Fluorescence *in situ* hybridization of genes in environmental microbiology

Cristina Moraru

Fluorescence *in situ* hybridization of genes in environmental microbiology

Dissertation zur

Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der

Universität Bremen vorgelegt von

Cristina Moraru

Bremen 2010

Table of contents

Summary

Our knowledge concerning the metabolic potentials of as yet uncultured microorganisms has increased tremendously with the advance of sequencing technologies and the consequent discovery of novel genes. On the other hand, it is often difficult to reliably assign a particular gene to a phylogenetic clade, because these sequences are usually found on genomic fragments that carry no direct marker of cell identity, such as rRNA genes. Therefore, the main objective of the present study was to develop geneFISH - a protocol for linking gene presence with cell identity in environmental samples. This protocol combines rRNA-targeted Catalyzed Reporter Deposition – Fluorescence *In Situ* Hybridization (CARD-FISH) and *in situ* gene detection. The method of rRNA-targeted CARD-FISH was previously developed (Pernthaler et al., 2002a). The gene detection method was adapted from Pernthaler and Amann (2004). It uses multiple digoxigenin labeled polynucleotide probes to target genes, followed by the binding of HRP-conjugated antibodies and catalyzed reporter deposition (CARD), to amplify and visualize the gene signal.

However, the specificity of polynucleotide probes has not been thoroughly investigated and a rational probe design concept is still missing, because the well established concept for oligonucleotide probe design cannot be transferred to polynucleotides. Therefore, we developed a concept and software (PolyPro) for rational design of polynucleotide probe mixes used to identify particular genes in defined taxa. PolyPro consists of three modules: a GenBank Taxonomy Extractor (GTE), a Polynucleotide Probe Designer (PPD) and a Hybridization Parameters Calculator (HPC). Applying this probe design concept to three metabolic marker genes revealed the following about the use of polynucleotide probes in FISH: (i) a single probe is not sufficient to detect all alleles of a gene; (ii) single probes can be used mostly at the genus level; (iii) probe mixes cannot be used to detect all alleles of a gene, because of differences in the melting temperature; (iv) probe mixes can be used for identifying a gene mostly at the genus and family level.

The newly developed concept for polynucleotide probe design was further applied to the probe design for the geneFISH experiments. The geneFISH protocol was first developed and tested in *Escherichia coli*. In a second phase, it was applied on seawater samples from Benguela upwelling system on the Namibian shelf, in which the presence of putative *amo*A gene was directly visualized in crenarchaeotal cells. This involved a specially designed polynucleotide probe mix (*amo*A-Nam) that targets the crenarchaeotal putative *amo*A alleles present in these environmental samples. Additionally, geneFISH was applied on two more systems, an enrichment sample, targeting *rdsrA* genes, and an eukaryotic host - bacterial

symbiont system, targeting *hyn*L and *apr*A genes in the symbionts. Further development of this method will in the direction of improving the gene detection efficiency, from less than 50% to 100%. This will allow a quantitative use of the geneFISH protocol.

Zusammenfassung

Unser Wissen über das metabolische Potenzial von bisher nicht kultivierten Mikroorganismen hat mit dem Fortschritt der Sequenzierungstechnologien und der Entdeckung neuer Gene gewaltig zugenommen. Andererseits ist es oft schwierig, zuverlässig ein bestimmtes Gen einer phylogenetischen Gruppe zuzuordnen, weil diese Sequenzen normalerweise auf genomischen Fragmenten gefunden werden, die keinen direkten Marker wie z.B. rRNA-Gene für die Zellidentität tragen.

Das Hauptziel der vorliegenden Studie war deshalb die Entwicklung von GenFISH - einer Methode, die den Nachweis von Genen mit der Identität von Zellen in Umweltproben verknüpft. Die Methode kombiniert die Fluoreszenz *in-situ* Hybridisierung mit spezifischen, rRNA-gerichtete und enzym-markierten Oligonukleotidsonden („catalyzed reporter deposition - fluoreszenz *in-situ* hybridisation - CARD-FISH) mit dem *in-situ* Gennachweis. Die Methode der rRNA-gerichteten CARD-FISH wurde bereits früher entwickelt (Pernthaler et al., 2002a). Das Verfahren des Gennachweises wurde von Pernthaler und Amann (2004) übernommen und angepasst. Die Gen-gerichteten Polynukleotidsonden sind mehrfach Digoxigenin-markiert, die anschließend in Verbindung mit Meerrettich-Peroxidase (HRP)-konjugierten Antikörpern fluoreszente Substrate in der Zelle ablagern (CARD), um das Gensignal in der Zelle zu amplifizieren und zu visualisieren.

Allerdings ist die Spezifität von Polynukleotidsonden bis jetzt noch nicht gründlich untersucht worden. Bis jetzt fehlte ein schlüssiges Konzept zur Entwicklung von Polynukleotidsonden, da das wohl-etablierte Konzept zum Design von Oligonukleotidsonden nicht direkt auf Polynukleotide übertragen werden kann. Deshalb entwickelten wir ein Konzept und eine Software (PolyPro) für das objektives Design von Polynukleotidsonden-Mischungen zur Identifizierung von spezifischen Genen in definierten Taxa.

PolyPro besteht aus drei Modulen: einem „GenBank Taxonomy Extractor“ (GTE) zur Extraktion der taxonomischen Information aus Genbank-Dateien, einem „Polynucleotidprobe Designer“ (PPD), dem eigentlichen Sonden-Designmodul und einem „Hybridisation Parameters Calculator“ (HPC) zur Berechnung der Hybridisierungsbedingungen. Der Einsatz dieses Sondendesign-Konzepts auf drei metabolische Markergene ergab folgendes in Bezug auf die Anwendung von Polynukleotidsonden in FISH: (i) eine einzige Sonde ist nicht ausreichend, um alle Allele eines Gens zu detektieren, (ii) einzelne Sonden können meistens auf Gattungsebene verwendet werden, (iii) wegen Unterschiede in den Schmelztemperaturen können Sondenmischungen nicht dazu verwendet werden, um alle

Allele eines Gens zu detektieren, (iv) Sondenmischungen für die Identifizierung eines Gens können hauptsächlich auf Gattungs- und Familienebene verwendet werden.

Das für Polynukleotidsonden-Design neu entwickelte Konzept wurde für das Sondendesign in GenFISH-Experimenten eingesetzt. Das GenFISH-Protokoll wurde zunächst an *Escherichia coli* entwickelt und getestet. In einer zweiten Phase wurde es in Meereswasserproben des Benguela-Auftriebsgebietes angewandt, in denen potentielle amoA-Gene direkt in Crenarchaeota-Zellen visualisiert werden konnten. Dazu wurde ein spezielles Polynukleotid-Sondenmix entwickelt (amoA-Nam), das gegen potentielle, zu den Crenarchaeota-gehörende amoA-Allele in diesen Umweltproben gerichtet ist. Darüber hinaus wurde GenFISH auf zwei weitere biologische Systeme angewandt mit spezifischen Polynukleotidsonden, die gegen die rdsrA Gene einer Anreicherungsprobe bzw. gegen die hynL und aprA Gene von symbiontischen Bakterien in einem eukaryotischen Wirt gerichtet waren.

Weitere Entwicklungen dieser Methode zielen auf eine Verbesserung der Effizienz des Gennachweises von unter 50% auf bis zu 100%, um eine quantitative Anwendung des GenFISH Protokolls zu ermöglichen.

List of abbreviations

%GC	Percent molar G + C
%MM	Percent Mismatches
*amo*A	Gene for the alpha subunit of ammonia monooxygenase
*apr*A	Gene for the alpha subunit of adenosine-5'-phosphosulfate reductase
bDNA	Branched DNA
bp	Base pairs
CARD	Catalyzed Reporter Deposition
CPRINS	Cycling Primed *In Situ* Amplification
DAPI	4', 6-diamidino-2-phenylindol
Dig	Digoxigenin
dsDNA	Double-stranded DNA
*dsr*A	Gene for the α subunit of the dissimilatory sulfite reductase
EDTA	Ethylenediaminetetraacetic acid
F	Forward
FA	Formamide
FISH	Fluorescence I*n Situ* Hybridization
F-RNA	Fluorinated RNA
HISH-SIMS	Halogen *In Situ* Hybridization- Secondary Ion Mass Spectroscopy
*hme*D	Gene for the subunit D of the Hdr-like menaquinol-oxidizing enzyme
HRP	Horseradish Peroxidase
*hyn*L	Gene for the large subunit of the [NiFe] hydrogenase
Kb	Kilo bases
kbp	Kilo base pairs
LAMP	Loop-Mediated Isothermal Amplification
LB	Luria-Bertani
mRNA	Messenger RNA
*nif*H	Gene for the subunit H of the dinitrogenase reductase gene
nt	Nucleotides
PBS	Phosphate Saline Buffer
PCR	Polymerase Chain Reaction
*pmo*A	Gene for the alpha subunit of particulate methane monooxygenase
Pr	Probe

PRINS	Primed I*n Situ* Labeling
R	Reverse
RCA	Rolling Circle Amplification
*rdsr*A	Gene for the α subunit of the reverse dissimilatory sulfite reductase
RING-FISH	Recognition of Individual Genes- FISH
*rpo*B	Gene for the ß subunit of the DNA dependent RNA polymerase
rRNA	Ribosomal RNA
SDS	Sodium Dodecyl Sulfate
sp.	Species
SSC	Saline-Sodium Citrate
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
sssDNA	Sheared Salmon Sperm DNA
TE	Tris-EDTA
TEACl	Tetraethylammonium chloride
Tg	Target
T_m	Melting temperature
WBR	Western Blocking Reagent

1. Introduction

In the past years microbial ecologists have increasingly used molecular techniques to improve the knowledge and understanding of the distribution and role of microorganisms in various ecosystems. The rapid advance of sequencing technologies has enabled the retrieval of many sequences from environmental samples, mostly as PCR based or metagenomic clone libraries. Hence, our knowledge about the metabolic potentials of the uncultivable microorganisms has increased tremendously. On the other hand, it is difficult to identify from which microorganism a particular sequence originates, because it is usually found on genomic fragments that carry no phylogenetic marker, such as rRNA genes. A promising technique to resolve this is fluorescence *in situ* hybridization (FISH), a frequently used technique in molecular microbial ecology.

In FISH, most of the times the target is rRNA (Amann and Fuchs, 2008), and short oligonucleotides(~20 nt) are used as probes to identify and quantify microorganisms in environmental samples. Pernthaler et Amann (2004) combined Catalyzed Reporter Deposition – Fluorescence *In Situ* Hybridization (CARD FISH) with mRNA FISH. Thereby, information on *in situ* transcription of a gene was linked to the identity of a cell. When present, mRNAs of a certain gene usually occur in multiple copies, improving thus, on the one hand, the detectability of the gene of interest. On the other hand, mRNAs might only be expressed when the cells need that function, and they are often short lived (Jain, 2002; Condon, 2003; Deutscher, 2006). Another way for linking identity and genomic potential would be a FISH based identification of single genes. DNA is more stable than mRNA, and yet, it is a challenge to detect genes, because of the low number of targets per cell, sometimes no more than a single copy.

The existent protocols for detection of protein-coding genes or of their mRNA use either oligonucleotide or polynucleotide probes. Polynucleotides have also been used to target rRNA (Trebesius et al., 1994; DeLong et al., 1999; Zimmermann et al., 2001; Pernthaler et al., 2002b). Few studies focused on the ability of rRNA polynucleotide probes to discriminate between phylogenetic clades (Ludwig et al., 1994; Trebesius et al., 1994; Zwirglmaier, 2005). However, as opposed to oligonucleotides, the specificity of polynucleotide probes has not been thoroughly investigated and a rational probe design concept is still missing.

1.1. Detection methods for *in situ* hybridizations

In situ hybridization represents the binding of a probe to its target nucleic acid which is found inside a cell. In situ hybridizations can be detected by various methods: autoradiography, bright field microscopy, fluorescence microscopy, electron microscopy and, lately, by NanoSIMS. Autoradiographic detection involves radioactively labeled probes. Although very sensitive, the disadvantages related with long exposure times, poor resolution and health risks have lead to a replacement of this method (Wilcox, 1993). Detection using bright field microscopy is based on deposition of chromogenic substrates (Amann et al., 1992), while detection using electron microscopy is based on deposition of gold particles (Mayer and Bendayan, 1999; Mayer et al., 2000; Gérard et al., 2005; Kenzaka et al., 2005b; Ehrhardt et al., 2009). In fluorescence microscopy the hybrids are detected by associating them with fluorescent molecules. Fluorescence microscopy offers a better signal to background ratio than bright field microscopy (Lichtman and Conchello, 2005), and is therefore more sensitive. It is also much easier to use than electron microscopy. Together with the possibility for multiplexing via different colors, this is making fluorescence microscopy the method of choice for detecting *in situ* hybridization events, the general term for these methods being Fluorescence *In Situ* Hybridization (FISH). In the last years a new detection method was developed: detection of probes associated with halogenated compounds using NanoSIMS (nano-scale secondary-ion mass spectrometry) (Behrens et al., 2008; Li et al., 2008; Musat et al., 2008). In this method, hybrid detection can be combined with the detection and quantification of different substrate incorporation, allowing a unique view of the ecophysiology at the single-cell level.

1.2. Gene detection protocols and signal amplification methods in FISH techniques

There are already several FISH protocols for visualization of genes in microorganisms, protocols which use different signal amplification methods. They can be classified in two categories. Firstly, there are methods that use oligonucleotide probes for detection by or upon amplification of the target sequence. This includes, e.g., *in situ* PCR (Hodson et al., 1995; Tani et al., 1998), Cycling Primed *In Situ* Amplification (CPRINS) (Kenzaka et al., 2005a), *In Situ* Loop-Mediated Isothermal Amplification (Maruyama et al., 2003), *In Situ* Rolling Circle Amplification (Maruyama et al., 2005), Peptide Nucleic Acid – Assisted Rolling Circle Amplification FISH (Smolina et al., 2007). Secondly, there are methods that use polynucleotide probes. The FISH-based detection of DNA fragments > 5 kbp is well established (Niki and Hiraga, 1997; Jensen and Shapiro, 1999). RING-FISH targets single genes by formation of a network of polynucleotide probes (Zwirglmaier et al.,

2004). Beside the methods mentioned above, there are more signal amplification methods to be considered, e.g. branched DNA or AT-polymerization.

1.2.1. Polynucleotide probes with multiple labels

The simplest way to amplify the signal is to add more than one fluorochrome to the probe molecule. This approach has been used in several studies to target plasmids or chromosome encoded regions of sizes between 5 and 13.5 kbp. The probes were dsDNA polynucleotides, labeled by random priming and sheared into smaller fragments by sonication (Niki and Hiraga, 1997, 1998). An alternative to the probe preparation was to add labeled deoxyribonucleotides to 20-300 bp restriction fragments by using terminal deoxynucleotidyltransferase (Jensen and Shapiro, 1999; Pogliano et al., 2001; Viollier et al., 2004).

Multiple labeled polynucleotide probes have been used also to target rRNA inside bacterial cells (Trebesius et al., 1994; DeLong et al., 1999; Zimmermann et al., 2001; Pernthaler et al., 2002b) with subsequent phylogenetic identification of the cells. Depending on the method variation used, the signal was either cell-wide or localized at the periphery of the cells, surrounding them as a halo. Zwirglmaier et al., (2003) explain the halo as a network of probe molecules partially hybridized to each other and fixed to the cells by the partial binding of some of the probe molecules to their rRNA targets (Figure 1). Taking advantage of this ability of the polynucleotide probes to form networks, Zwirglmaier and collaborators developed RING-FISH, a method for single copy gene detection (Zwirglmaier et al., 2004).

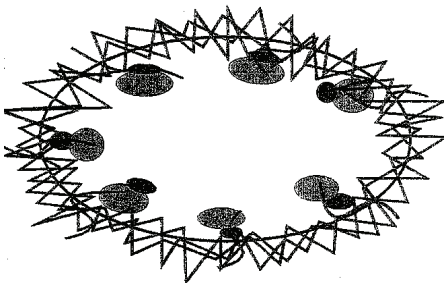


Figure 1: Schematic illustration of the network hypothesis (Zwirglmaier et al., 2003). The probes are forming a network around the cell. Some probes are binding the rRNA in ribosomes, serving as anchors.

1.2.2. PCR amplification

PCR amplification is used for *in situ* detection of single genes in microorganisms (Hodson et al., 1995; Tani et al., 1998; Hoshino et al., 2001). Detection of PCR amplicons can be done either by direct incorporation of labeled dNTPs or by hybridization with labeled

oligonucleotide probes. One main drawback associated with *in situ* PCR is the extensive permeabilization, which can result in amplicons leaking out of the cells and occurrence of false positives.

1.2.3. Primed in situ labeling (PRINS)

As opposed to *in situ* PCR, PRINS uses only one primer to form a long product, while incorporating labeled dNTPs. The length of the product can be several kb and depends on the length of the template, the enzyme used and on the elongation time. Compared to *in situ* PCR, the length of the product will make it less likely to diffuse out of the cells, thereby increasing specificity. Cycling primed *in situ* amplification (CPRINS) is a PRINS variant that uses more than one polymerization cycle. The accumulation of polymerization products is not exponential like in PCR, but linear. To avoid the problems associated with unspecific incorporation of labeled dNTPs, the CPRINS amplicons can be visualized by hybridization with multiple labeled oligonucleotide probes (Figure 2). Both PRINS and CPRINS have been applied successfully for low copy gene detection in microorganisms, in pure cultures (Jacobs et al., 1997; Kenzaka et al., 2005a) and in environmental samples (Kenzaka et al., 2007).

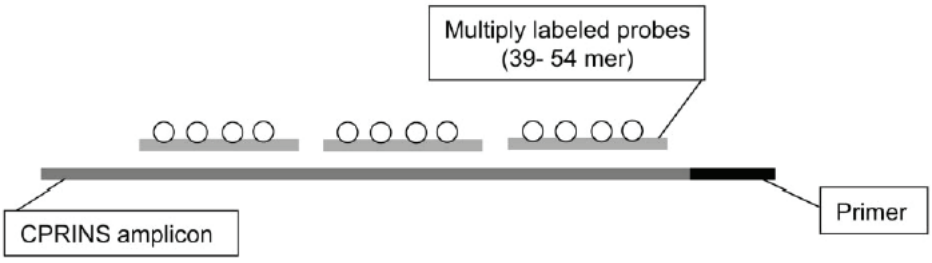


Figure 2: Detection of CPRINS amplicons with multiple labeled oligonucleotide probes. Figure adapted from Kenzaka et al., (2005a).

1.2.4. Loop-mediated isothermal amplification (LAMP)

LAMP is using a set of four primers (two inner and two outer primers), six recognition sequences on the target and a DNA polymerase with high strand displacement activity to generate multiple copies of the target region, bound in a repetitive stem-loop structure (Figure 3) (Notomi et al., 2000). Excluding the initial denaturation step, the process is isothermal, the new rounds of amplification being initiated without thermal denaturation. In the first phase, all primers will participate to form a dumb-bell DNA structure (structure 6). Initially, the elongation will start from the inner primer at one end of the template region. The newly synthesized strand will be displaced by elongation from the outer primer. The hang-over region of the inner primer is the reverse complement of a site from the newly synthesized strand to which it will hybridize. Thus, at the end of the single-stranded DNA molecule a loop will be formed. The same process at the other end of the strand will

generate a dumb-bell form DNA molecule (structure 6), with one loop at each end. The loops will further serve both as self priming sites and as sites for primer binding. The dumb-bell form DNA will be converted by self-primed DNA synthesis to a stem-loop structure (structure 7). Once this stem-loop structure will be formed, each template amplification round will produce new loops, which again will serve either as self priming sites or as binding sites for the inner primers. The final products are represented by a mixture of stem-loop structures (e.g. structures 13 and 15) with different stem lengths and cauliflower-like structures with multiple loops (e.g. structures 14, 17 and 18).

Applying a LAMP variant in which only the inner primers were used and the initial denaturation was eliminated, Maruyama et al., (2003) were able to detect low copy genes in *Escherichia coli* cells.

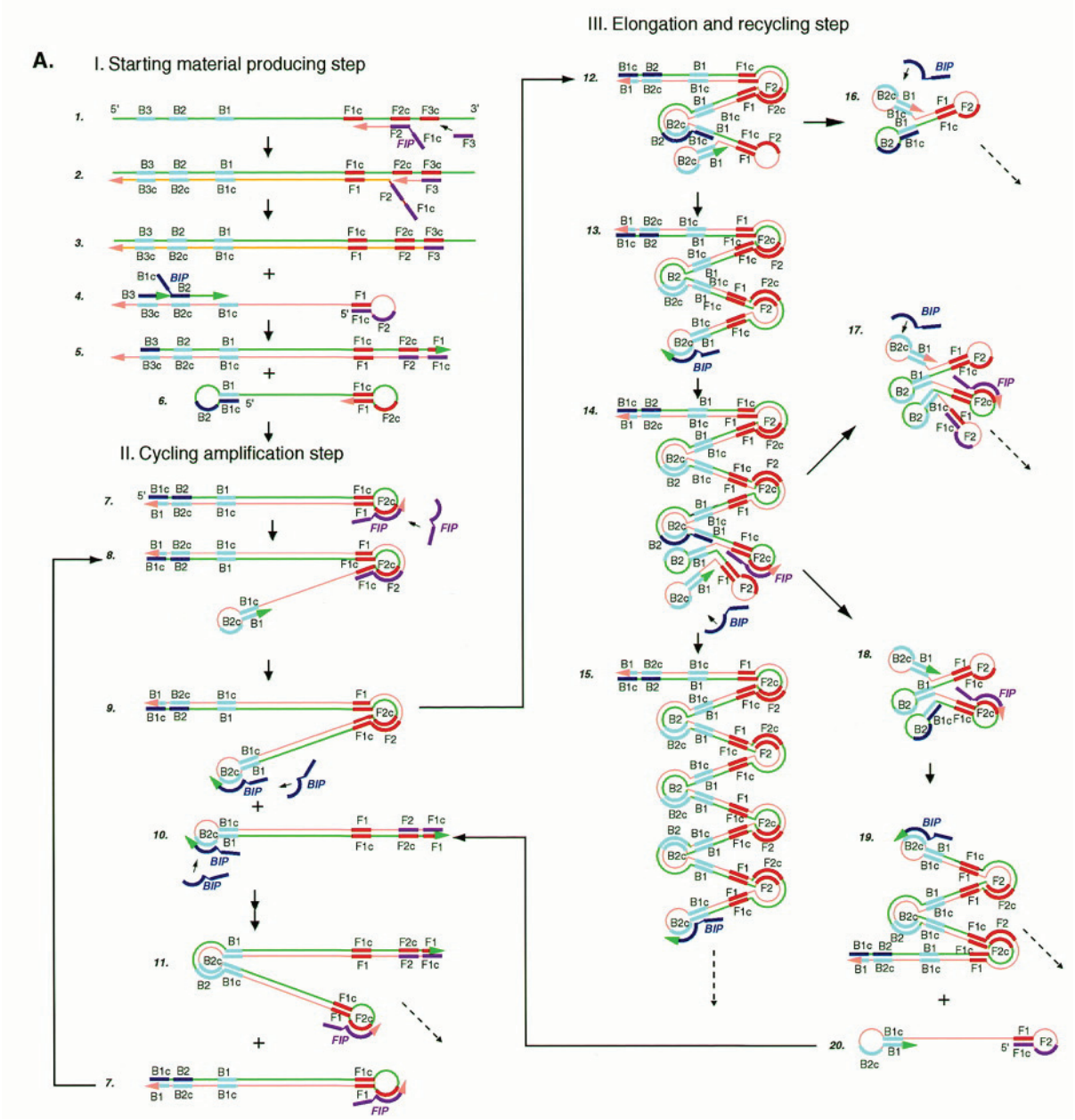


Figure 3: Schematic representation of the LAMP reaction (Notomi et al., 2000). I. Starting material producing step. In the first phase (1-3), the inner primer (FIP) binds to the target and elongation starts. A second elongation from the outer primer (F3) will displace the first strand. The hang-over region from the inner primer will hybridize to a complementary region from the newly synthesized strand, to form a loop at one of the ends (4). At the other end (5-6), the second inner and outer primers will bind and the elongation and strand displacement steps will take place. As a result, the resulted single strand will have two loops, one at each end, formed by the binding of the hang-over regions of the inner primers to complementary regions on the single strand (6). This dumb-bell DNA structure (6) will be converted to a stem-loop structure (7), by self-primed elongation from one of the loops. **II. Cycling amplification step.** The inner primers will bind to the loops from the stem-loop structures and will initiate elongation, displacing one of the strands in the stem. Due to the loop, the displaced strand will stay attached to the original molecule. Self-primed elongation will start from the newly formed loop. **III. Elongation and recycling step.** The loops will act further both as sites for self-primed elongation and for binding of the inner primers, elongating the stem and perpetuating the loop at one of the ends. The final products are represented by a mixture of stem-loop structures (e.g. structures 13 and 15) with different stem lengths and cauliflower-like structures with multiple loops (e.g. structures 14, 17 and 18).

