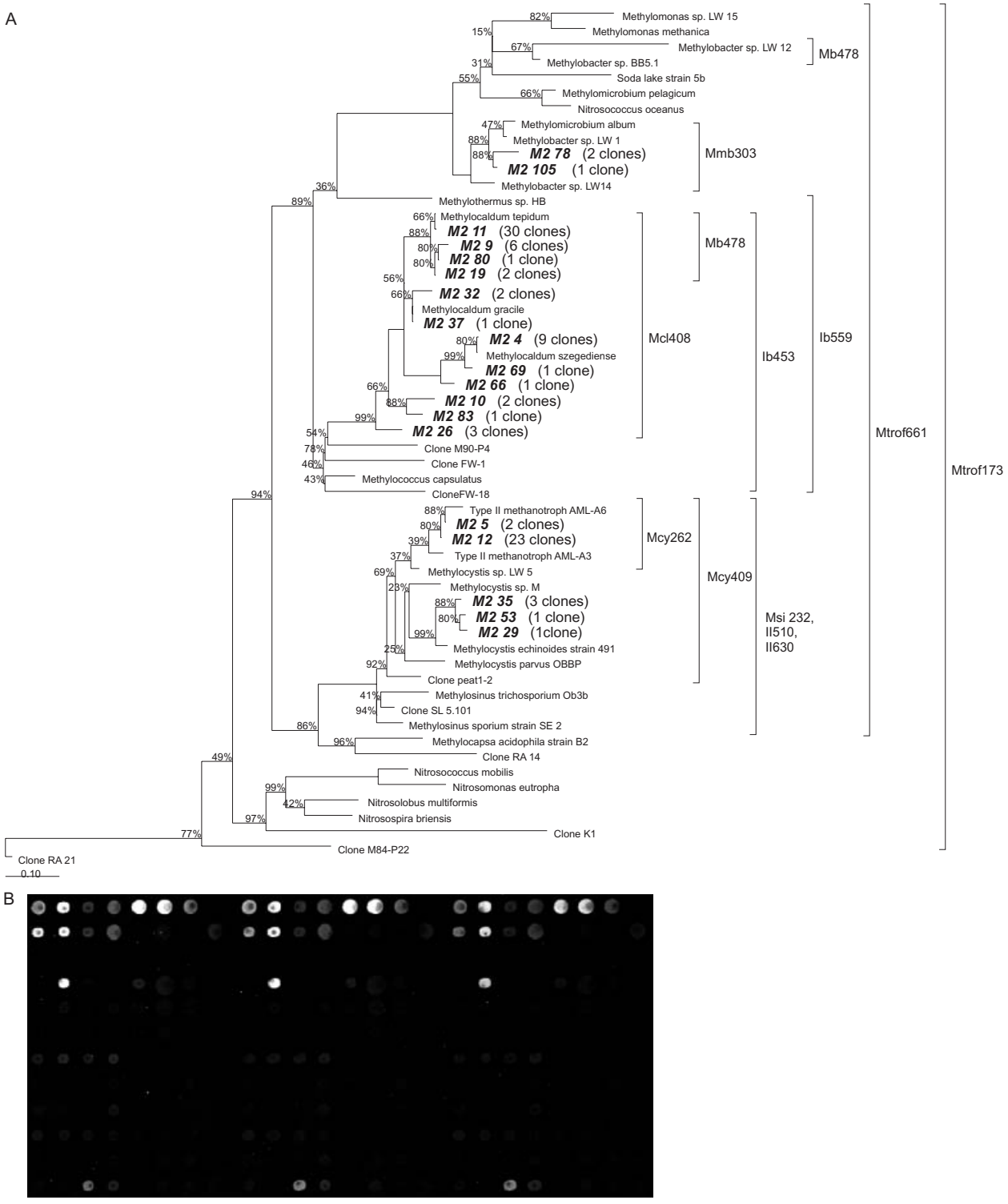


Fig. 3. A. *pmoA*/*amoA* DNA neighbour tree showing the phylogenetic positions of the clones obtained from the landfill site sample. The number of clones (out of 91 analysed) identical to the displayed representatives is indicated.
B. Application of the diagnostic microarray to analyse the diversity of methanotrophs in the landfill site sample. One slide contained three replicates of the array, thus each probe was printed in triplicate. For each microarray position, the name and the sequence of the probe is indicated in Fig. 1A and in Table 1 respectively. Probe spots having a normalized signal value (reference: mtrof173) greater than 5% of the maximum value obtained with reference sequences were considered as positive. Microarray image was adjusted for best viewing (quantitative conclusions drawn from the image may be misleading).