1.2.5. Rolling Circle Amplification (RCA)

RCA systems are using DNA polymerases with strand displacement activity to perform rolling circle amplification, provided that they have a circular template and a primer to initiate the polymerization. The result will be a long single-stranded DNA (ssDNA), in which the sequence of the circular template will be repeated many times. This ssDNA will serve as target for hybridization with fluorescently labeled oligonucleotides. The amplification factor is directly proportional with the number of repeats in the single-stranded amplicon. This, in turn, depends on the polymerization speed of the enzyme used and on the amplification time. When using ϕ29 DNA polymerase, Banér *et al.* (1998), found a mean incorporation rate of 1.0 x 10^3 nt/min. Assuming a circle size of 100 nt, each RCA reaction would result in 600 circle copies per hour.

In a hybridization reaction there are two ways to provide the circular template and a primer for the RCA reaction. One way is to use a hapten-labeled probe, to which, after hybridization, an anti-hapten antibody will bind (Figure 4). The antibody is carrying the RCA primer, to which the circular template for the RCA will bind by hybridization. The RCA generated amplicon will remain bound to the antibody. Another way is to use circularizable probes, consisting from two end regions which hybridize to the target DNA and one central region to which the RCA primer will bind (Banér et al., 1998). The second approach has been used for *in situ* detection of genes in microorganisms (Maruyama et al., 2005; Maruyama et al., 2006; Smolina et al., 2007).

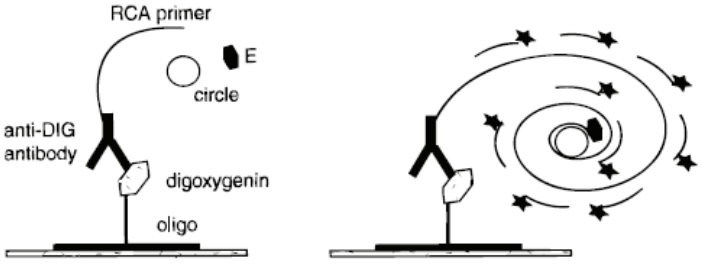


Figure 4: The principle of rolling circle amplification with the primer being conjugated with an antibody. An anti-DIG antibody with a covalently bound oligonucleotide primer binds to a Digoxigenin (Dig) labeled oligonucleotide probe. Then, a circular oligonucleotide binds to the primer and RCA is initiated by DNA polymerase (E). For detection, the amplified DNA product is hybridized with fluorochrome-labeled oligonucleotides. Figure adapted from Zhou et al., (2001).

1.2.6. AT-polymerization

The AT-polymerization system is based on the ability of some polymerases (the Stoffel fragment of *Taq*-polymerase, *ΔTth*-polymerase, the Klenow fragment of *E. coli DNA*-polymerase I and T_4 DNA-polymerase) to catalyze the polymerization of AT-tails without the need of a template (Hanaki et al., 1997; Hanaki et al., 1998). The reaction is isothermal. The

AT-polymerization with incorporation of labeled dUTP (Figure 5) was used for *in situ* detection of viral nucleic acids in mammalian cells (Nakajima et al., 1999; Nakajima et al., 2003). The amplification factor is given by the number of labeled dUTPs incorporated. Nakajima *et al.* reported the formation of a 50 kb long amplification product, but there was no information regarding the labeling rate (Nakajima et al., 1998).

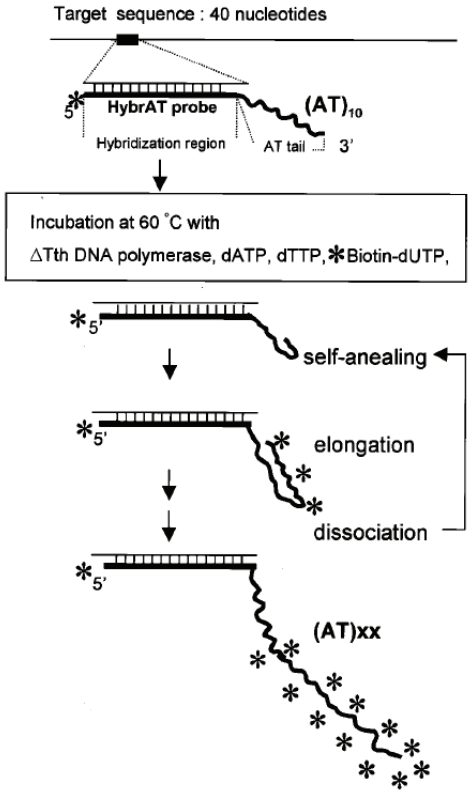


Figure 5: The principle of AT-polymerization. Figure adapted from Nakajima et al., (2003).

1.2.7. Catalyzed Reporter Deposition (CARD)

CARD uses the peroxidase activity of enzymes like horseradish peroxidase (HRP) to deposit labeled substrates (Figure 6). The most common substrates are tyramides conjugated either with fluorochromes or haptens for antibody recognition (Hopman et al., 1998). The tyramide part of the conjugate consists of a phenol ring, which, upon activation by free radicals (e.g. hydroxyl radicals), will covalently bind to tyrosine residues from proteins and fix the label into cells. There are two ways to specifically introduce HRP into cells: by covalently binding to nucleic acid probes or by covalently binding to antibodies that bind to hapten-labeled probes.

The CARD reaction is commonly used to enhance the signal during rRNA targeted FISH, especially on environmental samples (Pernthaler et al., 2002a; Teira et al., 2004; Amann and Fuchs, 2008). The amplification factor is between 26 to 42 fold when compared with regular FISH (Hoshino et al., 2008). Due to its sensitivity, the CARD reaction has been

also used in FISH protocols for mRNA detection (Wagner et al., 1998; Pernthaler and Amann, 2004; Pernthaler and Pernthaler, 2005; Pilhofer et al., 2009), both in pure cultures and in environmental samples. For a further increase of the sensitivity, a second CARD layer can be added (Kubota et al., 2006).

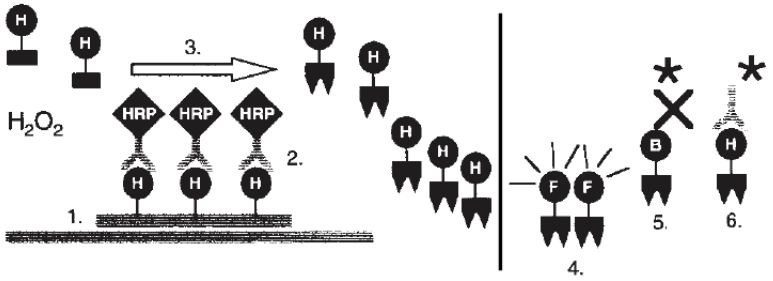


Figure 6: The principle of catalyzed reporter deposition of tyramides 1. In a first step, the hapten-labeled probe hybridizes to target nucleic acids (Hapten = H). 2. In a second step, the anti-hapten antibody conjugated with HRP binds to the hapten. 3. The last step is the catalyzed reporter deposition, in which the HRP will break the H_2O_2 into reactive hydroxyl radicals. The hydroxyl radicals will give one free radical to the phenol ring of the tyramide-conjugates, which, in turn, will bind to the tyrosine residues of cellular proteins. The tyramides can be conjugated with fluorochromes (F), and the signal is visualized directly (4), with biotin (B) and the signal is visualized by further binding of fluorochrome-streptavidin complexes (5) or with haptens, and the signal is visualized by further binding of fluorochrome-antibody complexes (6). Figure adapted from Speel et al., (1999).

1.2.8. Branched DNA (bDNA)

The branched DNA system is using a pre-amplification step to introduce more than one Catalyzed Signal Amplification (CSA) enzymes (alkaline phosphatase or HRP) per bound oligonucleotide probe (Collins et al., 1997). The pre-amplification is based on successive hybridizations of modified probes one to another and is depicted in detail in Figure 7. The modification consists in adding a tail to the probes which serves as binding sequence for the probe in the next level of hybridization. Since the tails consists of repetitive sequence elements, multiple probes will bind to them. The probes used in the last level do not possess any tail, but they are labeled with an enzyme. The total amplification factor is determined by the number of the repetitions in the tails and by the enzymatic reaction. The branched DNA amplification system was used for *in situ* detection of single copy viruses in mammalian cells (Player et al., 2001). The probes used were oligonucleotides covering ~90% of the viral genomes. Since the signal amplification system is based on successive probe hybridizations, the system is rather complicated and non-specific binding might be an additional problem.

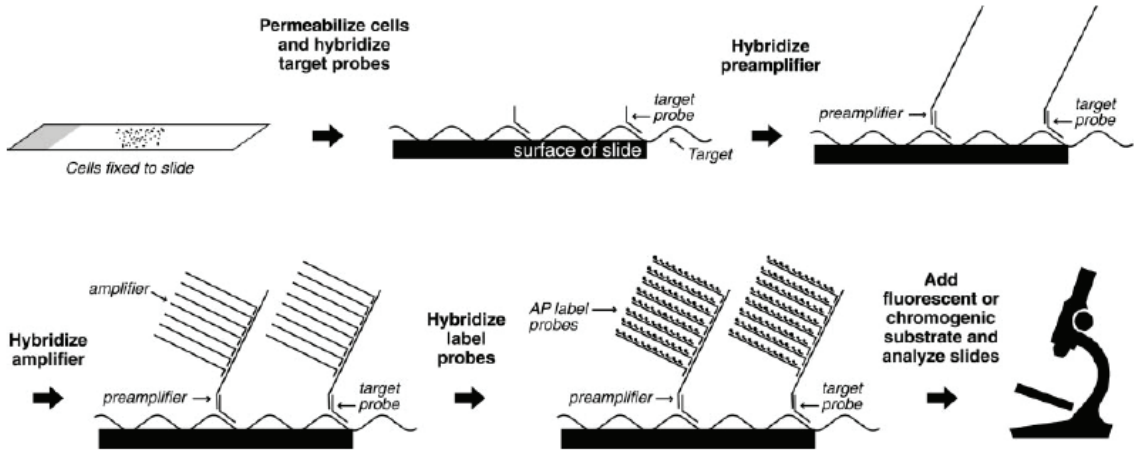


Figure 7: The principle of branched DNA amplification. Figure adapted from Player et al., (2001).

In order to evaluate the potential of the methods described above to detect low copy genes in environmental microorganisms, a summary of their characteristics is given in Table 1.

Table 1: Summary of *in situ* techniques used for low copy gene detection.

technique	probe type	minimum target size	amplification	detection	single-copy detection			main disadvantages
					applied to prokaryotes	combined with 16S rRNA FISH	with environmental samples	
Multiple polynucleotides FISH	Randomly shared polynucleotides	5 kbp	Multiple probes with multiple labels	Via labeled probes	yes	no	no	Target size much larger than the average size of prokaryotic genes.
RING-FISH	One polynucleotide	140 nt	Network formed by multiple labeled probes	Via labeled probes	yes	yes	yes	Probe network formation makes probe specificity difficult to predict. False positives and background due to the high probe concentrations used.
In situ PCR	primer oligonucleotides	500 bp	Target amplification via PCR, sometimes combined with CARD	Via incorporation of labeled dNTPs Via hybridization with labeled probes, followed or not by CARD	yes	no	yes	False positives due to leaking of the amplification product out of the cells and to unspecific incorporation of labeled dNTPs. Cell damage due to high temperatures during PCR.
CPRINS	primer oligonucleotides and reporter probes	600 nt	Target amplification via CPRINS, followed by hybridization with multiple labeled probes	Via incorporation of labeled dNTPs. Via hybridization with multiple labeled probes	yes	yes	yes	Cell damage due to high temperatures during amplification.
LAMP	2 oligonucleotides	140 bp	Isothermal target amplification via LAMP	Via incorporation of labeled dNTPs	yes	no	no	Too many oligonucleotide target regions per gene fragment, false positives due to unspecific incorporation of labeled dNTPs.

Table 1 - continued

technique	probe type	minimum target size	amplification	detection	single-copy detection			main disadvantages
					applied to prokaryotes	combined with 16S rRNA FISH	with environmental samples	
RCA	oligonucleotides	19 nt	Isothermal (reporter) probe amplification via RCA, followed by hybridization with labeled probes	Via hybridization with labeled probes	yes	no	yes	Too many reporter probes/primers needed for the amplification system.
AT-Polymerization	oligonucleotides	40 nt	Isothermal, AT-tailing of probe	Via incorporation of labeled dUTPs	no	no	no	False positives due to unspecific incorporation of labeled dUTPs.
CARD-FISH	randomly sheared polynucleotides	5.2 kbp	Enzymatic deposition of many labeled tyramides	Via deposition of labeled tyramides	no	no	no	Cell permeabilization, background given by the antibody step.
Branched DNA	up to 32 oligonucleotides	viral genome	Layers of hybridizations with reporter probes, followed by CARD	Via deposition of labeled tyramides	no	no	no	Many reporter probes needed in the signal amplification system.

1.2.9. Potential of different signal amplification methods for the detection of low copy genes in environmental microorganisms

The methods for single gene detection which use oligonucleotides amplify the target DNA by PCR-based technologies (e.g. *In Situ* PCR, CPRINS) or by isothermal amplification based technologies (e.g. *In Situ* Loop-Mediated Isothermal Amplification, *In Situ* Rolling Circle Amplification, Peptide Nucleic Acid – Assisted Rolling Circle Amplification FISH). The disadvantage of using PCR-based technologies is the need for repeated high temperature cycles, which often damage the cells. Another problem associated with PCR-based technologies is the appearance of false positives, either due to amplicon leaking from the cells or due to erroneous incorporation of labeled dNTPs. The disadvantage of using the isothermal amplification technologies is that, although they do not need high temperatures, they require many short reporter probes and/or primers which in complex environmental samples might raise specificity issues. The same consideration applies for the branched DNA amplification system.

The methods using polynucleotide probes have the advantage that one probe carries many labels, offering a simple signal amplification method. The disadvantage of RING-FISH is that the signal results from the formation of a molecular network by the probe. Since the formation of the probe network must be considered in probe design, it is difficult to predict the specificity of RING-FISH and this has hindered its routine application in complex environments. The FISH method of Niki and Hiraga (1997) targets genomic fragments > 5 kbp, a size much larger than the average bacterial and archaeal genes.

Therefore, to our knowledge, there is no robust protocol for identification of gene fragments of a size of < 0.5 kbp in complex environmental samples. On the other hand, considering that the amplification factor for CARD is between 26 to 42 fold and that multiple labeled polynucleotides, without amplification, enabled detection of about 5 kbp fragments, it seems feasible that application of polynucleotide probes followed by CARD amplification might lead to the identification of gene fragments of a size of < 0.5 kbp.

1.2.10. Probe choice: oligonucleotides versus polynucleotides

Another advantage of using polynucleotide probes is that, due to their large size, their specificity is less influenced by single nucleotide polymorphism than that of oligonucleotide probes. The main advantage of oligonucleotide probes in rRNA detection – the ability to discriminate at the single mismatch level (Wallace et al., 1981), is also a major disadvantage when used for detection of protein-coding genes in environmental samples. Oligonucleotide probes are short stretches of nucleic acid (~20 nt) that are perfectly matching the targeted region. On the other hand, polynucleotide probes are long stretches of nucleotides (>100 nt),

which tolerate a certain degree of mismatches. In the microbial world, there are many alleles of the same gene, belonging to different microorganisms. Because of their length, polynucleotides cannot be perfect match with all the alleles. The percent of mismatch (%MM) and the conditions of hybridization will determine how many mismatches are tolerated by a polynucleotide probe and, therefore, which of the alleles will be targeted.

1.3. Parameters influencing hybridization of polynucleotide nucleic acids

Hybridization represents the formation of sequence-specific, base-paired duplexes of nucleic acids (Britten and Davidson, 1985). Due to the specificity of the pairing, the ability to hybridize is used in a wide range of molecular biology tools, from Polymerase Chain Reaction (PCR) to Fluorescence *in situ* Hybridization (FISH).

There are two main types of hybridization formats: *in solution*, where both reactants are found in solution, and *mixed phase*, where one of the reactants is bound to a solid support (usually membrane filters or microarray slides) and the other is found in solution (Nakatsu and Forney, 2004). *In situ* hybridization represents a special type of the latter format. The sample (tissue or individual cells) is fixed either on membrane filters or glass slides and the probe is in solution. The target is represented by the nucleic acids in the cells, either DNA or different RNA types. Since they allow detection of specific nucleic acid sequences in whole fixed microbial cells, *in situ* hybridizations techniques have gained much importance in the field of microbiology.

The main factors affecting hybridizations are the ones which influence the stability of nucleic acid duplexes and the rate of hybridization. They have been mostly studied for the in solution hybridization format, but, except for the equations describing kinetics, they behave similarly also for the mixed phase hybridization format.

1.3.1. Temperature

The stability of nucleic acid duplexes decreases with increasing temperature. Hybrid stability is defined by the melting temperature (T_m). The T_m of polynucleotide hybrids is the temperature at which 50% of the base pairs are in single-stranded state and 50% of the base pairs are in double-stranded state (Wetmur, 1991). The T_m and the hybridization rate are correlated. At hybridization temperatures equal to T_m, probe binding and denaturation are in equilibrium. At temperatures above T_m, denaturation is favored. At temperatures below T_m, hybridization of a probe to its target is favored and the two nucleic acid strands bind to each other, i.e. they hybridize. With a further decrease of incubation temperature below T_m, the hybridization rate will increase until it reaches a maximum and then decreases (Figure

8), following a bell shaped curve (Wetmur and Davidson, 1968). Different types of hybrids reach the maximum rate at different temperatures: DNA:DNA at ~25°C, with a broad optimum at 20-30°C, below T_m (Wetmur and Davidson, 1968; Hutton, 1977), and RNA:DNA at 15-20°C below T_m (Birnstiel et al., 1972). For mismatched hybrids, both T_m and hybridization rate are decreasing with increasing %MM (Anderson and Young, 1985). The T_m of hybrids of identical length and sequence is higher for RNA:RNA than for DNA:DNA (Lesnik and Freier, 1995). At high formamide concentrations, RNA:DNA hybrids are more stable than their DNA:DNA counterparts (Casey and Davidson, 1977). Generally, the hybridization rate of RNA:DNA hybrids is lower than that of DNA:DNA hybrids (Bishop, 1972; Hutton and Wetmur, 1973a; Galau et al., 1977a; Galau et al., 1977b).

The bell shaped curve is maintained in the presence of different chemicals, e.g. Na^+, formamide and tetraalkylammonium salts (Chang et al., 1974; Hutton, 1977; Chien and Davidson, 1978), and also, for hybridization on solid support format (Anderson and Young, 1985). For chemicals like tetraalkylammonium salts, the curve is narrower, without the broad flat maximum between 20 and 30°C bellow T_m (Chang et al., 1974).

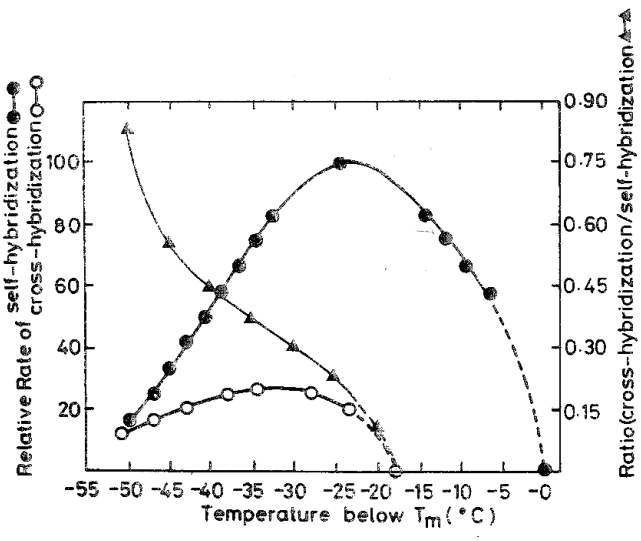


Figure 8: Relative renaturation rate versus temperature of DNA:DNA hybrids. (• - •) – renaturation of perfectly matched hybrids; (o - o) renaturation of mismatched hybrids (cross hybridization); (∇ - ∇) ratio of the renaturation rate of mismatched hybrids and the renaturation rate of perfectly matched hybrids. Figure from Wetmur and Davidson, (1968) .

1.3.2. Salt concentration

Since nucleic acids are highly charged polyanions, monovalent or divalent cations have a stabilizing effect on duplexes. They neutralize the negative charges of the phosphate groups present on the backbone of the nucleic acids and thus minimize the repulsive interaction between the two strands in a duplex. As a consequence, increasing concentration of cations will produce an increase in both T_m and hybridization rate.

The stabilizing effect of chloride salts of monovalent cation (Li^+, Na^+, K^+, Rb^+, Cs^+) on T_m is logarithmic and it has a pronounced strength in the range 0-0.2 M (Gordon, 1965; Schildkraut and Lifson, 1965). At high salt concentrations (1-1.2 M) the T_m will reach a plateau, after which it starts to decrease (≥ 3 M), due to a denaturation effect produced by the anion in the salt (Hamaguchi and Geiduschek, 1962; Schildkraut and Lifson, 1965; Gruenwedel et al., 1971; Hutton, 1977). The increase in the hybridization rate is highest in the interval 0-0.1 M, where a doubling of the salt concentration can increase the hybridization rate of DNA:DNA hybrids about 5 to 10 times. Above 1.2 M, the increase in the hybridization rate levels off (Wetmur and Davidson, 1968; Britten and Davidson, 1985). The influence of salt on the hybridization rate is less for RNA:DNA than for DNA:DNA hybrids (Ness and Hahn, 1982). For *in situ* hybridizations, the most used monovalent cation is Na^+.

Divalent cations have a similar, but much stronger effect than monovalent cations, both on T_m and hybridization rate. For example, a 1 M NaCl solution has the same effect on RNA secondary structure stability as a 10 mM $MgCl_2$ / 50 mM NaCl mixed solution (Jaeger et al., 1990). Due to this strong effect even at very low concentrations, the presence of Mg^{2+} in hybridization / washing buffers might prevent discrimination between specific and unspecific nucleic acid sequences. In such cases, a chelating agent, e.g. EDTA, has to be used to eliminate the effect of Mg^{2+} (Britten and Davidson, 1985).

1.3.3. Denaturants

In aqueous salt solutions the T_m of hybrids can be quite high, which leads to high hybridization / washing temperatures (above 60°C). Long incubations at such high temperatures can lead to thermal strand scission, depurination (Anderson and Young, 1985) and, in the case of *in situ* experiments, to a deterioration of cellular morphology and subsequent nucleic acid loss (Raap et al., 1986). To lower the incubation temperatures, the use of denaturants was introduced. Chemicals like formamide, urea, dimethylsulfoxide and ethylene glycol are destabilizing double-stranded nucleic acids, lowering their T_m and thereby, lowering the incubation temperatures (Bonner et al., 1967; Gillespie and Gillespie, 1971; Friedrich and Feix, 1972; Schmeckpeperl and Smith, 1972; Hutton, 1977).

For *in situ* hybridizations, the most used denaturant is formamide. With increasing formamide concentration, the T_m of nucleic acid duplexes decreases. For DNA:DNA hybrids the relation between formamide concentration and decrease in T_m is linear, with about 0.6-0.72°C per 1% formamide (McConaughy et al., 1969; Blüthmann et al., 1973; Casey and Davidson, 1977; Hutton, 1977). The value depends on the Na^+ concentration, being higher for concentrations >0.8 M Na^+. The effects of formamide and Na^+ cannot be treated separately. At 0% formamide, the T_m reaches a plateau at about 1 M Na^+. With increasing

the formamide concentration, not only the T_m is decreasing, but also, is reaching the plateau at much lower concentrations of Na^+. At 70% formamide, the T_m reaches the plateau at about 0.4 M Na^+. Within the range of 0-70% formamide, the following formula will give the Na^+ concentration (Molar) when the plateau in T_m is reached: $[Na^+]$ = - 0.0096 x percent formamide + 1.06. The same trend applies to the hybridization rate: with increasing formamide concentration, the hybridization rate not only decreases, but also, the Na^+ concentration at which the plateau is reached, is lower. The hybridization rate decreases with about 1.1% for each 1% formamide. The decrease in hybridization rate by adding formamide is more than compensated by the elimination of DNA degradation at high temperatures (Hutton, 1977).

In the case of RNA:DNA hybrids, the relation between T_m and formamide concentration is not linear, the decrease in T_m being smaller at high concentrations (above 50%) than at low formamide concentrations. As a consequence, at high formamide concentrations RNA:DNA hybrids are more stable than their DNA:DNA counterparts. This feature can be used to discriminate RNA:DNA from DNA:DNA hybrids (Casey and Davidson, 1977).

1.3.4. Strand concentration

Since in mixed phase hybridizations the concentration of either the probe or the target is usually much higher than that of the other, their kinetics is of pseudo-first order. Generally, the hybridization rate is expressed as the time necessary for the probe (found in solution) to bind to one half of the targets ($t_{1/2}$) and is described by the following equation: $t_{1/2} = \ln 2/(k \times C)$, where k is the hybridization rate constant and C is the concentration of the probe in solution (Nakatsu and Forney, 2004). As indicated by this equation, the half time for hybridization decreases with increasing probe concentration.

1.3.5. Probe length

The hybridization rate increases with increasing the length of the participating strands. However, with increasing length of the probe, the diffusion of the probe inside the cell will become more and more limiting in *in situ* hybridizations. Therefore, longer probes will have lower hybridization rates. Often, permeabilization is necessary to increase the diffusion of probe molecules inside cells (Brahic and Haase, 1978).

The strand length is positively correlated to the T_m. The length effect is given by the following formulae (Sambrook, 2001): D = 500/L for DNA:DNA hybrids, and D = 820/L for DNA:RNA and RNA:RNA hybrids, where D is the reduction in T_m (°C) and L is the length of the base-paired region of a duplex.

1.3.6. Base composition (%GC)

GC pairs form stronger hydrogen bonds and more energy is needed to break them, as compared to AT / AU pairs. As a result, in normal salt solutions (NaCl) they contribute more to T_m and, with increasing %GC , the T_m increases (Marmur and Doty, 1962). The effect of the base composition on the hybridization rate is controversial. Generally, it has been accepted that the base composition has a small effect on the rate of hybridization, which increases slightly with increasing %GC, as determined for DNA:DNA hybrids (Wetmur and Davidson, 1968).

There are, however, certain chemicals, e.g. betaine and tetraalkylammonium salts, that eliminate or even reverse the effect of %GC on T_m (Melchior and Hippel, 1973; Orosz and Wetmur, 1977; Rees et al., 1993). For example, 2.4 M tetraethylammonium chloride (TEACl) completely eliminates T_m dependence on %GC, while also destabilizing the duplex DNA and decreasing the T_m. Moreover, the melting transitions in TEACl are narrower than in standard salts, allowing discrimination between closely related hybrids (Melchior and Hippel, 1973; Chang et al., 1974).

1.3.7. Mismatches

Mismatches between the two hybridizing strands produce a decrease both in T_m and in hybridization rate. Early studies on chemically modified nucleic acids have estimated the decrease in T_m to be 0.5-1.5°C per 1% mismatch (Anderson and Young, 1985). Two recent studies, benefiting from the knowledge of the exact sequence of the hybridizing strands, found a 0.6°C and 0.85°C decrease per 1% mismatch (Caccone et al., 1988; Springer et al., 1992). The exact value of the decrease depends on the %GC and on the distribution of mismatches in the heteroduplex (Anderson and Young, 1985).

Different studies have quantified the decrease in hybridization rate with the increase in %MM (Bonner et al., 1973; Hutton and Wetmur, 1973b). Although the numbers obtained differed between studies, it is clear that with >25% mismatches the kinetics is no longer favorable for hybrid formation. At about 10% mismatches, one of the studies reported a ~50% reduction in the hybridization rate (Bonner et al., 1973), while the other reported only a ~20% reduction (Hutton and Wetmur, 1973b). For mismatched hybrids, the reduction in the optimum hybridization temperature is half the reduction in T_m (Bonner et al., 1973).

1.3.8. Inert polymers

Some inert polymers increase the hybridization rate. Such examples are dextran sulfate, Ficoll (Chang et al., 1974), polyethylene glycol and anionic polymers of polyacrylate and polymethylacrylate (Schwartz and Wash, 1989). The increase in the hybridization rate has been explained as a reduction by the polymer of the effective volume in which the nucleic acid molecules are found, leading to higher concentrations of nucleic acids.

A 10% dextran sulfate solution increases the DNA reassociation rate about 10 fold when the hybridization is taking place in solution (Wetmur, 1975). For hybridizations on solid support, the hybridization rate is increased 4-5 times for single-stranded probes and 12-100 for double-stranded probes (Wahl et al., 1979). This high increase in the hybridization rate for double-stranded probes could be due in part to the formation of probe networks resulting from the binding of partially complementary probe regions (for probes prepared by shearing or nick translation). Dextran sulfate itself has no influence on the thermal stability of nucleic acid duplexes. On the other hand, most often dextran sulfate is delivered as a sodium salt. In these cases, the dextran sulfate contribution to the total Na^+ concentration has to be taken into account when estimating T_m. For filter hybridization with single-stranded probes, polyethylene glycol enhances the hybridization rate much more than dextran sulfate (Amasino, 1986).

1.3.9. pH

Both T_m and hybridization rates are not affected by the pH in the range from pH 6 to pH 8. At alkaline and acidic pH, nucleic acids denature (Wetmur and Davidson, 1968).

1.3.10. Equations to estimate T_m

There are several equations which estimate T_m of polynucleotide hybrids. Most of them take into consideration only Na^+ concentration and %GC. The most useful formula for *in situ* hybridizations is that of Wetmur (1991), which takes into account the influence on T_m of Na^+ concentration, formamide concentration, %GC, strand length and degree of mismatch between probe and target.

- Schildkraut-Lifson (Schildkraut and Lifson, 1965) equation for DNA:DNA hybrids. It works for Na^+ concentrations in the range 0.01-0.2 M.

$T_m = 102 + 16.6 \times \log[Na^+]$

- Schildkraut-Lifson equation (Schildkraut and Lifson, 1965) including the term for %GC from Marmur and Doty (1962):

$T_m = 81.5 + 16.6 \times \log[Na^+] + 0.41 \times \%GC$

- Wetmur's extension of the Schildkraut-Lifson equation for DNA:DNA hybrids (Wetmur, 1991). The term for Na^+ contribution has been adjusted to fit Na^+ concentrations up to 1 M. The terms for formamide, strand length and mismatches have been also included.

$T_m = 81.5 + 16.6 \times \log([Na^+]/(1 + 0.7 \times [Na^+]) + 0.41 \times \%GC - 500/N - x\%MM - 0.63FA$

Additionally, Wetmur is giving two more equations:

for RNA:RNA hybrids:

$T_m = 78 + 16.6 \times \log10([Na^+]/(1+0.7 \times [Na^+]) + 0.7 \times \%GC - 500/N - x\%MM - 0.35FA;$

and for RNA:DNA hybrids:

$T_m = 67 + 16.6 \times \log10([Na^+]/(1+0.7 \times [Na^+]) + 0.8 \times \%GC - 500/N - x\%MM - 0.5FA;$

However, the last two equations have been obtained from much more reduced datasets than the one for DNA:DNA hybrids. Therefore, their predictive powers are lower. The terms for formamide have been added later (Sambrook, 2001), by taking into consideration the data on RNA:DNA hybrids from Casey and Davidson, (1977) and on RNA:RNA hybrids from Bodkin and Knudson, (1985).

In these formulae above, $[Na^+]$ is Na^+ concentration (molar), %GC is the percent molar G+C, N is the probe length, %MM is the percent mismatch of probe with the target, constant x is usually 0.5 – 1.5°C per %MM, FA is formamide concentration (% vol/vol).

For the particular purpose of predicting the parameters for hybridizations, the equations from Wetmur (1991) are the ones which take into consideration all factors important. Moreover, it is the only equation which applies to Na^+ concentrations up to 1 M. However, also this heuristic formula can give only approximate results, especially since it is very hard to separate one parameter from the effect of another parameter. For example, it has been shown that the influence of monovalent cations on the melting temperature is somehow dependent on the %GC and on the formamide concentration (Gruenwedel et al., 1971; Hutton, 1977). A more reliable way to calculate the T_m would be based on the thermodynamic properties of the hybrids. There are several such algorithms in use for oligonucleotides, but only few can be used also for the prediction of T_m values of polynucleotide hybrids and they do not consider the influence of formamide (Steger, 1994; Blake et al., 1999).

1.5. Thesis Objectives

Although there are several FISH protocols for gene detection in microorganisms, we are still missing a robust method for application in environmental samples. Such a FISH protocol should use polynucleotide probes for gene detection and thus, should take into

account the parameters influencing the hybridization of polynucleotides. Moreover, it should be combined with 16S rRNA-based identification of single cells.

Since oligonucleotide probes differ profoundly from polynucleotide probes, the well established concept for oligonucleotide probe design cannot be transferred to polynucleotides. Moreover, the specificity of polynucleotide probes has not been thoroughly investigated yet and, consequently, a rational probe design concept is still missing.

Therefore, a study aiming to develop a reliable protocol for gene detection in environmental microorganisms should address not only the sensitivity, i.e. signal amplification, but also the specificity, i.e. probe design and determination of hybridization parameters.

1.5.1. Objective 1: Development of concepts and software for a rational design of polynucleotide probes.

The first objective will address the probe design part of the specificity, by the development of a concept and software for rational design of polynucleotide probes. Such a concept and software should implement the knowledge on factors determining the binding of polynucleotides to fully matched and mismatched targets. To gain further insights into the potential and limitations of polynucleotides to be used as probes for gene detection in environmental microbes, this newly developed concept should be tested on several metabolic marker genes. This case studies will be focused on three commonly used "functional genes" encoding the alpha subunits of ammonia monooxygenase (*amo*A), particulate methane monooxygenase (*pmo*A) and adenosine-5'-phosphosulfate reductase (*apr*A).

1.5.2. Objective 2: Development of geneFISH – an *in situ* technique for linking gene presence and cell identity in environmental microorganisms

The second objective will address both the sensitivity, by developing a FISH protocol which detects low copy genes, and the specificity, by the determination of hybridization parameters. The *in situ* technique should allow linking the cell identity and the gene presence in environmental microorganisms. For this, CARD-FISH using rRNA-targeted oligonucleotide probes for single cell identification (Pernthaler et al., 2002a) should be combined with the detection of gene fragments by polynucleotide probes. The development of the gene detection method should start from protocols established by Pernthaler and Amann (2004) and Wagner et al. (1998) for mRNA detection. Those protocols used multiple Dig-labeled polynucleotide probes followed by the binding of HRP-conjugated antibodies and CARD. This new protocol, called geneFISH, should be developed and tested in *Escherichia coli*. In a second phase, it should be applied on different environmental samples. The parameters for hybridization with polynucleotide probes should be determined based on

the T_m of the probe-target hybrids and should take into account the factors influencing the hybridization rate.

1.5.3. Objective 3. Development of mRNA HISH-SIMS – the first steps toward single gene detection with NanoSIMS.

The last objective will be to initiate the development of a method for detection of mRNA with NanoSIMS at single cell level, and ultimately, of genes. The method should involve the use of fluorinated polynucleotide probes and should be applied for *nif*H mRNA detection in *Crocosphaera watsonii* cells. The main advantage of NanoSIMS in microbial ecology is that it measures substrate incorporation rates at single cell level. Combined with mRNA/gene HISH-SIMS, this would be a powerful tool for studying the physiology and cell biology of yet uncultivated microorganisms.

2. Materials and methods

Materials used in this thesis, together with their supplying company and catalog number, are listed in Appendix D – Materials.

2.1. PolyPro – software for a rational design of polynucleotide probes

2.1.1. Programming

PolyPro was written in Delphi programming language as a monolithic application which runs on Windows OS. PolyPro is freely available from http://www.mpi-bremen.de/en/method-development/PolyPro.zip.

2.1.2. Probe design module

To calculate the percent mismatches, polynucleotide sequences were aligned using a global, pairwise algorithm. The formula used for percent mismatch calculation is the following: %MM = (M×100)/N, where %MM = percent mismatch, M = number of mismatches (as resulted from a global pairwise alignment) and N = probe length.

For ΔT_m calculations, probe T_m was estimated based on the %GC with the following formulas: for RNA:RNA hybrids, T_m=0.7 (%GC), for RNA:DNA hybrids, T_m=0.8 (%GC) and for DNA:DNA hybrids T_m=0.41 (%GC) (Wetmur, 1991).

2.1.3. Hybridization Parameters Calculator (HPC) module

This module is calculating different parameters for in situ hybridization, based on the formulas from (Wetmur, 1991): i) for RNA:RNA hybrids, T_m = 78 + 16.6 log10([Na^+]/(1+0.7[Na^+]) + 0.7 (%GC) – 500/N – x%MM – 0.35FA; ii) for RNA:DNA hybrids, T_m = 67 + 16.6 log10([Na^+]/(1+0.7[Na^+]) + 0.8 (%GC) – 500/N – x%MM – 0.5FA; and iii) for DNA:DNA hybrids, T_m = 81.5 + 16.6 log10([Na^+]/(1+0.7[Na^+]) + 0.41 (%GC) – 500/N – x%MM – 0.63FA. In these formulas, [Na^+] = Na^+ concentration (molar), %GC = percent molar G+C, N = probe length, %MM = percent mismatch of probe with the target, constant x = usually 0.5 – 1.5°C per % mismatch, FA = formamide concentration (%).

2.2. Assessment of the specificity of polynucleotide probes

The sequences of *apr*A, *amo*A and *pmo*A genes belonging to cultivated (or otherwise identified) bacteria and archaea were retrieved from NCBI database. The sequences were

downloaded in GenBank format. For the *apr*A gene a 350 bp fragment was selected for probe design, while for *amo*A and *pmo*A a common fragment of 270 bp was selected. For analysis, PolyPro software was used. The parameters used in Polynucleotide Probe Designer (PPD) module were the following: DNA:DNA hybridization, Th1=10 (or when Th2<10, Th1=5) and T_m difference of 0.05. Th1, Th2 and T_m difference are explained inn section 3.1.1..

2.3. Cell cultures – cultivation and fixation

2.3.1. Escherichia coli cultures

The *E. coli* cultures used in this study were: *E. coli* strain K12, *E.coli* copy control plasmid clones 1E3, 3G4 and 1E7, *E. coli* fosmid clones WS 39F7, WS 7F8, NK54 and FOS_101A12. The copy control cultures were *E. coli* EPI300 transformed with a pCC1 vector (Epicentre) that contained three closely related crenarchaeotal *amo*A genes. The fosmid clones came from a metagenomic clone library prepared by Marc Mussmann. The first two fosmid clones had been described by Mussmann et al., 2005. The cultures were grown overnight starting from single colonies on LB media (1% Tryptone, 0.5% Yeast Extract and 1% NaCl) with 12.5 µg ml^{-1} chloramphenicol. In order to increase the number of vector copies per cell, the 1E7 culture was amended with Induction Solution (Epicentre) to a final concentration of 2x. The cells were harvested by centrifugation and then fixed in 1% paraformaldehyde in 1x PBS, for 1 h at room temperature. The paraformaldehyde solution was removed by centrifugation and aspiration, followed by one 1x PBS wash. The cells were resuspended in 1x PBS and 96% ethanol was added to a final concentration of 50%. The cells were stored at -20 °C.

2.3.2. Crocosphaera watsonii cultures

Axenic batch cultures of *Crocosphaera watsonii* WH8501 were used. They were grown at 28°C in YBCII medium (Chen et al., 1996) in dark - and temperature controlled chambers. For fixation, paraformaldehyde was added at a final concentration of 1% directly to a subsample of the culture and incubated 1 h at room temperature. The fixation was stopped by centrifugation and washing of the cell pellet two times with 1x PBS, 50% ethanol. The cells were stored at -80 °C in 96% ethanol.

2.3.3 AK199 enrichment

The enrichment was obtained using as inoculum from the anoxic sediment layer (app. 2-3 cm depth) from Sylt intertidal sediments. The cells were harvested by centrifugation and then fixed in 1% paraformaldehyde in 1x PBS, for 1 h at room

temperature. The paraformaldehyde solution was removed by centrifugation and aspiration, followed by one 1x PBS wash. The cells were resuspended in 1x PBS and 96% ethanol was added to a final concentration of 50%. The cells were stored at -20 °C.

2.4. Environmental samples – sampling and fixation

2.4.1. Seawater samples from Benguela upwelling system Namibia

The seawater samples were collected from the Benguela upwelling system on the Namibian shelf, onboard the R/V Meteor in May/June 2008 during the M76/2 cruise. Two stations were sampled: station 249 (22.99° S, 14.04° E), at a depth of 128 m, and station 213 (20.99° S, 13.36° E) at depths of 65 m and 76 m. Samples for DNA extraction were filtered through 0.2 µm Sterivex GS filters (Millipore), and stored at -80 °C until extraction. Samples for FISH were fixed by adding paraformaldehyde at a final concentration of 1% to the freshly collected seawater and incubating at 4 °C for 8-14 h. The fixation was stopped by filtration and the filters were stored at -20 °C or -80 °C.

2.4.2. Seawater samples from the Baltic Sea

Seawater samples enriched in *Aphanizomenon* colonies were sampled vertically in the upper 10 m of the water column at station B1 (58.80° N, 17.62° E) in the Stockholm archipelago using a plankton net (Hydrobios, 0.5 m diameter, mesh size: 90 µm). Samples were immediately brought to the laboratory where they were filtered on 0.22 µM GTTP filters (Millipore) and immediately stored at -20°C.

2.4.3. Bathymodiolus puteoserpentis

The *Bathymodiolus puteoserpentis* samples were collected from the Logatchev hydrothermal vent field (14.75° N, 44.98° W), on board the research vessel Meteor, during the M64/2 cruise. The samples were fixed in 4% formaldehyde in 1x PBS, for 10-16 h at +4°C (Pernthaler and Pernthaler, 2005).

2.5. DNA extraction from environmental samples

DNA was extracted from the seawater samples of the organic phases of RNA extraction via the Totally RNA Kit (Ambion). Cell lyses were carried out within a Sterivex filter cartridge according to (Somerville et al., 1989), with an additional 10 units ml^{-1} of RNase inhibitor (SUPERaseIn, Ambion). DNA was extracted from the *Bathymodiolus puteoserpentis* samples according to (Zhou et al., 1996).

2.6. Cloning, screening and sequencing

Clone libraries were constructed and sequenced to assess the diversity of several genes in environmental samples. The *amo*A gene was cloned from the DNA extracted from the Namibian seawater samples. The *nif*H gene was cloned from the DNA extracted from the *Aphanizomenon* concentrate. The *hyn*L and *apr*A gene was cloned from the DNA extracted from the *Bathymodiolus puteoserpentis* samples.

2.6.1. PCR amplification

The following primers were used: (i) Arch-amoAF and Arch-amoAR for the *amo*A fragment, (ii) *nif*H1 and *nif*H2 for the *nif*H fragment (see Appendix A: primers), (iii) HUPLX1 and HUPLW2 for the *hyn*L fragment (Csákia et al., 2001), and (iv) AprA-1-FW and AprA-5-RV for the *apr*A fragment (Meyer and Kuever, 2007a). Initially, different PCR cycle numbers were tested for each sample. The thermal protocol with the lowest cycle number where a product was visible in the gel was chosen. The number of replicates performed during the PCR for cloning was five for *amo*A, 14 for *nif*H, 10 for *hyn*L and seven for *apr*A. The replicates were poled during the purification step. The components of the PCR reaction are described in Table 2 and Table 4, respectively. The thermocycling programs are described in Table 3 and Table 5, respectively. As template was used DNA extracted from the respective environmental sample.

Table 2: PCR reaction for *nif*H and *amo*A amplicons, fragment for cloning:

Reagent	***nif*H amplicons**		***amo*A amplicons**	
	volume (µl)	**Final concentration**	**volume (µl)**	**Final concentration**
10x Reaction Buffer	2.5	1x (1.5 mM Mg^{2+})	2.5	1x (1.5 mM Mg^{2+})
dNTPs (2.5 mM each)	2	200 µM	2	200 µM
10x BSA	2.5	1x	2.5	1x
F primer (100 µM)	0.5	2 µM	0.25	1 µM
R primer (100 µM)	0.5	2 µM	0.25	1 µM
Eppendorf Taq Polymerase (5 U/µl)	0.25	1.25 U/rxn	0.2	1 U/rxn
$MgCl_2$ (25 mM)	1	1 mM	-	-
PCR water	14.75		16.8	
Template	1		0.5	
Total volume	**25**		**25**	

Table 3: Thermocycling program for *nif*H and *amo*A amplicons (for cloning):

Step no.	*nif*H amplicons		*amo*A amplicons		process
	Temperature (°C)	time	Temperature (°C)	time	
1	94	5 min	95	5 min	Initial denaturation
2	94	1 min	94	45 sec	Denaturation
3	57	1 min	53	1 min	Annealing
4	72	1 min	72	1 min	Elongation
5			Go to 2 repeat x*		
6	72	15 min	72	30 min	Final elongation
7	4	Hold	4	Hold	store

* x = 25 cycles for the Namibia Station 213, 65 m and Station 213, 76 m and for the Baltic Sea samples; 27 cycles for the Namibia sample Station 249.

Table 4: PCR reaction for *hyn*L and *apr*A amplicons, fragment for cloning

Reagent	*hyn*L amplicons		*aprA* amplicons	
	volume (µl)	Final concentration	volume (µl)	Final concentration
10x Reaction Buffer	5	1x (1.5 mM Mg^{2+})	2	1x (1.5 mM Mg^{2+})
dNTPs (2.5 mM each)	5	250 µM	2	250 µM
10x BSA	-	-	2	1x
F primer (50 µM)	0.5	0.5 µM	0.2	0.5 µM
R primer (50 µM)	0.5	0.5 µM	0.2	0.5 µM
Eppendorf Taq Polymerase (5 U/µl)	0.2	1 U/rxn	0.1	0.5 U/rxn
Water	37.8		12.5	
Template	1		1	
Total volume	**50**		**20**	

Table 5: Thermocycling program for *hyn*L and *apr*A amplicons (for cloning)

Step no.	*hyn*L amplicons		*aprA* amplicons		process
	Temp (°C)	time	Temp (°C)	time	
1	95	5 min	95	5 min	Initial denaturation
2	95	1 min	95	1 min	Denaturation
3	52.5	1 min 30 sec	56	1 min 30 sec	Annealing
4	72	2 min	72	2 min	Elongation
5			Go to 2 repeat x*		
6	72	10 min	72	30 min	Final elongation
7	4	Hold	12	Hold	store

* 32 cycles for *hyn*L and 28 cycles for *apr*A

2.6.2. Purification of amplicons used for cloning

The *amo*A amplicons used for cloning were purified with the Gene Clean Turbo kit (see Appendix C). The *nif*H amplicons were first concentrated using Microcon YM100 columns, then the specific band was excised from a 3% agarose gel and purified with the Pure Link Quick Gel extraction kit (see Appendix C). The *hyn*L and *apr*A amplicons were purified using the QIAquick PCR purification kit (QIAGEN, Hilden).

2.6.3. Cloning

The a*mo*A and *nif*H amplicons were cloned with TopoTA cloning for sequencing kit (see Appendix D: Materials), including the pCR4 vector and One Shot TOP10 competent. Additionally, *amo*A was cloned in a copy control vector-host system, using the Copy Control cDNA, Gene and PCR Cloning Kit (see Appendix D: Materials), including the pCC1 copy control vector and Epi300 competent cells. The *hyn*L and *apr*A amplicons were cloned with the pGEMT-Easy vector and One Shot TOP10 competent cells.

TopoTA cloning

The TopoTA cloning reaction was set up as described in Table 6 and the pGEMT-Easy ligation reaction as described in Table 7, followed by 30 min incubation at room temperature. The transformation reaction was set up by adding 2 µl of the cloning reaction to a vial of One Shot TOP10 chemically competent *E. coli* cells. The transformation reaction was incubated for 30 min of ice, then for 40 sec at 42°C (water bath) for heat shocking, followed by 5 min on ice. Next, 250 µl of SOC medium (prewarmed at RT) were added and the reaction was incubated for 1 h at 37°C, with shaking at 200 rpm. Different volumes of the transformation reaction were platted on LB with 100 µg/ml Ampicillin plates, followed by overnight incubation at 37°C. Single colonies were picked and inoculated in 96 well plates, each well having 100 µl LB with 100 µg/ml Ampicillin. Archive plates were created by adding 50 µl 87% glycerol to a 100 µl culture, and then stored at -80°C.

Table 6: TopoTA cloning reaction

component	**volume**
Purified amplicons	0.5 - 4 µl
Salt solution	1 µl
Water	up to 6 µl
vector	1 µl
Total volume	**6 µl**

Table 7: PGEM-T Easy ligation reaction

component	volume
Purified amplicons	2 µl
2X Ligation buffer	5 µl
Water	1 µl
Ligase	1 µl
vector	1 µl
Total volume	**10 µl**

Copy control cloning

End-repair to generate blunt-ended and 5'-phosphorylated PCR product

The amount of PCR product needed in each reaction was calculated as follows: **ng of PCR product** = [75 ng (amount of pCC1 vector in the ligation) x 630 (Length of PCR product in bp)]/8139 (size in bp of the CopyControl pCC1 vector) = 5.8 ng. The end repair reaction was set up as described in Table 8 and then incubated at room temperature for 30 min, followed by 10 min at 70°C and 5 min on ice.