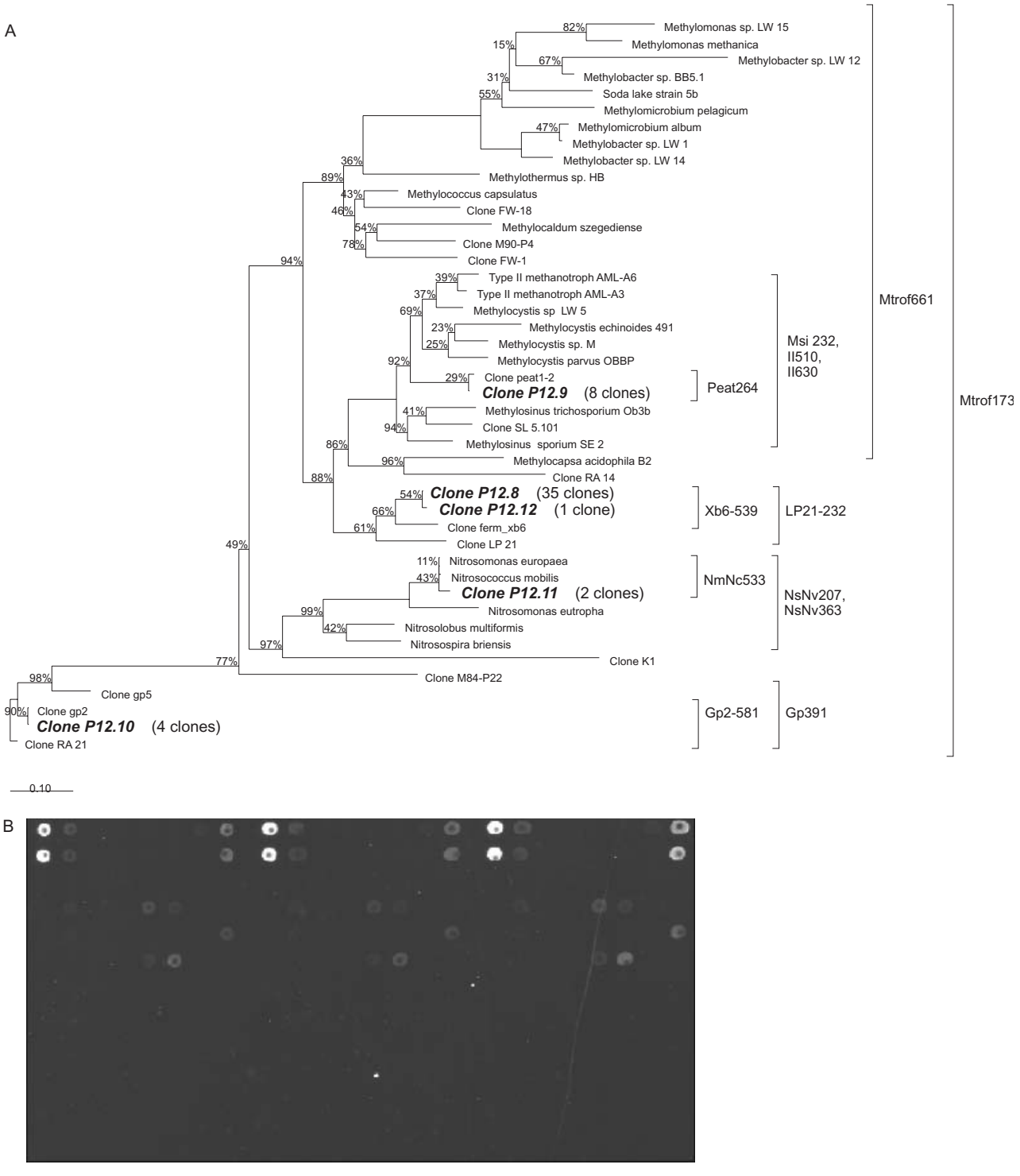
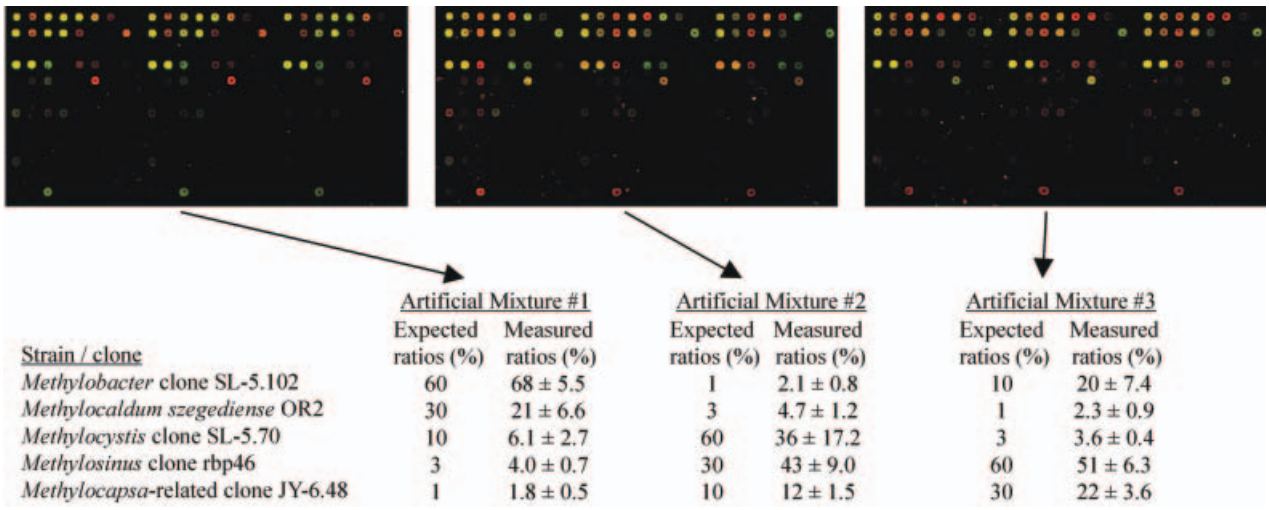