Table 8: Copy Control end-repair reaction

component	volume
Purified amplicons	x µl with 5.8 ng DNA
Water	up to 10 µl
PCR 10x Cloning Buffer	1 µl
PCR End-Repair Enzyme Mix	1 µl
Total volume	**10 µl**

Ligation of end-repaired PCR product into CopyControl pCC1 (blunt cloning-ready) vector

To the 10 µl end-repair reaction were added 1 µl of the CopyControl pCC1 Vector and 1 µl of Fast-Link DNA Ligase. The reaction was incubated for 2 h at RT, followed by 10 min at 70°C and 5 min on ice.

Transformation of the chemically competent TransforMax EPI300 E. coli

The chemically competent TransforMax EPI300 *E. coli* was thawed on ice. To 50 µl of the competent cells were added 5 µl of the Ligation Reaction. The mixture was incubated on ice for 30 min and then heat shocked at 42°C for 30 seconds, followed by 2 min on ice. To the mixture 250 µl of SOC medium were added, followed by incubation at 37°C for 1 h, with gentle shaking. Different volumes of the transformation reaction were platted on LB with 12.5 µg/ml Chloramphenicol plates, followed by overnight incubation at 37°C. Single colonies were picked and inoculated in 96 deep well plates, each well having 500 µl LB with

12.5 µg/ml Chloramphenicol. Archive plates were created by adding 100 µl 87% glycerol to a 100 µl culture, and then stored at -80°C.

2.6.4. Screening of clone libraries

All clone libraries prepared by TopoTA cloning or PGEM-T Easy were PCR screened with M13 F and M13 R primers. See Table 9 for screening PCR reaction and Table 10 for thermocycling program. The template was represented by overnight cultures of *E. coli* clones. The presence of the insert was checked by agarose gel electrophoresis. The screening PCR products were purified on Sephadex columns (see Appendix C, Sephadex purification) and then used as template for sequencing PCR.

Table 9: PCR reaction for screening PCR

Reagent	volume (µl)	Final concentration
10x Reaction Buffer	2	1x (1.5 mM Mg^{2+})
dNTPs (2.5 mM each)	2	250 µM
10x BSA	2	1x
F primer (200 µM)	0.05	0.5 µM
R primer (200 µM)	0.05	0.5 µM
Eppendorf Taq Polymerase (5 U/µl)	0.08	0.4 U/rxn
PCR Water	13.32	
Template	0.5	
Total volume	**20**	

Table 10: Thermocycling program for screening PCR

Step no.	Temperature (°C)	time
1	94	10 min
2	94	1 min
3	50	1 min
4	72	2 min
5	Go to 2 repeat 27×	
6	72	10 min
7	15	Hold

The *amo*A clone libraries created with the Copy Control cloning were screened by vector size. For screening, 2 µl 100x Copy Control Induction solution were added to 200 µl of

2.5 h old cultures. The cultures were further incubated at 37°C for 2 h, shaking at 600 rpm. Then, the cultures were centrifuged for 10 min at 3000 × g and the pellet was resuspended in 20 µl Epi Blue solution, followed by 20 µl Epi Lyse solution. The cell lysate was then run on an agarose gel to estimate the size of the vector plus insert.

2.6.5. Plasmid extraction from Copy Control clones

For plasmid extraction, the copy control clones were induced to a high copy number of vector per cell. The induction was done as follows: 4.5 ml fresh LB with 12.5 µg/µl Chloramphenicol and Copy Control Induction Solution to a final concentration of 1x were added to 500 µl overnight culture (37°C, 250 rpm). The cultures were further incubated for 5 h at 37°C, shaking at 250 rpm. The plasmid DNA was extracted then from the cell pellets using the Plasmid Maxi kit.

2.6.6. Sequencing of clone libraries

The *amo*A and *nif*H clone libraries were sequenced using T7 or T3 primers (see Appendix A: primers). The *hyn*L clone library was sequenced using T7, SP6 or HUPL primers (see Appendix A: primers). The Big Dye Terminator Sequencing Kit (Applied Biosystems) was used for sequencing PCR. For the TopoTA and pGEM-T Easy clone libraries, the template used was purified screening PCR, diluted to a concentration of 0.8-3.3 ng/µl. For the copy control clone libraries, the template was purified plasmid DNA of a concentration of 100-200 ng/µl. The reaction was set up as described in Table 11 and the thermocycling as in Table 12. After sequencing PCR, the reactions were purified on Sephadex columns (see Appendix C, Sephadex purification) and the capillary electrophoresis was run on an ABI 3700 instrument.

Table 11: Sequencing PCR reaction

Components	Volume (µl)	Final concentration
Big Dye Mix	1	
5x Sequencing Buffer	0.5	1 x
Primer (200 µM)	0.025	1 µM
PCR W	0.475	
template	3	
Total volume	**5**	

Table 12: Thermocycling program for sequencing PCR

Step no.*	Temperature (°C)	time
1	96	5 min
2	96	10 sec
3	X**	5 sec
4	60	4 min
5	Go to 2 repeat 59×	
7	4	Hold

* for steps 1-4, the ramping was 1°C/sec
** see Appendix A: primers for the annealing temperature for each primer used

2.6.7. Sequence analysis

Nucleotide sequences were cleaned and assembled using DNA Baser software (HeracleSoftware, Germany, www.DnaBaser.com). The identity of the sequences was verified via BlastN searches (Johnson et al., 2008). The sequences were aligned and phylogenetically analyzed using ARB (Ludwig et al., 2004). The distance matrices generated therein were further used for statistical analysis using DOTUR (OTU grouping and distribution, Chao and Jackknife diversity estimators) (Schloss and Handelsman, 2005) and MOTHUR (Venn diagrams) (Schloss et al., 2009).

2.6.8. Nucleotide sequence accession numbers.

All partial putative *amo*A gene sequences determined in this study were deposited in the GenBank database under the following accession numbers: GQ500142 - GQ500577.

2.7. Probe design

All probes used for gene or mRNA detection were polynucleotides of about 350 nt or bp length. The probes applied on pure cultures were designed based on the gene sequence of the targeted organisms. The probes applied on environmental samples were designed based on sequences from the clone libraries, using PolyPro software.

2.7.1. NonPolyPr350: the negative control probe for gene detection.

NonPolyPr350 was synthesized from a bigger template, named NonPolyPr (GenBank accession number GU583840). **NonPolyPr** was designed not to give any significant Blast hits with the nucleotide datasets for microorganisms (*Bacteria* and *Archaea*). The sequence was synthesized by the BlueHeron company, introduced into a pUCminusMCS vector and further transformed into E. coli cells. The NonPolyPr template has a T7 RNA polymerase promoter and multiple restriction sites (see below), so that it can be transcribed *in vitro* into RNA probes of different sizes (Table 13). Alternatively, primers can be designed to

synthesize dsDNA probes of different sizes. **NonPolyPr350** was the probe used in this study. It was a dsDNA, 350 bp in length and PCR synthesized using the following primes: NonPolyPr350-F (5'-ACAGTCGAATGTCTACCTAC-3') and NonPolyPr350-R (5'-AATATTGTGCAGTCGGATC-3').

> NonPolyPr sequence. In bold, lower caps, the T7 RNA polymerase promoter. In colors, different restriction sites.

GTTTAAAC**gccagtgaattgtaatacgactcactataggg**ACAGTCGAATGTCTACCTACACAGTCGATCTGG
TCCACAGTCGAATGTCTACCTCGACAGTTGATCTGACAGACTGTCTGACAGTCGAATGTCTACCC
AGCTGACAGTCGATCTGTGTAACAATCTATCCGACTACATGACTGACTATTTAAATACAATCGATC
CGACAGTCGATCTGGTGAACAATCGATCCGACTTTATAACAGTCGATCTGCATGACAGTCGATC
TAGCTGACAGTCGATCTGCACGTCACAGTCGATCAAACAGTTGATCTGACAGACTGGACAGTCG
ATCTCCGCTCACAGTCGATCTGTGACAGTCGATCTACAATCGATCCGACTGCACAATATTACTGA
CGACTGACTACAGTCGAATGTCTACCAGACTGACTGTACGTTAAC

Table 13: NonPolyPr template - restriction enzymes with the corresponding restriction sites and resulting fragments

Enzyme		Cut position in the insert	Virtual double digestion between PmeI and a second enzyme		Probe length	%GC
Name	Recognition site (^= cut)		Fragment length (bp)	No. of fragments with the size of interest	(bp)	
PmeI	GTTT^AAAC	4				
PvuII	CAG^CTG	140	136	1	100	49
SwaI	ATTT^AAAT	190	186	1	150	46
PsiI	TTA^TAA	240	236	1	200	45
BmgBI	CAC^GTC	290	286	1	250	45.6
BsrBI	CCG^CTC	340	336	1	300	46.3
SspI	AAT^ATT	390	386	2	350	46.6
HincII	GTY^RAC	440	436	1	400	46.2

2.8. Probe synthesis

2.8.1. Single-stranded RNA (ssRNA) probes

PCR amplification

The following primers were used to amplify the template for *in vitro* transcription: i) for *rpo*B amplicon - *rpoB R* and *T7-rpoB F*; ii) for *hmeD* amplicon - *T7 HmeD_384* for and *HmeD_704 rev*; iii) for *dsrA* amplicon - *DsrA_520R* and *T7 MM Dsr1F deg*; iv) for *nif*H antisense amplicon from *Crocosphaera* - *T7-nifHCr-R* and *nifHCr-F*; v) for *nif*H sense amplicon from *Crocosphaera* - *nifHCr-R* and *T7-nifHCr-F*, vi) for *nif*H antisense amplicon from Baltic Sea - *T7_nifH_D-R* and *nifH_B-F* and vii) for *nif*H sense amplicon from Baltic Sea – *nifH_D-R* and *T7-nifH_B-F*. Initially, for each primer pair and its respective template, a gradient PCR was performed, to establish the annealing temperature. For primer sequences and annealing temperatures, see Appendix A: primers. Afterwards, replicate PCR reactions were set up as described in Table 14 and Table 16. The replicates were pooled during the purification step. The programs for thermocycling are described in Table 15 and Table 17. The templates used for PCR were: i) for *rpo*B amplicon – *E. coli* K12 cells, 1% PFA fixed (1:1000 dilution); ii) for *dsrA* and *hme*D amplicons – *E. coli* culture (clone WS39F7); iii) for nif*H* amplicons – *Crocosphaera watsonii* cell suspension or DNA extracted from Baltic Sea samples.

Table 14: PCR reaction for *rpo*B, *dsrA* and *hme*D amplicons (preparing template for *in vitro* transcription)

Reagent	volume (µl)	Final concentration
10x Reaction Buffer	5	1x
25 mM $MgCl_2$	3	1.5 mM
dNTPs (10 mM each)	1	200 µM
F primer (50 µM)	0.5	0.5 µM
R primer (50 µM)	0.5	0.5 µM
Ampli Taq Gold Polymerase (5 U/µl)	0.4	2U/rxn
PCR Water	37.6	
Template	2	
Total volume	**50**	

Table 15: Thermocycling program for *rpo*B, *dsrA* and *hme*D amplicons (preparing template for *in vitro* transcription)

Step no.	*rpo*B amplicons		*dsrA* and *hme*D amplicons		process
	Temperature (°C)	time	Temperature (°C)	time	
1	96	10 min	94	4 min	Initial denaturation
2	96	1 min	94	30 sec	Denaturation
3	61	1 min	61	30 sec	Annealing
4	72	30 sec	72	40 sec	Elongation
5	Go to 2 repeat 27x		Go to 2 repeat 31x		
6	72	10 min	72	10 min	Final elongation
7	15	Hold	15	Hold	store

Table 16: PCR reaction for *nif*H amplicons (preparing template for *in vitro* transcription)

Reagent	volume (µl)	Final concentration
10x Reaction Buffer	5	1x
10x BSA	5	1x
dNTPs (2.5 mM each)	4	200 µM
F primer (100 µM)	0.5	1 µM
R primer (100 µM)	0.5	1 µM
Eppendorf Taq Polymerase (5 U/µl)	0.4	2U/rxn
PCR Water	32.1	
Template	2.5	
Total volume	**50**	

Table 17: Thermocycling program for *nif*H amplicons (preparing template for *in vitro* transcription)

Step no.	Temperature (°C)	time
1	94	10 min
2	94	1 min
3	48	1 min
4	72	45 sec
5	Go to 2 repeat 29×	
6	72	10 min
7	4	Hold

Purification of amplicons used for in vitro transcription

The *rpo*B amplicons were purified with the Gene Clean Turbo kit (see Appendix C: Nucleic acid purification protocols). The *dsrA*, *hme*D and *nif*H amplicons were first concentrated using Microcon YM100 columns, than the specific band were excised from a

3% agarose gel and purified with the Pure Link Quick Gel extraction kit (see Appendix C: Nucleic acid purification protocols).

In vitro transcription to produce ssRNA probes and targets

The *rpo*B, *dsrA* and *hme*D probes and the *nif*H targets were prepared using the Ampliscribe T7 High Yield Transcription kit (see Table 18 and Appendix D: Materials). The *nif*H probes were prepared using the DuraScribe kit (see Table 19 and Appendix D: Materials). After the reactions were set up, they were incubated for 4-6 h at 42°C (AmpliScribe) and for 6 h at 37°C (DuraScribe). At the end, 1-2 µl of RNase free DNase were added to the reaction, to digest the template DNA, followed by 30 min incubation at 37°C.

Table 18: AmpliScribe *in vitro* transcription reaction

component	Probes		Targets	
	Volume (µl)	**Final concentration**	**Volume (µl)**	**Final concentration**
Template	3.5		8	
10x Reaction Buffer	2	1x	2	1x
100 mM ATP	1.5	7.5 mM	1.5	7.5 mM
100 mM CTP	1.5	7.5 mM	1.5	7.5 mM
100 mM GTP	1.5	7.5 mM	1.5	7.5 mM
100 mM UTP	1	5 mM	1.5	7.5 mM
10 mM Dig 11-UTP	5	2.5 mM	-	-
100 mM DTT	2	10 mM	2	10 mM
T7 enzyme solution	2		2	
Total volume	**20 µl**		**20 µl**	

Table 19: DuraScribe *in vitro* transcription reaction

component	Probes	
	Volume (µl)	**Final concentration**
Template	6	
10x Reaction Buffer	2	1x
50 mM ATP	2	5 mM
50 mM GTP	2	5 mM
50 mM 2'-F-dCTP	2	5 mM
50 mM 2'-F-dUTP	2	5 mM
100 mM DTT	2	10 mM
DuraScribe T7 enzyme solution	2	
Total volume	**20 µl**	

Purification of ssRNA probes and targets

The probes and targets were purified using the RNeasy Min Elute Cleanup kit (see Appendix C: Nucleic acid purification protocols and Appendix D: Materials). Additionally, after this purification, the *rpo*B, *dsrA* and *hme*D probes were further purified with the NucAwaySpin Column kit (see Appendix C: Nucleic acid purification protocols and Appendix D: Materials).

2.8.2. Double-stranded DNA (dsDNA) probes

Probe synthesis PCR

The *amo*A1-E3, *amo*A-Nam, *FOS-rdsrA*, *hyn*L, *apr*A-Bath and *NonPolyPr350* probes were produced by incorporating Dig-dUTP into dsDNA via PCR, using PCR Dig Probe Synthesis Kit (see Table 20 and Appendix D: Materials). The targets for T_m measurements were produced using the same kit, without Dig incorporation. The primers used for synthesis were: i) for *amo*A probes and target – from 1f to 9f and from 1r to 14r (see Table 21); ii) for *FOS-rdsrA* probe – rsdr1-R and rdsr1-F; iii) for *hyn*L probes and targets - hynl1-F to -4-F and hynl1-R to -4-R and (iv) for *apr*A-bath probe - aprA-Bath F and aprA-Bath R. The templates were either whole *E. coli* cells, plasmid DNA extracted with the Plasmid Maxi Kit or fosmid DNA extracted with the Plasmid MiniPrep Kit. Initially, for each primer pair and its respective template, a gradient PCR was performed, to establish the annealing temperatures. For primer sequences and annealing temperatures, see Appendix A: primers. The components of the PCR reactions are described in Table 20. The programs for thermocycling are described in Table 22.

Table 20: Probe synthesis PCR

Reagent	Probes		Targets	
	Volume (µl)	Final concentration	Volume (µl)	Final concentration
10x Reaction Buffer	10	1x (1.5 mM Mg^{2+})	10	1x (1.5 mM Mg^{2+})
PCR Dig Probe Synthesis Mix	10	200 µM dATP, dCTP, dGTP 130 µM dTTP 70 µM dUTP	-	-
dNTP stock	-	-	10	200 µM each dNTP
F primer (100 µM)	1	1 µM	1	1 µM
R primer (100 µM)	1	1 µM	1	1 µM
Enzyme mix	1.5		1.5	
PCR Water	74		74	
Template	2.5		2.5	
Total volume	**100**		**100**	

Table 21: Annealing temperatures for *amo*A probes and targets

Clone / Probe name	F primer	R primer	Annealing temperature
*amo*A-1E3	5f	7r	48°C
*amo*A clib1-pl1 cl10	1f	9r	57°C
*amo*A clib1-pl1 cl34	2f	8r	57°C
*amo*A clib1-pl1 cl53	3f	3r	57°C
*amo*A clib1-pl1 cl54	3f	10r	48°C
*amo*A clib1-pl1 cl56	4f	3r	57°C
*amo*A clib1-pl1 cl59	5f	7r	48°C
*amo*A clib1-pl1 cl75	7f	5r	48°C
*amo*A clib1-pl2 cl106	3f	8r	57°C
*amo*A clib1-pl2 cl145	6f	6r	57°C
*amo*A clib2-pl2 cl66	5f	7r	57°C
*amo*A clib2-pl3 cl105	8f	4r	57°C
*amo*A clib3-pl1 cl74	9f	8r	48°C

Table 22: Thermocycling program for dsDNA probes and targets

Step no.	*amo*A *and* NonPolyPr350		*FOS-rdsrA*		*hyn*L and *apr*A	
	Temp (°C)	time	Temp (°C)	time	Temp (°C)	time
1	94	10 min	96	5 min	96	5 min
2	94	1 min	96	1 min	94	1 min
3	*	1 min	56	1 min	**	1 min
4	72	45 sec	72	45 sec	72	45 sec
5			Go to 2 repeat 29x			
6	72	10 min	72	10 min	72	10 min
7	4	Hold	15	Hold	4	Hold

* annealing temperature varied: see Table 21 *amo*A probes and targets. For NonPolyPr350, the annealing temperature is 58°C.
** 57°C for *hyn*L and 48°C for *apr*A

Purification of dsDNA probes and targets

The *dsDNA* probes and targets were purified with the Gene Clean Turbo kit (see Appendix C: Nucleic acid purification protocols).

2.8.3. Checking probe quality

The probes and targets were checked electrophoretically in 2.5-3% agarose gels, for the right size and for Dig incorporation. The concentration was determined spectrophotometrically, using a NanoDrop instrument. The probes were stored at -20°C.

2.9. Determination of T_m and of the stringency conditions

First, the T_m of the probe-target hybrids was calculated using the formulae from Wetmur, (1991), which were integrated in the PolyPro software. Second, the T_m of the perfectly matched probe-target hybrids was measured in a hybridization-like buffer and washing-like buffer II, *in vitro*, using a real time PCR machine (IQ5, Biorad) and Syto9 dye.

2.9.1. T_m measurements in hybridization-like buffer

For measurements in hybridization-like buffers, first the buffers were prepared as described below. Then, 3 µl of 5 mM Syto9 were added to 1.5 ml of hybridization-like buffer, to get a final concentration of 10 µM Syto9. To 100 µl of the latter mixture 6 µl of dsDNA (230-350 ng) were added, and the resulting solution was aliquoted in 25 µl per well and used for T_m determinations. The T_m was measured for probe dsDNA (with Dig-dUTP), for target dsDNA and for hybrid dsDNA. The thermal protocol used for the T_m determination in

hybridization-like buffer was the following: denaturation at 80°C for 5 min, hybridization at 42°C for 25 min and melting from 50°C to 75°C, +0.2°C per 1.5 min, minimum ramp rate.

Preparation of 20 ml of 35% formamide hybridization-like buffer (for geneFISH)

- in a 50 ml Falcon tube, add the following:

Component	**Volume**	**Final concentration**	**Na^+ contribution (mM)**
20x SSC	5 ml	5x	975
Dextran sulfate	2 g	10%	660
0.5 M EDTA	0.8 ml	20 mM	80
water	5.2 ml		

- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add the following components:

Component	**Volume**	**Final concentration**	**Na^+ contribution (mM)**
100% formamide	7 ml	35%	
20% SDS	100 µl	0.1%	3.45
			Total Na^+ = 1718 mM

- Vortex, spin down
- Aliquot and store at -20°C

Preparation of 20 ml of 45% formamide hybridization-like buffer (for geneFISH)

- in a 50 ml Falcon tube, add the following:

Component	**Volume**	**Final concentration**	**Na^+ contribution (mM)**
20x SSC	5 ml	5x	975
Dextran sulfate	2 g	10%	660*
0.5 M EDTA	0.8 ml	20 mM	80
water	3.2 ml		

- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add the following components:

Component	Volume	Final concentration	Na^+ contribution (mM)
100% formamide	9 ml	45%	
20% SDS	100 µl	0.1%	3.45
			Total Na^+ = 1718 mM

- Vortex, spin down
- Aliquot and store at -20°C

* the contribution of dextran sulfate to the Na^+ concentration was calculated from the molecular formula of dextran sulfate 500 000, assuming 3 Na^+ per glucosyl residue.

2.9.2. T_m measurements in washing-like buffer

For geneFISH experiments, the composition of the washing-like buffer II was 0.1x SSC, 0.1% SDS, 10 µM Syto9 dye and dsDNA (~240 ng per 25 µl reaction). The T_m was measured for the probe dsDNA (with Dig-dUTP) and for target dsDNA. The thermal protocol used for T_m in washing-like buffer was: from 50°C to 75°C, +0.2°C per 1.5 min, minimum ramp rate.

For mRNA HISH-SIMS experiments, the composition of the washing-like buffer II was 0.1x SSC, 0.1% SDS, 10 µM Syto9 dye, 0 to 60% formamide and Fluorinated (F)-RNA:RNA hybrids. The thermal protocol used was: from 40°C to 95°C, +0.2°C per 0.5 min, minimum ramp rate. The F-RNA:RNA hybrids were obtained by hybridizing the fluorine labeled antisense strand with the unlabeled sense strand in a solution containing 5x SSC and 50% formamide. The thermal protocol for hybridization consisted of 3 min denaturation at 95°C, followed by 2 h hybridization at 68°C. The whole hybridization reaction was loaded on 4% agarose gel and the hybrids were separated electrophoretically from the single-stranded molecules. The gel loading buffer was 66% formamide in glycerol. The electrophoresis buffer was 1x MOPS. After migration, the gel was stained with 1x SybrGreen II and observed under a Dark Reader Transilluminator (Clare Chemical Research). Since they are double-stranded, the hybrid molecules have a slower electrophoretical mobility than the single-stranded molecules and their bands can be clearly distinguished. The hybrid bands were cut out from the gels and purified using the Pure Link Quick Gel Extraction kit (see Appendix C: Nucleic acid purification protocols).

2.10. rRNA CARD-FISH protocol

All the water used during the protocol was autoclaved and 0.22 µm filtered, MilliQ water. Unless stated otherwise, the incubations were performed at room temperature (RT). All washing steps were carried out in 50 ml volume.

2.10.1. Sample immobilization

Different volumes of fixed *E. coli* or AK199 enrichment cells were mixed with 10 ml 1x PBS and filtered on 0.2 µm polycarbonate filters (GTTP, Millipore). The filters were previously coated with a Pd/Au alloy (Musat et al., 2008) to give a black background, although this is not essential for the geneFISH protocol. The filters were then washed with 10 ml 1x PBS and with 15 ml water, air dried and stored at -20°C. The Namibian seawater samples were directly filtered after fixation, washed once with water, air dried and stored at -80°C. After dehydration and Steedman's Wax embedding, the *Bathymodiolus puteoserpentis* samples were sectioned, mounted on glass-slides and stored at -20°C (Pernthaler and Pernthaler, 2005). Immediately before use, the sections where dewaxed in absolute ethanol (3 x 10 min), air dried, encircled with a PapPen and rehydrated (2 min in 80% ethanol, 2 min in 70% ethanol and 2 min in 50% ethanol).

2.10.2. Inactivation of endogenous peroxidases

The inactivation was performed by overlaying the filters with 0.01 M HCl for 10 min, followed by washing with 1x PBS for 5 min and with water for 1 min. For the *Bathymodiolus puteoserpentis* samples the inactivation was done with 0.1 M HCl for 10 min, followed by washing with 0.02 M Tris-HCl pH 8 for 5 min and with water for 2 x 5 min. From here on, all washing steps were done in 50 ml.