Fig. 4. A. *pmoA*/*amoA* DNA neighbour tree showing the phylogenetic positions of the clones obtained from the microcosm sample. The number of clones (out of 50 analysed) identical to the displayed representatives is indicated.
B. Application of the diagnostic microarray to analyse the diversity of methanotrophs in the microcosm sample. One slide contained three replicates of the array, thus each probe was printed in triplicate. For each microarray position, the name and the sequence of the probe is indicated in Fig. 1A and in Table 1 respectively. Probe spots having a normalized signal value (reference: mtrof173) greater than 5% of the maximum value obtained with reference sequences were considered as positive. Microarray image was adjusted for best viewing (quantitative conclusions drawn from the image may be misleading).

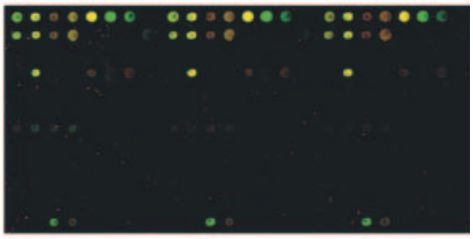


Strain / clone	Artificial Mixture #1 Expected ratios (%)	Artificial Mixture #1 Measured ratios (%)	Artificial Mixture #2 Expected ratios (%)	Artificial Mixture #2 Measured ratios (%)	Artificial Mixture #3 Expected ratios (%)	Artificial Mixture #3 Measured ratios (%)
Methylobacter clone SL-5.102	60	68 ± 5.5	1	2.1 ± 0.8	10	20 ± 7.4
Methylocaldum szegediense OR2	30	21 ± 6.6	3	4.7 ± 1.2	1	2.3 ± 0.9
Methylocystis clone SL-5.70	10	6.1 ± 2.7	60	36 ± 17.2	3	3.6 ± 0.4
Methylosinus clone rbp46	3	4.0 ± 0.7	30	43 ± 9.0	60	51 ± 6.3
Methylocapsa-related clone JY-6.48	1	1.8 ± 0.5	10	12 ± 1.5	30	22 ± 3.6

Fig. 5. Quantification of artificial mixtures of *pmoA* sequences using the diagnostic microarray. Relative abundance values were calculated from each positive probe the two-colour calculation method: Cy3/Cy5 ratios were used where the Cy3 signal arose from artificial mixtures of varying composition (column 'expected ratios') and the Cy5 signal from the artificial reference mixture containing all 5 sequences in equal amount (20%). Standard deviations are indicated next to measured ratios. Microarray images were adjusted for best viewing (quantitative conclusions drawn from the image may be misleading).

A

	Probes									
	Mb478	Mcl408	N_lb453	lb559	Mcy262	Mcy409	Msi232	ll510	ll630	mtrof173
Max. value	353	82	162	497	238	61	222	55	61	100
1-colour: Cy3 values	149	63	24	234	151	54	14	15	10	100
Procent predicted	42	77	15	47	63	88	6	27	16	100
2-colour: Cy3/Cy5 ratios	48	56	29	61	N.A.	N.A.	56	43	57	100
Procent predicted	24	28	14	30	N.A.	N.A.	28	21	28	100



B

	Probes						
	Peal264	Msi232	ll510	ll630	N_xb6-539	LP21-232	mtrof173
Max. value	101	222	17	16	192	306	100
1-colour: Cy3 values	23	28	7	3	8	14	100
Procent predicted	23	12	41	19	4	5	100
2-colour: Cy3/Cy5 ratios	51	71	18	21	6	10	100
Procent predicted	25	35	9	10	3	5	100

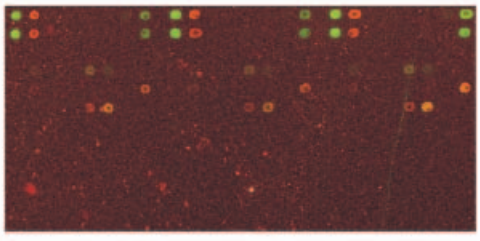


Fig. 6. Quantitative analysis of environmental samples using the diagnostic microarray.
A. Landfill site sample.
B. Microcosm sample.
Images of two-colour hybridization results are shown. Relative abundance values were calculated from each positive probe both using the single-colour and the two-colour calculation methods. In single colour calculations, normalized ratios were divided by the highest values obtained with reference sequences and the resulting values were taken as indications for the relative abundance of the carrying bacteria. In two-colour hybridizations, the Cy3/Cy5 ratios were used where the Cy3 signal arose from the environmental sample and the Cy5 signal from artificial reference mixtures. Microarray images were adjusted for best viewing (quantitative conclusions drawn from the image may be misleading).

31% in the clone library respectively; microcosm: 25–35% predicted for the 'peat clones' group and only 3–5% for the LP21 group versus 18% and 70% in the clone library respectively) it becomes obvious that further work is needed before diagnostic microbial microarrays can be used for quantitative environmental work. Obvious culprits for these discrepancies are biases in the PCR and in the cloning of the PCR mixes.

Conclusions

The set of techniques presented here enables the design, production and application of diagnostic microbial microarrays by an average microbiology laboratory with access to standard molecular biology equipment, a commercially available spotter and scanner. Even though the pilot array developed is targeting a functional gene (*pmoA*), the techniques and experience described here are also directly applicable to 16S rRNA based arrays.

The current detection limit of the technology is about 5% of the total population analysed. If cells are present at a lower ratio, they may be missed – this depends on the narrow specificity probes targeting them. This bottleneck can be improved by employing more stringent hybridization conditions, however, that way one gets limited to PM probes, thereby losing the potential for designing wider specificity probes. New, higher binding capacity surfaces will improve the performance of the approach. Finally, alternative labelling techniques (Small *et al*., 2001; Rudi *et al*., 2002) – where only a short oligonucleotide is labelled instead of the entire length of the PCR product – may decrease background arising from non-specific hybridization, thus improving the system.

There are many advances taking place in the field of microarrays which will result in novel technologies that may significantly improve the power of this technology. Novel platforms, such as new, three dimensional slide surfaces, electrically addressed microarrays, bead-arrays and lab-on-the-chip techniques are being developed. The cornerstone of diagnostic microbial microarrays, the design and behaviour of the oligonucleotide probes will, however, not differ too much between the different platforms. Thus most of the techniques and guidelines presented here will be easily transferable to emerging novel technologies.

A semi-quantitative analysis of environmental samples is possible in two stages: a first, single-colour hybridization is used to develop a rough estimation of the community structure, followed by a second, two-colour hybridization with a custom-made reference set based on the initial results. Biases inherent in the preceding molecular biology procedures impose limitations upon this approach. Even after a careful optimization of these steps, such results must be interpreted with caution. Perhaps the most promising field for diagnostic microbial microarray based quantification is the analysis of temporal and spatial changes within one environment. Before applying the developed *pmoA* microarray in high-throughput analysis of environmental samples for methanotroph diversity, it is necessary to optimize DNA extraction, purification and 'universal' *pmoA*/*amoA* PCR protocols to minimize the bias introduced. This aim will also be made easier by this microarray.

Experimental procedures

Environmental samples

A landfill site sample (Seibersdorf, Austria) was collected at the end of August, 2002. The sample collected from the top 20 cm of the landfill site cover soil had a pH of 7.05 and a temperature of 10°C above ambient air temperature. The sample was lyophilized and stored at – 80°C until use. A soil microcosm incubated with ^{12}C-CH_4 was also used to evaluate the microarray. Details of this microcosm are already published (Radajewski *et al*., 2002).

Oligonucleotide probe design

Database and phylogenetic trees were constructed and oligonucleotide probes were designed using the phylogenetic software package ARB (Strunk *et al*., 2000). A comprehensive database containing all published *pmoA*/*amoA* and related sequences, as well as many unpublished ones was established. Alignments were made using Old Aligner function in ARB_EDIT. Parsimony DNA and protein trees were constructed and used to guide the probe design process. Probes were designed using the Probe Design and Probe Match functions, accessing a PT-server database created from the above ARB database. Outputs of the Probe Match function were imported into Excel and a pivot table was constructed indicating the number of mismatches between each probe-target pair. This table was refined by applying a set of empirical rules as described in the relevant section of *Results and discussion*. Melting temperatures of the probes were predicted using the nearest neighbour method, using the public web site: http://biotools.idtdna.com/analyzer/. Weighed mismatch values were calculated from the number of mismatches as described in the *Results* section.

Microarray preparation

Oligonucleotides for immobilization were custom synthesized (VBC Genomics, Vienna, Austria) with a 5′ NH_2 group, followed by a C_{12} spacer and five thymidines residues preceding the probe sequence. A 384-well flat bottom plate was prepared with 30 μl of 50 μM oligonucleotide solutions in 50% DMSO. Samples were spotted with an OmniGrid spotter (1 TeleChem SMP3 pin) at 50% relative humidity (using the humidity controller of the spotter) and 22°C, onto silylated slides (with aldehyde chemistry, Cel Associates, Houston). Arrays were always spotted in triplicate to enable a statistical correction for errors. Spotted slides were incubated overnight

at room temperature at <30% relative humidity, rinsed twice in 0.2% (w/v) SDS for 2 min at room temperature with vigorous agitation to remove the unbound DNA. Slides were then rinsed twice in distilled water (dH_2O) for 2 min at room temperature with vigorous agitation, transferred into dH_2O, preheated to 95–100°C for 2 min, and allowed to cool at room temperature (~ 5 min). Slides were treated in a freshly (immediately before use) prepared sodium borohydride solution for 5 min at room temperature to reduce free aldehydes. Preparation of the sodium borohydride solution: 0.5 g $NaBH_4$ was dissolved in 150 ml phosphate-buffered saline (PBS; 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, in 1000 ml H_2O, pH 7.4, autoclaved) then 44 ml of 100% ethanol was added to reduce bubbling. Slides were rinsed three times in 0.2% (w/v) SDS and once in dH_2O for 1 min each at room temperature. Finally, slides were dried individually using an airgun fitted with a cotton wool filter inside (to keep oil microdroplets away from the slide surface). Dried slides were stored at room temperature in the dark before use.