2.10.3. Permeabilization

For the *E coli* samples, the permeabilization was done in 0.5 mg ml^{-1} lysozyme (AppliChem,), 1x PBS pH 7.4, 0.1 M Tris-HCl pH 8.0 and 0.05 M EDTA pH 8.0, for one h, on ice. The wash consisted of 5 min with 1x PBS, 1 min with water, and 1 min with 96% ethanol, followed by air-drying. For the Namibian seawater samples the permeabilization consisted in 1 min with 0.1 M HCl, followed by 1x PBS (1 and 5 min), 1 min with water, 1 min with 96% ethanol and air-drying. The Ak199 enrichment samples were permeabilized as described for the *E. coli* samples, except that the lysozyme (Fluka) concentration was 10 mg ml^{-1} and the incubation was 1 h at 37°C.

2.10.4. rRNA hybridization

E. coli samples were hybridized with the EUB338 probe (Amann et al., 1990), while the Namibian seawater samples were hybridized with the Cren554 probe (Massana et al., 1997), which is targeting MGI *Crenarchaeota*. The *Bathymodiolus puteoserpentis* samples where hybridized with BMARt-193 (for the Thiotrophic symbionts - (Duperron et al., 2006)) and BangM-138 (for the Methanotrophic symbionts -(Duperron et al., 2005)) probes. The AK199 samples where hybridized with Ros537 (for marine *Roseobacter* clade - (Eilers et al., 2001)) and GAM42a (for G*ammaproteobacteria* –(Manz et al, 1992)) probes. The hybridization took place for 3 h or overnight at 46°C, followed by 15 min of washing at 48°C. The hybridization buffers and the corresponding washing buffers were prepared as describe previously (Pernthaler et al., 2002a). The formamide concentrations used in the hybridization buffers were 35% for the EUB338, Ros537 and GAM42a probes, 30% for BMARt-193 and BangM-138 probes and 0% for Cren554 probe.

2.10.5. CARD for rRNA detection

All the samples were equilibrated for 20 min in 1x PBS. Next, the *E. coli* samples were incubated for 10 min at 37°C in a solution containing 1x PBS, 0.0015% H_2O_2 and 0.25 µg ml^{-1} $Alexa_{488}$-labeled tyramide (prepared as described by Pernthaler and Pernthaler, (2005)). The wash consisted of 2 and 15 min with 1x PBS, 1 min with water, and 1 min with 96% ethanol, followed by air-drying. The Namibian seawater samples were incubated for 40 min at 46°C and the *Bathymodiolus puteoserpentis* samples were incubated for 20 min at 37°C in a solution containing amplification buffer (see below), 0.0015% H_2O_2 and 1 µg/ml $Alexa_{488}$-labeled tyramide. The wash steps for Namibian samples were 1, 2 and 2x 10 min with 1x PBS at 46°C, 1 min with water, 1 min with 96% ethanol, followed by air-drying. The wash steps for *Bathymodiolus puteoserpentis* samples were 1 and 5 min in 1x PBS, 3x 1 min water, 1 min in 70% ethanol, followed by air drying. The Ak199 enrichment samples were incubated for 20 min at 46°C in a solution containing amplification buffer, 0.0015% H_2O_2 and 2 µg/ml $Alexa_{488}$-labeled tyramide. The wash steps were 15 min with 1x PBS, 1 min with water, 1 min with 96% ethanol, followed by air-drying. Starting with the CARD incubations, all steps were performed in dark. Overnight, until the next step, the filters were stored at -20°C and the thin sections at +4°C.

Preparation of 40 ml amplification buffer for the rRNA CARD step

Component	Volume	Final concentration
10x PBS pH 7.4	4 ml	1x
5 M NaCl	16 ml	2 M
10% Blocking Reagent for nucleic acid hybridizations (BR)	400 µl	0.1%
Dextran Sulfate	4 g	10%
water	Up to 40 ml	

- in a 50 ml Falcon tube add all, except the blocking reagent
- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add 400 µl Blocking reagent (final concentration 0.1%)
- Vortex, spin
- Filter through 0.22 µm filter, Millipore
- Store @ +4°C

2.11. GeneFISH protocol

All the water used during the protocol was autoclaved and 0.22 µm filtered, MilliQ water. Unless stated otherwise, the incubations were performed at room temperature (RT).

2.11.1 rRNA CARD –FISH

The rRNA CARD-FISH step was performed as described at point 2.10..

2.11.2. RNase treatment

Before RNase digestion, some of the Namibian and all *Bathymodiolus puteoserpentis* samples were denatured as follows: 20 min at 60°C and 48°C respectively, in denaturation mix (90% formamide, 0.1x SSC and 0.1% SDS). The wash consisted of 3x 2 min ice-cold water, then 2x 2 min ice cold 96% ethanol, followed by air-drying. For RNase digestion, the samples were incubated at 37°C, 4 - 5 h, in RNase solution - 0.5 U μl^{-1} RNase I, 30 µg ml^{-1} RNase A, 0.1 M Tris-HCl pH 8. The wash consisted of 3x 7 min with 1x PBS, 1 min with water, and 1 min with 96% ethanol, followed by air-drying. For geneFISH with ssRNA probes only RNaseI was used, in a 10 mM Tris-HCl, 20 mM $MgCl_2$ and 4 mM $CaCl_2$ buffer. The RNaseI was inactivated by incubating for 20 min in 1x PBS at 70°C, 3x 5 min washing in 1x PBS, 0.1 M EDTA, 0.1% SDS, then rinsing filters in funnel with water.

2.11.3. DNase treatment

One of the controls for the Namibian seawater samples consisted of DNase digestion. In the previous step, these samples were treated only with RNaseI and without RNaseA. The RNase I was inactivated by incubating for 20 min in 1x PBS at 70°C, then washing for 1 min in water. The DNase treatment was performed at 37°C overnight, in a solution containing 0.2 U μl^{-1} DNaseI, 20 mM $MgCl_2$, 4 mM $CaCl_2$ and 10 mM Tris-HCl pH 8.0. The washing and inactivation of DNaseI consisted of 25 min at 75°C, 2x 20 min with 1x PBS, 0.1 M EDTA and 1 min with water.

2.11.4. Inactivation of HRP introduced with the rRNA probe

The inactivation of HRP consisted of 10 min with 0.2 M HCl (for *E. coli* samples*)* or 1 min with 0.1 M HCl (for Namibian water samples), followed by washing with 1x PBS for 1 and 5 min, then 1 min with water, 1 min with 96% ethanol and air-drying. The AK199 enrichment samples were inactivated with 3% H_2O_2 in 1x PBS for 30 min, 0.1M HCl for 10 min, washed for 1 min and 2x 5 min in 1x PBS, 1 min in water and 1 min in 96% ethanol and air dried. The *Bathymodiolus puteoserpentis* samples were inactivated with 3% H_2O_2 in methanol for 30 min, 1 min in water, 0.5 M HCl for 10 min, washed for 2x 10 min in 1x PBS, 1 min in water and 1 min in 96% ethanol and air dried.

2.11.5. Gene hybridization

For prehybridization and hybridization, the samples were covered in hybridization buffer. Prehybridization was done in the same buffer as hybridization, but without the probe, for 5 h at the hybridization temperature (shaking water bath for filters, oven for sections). The filters were placed in 2 ml eppendorf tubes, while the sections were placed in humid, isotonic chambers. The *rpo*B probe was used for *E. coli* K12, the *dsr*A and *hme*D probes were used for fosmid containing *E. coli* clones, *amo*A-1E3 was used for the *E. coli* clones with *amo*A inserts, the *amo*A-Nam probe for the Namibian seawater samples, the *hyn*L probe for the *Bathymodiolus puteoserpentis* samples and the *FOS-rdsrA* probe for the AK199 enrichment. The hybridization buffer for the *rpo*B, *dsr*A and *hme*D probes had 50% formamide, for the *amo*A and *hyn*L probes had 35% formamide, for *FOS-rdsrA* probe had 45% formamide and for NonPolyPr350 it had either 35% (when used as negative control for *amo*A and *hyn*L probes) or 45% (when used as negative control for the *rdsrA* probe). The preparation of hybridization buffers is described below. Additionally, two more hybridization buffers were tested with the *rpo*B probe: UltraHyb Ultrasensitive Hybridization Buffer and Northern Max Hybridization Buffer (see Appendix D). The probe concentrations tested were 0.25 pg μl^{-1}, 2.5 pg μl^{-1}, 25 pg μl^{-1} and 250 pg μl^{-1}. When multiple polynucleotides were used in the same mix, each of them had a concentration of 2.5 pg $\mu l^{-}1$. After adding the probes to the hybridization buffer, the samples were denatured for 25 min at 85°C for the AK199

enrichment samples and at 75°C for all other samples. The samples were transferred immediately to the hybridization temperature and hybridization took place for 18-22 h or 41 h. The hybridization temperature was 42°C for all probes and 50°C for part of the *FOS-rdsrA* samples). The washes were first performed with washing buffer I (WBI) (2x SSC, 0.1% SDS) for 5 min and 30 min at 42°C, followed by washing buffer II (WBII) (0.1X SSC, 0.1% SDS) for 3x 1 min at RT and 1.5 h at hybridization temperature, in a slow shaking water bath, and finally 2x 1 min with 1x PBS.

Preparation of 20 ml of 35% formamide gene hybridization buffer

- in a 50 ml Falcon tube, add the following:

Component	**Volume**	**Final concentration**
20x SSC	5 ml	5x
Dextran sulfate	2 g	10%
0.5 M EDTA	0.8 ml	20 mM
water	2.2 ml	

- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add the following components:

Component	**Volume**	**Final concentration**
100% formamide	7 ml	35%
20% SDS	100 µl	0.1%

- Vortex, spin down
- Add the following components:

Component	**Volume**	**Final concentration**
sssDNA (10 mg/ml)	500 µl	0.25 mg/ml
Yeast RNA (10 mg/ml)	500 µl	0.25 mg/ml
10% Blocking Reagent	2 ml	1%

- Vortex, spin down
- Aliquot and store at -20°C

Preparation of 20 ml of 45% formamide gene hybridization buffer

- in a 50 ml Falcon tube, add the following:

Component	Volume	Final concentration
20x SSC	5 ml	5x
Dextran sulfate	2 g	10%
0.5 M EDTA	0.8 ml	20 mM
water	0.2 ml	

- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add the following components:

Component	Volume	Final concentration
100% formamide	9 ml	45%
20% SDS	100 µl	0.1%

- Vortex, spin down
- Add the following components:

Component	Volume	Final concentration
sssDNA (10 mg/ml)	500 µl	0.25 mg/ml
Yeast RNA (10 mg/ml)	500 µl	0.25 mg/ml
10% Blocking Reagent	2 ml	1%

- Vortex, spin down
- Aliquot and store at -20°C

2.11.6. Antibody binding

The samples were blocked for 1 h in 1x PBS and 0.5% Western Blocking Reagent (WBR). The antibody binding took place for 1.5 h in a solution containing 1x PBS, 1% WBR and 0.3 U ml^{-1} anti-Dig HRP-conjugated antibody (Fab fragments). The wash was done in a 1x PBS, 0.5% WBR solution for 1, 5 and 2x 10 min. Before each use, the antibody stock solution was centrifuged 10 min @ 10000xg @ 4°C, to pellet eventual precipitates. The solutions containing antibodies were not vortexed. All steps were carried on a shaker at 50 rpm.

2.11.7. CARD for gene detection

The samples were equilibrated for 20 min in 1x PBS. Then, they were incubated for 40 min at 37°C in amplification buffer containing (1x PBS, 20% dextran sulfate, 0.1% blocking reagent, 2 M NaCl) with 0.0015% H_2O_2 and 2 µg ml^{-1} $Alexa_{594}$-labeled tyramide. For

preparation of the amplification buffer, see below, and of the labeled tyramides see Pernthaler and Pernthaler, (2005). They were then washed for 1, 5 and 2x 10 min with 1x PBS, in a 46°C oven, slow shaking, 1 min with water, 1 min with 96% ethanol, followed by air-drying.

Preparation of 40 ml amplification buffer for the gene CARD step

Component	Volume	Final concentration
10x PBS pH 7.4	4 ml	1x
5 M NaCl	16 ml	2 M
10% Blocking Reagent for nucleic acid hybridizations (BR)	400 µl	0.1%
Dextran Sulfate	8 g	20%
water	15.6 ml	

- Add all in a 50 ml tube, except the blocking reagent
- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add 400 µl Blocking reagent (final concentration 0.1%)
- Vortex, spin
- Filter through 0.22 µm filter, Millipore
- Store @ +4°C

2.11.8. Embedding and counterstaining

The samples were embedded either in ProLongGold antifade reagent or SlowFadeGold antifade reagent, containing 1 µg/ml 4´,6-diamidino-2-phenylindole (DAPI).

2.11.9. Microscopy

Microscopy was performed on a Axio epifluorescence Microscope (Carl Zeiss), equipped with the following fluorescence filters: DAPI (365/10 nm excitation, 420 LP emission, FT 395 Beam Splitter), $Alexa_{488}$ (472/30 excitation, 520/35 emission, 495 Beam Splitter) and $Alexa_{594}$ (562/40 excitation, 624/40 emission, 593 Beam Splitter). The pictures were taken with a black and white camera and the exposure times for Alexa 594 were 200 ms (*E. coli* samples) and 250 ms (Namibian seawater samples). The counts for the gene FISH hybridization efficiency were done from the pictures in the Axio Visio software.

2.12. mRNA HISH-SIMS

For all the steps performed during this protocol special precautions to eliminate RNase contamination were taken. All the water, solutions and plasticware were RNase free. All the glassware and metallic instruments were baked to remove eventual RNase contamination.

2.12.1. Sample immobilization

Different volumes of fixed *Crocosphaera watsonii* cells were mixed with 10 ml 1x PBS and filtered on 0.2 µm polycarbonate filters. The filters were previously coated with a Pd/Au alloy (Musat et al., 2008), as required for NanoSIMS analysis. The filters were then washed with 10 ml 1x PBS, 15 ml water, air dried and stored at -20°C.

2.12.3. Permeabilization

The permeabilization was done in 25 mg ml^{-1} lysozyme (AppliChem), 1x PBS pH 7.4, 0.1 M Tris-HCl pH 8.0 and 0.05 M EDTA pH 8.0, for one hour, on ice. The wash consisted of 5 min with 1x PBS, 1 min with water, and 1 min with 96% ethanol, followed by air-drying.

2.12.3. mRNA hybridization

For prehybridization, the samples were incubated in hybridization buffer (for preparation, see below) without probe, for 30 min at 55°C. After that, the samples were transferred in the corresponding hybridization buffer – probe mix and incubated for 5 h at 55°C. The probes were ssRNA labeled with fluorine (F-RNA) and their concentration was 5 ng $µl^{-1}$. The washes were first performed with washing buffer I (WBI) (2x SSC, 0.1% SDS) for 2x 1 min and 10 min at 42°C, followed by washing buffer II (WBII) (0.1X SSC, 0.1% SDS, 60% formamide) for 2x 1 min at RT and 1 h at 57°C, in a shaking water bath. At the end, the filters were treated for 2x 1 min with water, 1 min with 96% ethanol, air dried and stored at -20°C until analysis.

Preparation of 10 ml of 50% formamide mRNA hybridization buffer

- in a 50 ml Falcon tube, add the following:

Component	**Volume**	**Final concentration**
5 M NaCl	1.5 ml	750 mM
Dextran sulfate	1 g	10%
0.5 M EDTA	100 µl	5 mM
1 M Tris-HCl pH 8.0	1 ml	100 mM
water	2 ml	

- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add the following components:

Component	Volume	Final concentration
100% formamide	5 ml	50%
20% SDS	5 µl	0.01%

- Vortex, spin down
- Add the following components:

Component	Volume	Final concentration
sssDNA (10 mg/ml)	200 µl	0.2 mg/ml
Yeast RNA (10 mg/ml)	200 µl	0.2 mg/ml

- Vortex, spin down
- Aliquot and store at -20°C

Preparation of 10 ml of 70% formamide mRNA hybridization buffer

- in a 50 ml Falcon tube, add the following:

Component	Volume	Final concentration
5 M NaCl	1.5 ml	750 mM
Dextran sulfate	1 g	10%
0.5 M EDTA	100 µl	5 mM
1 M Tris-HCl pH 8.0	1 ml	100 mM

- Shake to dissolve dextran sulfate and incubate at 48°C (water bath) until dextran sulfate is dissolved
- Cool to room temperature
- Add the following components:

Component	Volume	Final concentration
100% formamide	7 ml	70%
20% SDS	5 µl	0.01%

- Vortex, spin down
- Add the following components:

Component	Volume	Final concentration
sssDNA (10 mg/ml)	200 µl	0.2 mg/ml
Yeast RNA (10 mg/ml)	200 µl	0.2 mg/ml

- Vortex, spin down

- Aliquot and store at -20°C

2.12.4. NanoSIMS analysis

The hybridized filters were analyzed using a NanoSIMS 50L manufactured by Cameca. For each individual cell, simultaneously secondary ion images of $^{12}C^-$, $^{12}C^{14}N^-$ and $^{19}F^-$ were recorded in parallel. Images and data were processed using the proprietary CAMECA Win-Image processing software working under PC Windows XP environment and new software, developed by Lubos Polerecky.

3. Results and Discussions

3.1. Development of concepts and software for a rational design of polynucleotide probes

3.1.1. PolyPro - Software for Polynucleotide Probe Design

The PolyPro software for polynucleotide probe design was built of three modules, a GenBank Taxonomy Extractor (GTE), a Polynucleotide Probe Designer (PPD) and a Hybridization Parameters Calculator (HPC). The work flow for probe design is outlined in Figure 9, and the individual steps are described in detail in the next paragraphs. The output is represented by polynucleotide probe mixes which cover all selected targets.

Defining the probe-target region

The first step is to select the gene region to be targeted by the probe, based on a multiple sequence alignment (Figure 10) done in an external program, e.g. ARB, Bioedit, ClustalW (Thompson et al., 1994; Hall, 1999; Ludwig et al., 2004). Trimming of this region will result in one polynucleotide for each allele. The file format in which the polynucleotides should be exported from such software is GenBank or FASTA. The GenBank format is necessary if the probe design is targeting certain phylogenetic clades, in which case the GenBank files must contain taxonomic information.

The size of the probe-target region should be between 100 – 450 nt (DeLong et al., 1999; Pernthaler et al., 2002b; Pernthaler and Amann, 2004). The optimum size depends on various issues, like sensitivity of the signal detection method and probe penetration, the last being influenced by cell wall composition of targeted microorganisms, fixation and cell wall permeabilization protocol. The larger a probe, the more label can be attached, but the less easy it will enter the cells. It is therefore a common procedure to include a shearing step, which breaks long polynucleotide probes in a random mixture of shorter fragments of different lengths and sequences (Niki and Hiraga, 1997; DeLong et al., 1999). In this study we refrained from shearing the polynucleotides, as a defined size and sequence are prerequisites for rational design and perfect reproducibility. If, for sensitivity reasons, a longer probe-target region is required, a second or even a third un-sheared polynucleotide probe can be applied.

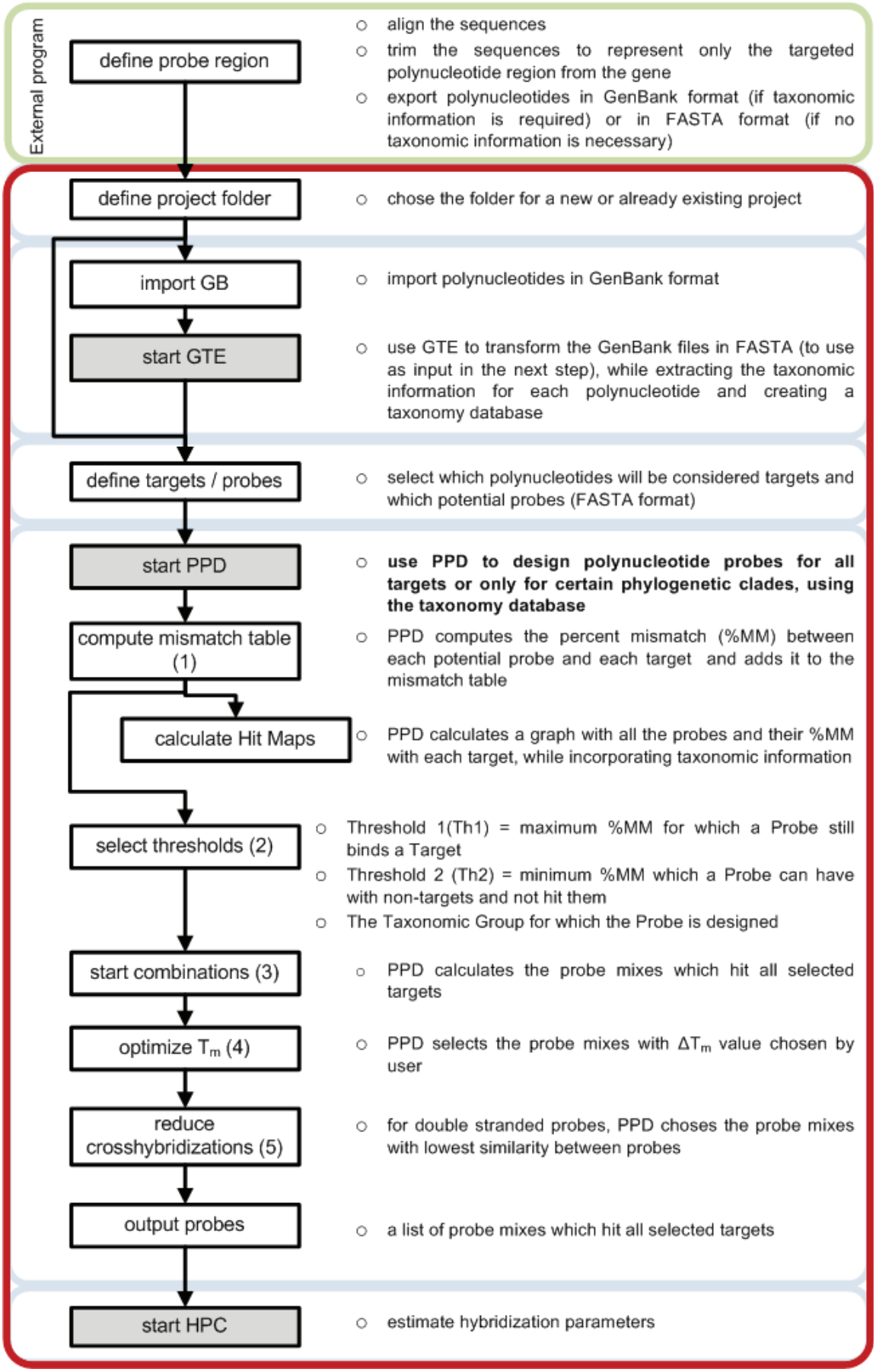


Figure 9: Work flow for polynucleotide probe design.

The probe-target region should be chosen so that the multiple alignment is free of long gaps and there is a homogenous distribution of mismatches. Both long gaps and non-homogenous mismatch distribution will decrease the predictive power of the PolyPro, which uses the %MM as an approximation of the T_m differences between the hybrids formed by a probe with its targets. Accumulation of mismatches or gaps in certain parts of the probe-target hybrid promotes the formation of multiple melting domains (Wartell and Benight, 1985) These domains will have different T_m, which cannot be estimated by a %MM for the whole hybrid. When the suitable alignment region is long enough, various probe-target regions should be evaluated with PolyPro before the probe mix best suited for the purpose of the study is chosen. In many cases, though, there is only one probe-target region available, because the PCR primers commonly used retrieve gene fragments of a length of several hundred nucleotides. The primer regions should not be included in the target region, as they do not always reflect the true base sequence of the respective allele.

Defining the project

A project is a folder selected by the user, which contains all data for the project (Figure 11). In this folder, PolyPro will create the following subfolders: “GBK samples”, “Converted fasta”, “Taxonomy”, “Probes”, “Targets” and “HPC Reports”. The “GBK samples” subfolder will contain the input polynucleotides for GTE, while the output will be saved in “Taxonomy” (the taxonomy database - Taxonomy.TAX) and “Converted fasta” (sequences converted to FASTA format). The subfolders “Targets” and “Probes” will contain the targets and potential probes, respectively. All the files generated by the multiple probes submodule of PPD will be saved in the main folder. The files generated by the hit maps submodule will be saved in the “Taxonomy” subfolder and the ones generated by HPC in “HPC Reports”.

The GenBank Taxonomy Extractor (GTE) module

GTE module must be used when designing probes for certain taxonomic clades. This module is extracting taxonomic information from GenBank files (NCBI format and ARB format), and is converting the polynucleotides to FASTA format. The taxonomic information (Figure 12) is stored in a database (Taxonomy.TAX). The correlation between the polynucleotide and the taxonomic information is maintained through the name of the corresponding FASTA file. The taxonomic information consists of: Domain, Phylum, Class, Order, Family, Genus, Species, and Strain. In the case of environmental sequences, where the microorganisms are not identified, it can include only partial taxonomic information, as revealed by phylogenetic trees. Both the database and the FASTA files generated by GTE are used as input in the next steps.