Environmental DNA preparation

The DNA isolation was performed as previously described (Sessitsch *et al*., 2001). Briefly: freeze-dried soil was resuspended in 0.12 M sodium phosphate buffer (pH 8.0) and treated with lysozyme. Cells were disrupted by bead beating and cell lysis was completed by SDS treatment. Impurities were removed by phenol-chloroform extraction and precipitation of humic acids by potassium acetate. DNA was precipitated by isopropanol, washed in 70% ethanol, dried and resuspended. For final purification, spin columns that contained Sepharose CL-6B (Pharmacia) and polyvinylpyrrolidone (Sigma, 20 mg ml^{-1}) were used.

Target preparation

pmoA/*amoA* genes were amplified using the forward primer *pmoA*189 (5′-GGBGACTGGGACTTCTGG-3′) and either one of the reverse primers T7-mb661 (5′-TAATACGACTCACTATA GCCGGMGCAACGTCYTTACC-3′) or T7-A682 (5′-TAATACG ACTCACTATAGGAASGCNGAGAAGAASGC-3′) where B = (CGT), M = (AC), Y = (CT), S = (CG) and N = (ACGT). Primers T7-mb661 and T7-A682 were specific for methanotrophs and for methanotrophs/AOBs/homologous genes from environmental libraries respectively (Bourne *et al*., 2001). The reverse primers contained the T7 promoter site (5′-TAATAC GACTCACTATAG-3′) at their 5′ end, which enabled T7 RNA polymerase mediated *in vitro* transcription using the PCR products as templates. For each target, three PCR reactions of 50 μl volume each, consisting of 1× PCR buffer, 1.5 mM $MgCl_2$, 50 nM for each four dNTPs, 15 pmoles of both primers, 1 ng genomic/environmental DNA or 0.1 ng cloned PCR product as template, and 1 U of Taq polymerase (Invitrogen), were performed in a Hybaid Combi Thermal Reactor TR2 using *Taq* DNA polymerase in accordance with the manufacturer's instructions. Amplification conditions were: 95°C for 5 min before template was added, then 32 cycles of: 1 min at 95°C, 1 min at the annealing temperature, 1 min at 72°C; followed by a final elongation step of 10 min at 72°C. Polymerase chain reaction products were pooled and purified using the HighPure PCR purification kit (Macherey-Nagel), according to manufacturer's instructions. Purified DNA was dissolved in ultrapure water to a DNA concentration of 50 ng/μl and stored at – 20°C.

Working under RNAse-free conditions, *in vitro* transcription was carried out as follows: 8 μl purified PCR product (50 ng μl^{-1}), 4 μl 5× T7 RNA polymerase buffer, 2 μl DTT (100 mM), 0.5 μl RNAsin (40 U μl^{-1}) (Promega), 1 μl each of ATP, CTP, GTP (10 mM), 0.5 μl UTP (10 mM), 1 μl T7 RNA polymerase (40 U μl^{-1}) (Gibco BRL) and 1 μl Cy3 or Cy5-UTP (5 mM) were added into a 1.5 ml microcentrifuge tube and incubated at 37°C for 4 h. RNA was purified immediately using the Quiagen RNeasy kit according to manufacturer's instructions. Purified RNA was eluted into 50 μl dH_2O. RNA yields and dye incorporation rates were measured by spectrophotometry.

Purified RNA was fragmented by incubating with 10 mM $ZnCl_2$ and 20 mM Tris.Cl (pH 7.4) at 60°C for 30 min. Fragmentation was stopped by the addition of 10 mM EDTA pH 8.0 to the reaction and putting it on ice. RNAsin (1 μl 40 U μl^{-1}) was added to the fragmented target. Fragmented, labelled RNA targets were stored at –20°C. Length of the fragmented RNA target was measured by running the sample on an ABI capillary sequencer as well as running on a thin, 2% agarose gel applied onto a standard microscope slide and subsequent scanning in a GenePix 4000 A scanner.

Reference targets and artificial target mixtures for testing the quantification potential were synthesized by mixing known amounts of purified PCR products and carrying out *in vitro* transcription and target fragmentation as described above.