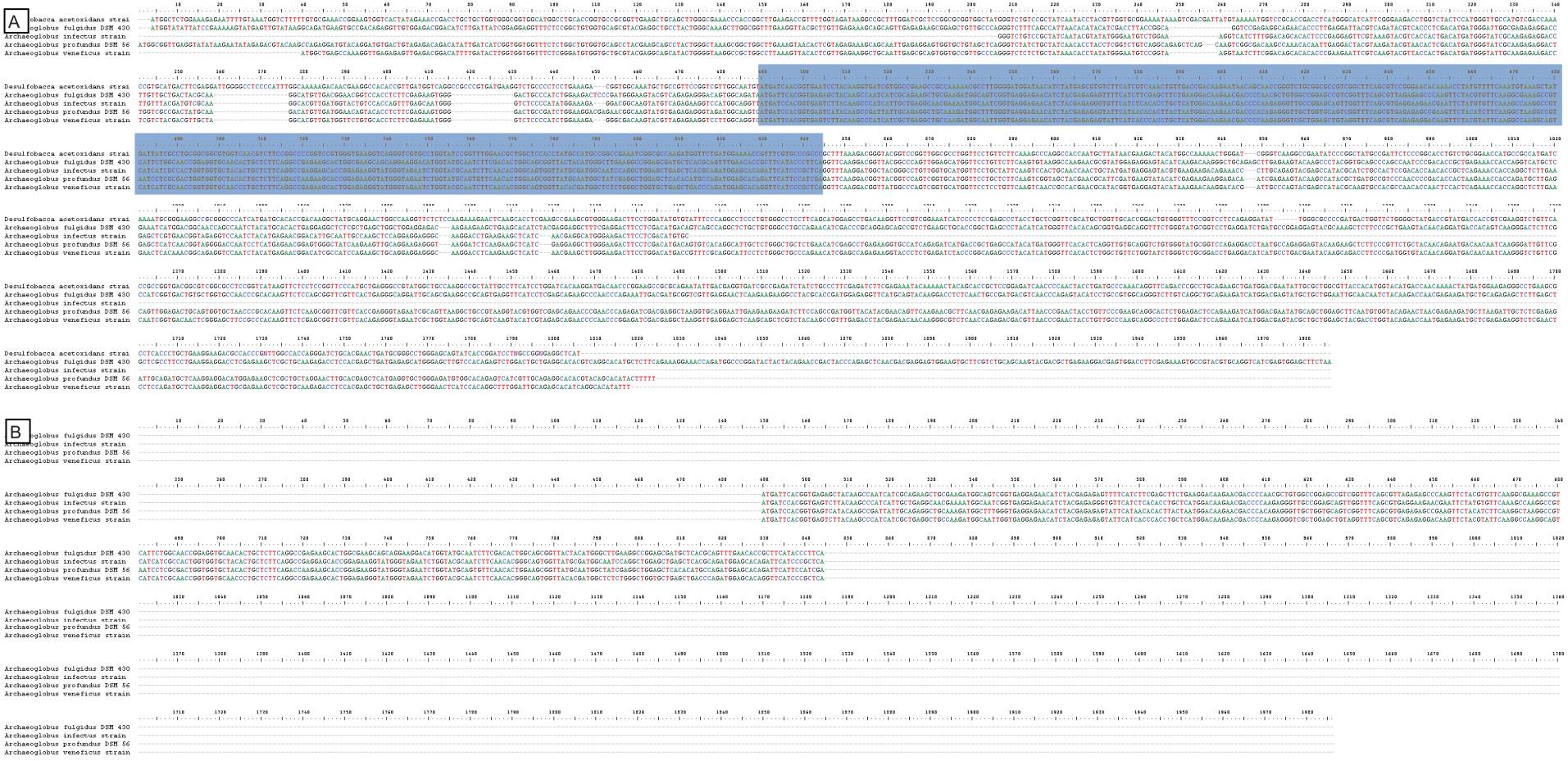


Figure 10: An example of defining the target region for polynucleotide probe design. A. First, all the alleles of the gene are aligned. Second, a region of a few hundred bases is selected (in the blue rectangle). Long gaps in the alignment should be avoided. **B.** The selected region is trimmed and the sequences are exported as GenBank or FASTA format. These sequences will serve as probes / targets in the probe design process.

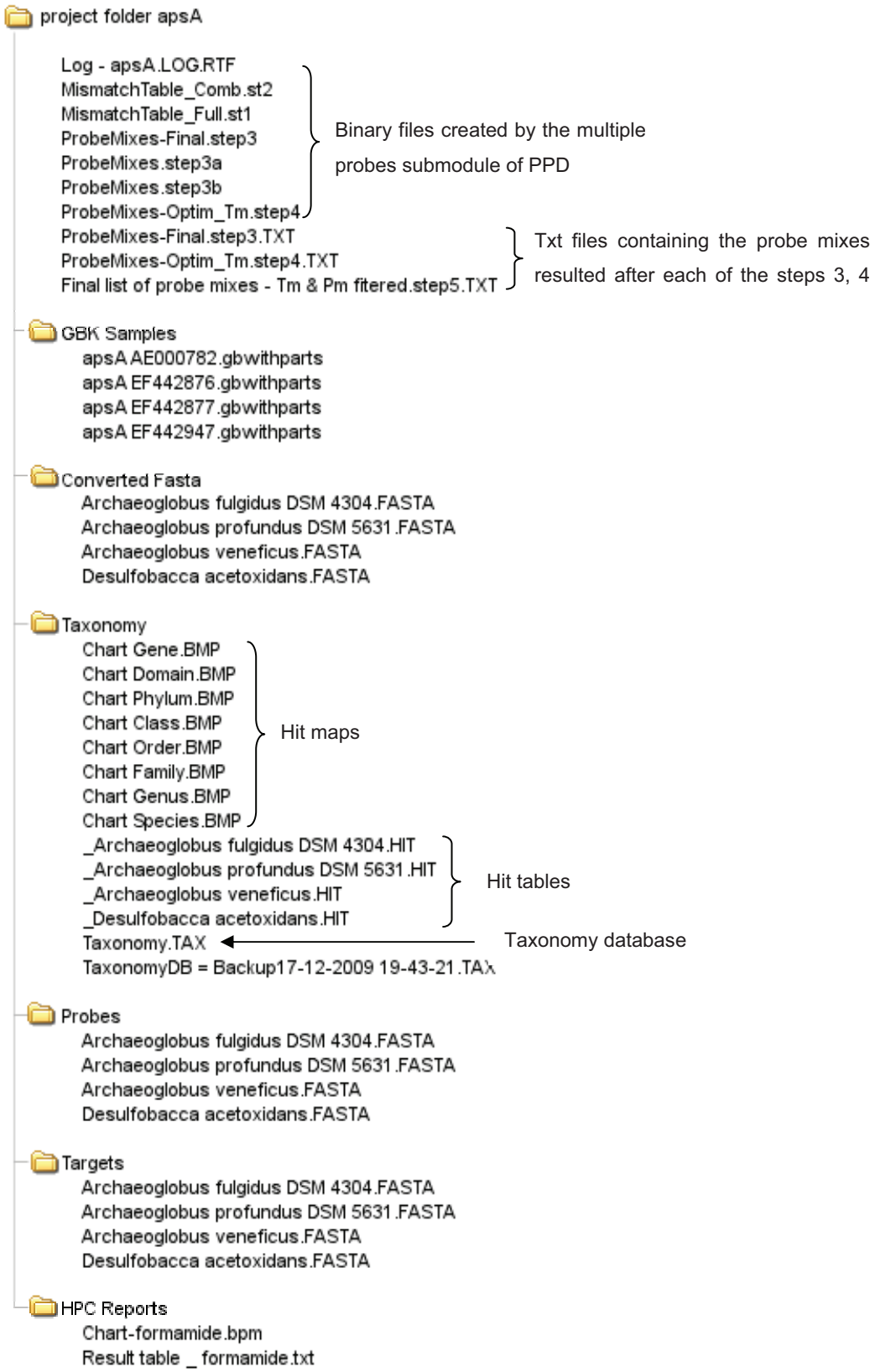


Figure 11: Example of a project folder and its components. The "GBK samples" subfolder contains the GenBank files, and the taxonomic information is extracted from these files. Then, they are converted into FASTA files, which are saved in the "Converted FASTA" subfolder. The "Taxonomy" subfolder contains the "Taxonomy.TAX' file (which stores the taxonomic information extracted by GTE), plus a series of hit maps (saved as BMP) and hit tables (saved as HIT) generated

by the Hit Map submodule. A hit map is created for each taxonomic level, plus one for the gene (when multiple genes are analyzed). A hit table is created for each probe and it is basically a taxonomic database that contains the percentage mismatches of the probe with each target. The “Targets” and “Probes” subfolders contain the sequences of the targets and potential probes, respectively. In the main folder (project folder) there are several files generated by the Multiple Probes sub-module. Important for the user are: “ProbeMixes-Final.step3.TXT”, “ProbeMixes-Optim_Tm.step4.TXT” and “Final list of probe mixes-Tm & Pm fitered.step5.TXT” files, which contain the probe mixes generated at steps 3, 4 and 5, respectively.

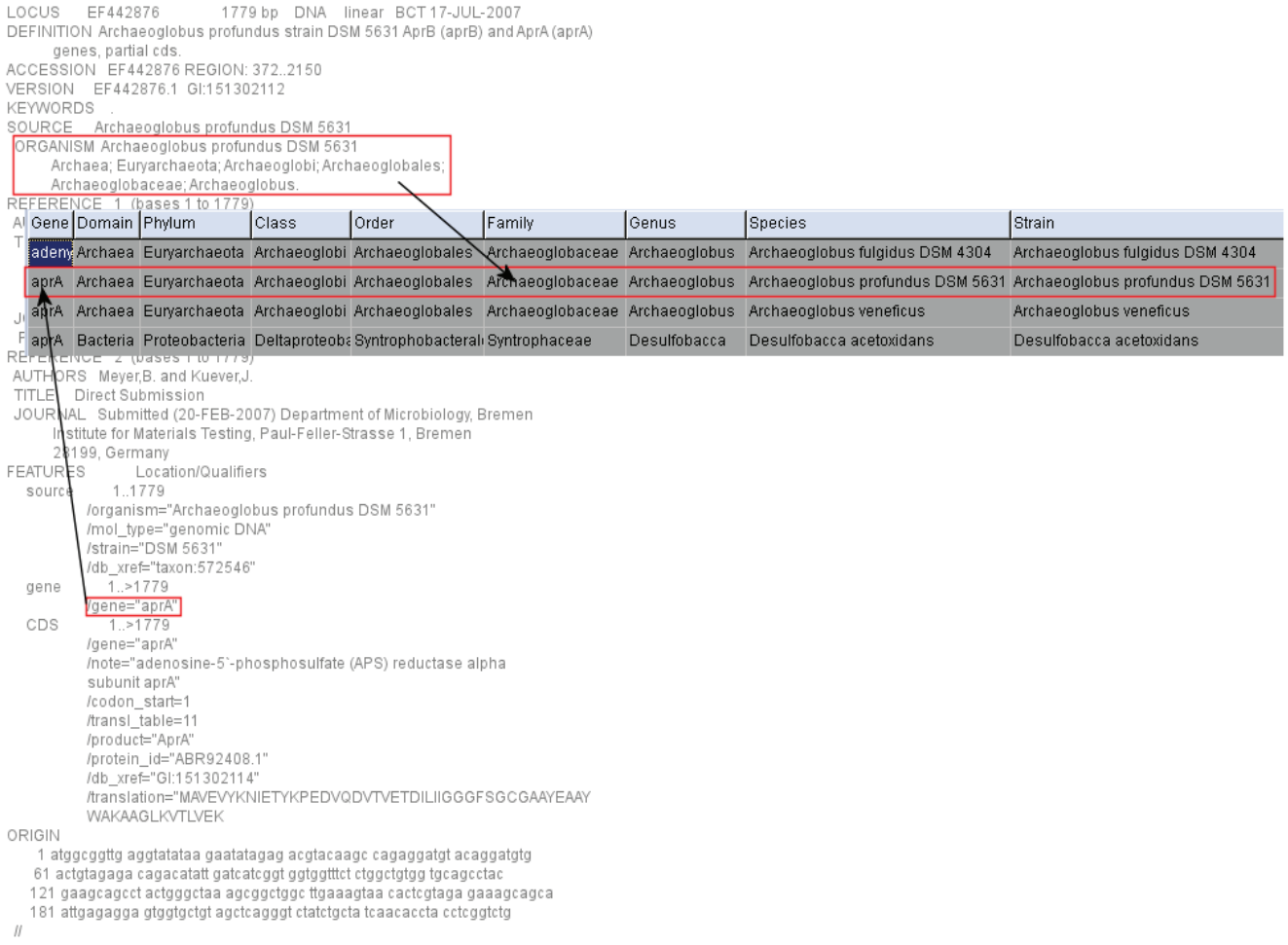


Figure 12: Example of a taxonomy database with the information extracted from a GenBank file. The first column in the database is the gene name. The subsequent columns contain the taxonomic information as given in the GenBank source/organism field. The information in the strain field is a composite of several GenBank fields, as selected by the user. In this example, it consists of the information in the organism field.

Defining probes and targets for PPD and HPC modules

In this step the user defines which of the polynucleotides (corresponding to one probe-target region) will be potential probes and which targets in PPD and HPC modules. The accepted file format is FASTA. Ideally, all alleles should be given both as probes and

targets. However, because polynucleotide probes have to be synthesized from a template DNA molecule, template availability has to be considered. Polynucleotides for which a template is missing have to be excluded from the potential probes.

The Polynucleotide Probe Designer (PPD) module

The PPD module consists of two submodules: the multiple probes submodule and the hit maps submodule.

Multiple probes submodule

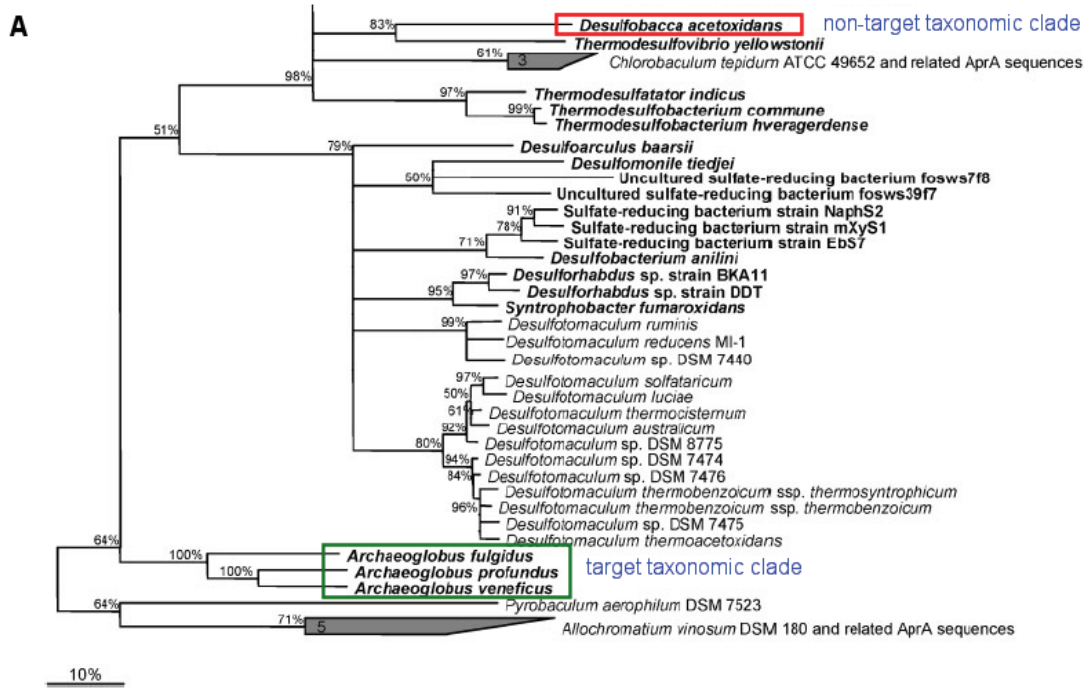
The multiple probes submodule yields as an output probes or mixes of probes which bind to the selected targets. The workflow has five steps (Figure 9).

In step 1 the percent mismatch (%MM) value between each potential probe and each target is calculated and added to a mismatch table. In step 2 the user defines the conditions for probe design, by choosing the values of the following parameters: (i) mismatch threshold 1 (Th1), (ii) T_m difference, (iii) target taxon, i.e. clade of interest, and (iv) mismatch threshold 2 (Th2). Using the given parameters, the software does the following (Figure 13), panels B-D): (i) it marks all probe-target pairs with %MM smaller or equal with Th1; (ii) it removes all probes which belong to other taxonomic clades than the selected one (non-target taxonomic clades); (iii) it removes all probes that are having hits in non-target taxonomic clades, as defined by Th2; (iv) it removes all non-targets (targets belonging to non-target taxonomic clades) v) it removes targets which are not hit by any of the remaining probes, (vi) it removes “replicate probes” of identical target range from the mismatch table. Points (ii),(iii) and (iv) and (v) are performed only when probe design is targeting a certain taxonomic clade. The result of step 2 will be a hit matrix, where the ‘#’ sign means that the probe hits the respective target (Figure 13D). This hit matrix represents the input for step 3.

The T_m difference parameter is used to find replicate probes. Probes are replicates when they have identical target range and a T_m difference lower or equal with the one set by user (default is 0.05°C). The T_m difference is calculated based on the %GC of each probe, as described in materials and methods.

Th1 is a threshold of the %MM up to which targets are detected by a probe. Polynucleotide probes can bind to mismatched sequences up to a certain %MM, depending on the hybridization stringency. Th1 is the maximum %MM a probe is allowed to have with a target and still hybridize. So far, there is no systematic study regarding Th1 value for different FISH methods. Different studies have quantified the decrease in hybridization rate with the increase in percent mismatches (Bonner et al., 1973; Hutton and Wetmur, 1973b). Although the numbers obtained differed between studies, it is clear that above 25%

mismatches the kinetics is no longer favorable for hybrid formation. At about 10% mismatches, one of the studies reported ~50% reduction in the hybridization rate (Bonner et al., 1973), while the other reported only ~20% reduction (Hutton and Wetmur, 1973b).



Set **Th1** to 10%

B

targets \ probes	Archaeoglobus fulgidus	Archaeoglobus profundus	Archaeoglobus veneficus	Desulfobacca acetoxidans
Archaeoglobus fulgidus	0	23.9	20.5	38.1
Archaeoglobus profundus	23.9	0	17.7	37.5
Archaeoglobus veneficus	20.5	17.7	0	37.8
Desulfobacca acetoxidans	38.1	37.5	37.8	0

Set the **target taxonomic clade** to Genus *Archaeoglobus* and **Th2** to 25%

C

	Archaeoglobus fulgidus	Archaeoglobus profundus	Archaeoglobus veneficus	Desulfobacca acetoxidans
Archaeoglobus fulgidus	0	23.9	20.5	38.1
Archaeoglobus profundus	23.9	0	17.7	37.5
Archaeoglobus veneficus	20.5	17.7	0	37.8
Desulfobacca acetoxidans	38.1	37.5	37.8	0

Apply thresholds

D

	Archaeoglobus fulgidus	Archaeoglobus profundus	Archaeoglobus veneficus
Archaeoglobus fulgidus	#		
Archaeoglobus profundus		#	
Archaeoglobus veneficus			#

Figure 13: An example of the operations carried out during step 2 from the multiple probes submodule - designing polynucleotide probes to discriminate the *Archaeoglobus* genus from *Desulfobacca acetoxidans*. Panel A: an *apr*A gene phylogenetic tree adapted from (Meyer and Kuever, 2007b), showing the 3 *Archaeoglobus* alleles grouping in a taxonomic clade close to each other and distant from the *Desulfobacca acetoxidans* allele. Panels B, C and D: the mismatch table during step 2. The probes are in the first row and the targets in the first column. The numerical values in the table represent %MM between probes and targets. **Panel B:** Th1 is set to 10%. Based on this value, the hits for each probe are marked in green. **Panel C:** The target taxonomic clade is set to genus *Archaeoglobus*. All probes not belonging to genus *Archaeoglobus* are marked (in dark grey) for deletion (in this example – *Desulfobacca acetoxidans*). Th2 is set to 25%. All %MM values lower than Th2 are marked in red (the overlay with Th1 is orange). **Panel D:** Three probes remained after the thresholds were applied. The non-target *Desulfobacca acetoxidans* has been removed. The targets hit by each of the remaining probes are marked with a # sign in the Hit Matrix.

For *in situ* hybridizations, the influence of hybridization rate on the detection efficiency is more pronounced for methods targeting low number of molecules per cells (e.g. low copy mRNA or genes) than for those targeting high numbers of molecules per cell (e.g. rRNA), due to the signal detection sensitivity issues. Therefore, most probably, Th1 value will vary with the FISH method, increasing with the increase in the number of target molecules per cell, from genes to mRNA and to rRNA. Ludwig et al. (1994), doing rRNA FISH, placed the %MM at which no hybridization signal was visible between 15% and 21.3% mismatches. Trebesius et al, 1994, showed that, by increasing the stringency of the hybridization, the %MM which can be detected decreases. Although further experiments are needed to establish the maximum %MM at which the hybrids are still forming, meanwhile the probe design can be done using lower Th1 values (e.g., in the case of rRNA, ≤15%) at which it is known that hybridization is taking place. This will most likely result in having a higher number of probes in the mix. On the other hand, using a too high Th1 can result in probe mixes which do not cover all targets.

Th2 is needed when designing probes specific for a certain taxonomic clade. Its purpose is to select for probes which do not hybridize with non-target taxonomic clades. Th2 is the minimum %MM a specific probe can have to a non-target polynucleotide, without hybridizing to it. The best values for Th2 correspond to the %MM at which hybridization is not kinetically favorable, as it has been discussed above for Th1. For such Th2 values, the hybridization can be performed at temperatures at which the hybridization rate is highest (e.g. 25°C below T_m for DNA-DNA hybrids) and the probes will not bind to non-target polynucleotides. On the other hand, the %MM between probes and non-targets can be low enough for hybrids to form. In this case, as long as the T_m of the target hybrids is higher than the T_m of the non-target hybrids, the discrimination can be made by varying the stringency of hybridization (Trebesius et al., 1994) and hybridizing at temperatures higher than the T_m of the non-target hybrids and lower than the T_m of the target hybrids (Figure 14A). In this case, the fluorescent signal of target cells is significantly decreased (Trebesius et al., 1994).

The choice of a particular hybridization temperature to discriminate a target group depends on the two %MM thresholds, Th1 and Th2. Different studies have found that 1% mismatch in a polynucleotide will produce a decrease in the T_m of mismatched hybrids of 0.5-1.4°C (Anderson and Young, 1985; Caccone et al., 1988; Springer et al., 1992), compared with the T_m of the perfectly matched hybrids. Therefore, %MM can be used to estimate the decrease in T_m of the mismatched hybrids (both target and non-target). Accordingly, Th1 is set in a way that allows for the decrease in T_m caused by mismatches in imperfect probe-target hybrids, and Th2 is set so that hybridization to non-target sequences should not occur. If Th1 is close to Th2, the hybridization temperature will be near the T_m of

mismatched target hybrids (Figure 14B), resulting in unfavorable hybridization kinetics. Doing rRNA FISH, Trebesius et al., 1994, were able to discriminate by very stringent hybridization a non-target with only 4% mismatches, although at a price of a significant loss of hybridization efficiency. For probe mixes, the T_m difference between probes has to be considered, because it can result in overlapping T_m of target and non-target hybrids formed by the different probes in the mix (see step 3, T_m difference section). As soon as the user has set the Th1 and Th2 parameters, PolyPro will select probes which are likely to discriminate between target and non-target hybrids.

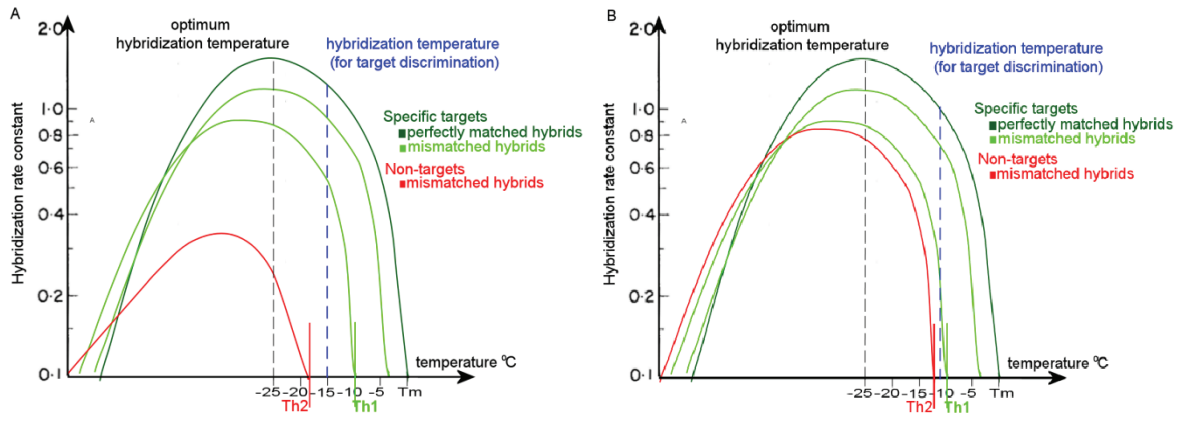


Figure 14: Model for the discrimination between targets and non-targets using polynucleotide probes (adapted from Bonner et al. 1973). Panel A: Targets are distant from non-targets. Th1 is set to 10% and Th2 is set to 18%. Therefore, assuming a 1 °C decrease in T_m for a 1% mismatch, there is an 8 °C gap between the last target hybrid and the first non-target hybrid. The hybridization temperature can be set at about -15 °C below the T_m of the perfectly matched hybrid, allowing a reasonable hybridization rate for all target hybrids, both perfectly matched and mismatched. **Panel B**: Targets are close to non-targets. Th1 is set to 10% and Th2 is set to 12%. Assuming a 1 °C decrease in T_m for a 1% mismatch, there is a 2 °C gap between the last target hybrid and the first non-target hybrid. To discriminate between the hybrids, the hybridization temperature would have to be about -1 °C below the T_m of the last mismatched hybrid. Therefore, in these hybridization conditions, part of the targets would not be hybridized, due to a very low hybridization rate so close to the T_m. The graphs were adapted from Bonner et al. 1973 and the lower part of the hybridization curves represent an extrapolation of the original experimental data.

In step 3, the program uses the hit matrix generated in step 2 to search for combinations of probes ("probe mixes") that cover all alleles in a target group (Figure 15A). Two types of algorithms can be used: standard combinations and optimized combinations. The first one is returning all possible probe mixes. It starts searching for two-probe combinations and if none of those hits all targets it goes to combinations of three probes, up to combination of five probes. The computational demands (CPU and memory usage) are increasing with the number of elements in a combination and with the number of total probes. To reduce the computational demands, we developed the optimized combinations algorithm (Figure 15B). This algorithm is initiated by a standard combination of two or more probes. Then, before going to the next combination step, it does the following: (i) it keeps the

probe mixes with the highest number of hits from the previous step; and (ii) it does a second selection of the probe mixes, based on the T_m tolerance parameter (see below). At the end, for each remaining probe mix, it reunites the hits of all probes in the mix and combines them further with each probe in the hit matrix. The reunion reduces the algorithm to a combination of two elements, with important gain of memory and CPU. These steps are repeated until a probe mix which hits all the targets is found.

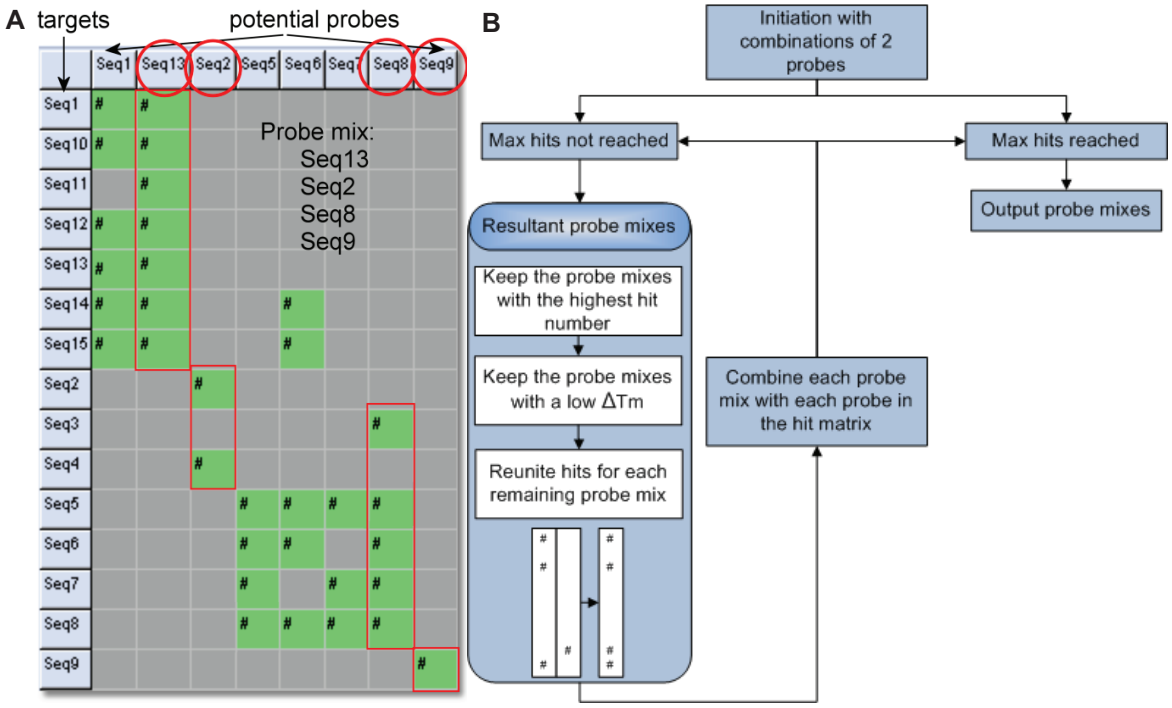


Figure 15: A. Example of how probe mixes are calculated from a hit matrix. In a hit matrix the potential probes are in the first row and the targets in the first column. The "#" sign represents the hits for each probe. The combination algorithm calculates the mix with a minimum number of potential probes which has hits with all the targets. In the example here, the probe mix is formed by Seq13, Seq2, Seq8 and Seq9. **B. Optimized combinations algorithm.**

The T_m tolerance parameter is correlated with the T_m difference between the probes in a mix (ΔT_m), which is mostly caused by difference in the DNA %GC (see materials and methods section for how ΔT_m is calculated). A T_m tolerance of zero means that PolyPro will keep only the probe mixes with the lowest ΔT_m from all probe mixes. For T_m tolerance higher than zero, PolyPro will keep all probe mixes with a ΔT_m between the ΔT_m lowest value and the sum of the ΔT_m lowest value and T_m tolerance value. In step 4 the user can extract the probe mixes which are below a certain ΔT_m. Therefore, the software allows to select the probe mixes with the lowest ΔT_m, either during the process of combinations (step 3), or, afterwards, in step 4. This is necessary because multiple polynucleotides can be used as a probe mix only if their T_m are close to each other, i.e. if they have a low ΔT_m. If ΔT_m is too

high, then the probes will have different optimum hybridization temperatures. As a result, some of the probes will hybridize at optimum hybridization temperature, and some at suboptimum temperature (Figure 16A). The deviation from the optimum will be even higher for the mismatched hybrids. A theoretical calculation of this deviation was done by assuming that 1% mismatches produce a 1°C decrease in T_m and that the reduction in the optimum hybridization temperature for mismatched hybrids is half the reduction in T_m (Bonner et al., 1973). As expected, the deviation from the optimum temperature became larger with increasing both ΔT_m and Th1 (Figure 16B). For DNA:DNA hybrids, the maximum ΔT_m values should be 7°C for Th1=5, 5°C for Th1=10 and 2°C for Th1=15. For these values, the maximum deviation around the optimum temperature is ±5 °C, still placing the hybridization temperature in a optimum interval, for both perfectly matched and mismatched hybrids.

A T_m difference between the probes in a mix will displace the Th1 of the probes with lower T_m closer to the Th2 of the probes with higher T_m (Figure 17). For this reason, an option for an additional selection of the probe mixes based on a recalculated Th2-Th1 value was introduced. For each probe in a mix, all the %MM values are normalized for the maximum T_m in the probe mix, by addition of a factor equal with $T_m$max - T_mPr, where $T_m$max is the maximum T_m in the probe mix and the T_mPr is the T_m of the respective probe. After this correction, Th1 and Th2 will take different values for each probe in a mix. The difference between minimum Th2 and maximum Th1 will be calculated for each probe group and, at the end of this step, only the probe groups with maximum Th2-Th1 will be kept.

When using double-stranded DNA (dsDNA) probes, probes in the mix can crosshybridize if they are too similar. In step 5 the user can select the probe mixes with the lowest similarity between the probes, by calculating the %MM between the probes in a mix and than the minimum %MM for each mix. At the end, it keeps only the probe mixes which have the highest minimum %MM. Additionally, it calculates the average %MM per mix and the coefficient of variation, which are reported in the final result file, for a further selection by the user.

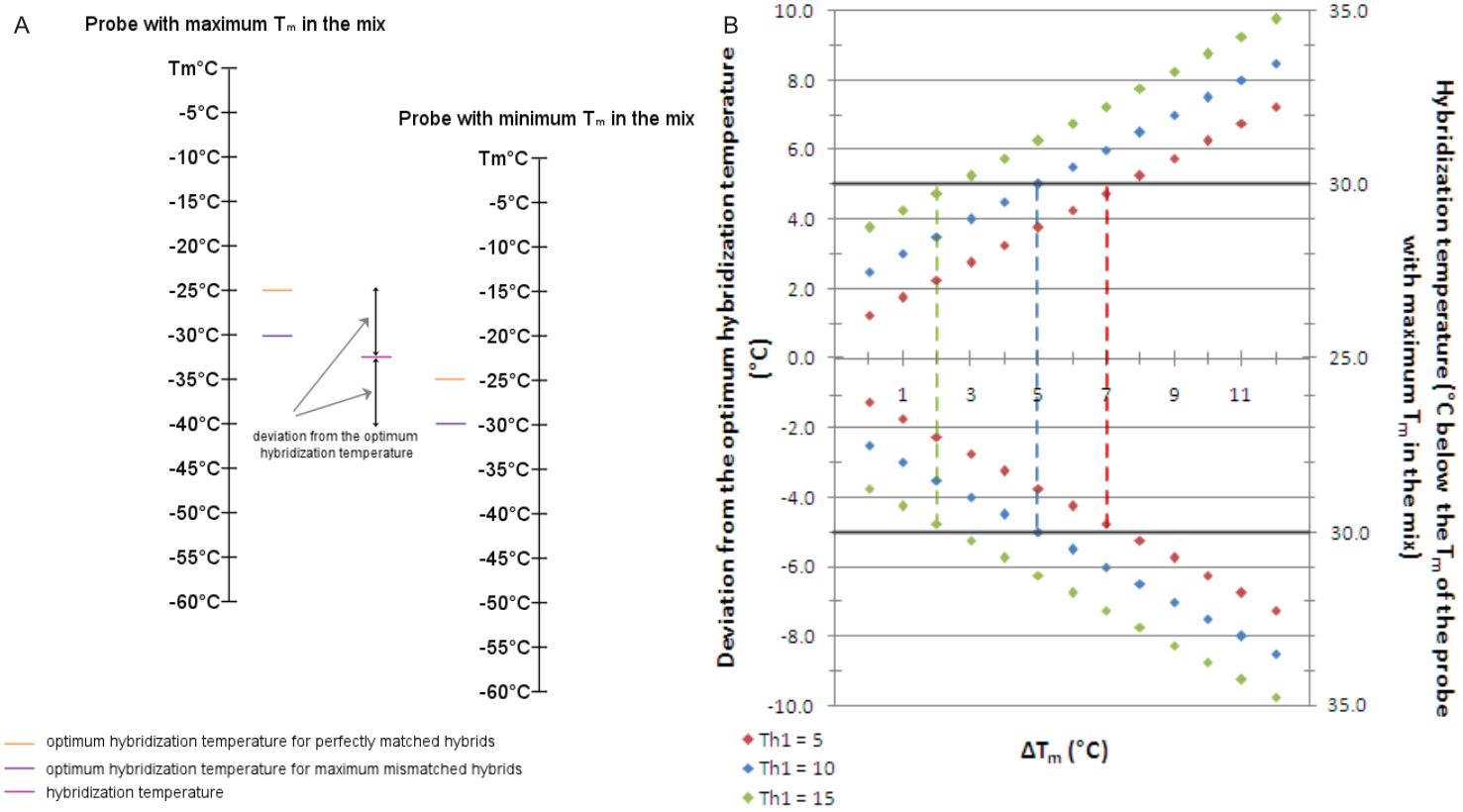


Figure 16: Theoretical calculation of the deviation from the optimum hybridization temperature for DNA:DNA probe–target hybrids at different ΔT_m and Th1 values. A. Example of how the calculation was performed for a probe mix with Th1=10% and ΔT_m=10 °C. It was assumed that a 1% mismatch produces a 1 °C decrease in T_m and that the reduction in the optimum hybridization temperature for mismatched hybrids is half the reduction in T_m (Bonner et al., 1973). The optimum hybridization temperature for mismatched hybrids is represented relative to the T_m of the perfectly matched hybrids. The hybridization temperature was chosen to be the middle point between the minimum and maximum optimum hybridization temperatures in the probe mix. **B. Calculated deviations for different ΔT_m and Th1 values.** The left y axis represents the deviation in °C from the optimum hybridization temperature of each hybrid. The right y axis represents the hybridization temperature relative to the maximum T_m in the probe mix. The three thresholds used for calculations are color coded. The dots represent the maximum deviation for the probe mix. The dark grey, continuous lines indicate the maximum deviation, after which the hybridization temperature will be too high for the probes with maximum T_m in the mix (unspecific hybrids start to form over 30 °C, (Wetmur, 1991)). The discontinuous lines represent the maximum ΔT_m for different Th1 values: red for Th1=5, blue for Th1=10 and green for Th1=15.

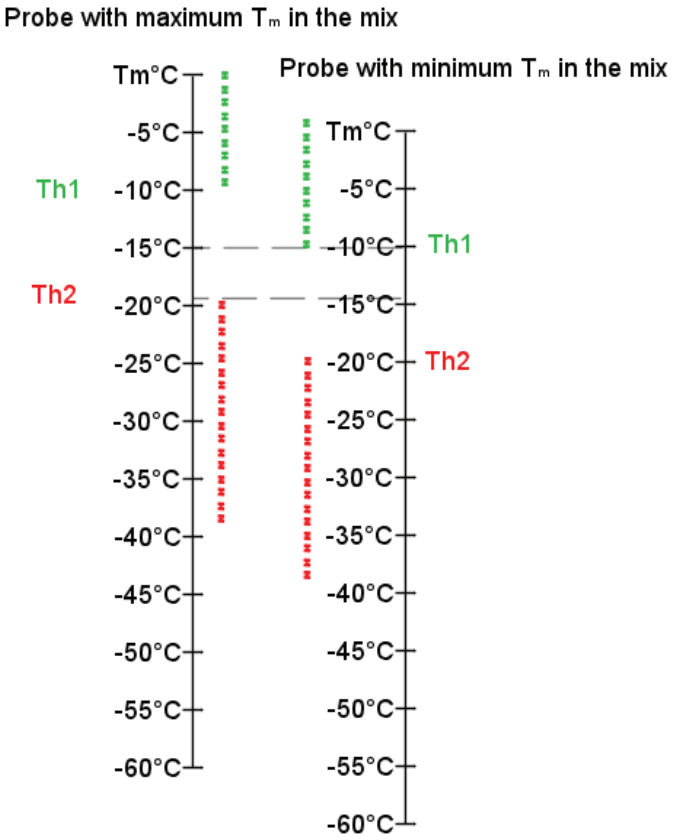


Figure 17: Illustration of the ΔT_m influence on the Th2-Th1 value. The Th1 and Th2 values of the probe with the minimum T_m in the mix have been normalized to the probe with the maximum T_m in the mix. As a result, the Th2-Th1 has decreased from 10 °C to 5 °C.

Hit Maps Submodule

The hit maps submodule creates graphic representations (hit maps) of each probe with all the polynucleotides given as targets, while incorporating their taxonomic affiliation. A hit map (Figure 18) is defined as a plot which has on the x axis the probes and on the y axis the %MM with the targets. Therefore, a dot in the map represents the %MM of a probe with a target. The color of the dot encodes taxonomic information: green means the probe and the target belong to the same taxonomic clade, red means they belong to different taxonomic clades, while blue means no taxonomic information is available. The software is creating hit maps for each of the following taxonomic levels: domain, phylum, class, order, family, genus and species. Additionally, the software creates hit tables for each probe. A hit table is essentially a taxonomy database to which a column with the %MM of the respective probe with all targets has been added. The role of hit maps is to give an overview of the target distribution across probes and hit tables can be used for gaining specific information on each probe. Moreover, when designing probe mixes for taxonomic clades, hit maps can be used for a better visualization of the target versus non-target polynucleotides and of the

Th2-Th1 difference for the selected probe mix (hit maps can be generated also for T_m normalized %MM values, as described above, in T_m difference section).

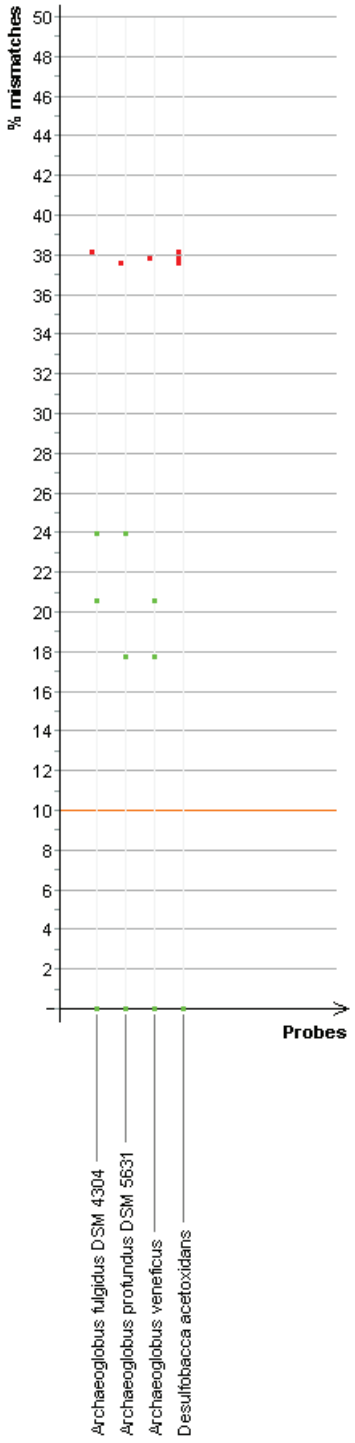


Figure 18: Example of a hit map. The probes are on the x axis. For each probe, the percentage mismatch with each target is represented as a dot (on the y axis). The color of the dot is green when the respective target belongs to the same phylogenetic clade as the probe and red when it belongs to a different clade. In this example, the %MM for *Archaeoglobus* probes with *Archaeoglobus* targets are marked in green (they all belong to *Euryarchaeota*) and with the *Desulfobacca* target they are marked in red.

The Hybridization Parameters Calculator (HPC) module

HPC (Figure 19) is estimating the optimal hybridization parameters of a particular polynucleotide probe mix to defined targets, using T_m formulae from Wetmur (1991). These formulae consider the influence of the concentrations of Na^+ and formamide, %GC, the probe length and degree of mismatch.

HPC receives the input probe and target sequences in FASTA format. The input parameters are the following: (i) Na^+ concentration, (ii) formamide concentration, (iii) criterion (the difference between the T_m and hybridization temperature), (iv) °C decrease in T_m per 1% mismatch, (v) hybridization temperature and (vi) hybrid type (DNA:DNA, RNA:DNA or RNA:RNA). The results are given as single values or as graphs of the dependence of one parameter from another. In the first case, it can calculate formamide concentration, Na^+ concentration, criterion and hybridization temperature. In the second case, it can depict the following dependencies: (i) hybridization temperature as function of formamide concentration, (ii) hybridization temperature as function of Na^+ concentration, (iii) hybridization temperature as function of criterion, (iv) formamide concentration as function of Na concentration, and (v) formamide concentration as function of criterion.

The T_m formulae can only approximate real T_m values, for the following reasons: i) the effect of Na^+ is dependent on the %(G+C) and the formamide concentration (Gruenwedel et al., 1971; Hutton, 1977), but this is not modeled into the T_m formulas; ii) the effect of formamide concentration on the T_m of RNA:DNA hybrids is not linear (Casey and Davidson, 1977), but in the formula is expressed as linear; and iii) the effect of %MM is predicted to be between 0.5 and 1.4°C decrease in T_m for 1% mismatch, which is giving a high range of possible T_m for the mismatched hybrids. Moreover, the influence of the nucleic acid sequence itself is considered only in an averaging way, in the terms for %GC and %MM, and multiple melting domains are not revealed by these formulas. Therefore, the T_m formulas are a good starting point, but additional lab experiments (e.g. *in vitro* T_m measurements) are needed to establish specific hybridization conditions.

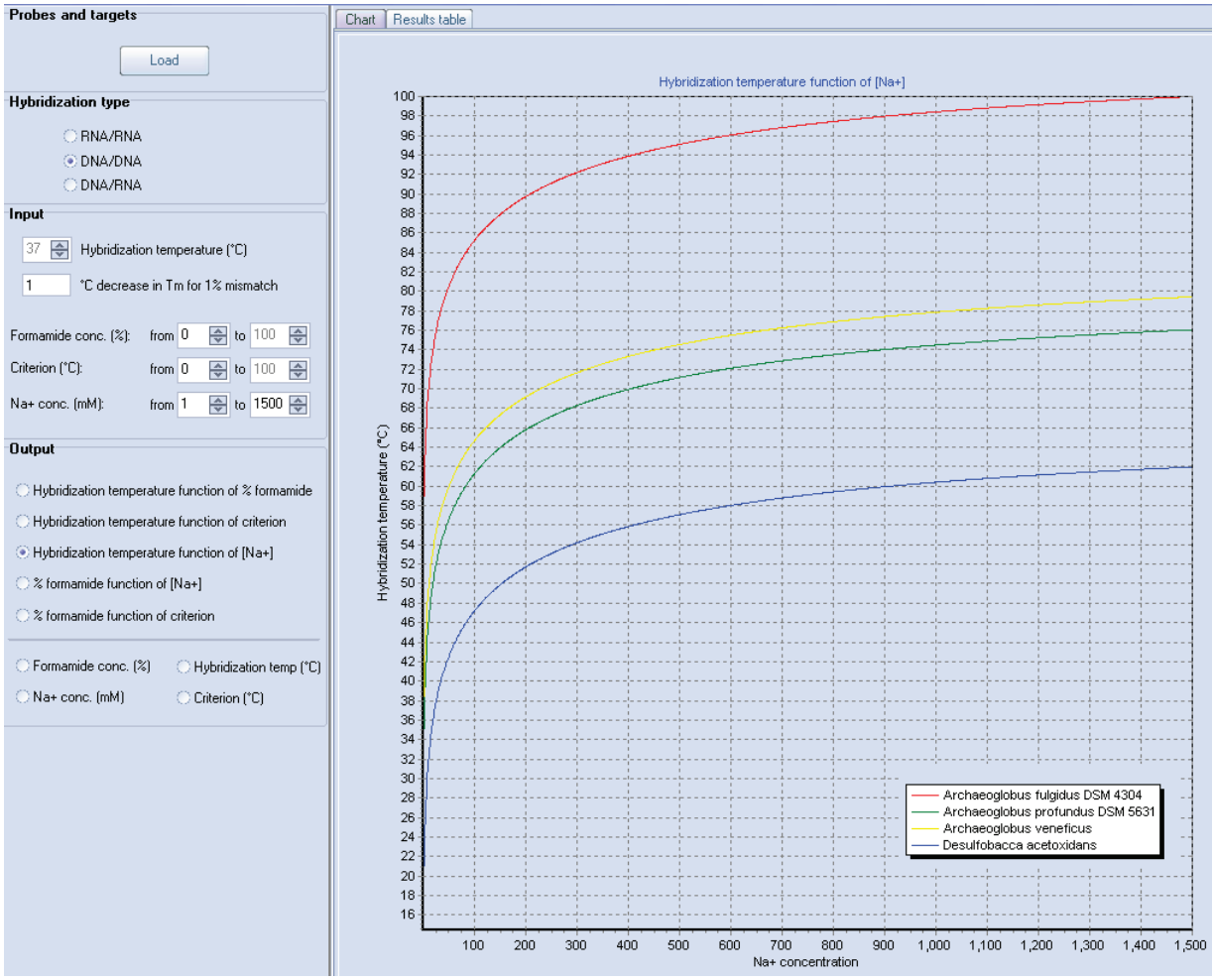


Figure 19: HPC screenshot. In the left panel, the input (hybridization type, formamide and salt concentration, criterion, hybridization temperature and the decrease in T_m for each percent mismatch) and output (formamide and salt concentration, hybridization temperature, criterion) parameters can be selected. In the right panel, the results can be displayed either as graphs or as tables.