Hybridization

No prehybridization was done. Hybridization was carried out in a custom tailored aluminum block used as an insert for a temperature controlled Belly Dancer (Stovall Life Sciences, Greensboro, NC) set at maximum bending (about 10°). The hybridization block was preheated to 55°C for at least 30 min to allow the temperature to stabilize. An Eppendorf incubator was also preheated to 65°C. HybriWell (Grace BioLabs) stick-on hybridization chambers (200 μl in volume) were applied onto the slides containing the arrays. Assembled slides were preheated on top of the hybridization block. For each hybridization, 124 μl DEPC-treated water, 2 μl 10% SDS, 4 μl 50× Denhardt's reagent (Sigma), 60 μl 20× SSC and 10 μl target RNA were added into a 1.5 ml Eppendorf tube and incubated at 65°C for 1–15 min. Preheated hybridization mixtures were applied onto assembled slides via the port in the lower positions (to minimize risk of air bubbles being trapped within the chamber). Chambers were sealed with seal spots (Grace BioLabs) and incubated overnight at 55°C at 30–40 r.p.m. circulation and maximum bending.

Following hybridization, HybriWell chambers were removed individually and slides were immersed immediately into 2× SSC, 0.1% (w/v) SDS at room temperature. Slides were washed by shaking at room temperature for 5 min in 2× SSC, 0.1% (w/v) SDS; twice for 5 min in 0.2× SSC and finally for 5 min in 0.1× SSC. Slides were dried individually using an airgun with a cotton wool filter inside. Slides were stored at room temperature in the dark and scanned the same day.

Scanning and data analysis

Hybridized slides were scanned at three lines to average, 10 μm resolution with a GenePix 4000 A laser scanner (Axon, Foster City, CA) at wavelengths of 532 nm and 635 nm for Cy3 and Cy5 respectively. Fluorescent images were captured as multilayer tiff images and analysed with the GENEPIX PRO 3.0 software (Axon). Microsoft Excel was used for statistical analysis and presentation of results.

Results were normalized to a positive control. Hybridization signal for each probe was expressed as percentage of the signal (median of signal minus background) of the positive control probe mtrof173 on the same array. As each slide contained triplicate arrays, normalized signal intensities of the triplicate spots on a slide were used to determine average results and standard deviations. Non-specific signals reached in some cases a value of up to 7 (% of that of mtrof173). Such high non-specific signals were found for probes which had a high binding capacity as reflected by the signals obtained with perfect match targets. Thus, a cutoff value of 5% in relation to the highest signal (obtained with perfect match targets) was chosen to define positive and negative calls.

For assays where a (Cy3)-labelled target and a reference target (labelled with Cy5) were applied ('two-colour hybridizations'), the median of ratios were used. Results were normalized to mtrof173 and corrected for the relative amounts of different sequences added into the reference target mixture.

Even though no dedicated negative controls were applied, for each individual hybridization over 70% of all probes present on the array were negative controls.

All results reported represent the average of at least three replicates.

Quantification potential assessment

Mixtures with known compositions were made from reference sequences (PCR products ready for IVT labelling) and were labelled with Cy3 and Cy5. Cy5-labelled mixtures were used as reference. Median of ratios (Cy3/Cy5) values from GenePix were taken and normalized to that of the probe mtrof 173. For each reference sequence, the average of these values from all the positive spots were taken as an estimate of the relative abundance of the given sequence in the artificial mixture.

Cloning and sequencing of PCR products

pmoA PCR amplicons obtained from environmental samples were ligated into the pGEM-T plasmid vector (Promega) and transformed into *E. coli* DH5α competent cells. Screening for positive clones was done by the α-complementation as described by Sambrook *et al*. (1989). Plasmid DNA was prepared from positive clones using the NucleoTrap kit (Macherey-Nagel) according to the manufacturer's instructions and used as a template in sequencing reactions. DNA sequencing was performed with an ABI 373 A automated DNA sequencer (PE Applied Biosystems, Foster City, CA) and the ABI PRISM Big Dye terminator cycle sequencing kit (Perkin-Elmer). All DNA sequences reported were sequenced on both strands.

Nucleotide sequence accession numbers

The partial *pmoA* sequences used in this study to validate the probe set are available under accession nos. AB484595, AB484597, AB484601, AB480948, AF148521, AF148522, AF150764, AF177325, AF211872, AF211879, AF211883, AF211889, AF264115, AF239884, AF264136, AF358040, AF358041, AF358045, AF358050, AF358053, AF358054, AF358055, AF368358, AF368373, AJ278727, AJ299947, AJ299948, AJ299951, AJ299955, AJ299957, AJ299963, AJ299964, AJ459006, AY080942, AY080950, AY080955, AY236074, AY236075, AY236076, AY236077, AY236078, AY236079, AY236080, AY236081, AY236082, AY236083, AY236084, AY236085, AY236086, AY236087, AY236517, AY236518, U31650, U31654, U72670, U81596, U89302, U89304 and U94337. M2 + *pmoA* sequences have been deposited with GenBank (accession number AY195653-AY195672).

Acknowledgements

The authors thank Ann Auman, Svetlana Dedysh, Peter Dunfield, Werner Liesack, Ian McDonald, Samantha Morris and Stephen Nold for unpublished sequences, *pmoA* clones and methanotroph isolates. We thank Sándor Bottka, József Csontos, Fodor Szilvia, Andrew J. Holmes and many active members of the Yahoo microarray discussion group (http://www.groups.yahoo.com/group/microarray/) for their help and suggestions concerning various phases of the work presented here. Helpful comments and suggestions by two anonymous reviewers are gratefully acknowledged. Work at ARC was supported by the Fonds zur Förderung der wissenschaftlichen Forschung, Austria (Project number P15044), and funding through the EU 6th framework quality of life and management of living resources grant QLK-3 CT-2000-01528.

References

Anthony, R.M., Brown, T.J., and French, G.L. (2000) Rapid diagnosis of bacteremia by universal amplification of 23S ribosomal DNA followed by hybridization to an oligonucleotide array. *J Clin Microbiol* **38:** 781–788.

Behr, T., Koob, C., Schedl, M., Mehlen, A., Meier, H., Knopp, D., *et al.* (2000) A nested array of rRNA targeted probes for the detection and identification of enterococci by reverse hybridization. *Syst Appl Microbiol* **23:** 563–572.

Bourne, D.G., McDonald, I.R., and Murrell, J.C. (2001) Comparison of *pmoA* PCR primer sets as tools for investigating methanotroph diversity in three Danish soils. *Appl Environ Microbiol* **67:** 3802–3809.

Breslauer, K.J., Frank, R., Blocker, H., and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA* **83:** 3746–3750.

Brown, T.J., and Anthony, R.M. (2000) The addition of low numbers of 3′ thymine bases can be used to improve the hybridization signal of oligonucleotides for use within arrays on nylon supports. *J Microbiol Meth* **42:** 203–207.

Chizhikov, V., Wagner, M., Ivshina, A., Hoshino, Y., Kapikian, A.Z., and Chumakov, K. (2002) Detection and genotyping

of human group a rotaviruses by oligonucleotide microarray hybridization. *J Clin Microbiol* **40:** 2398–2407.

Cho, J.C., and Tiedje, J.M. (2001) Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl Environ Microbiol* **67:** 3677–3682.

Cho, J.C., and Tiedje, J.M. (2002) Quantitative detection of microbial genes by using DNA microarrays. *Appl Environ Microbiol* **68:** 1425–1430.

Dedysh, N., Liesack, W., Khmelenina, V.N., Suzina, N.E., Trotsenko, Y.A., Semrau, J.D., *et al.* (2000) *Methylocella palustris General 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.

Dunfield, P.F., Yimga, M.T., Dedysh, S.N., Berger, U., Liesack, W., and Heyer, J. (2002) Isolation of a *Methylocystis* strain containing a novel *pmoA*-like gene. FEMS. *Microbiol Ecol* **41:** 17–26.

Fotin, A.V., Drobyshev, A.L., Proudnikov, D.Y., Perov, A.N., and Mirzabekov, A.D. (1998) Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Res* **26:** 1515–1521.

Gulledge, J., Ahmad, A., Steudler, P.A., Pomerantz, W.J., and Cavanaugh, C.M. (2001) Genus- and family-level 16S rRNA-targeted oligonucleotide probes for ecological studies on methanotrophic bacteria. *Appl Environ Microbiol* **67:** 4726–4733.

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., and Smith, L.M. (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res* **22:** 5456–5465.

Hamels, S., Gala, J.L., Dufour, S., Vannuffel, P., Zammatteo, N., and Remacle, J. (2001) Consensus PCR and microarray for diagnosis of the genus *Staphylococcus*, species, and methicillin resistance. *Biotechniques* **31:** 1364–1362.

Hanson, R.S., and Hanson, T.E. (1996) Methanotrophic bacteria. *Microbiol Rev* **60:** 439–471.

Henckel, T., Jackel, U., Schnell, S., and Conrad, R. (2000a) Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane. *Appl Environ Microbiol* **66:** 1801–1808.

Henckel, T., Roslev, P., and Conrad, R. (2000b) Effects of O_2 and CH_4 on presence and activity of the indigenous methanotrophic community in rice field soil. *Environ Microbiol* **2:** 666–679.

Henckel, T., Jackel, U., and Conrad, R. (2001) Vertical distribution of the methanotrophic community after drainage of rice field soil. *FEMS Microbiol Ecol* **34:** 279–291.

Holmes, A.J., Roslev, P., McDonald, I.R., Iversen, N., Henriksen, K., and Murrell, J.C. (1999) Characterization of methanotrophic bacterial populations in soils showing atmospheric methane uptake. *Appl Environ Microbiol* **65:** 3312–3318.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., *et al.* (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19:** 342–347.

Hung, S.H., Gray, Q.Yu, D.M., and Ratliff, R.L. (1994) Evidence from CD spectra that d (purine) r (pyrimidine) and r (purine) d (pyrimidine) hybrids are in different structural classes. *Nucleic Acids Res* **22:** 4326–4334.

Ishii, K., and Fukui, M. (2001) Optimization of annealing temperature to reduce bias caused by a primer mismatch in multi-template PCR. *Appl Environ Microbiol* **67:** 3753–3755.

Koizumi, Y., Kelly, J.J., Nakagawa, T., Urakawa, H., El Fantroussi, S., Al Muzaini, S., *et al.* (2002) Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization and DNA microarray technology. *Appl Environ Microbiol* **68:** 3215–3225.

Liu, W.T., Mirzabekov, A.D., and Stahl, D.A. (2001) Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ Microbiol* **3:** 619–629.

Lovell, C.R., Friez, M.J., Longshore, J.W., and Bagwell, C.E. (2001) Recovery and phylogenetic analysis of *nif*H sequences from diazotrophic bacteria associated with dead aboveground biomass of *Spartina alterniflora*. *Appl Environ Microbiol* **67:** 5308–5314.

Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., *et al.* (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68:** 5064–5508.

Lueders, T., Chin, K.J., Conrad, R., and Friedrich, M. (2001) Molecular analyses of methyl-coenzyme M reductase alpha-subunit (*mcr*A) genes in rice field soil and enrichment cultures reveal the methanogenic phenotype of a novel archaeal lineage. *Environ Microbiol* **3:** 194–204.

McDonald, I.R., and Murrell, J.C. (1997) The particulate methane monooxygenase gene *pmoA* and its use as a functional gene probe for methanotrophs. *FEMS Microbiol Lett* **156:** 205–210.

McTavish, H., Fuchs, J.A., and Hooper, A.B. (1993) Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. *J Bacteriol* **175:** 2436–2444.

Meroueh, M., and Chow, C.S. (1999) Thermodynamics of RNA hairpins containing single internal mismatches. *Nucleic Acids Res* **27:** 1118–1125.

Mikhailovich, V., Lapa, S., Gryadunov, D., Sobolev, A., Strizhkov, B., Chernyh, N., *et al.* (2001) Identification of rifampin-resistant *Mycobacterium tuberculosis* strains by hybridization, PCR, and ligase detection reaction on oligonucleotide microchips. *J Clin Microbiol* **39:** 2531–2540.

Polz, M.F., and Cavanaugh, C.M. (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64:** 3724–3730.

Purkhold, U., Pommerening-Roser, A., Juretschko, S., Schmid, M.C., Koops, H.P., and Wagner, M. (2000) Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* **66:** 5368–5382.

Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403:** 646–649.

Radajewski, S., Webster, G., Reay, D.S., Morris, S.A., Ineson, P., Nedwell, D.B., *et al.* (2002) Identification of active

methylotroph populations in an acidic forest soil by stable-isotope probing. *Microbiology* **148:** 2331–2342.

Raskin, L., Stromley, J.M., Rittmann, B.E., and Stahl, D.A. (1994) Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl Environ Microbiol* **60:** 1232–1240.

Reay, D.S., Radajewski, S., Murrell, J.C., McNamara, N., and Nedwell, D.B. (2001) Effects of land-use on the activity and diversity of methane oxidizing bacteria in forest soils. *Soil Biol Biochem* **33:** 1613–1623.

Reysenbach, A.L., Giver, L.J., Wickham, G.S., and Pace, N.R. (1992) Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol* **58:** 3417–3418.

Rudi, K., Flateland, S.L., Hanssen, J.F., Bengtsson, G., and Nissen, H. (2002) Development and evaluation of a 16S ribosomal DNA array-based approach for describing complex microbial communities in ready-to-eat vegetable salads packed in a modified atmosphere. *Appl Environ Microbiol* **68:** 1146–1156.

Rudi, K., Skulberg, O.M., Skulberg, R., and Jakobsen, K.S. (2000) Application of sequence-specific labeled 16S rRNA gene oligonucleotide probes for genetic profiling of cyanobacterial abundance and diversity by array hybridization. *Appl Environ Microbiol* **66:** 4004–4011.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* **93:** 10614–10619.

Semrau, J.D., Chistoserdov, A., Lebron, J., Costello, A., Davagnino, J., Kenna, E., *et al.* (1995) Particulate methane monooxygenase genes in methanotrophs. *J Bacteriol* **177:** 3071–3079.

Sessitsch, A., Weilherter, A., Gerzabek, M.H., Kirchmann, H., and Kandeler, E. (2001) Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl Environ Microbiol* **67:** 4215–4224.

Shchepinov, M.S., Case-Green, S.C., and Southern, E.M. (1997) Steric factors influencing hybridization of nucleic acids to oligonucleotide arrays. *Nucleic Acids Res* **25:** 1155–1161.

Small, J., Call, D.R., Brockman, F.J., Straub, T.M., and Chandler, D.P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol* **67:** 4708–4716.

Spiro, A., Lowe, M., and Brown, D. (2000) A bead-based method for multiplexed identification and quantitation of DNA sequences using flow cytometry. *Appl Environ Microbiol* **66:** 4258–4265.

Strunk, O., Gross, O., Reichel, B., May, M., Hermann, S., Stuckman, N., *et al.* (2000) ARB: A software environment for for sequence data. [WWW document]. URL http://www.mikro.biologie.tu-muenchen.de.

Sugimoto, N., Nakano, M., and Nakano, S. (2000) Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes. *Biochemistry* **39:** 11270–11281.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181:** 6425–6440.

Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A., and Andersen, G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol* **68:** 2535–2541.

Witzingerode, F., Gobel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Rev* **21:** 213–229.

Wood, W.I., Gitschier, J., Lasky, L.A., and Lawn, R.M. (1985) Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc Natl Acad Sci USA* **82:** 1585–1588.