3.1.2. Testing how polynucleotides can be used as probes for metabolic marker genes

The *apr*A, *amo*A and *pmo*A genes were used to test the potential of polynucleotide probes to identify cells containing the marker genes for metabolic functions. They were chosen to test different scenarios: i) *apr*A is found in a broad range of sulfide oxidizing and sulfate reducing taxa and is subjected to lateral gene transfer (LTG) (Friedrich, 2002; Meyer and Kuever, 2007a; Mussmann et al., 2007), ii) *amo*A and *pmo*A are restricted to a few taxonomic clades (Bothe et al., 2000; Könneke et al., 2005; Dunfield et al., 2007; Pol et al., 2007; Torre et al., 2008), and they seem to be less subjected to lateral gene transfer. Only sequences from cultivated or otherwise identified microorganisms were used. The *amo*A and *pmo*A genes were analyzed together, because they are homologous (Holmes et al., 1995;

Dunfield et al., 2007; Pol et al., 2007). The *pmo*A and the bacterial autotrophic *amo*A genes are closely related with each other and so distant from the crenarchaeal *amo*A (Könneke et al., 2005; Torre et al., 2008) and the heterotrophic *amo*A (Moir et al., 1996; Daum et al., 1998), that the later two were excluded from this analysis. In the case of the *apr*A gene, the two homologues, the one in Sulfate Reducing Prokaryotes (SRP) and the one in Sulfur Oxidizing Prokaryotes (SOP) (Fritz et al., 2000; Meyer and Kuever, 2008), were analyzed together.

The sequences were analyzed with PolyPro software and the work flow described above. Each sequence was considered a potential probe and its specificity tested against all sequences of the respective gene, using as parameters 10% mismatches for Th1 and DNA:DNA for hybridization type. The results are detailed in Appendix E: Polynucleotide probe design – results table, Table 38 for the *amo*A and *pmo*A genes, and in Table 39 for the *apr*A gene.

The first question was whether all alleles of a certain gene can be identified in a single hybridization experiment. Due to the high variability of the alleles, a single probe cannot cover all of them, as observed also by previous studies (Pernthaler and Amann, 2004; Pilhofer et al., 2009). The alternative would be to use mixes of polynucleotide probes. This is possible when ΔT_m is low enough for the probe mix to behave as one probe, i.e. for Th1 of 10%, ΔT_m should be ≤5°C. High computational demands made it impractical to search probe mixes that hit all sequences from one gene, as well as some of the high level taxonomic clades (e.g. the *amo*A probe-mix for *Proteobacteria*, the *apr*A probe-mixes for *Bacteria, Proteobacteria* and *Deltaproteobacteria*). In these cases, we calculated the probe mixes by putting together probe mixes from lower taxonomic levels (Appendix E: Polynucleotide probe design – results table, Table 38 and Table 39). The probe mix for all *amo*A had 35 probes and a ΔT_m of 7.9 °C, for all *pmo*A had 31 probes and a ΔT_m of 12.4°C, and for all *apr*A had 115 probes and a ΔT_m of 10.4 °C. The probe mix for *apr*A from SRP had 85 probes and a ΔT_m of 9.8 °C and for *apr*A from SOP had 26 probes and a ΔT_m of 9.4 °C. In all cases, the ΔT_m was higher than 5°C. Therefore, general gene probes could not be designed. We do not exclude the possibility that, for highly conserved genes, ΔT_m would be favorable and the probe mixes could be used for hybridization.

The next question was if and how polynucleotide probes can be used to target different taxonomic clades, by designing probe mixes from the domain to genus level. The resulted probe mixes had between 1 and 107 probes (Figure 20A). The number of probes in a mix increased with the number of targets in a clade, the linear correlation being strong for *amo*A and *apr*A genes (R^2 of 0.94 and 0.99, respectively) and only a moderate for *pmo*A gene (R^2 of 0.68). The increase correlated with the variability of the respective gene. For example, the *apr*A gene, which is found in a wide number of taxa, needed a high number of

probes for a low number of targets. In the same time, the *amo*A and *pmo*A genes, which are less variable and found in few taxa only, needed fewer probes for a higher number of targets. The same trend can be seen when looking at how the number of probes related to the taxonomic level: for lower taxonomic levels, with low allele variability, the number of probes was smaller than for higher taxonomic levels, with high allele variability (Figure 20B). At the genus level, in 63% of the cases, one probe was enough to hit all targets. Therefore, the number of probes in a mix depends on the evolutionary distance between the alleles of the gene. The more variable a gene is, the more probes in a mix will be needed.

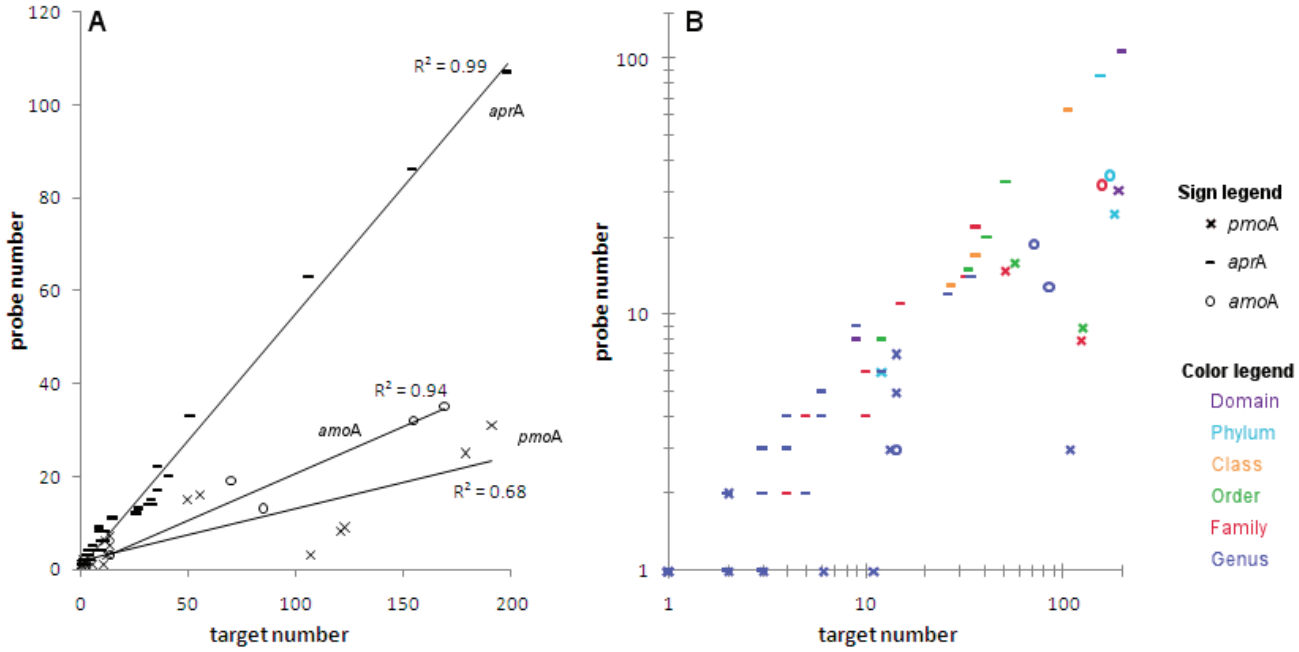


Figure 20: The relation between the target number in a taxonomic clade and the probe number in a mix that is hitting all targets of that particular taxonomic clade. A. The three genes – *amo*A, *pmo*A and *apr*A – are analyzed separately. The *amo*A and *apr*A genes show a strong linear correlation between the target number and the probe number. The *pmo*A gene shows a moderate linear correlation. **B.** The taxonomic level for each data point is color coded (see color legend). The main trend is that low taxonomic levels (genus, family) have fewer probes in a mix and higher taxonomic levels have more probes in a mix.

ΔT_m for all probe mixes was between 0.35 and 12.4°C (Figure 21A). It increased with the probe number in a mix, the linear correlation being strong for *amo*A and *pmo*A genes (R^2 of 0.98 and 0.92, respectively) and only moderate for *apr*A gene (R^2 = 0.51). Generally, low taxonomic levels had low ΔT_m values, most of the probe mixes for genera and families having values ≤5°C (Figure 21B), e.g. the *amo*A probe mix for *Nitrospira* (ΔT_m = 2.9°C), the *pmo*A probe mix for *Methylocaldum* (ΔT_m = 1.5°C), the *apr*A probe mixes for *Thermodesulfobacteriaceae* (ΔT_m = 2.1°C), *Pyrobaculum* (ΔT_m = 1.8°C), *Syntrophaceae* (ΔT_m = 0.5°C) and *Desulfococcus* (ΔT_m = 0.35°C). The more scattered distribution of the ΔT_m

values for the *apr*A gene might be explained by the acquirement of alleles from other clades, by LTG. The new alleles can have different %GC from the alleles of the clade, resulting in a higher ΔT_m.

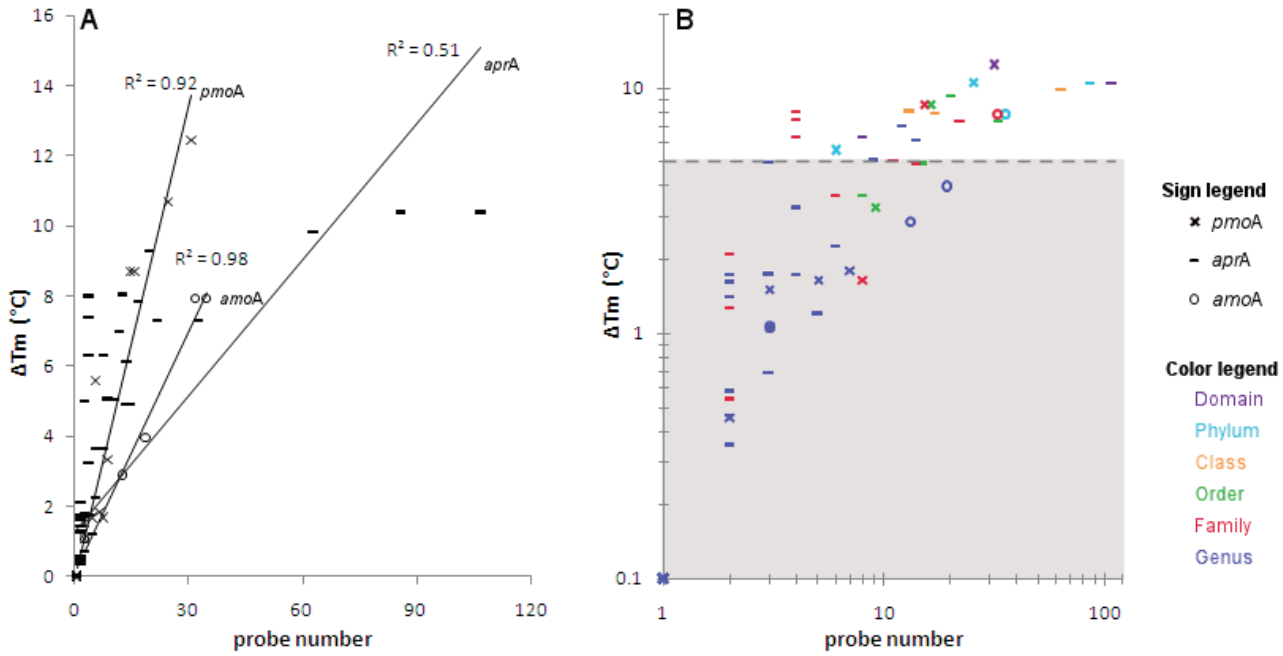


Figure 21: The relation between the probe number in a mix and ΔT_m. A. *amo*A, *pmo*A and *apr*A were again analyzed separately. The *amo*A and *pmo*A genes show a strong linear correlation between the probe number and ΔT_m. The *apr*A gene shows a moderate linear correlation. **B.** The taxonomic level for each data point is color coded (see color legend). The main trend is that low taxonomic levels (genus, family) have smaller ΔT_m than higher taxonomic levels. Most of the probe mixes for genus and family have a ΔT_m within the 5°C limit for Th1=10 (grey area), and thus, they can be used as a single probe.

To test the potential for specificity of the polynucleotide probes, the Th2 value needed to be considered. The Th2 was above 25% mismatches for most of the higher taxonomic clades and for some of the genus and family clades (Figure 22A), value at which the probe mixes should not bind to non-targets (e.g. the *amo*A probe mix for *Betaproteobacteria* with Th2 = 28). For the probe mixes with Th2 values lower than 25, the difference between Th2 and Th1 was calculated, taking into consideration ΔT_m. Most of the genera and families had Th2-Th1 values larger than 4°C (with an average of 11.6±5.6°C) (Figure 22B), which allows for discrimination during hybridization (e.g. the *pmo*A probe mix for *Methylohalobius*, with Th2-Th1 = 14.4 °C).

For *apr*A gene, low Th2 values can be correlated with LTG events, and this might explain why some of the probe mixes for higher taxonomic clades had low Th2-Th1 value (Figure 22B). One example of LTG is between Phyla *Nitrospirae* and *Firmicutes* (Friedrich, 2002), the *apr*A sequence from *Thermacetogenium phaeum* grouping closer to *Nitrospirae* than to *Firmicutes*. As a result, Th2 decreased to 13% for *Nitrospirae* and to 12% for

Firmicutes, instead of 32% and 19% as it would have been excluding the *Thermacetogenium phaeum* sequence. Putative LTG events have been noticed also inside *Chromatiaceae* (Meyer and Kuever, 2007a). This might be why we could not design specific probes for the genera *Allochromatium* and *Lamprocystis* (*Allochromatium minutissimum* and *Lamprocystis roseopercina* have a %MM of 0%), or for the *Thiodictyon* and *Thiocapsa* genera (*Thiodictyon bacilossum* and *Thiocapsa roseopercina* have a %MM of 0%) (Table 39).

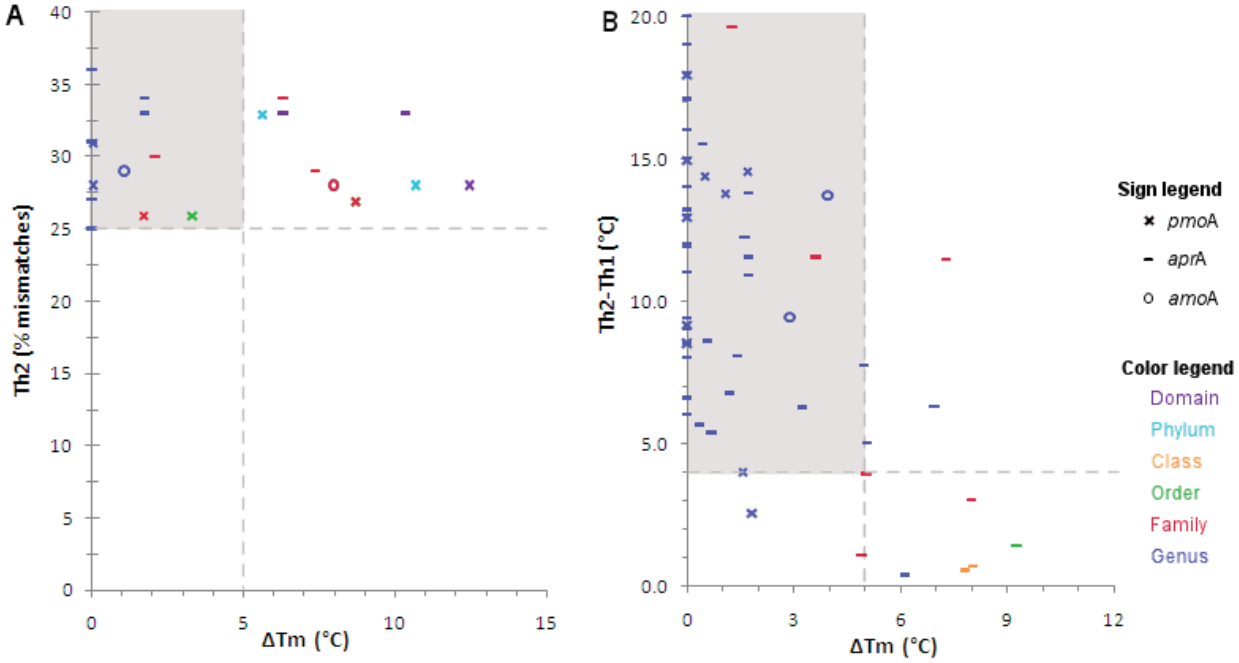


Figure 22: Th2 distribution across taxonomic levels for *amo*A, *pmo*A and *apr*A genes. The taxonomic level is color coded and the gene is sign coded (see legend). A. The probe mixes with Th2 ≥ 25% mismatches are represented here. This probe mixes should not bind to non-target alleles when hybridized at optimum hybridization temperature. **B.** For the probe mixes with Th2 ≤ 25 mismatches, Th2-Th1 was calculated applying the correction for the T_m difference in the mix. All the probe mixes with Th2-Th1 above 4°C will allow discrimination of the targets from the non-targets, by modulating the hybridization / washing parameters. The shaded areas represent the probe mixes which have both discrimination ability and favorable ΔT_m.

The analysis of the three genes indicated that polynucleotide probes can be used mostly for genus and family levels. For these taxonomic levels, the ΔT_m is low enough so that the probe mixes can be used like individual probes in a single hybridization. Moreover, the Th2-Th1 values are mostly high enough to allow for specific hybridization of only the clades of interest. Examples of such probe mixes are the *amo*A probe mixes for *Nitrosomonas* (ΔT_m=4 °C, Th2-Th1 =13.8 °C) and for *Nitrosospira* (ΔT_m=2.9 °C, Th2-Th1=9.5 °C). The respective probe mixes could be used to discriminate between these two ammonia oxidizing clades. Another example would be the *pmo*A probe mixes for *Alphaproteobacteria* (ΔT_m=3.3°C, Th2 =26% mismatches), which could be used to differentiate type II methanotrophs from type I and X. However, because often the whole

diversity of a gene is not covered by the cultivated representatives, it is not recommended to design probes only based on the information from cultivated microorganisms, but rather on the information specific for the analyzed sample.

Having general probes for genes would be very useful for many FISH studies, e.g. when asking if a certain uncultivated microorganism has a certain gene or, when monitoring mRNA expression. Since such general polynucleotide probes cannot be designed, other ways are required to address these questions. One way would be to split a general probe mix containing probes that cannot be hybridized under the same conditions in sub-mixes with lower ΔT_m. All the alleles, both from cultivated and uncultivated microorganisms should enter the probe design. For highly variable genes this would result in a high number of probes in the general mix, and therefore, in many sub-mixes. Working with many sub-mixes, each in a separate hybridization reaction, can be impractical. On the other hand, the diversity at one particular environmental site is presumably lower than the whole diversity of the gene. Therefore, another way around the inability to design general probes for genes would be to design sample specific probes.

The probe mix approach is certainly demanding. There are other much simpler approaches for using polynucleotide probes: using only one allele as probe, as it has been done in previous studies, or, using as probe the PCR product obtained from the environmental sample. The first has the obvious disadvantage that a single probe in many cases is not enough to cover the diversity of the clade of interest or, alternatively, it might cover quite diverse target groups. The second has the disadvantage of using an allele mixture of uncharacterized diversity (including unspecific PCR products) and ΔT_m. Rationally designed probe mixes, in contrast, have defined target groups and optimized hybridization conditions. Our concept requires knowledge of the allele diversity in the sample of interest. This can be obtained by clone libraries or in depth metagenomics. Probes could be designed for all the alleles in the sample, or, alternatively, phylogenetic trees can be constructed and probes designed for the clades of interest.

3.2. GeneFISH – protocol development and applications

The geneFISH protocol was developed on *E. coli* cultures, using first ssRNA probes and then dsDNA probes. Later, the protocol was applied on an enrichment sample and two types of environmental samples, using dsDNA probes. The genes targeted with ssRNA probes were: (i) *rpo*B (the ß subunit of the DNA dependent RNA polymerase), with a chromosomal localization in *E. coli* K12, and (ii) *dsr*A (the α subunit of the dissimilatory sulfite reductase) and *hme*D (subunit D of the Hdr-like menaquinol-oxidizing enzyme), with a fosmid localization in *E. coli* clones (copy control fosmids with 1-2 copies per cell). The genes targeted with dsDNA probes were: (i) *amo*A, with a plasmidial localization in *E. coli* clones (copy control plasmids with 1-2 copies per cell), (ii) *amo*A, with a chromosomal localization, detected in picoplankton cells from Benguela upwelling system, (iii) *rdsr*A (the α subunit of the reverse dissimilatory sulfite reductase), with a chromosomal localization, detected in an enrichment sample, (iv) *hyn*L (the large subunit of the [NiFe] hydrogenase) and (v) *apr*A, both with a chromosomal localization, detected in bacterial symbionts of *Bathymodiolus puteoserpentis*. The main focus was the detection of single cells containing the crenarchaeotal *amo*A gene in seawater samples from Benguela upwelling system.

3.2.1. Probe design

The *in silico* studies described in chapter 3.1.2. of this thesis showed that the diversity of functional genes, as inferred from cultured microorganisms, is too high to allow the design of a single polynucleotide probe mix which targets all alleles of a gene. Therefore, probes specific for particular samples were designed, ranging from pure cultures, enrichments or environmental samples.

Tables listing the probes applied in this study to pure cultures (Table 23) and mixed samples (Table 24) are shown below.

The *rpo*B probe was designed to perfectly match a ~350 bases region of the respective gene from *E. coli* K12. The *dsrA* and *hme*D probes were designed to be perfect matches of genes in *E. coli* clone WS 39F7, and it had 19%, respectively 18% mismatches with the genes in *E. coli* clone WS 7F8.

Table 23: Polynucleotide probes for pure cultures

Probe	Probe type	% MM with target		Size	Targeted gene region
		% MM	Target organism		
*rpo*B	ssRNA	0%	*E. coli* K12	359 nt	position 1702-2060, *E. coli* K12 numbering
*dsr*A	ssRNA	0%	*E. coli* clone WS 39F7	365 nt	position 154-518, clone WS 39F7 numbering
		19%	*E. coli* clone WS 7F8		
*hme*D	ssRNA	0%	*E. coli* clone WS 39F7	370 nt	position 331-700, clone WS 39F7 numbering
		18%	*E. coli* clone WS 7F8		
*amo*A 1E3	dsDNA	0%	*E. coli* clone 1E3	351 bp	position 95-445, *Cenarchaeum symbiosum* numbering
		1%	*E. coli* clone 1E7		
		5%	*E. coli* clone 3G4		

Table 24: Polynucleotide probes for mixed samples

Probe	Probe type	No. of targeted alleles	No. of probes	%MM with target	Size (bp)	Targeted gene region
*amo*A-Nam	dsDNA	96	12	0%-5%	351	position 95-445, *Cenarchaeum symbiosum* numbering
*FOS-rdsr*A	dsDNA	1	1	7.8%	359	position 246-606, *Magnetospirillum magneticum* AMB-1 numbering
*hyn*L-Mix	dsDNA	1	4	0-1.1%	349-356	positions 121-480 (*hyn*L1), 533-886 (*hyn*L2), 902-1256 (*hyn*L3), 1329-1679 (*hyn*L4), *Beijerinckia indica* ATCC 9039 numbering
*apr*A-Bath	dsDNA	1	1	0%	352	position 411-764, Candidatus *Ruthia magnifica* numbering

Polynucleotide probe design for amoA gene

The design of appropriate polynucleotide probes targeting crenarchaeotal putative *amo*A genes started with the analysis of clone libraries of PCR-amplified *amo*A gene fragments. Three samples were collected from two stations in the Benguela upwelling system on the Namibian shelf onboard R/V Meteor in May/June 2008 during the M76/2 cruise: station 249, 128 m depth, station 213, 65 m depth and station 213, 76 m depth. For each *amo*A clone library, 163-198 clones were selected for sequencing. The resulting sequences were aligned and a region of 351 bp (position 95-445 of *amo*A, *Cenarchaeum symbiosum* numbering) was selected for probe design. For protocol development, we first tried to detect the *amo*A gene fragment in three *E. coli* clones. A single polynucleotide probe was designed, *amo*A-1E3, based on clone 1E3 (clone library from station 213, 76 m). This 351 bp probe (see Appendix F) had the following percent mismatches with the *amo*A clone inserts: 0% with 1E3 clone, 1% with 1E7 clone and 5% with 3G4 clone.

A maximum likelihood tree was constructed that included all *amo*A sequences from this study (Figure 23). Most of the sequences grouped within the previously described marine clusters A (here denoted A1) and B, with a lower number grouping close to *Nitrosopumilus maritimus* and in a new marine cluster, A2, which was well separated from cluster A1 (Figure 23). For detection of the *amo*A gene in the Namibian seawater samples, the diversity of the 351 bp region used for probe design needed to be analyzed. The analysis revealed the presence of a high number of alleles, with more than 100 unique sequences retrieved from each clone library. Operational taxonomic unit (OTU) grouping reduced the number of sequences to ~50 (99% identity cutoff) and to ~14 OTUs (95% identity cutoff) (Table 25) per clone library. In total, 286 OTUs were retrieved when using a 99% identity cutoff to allow for *Taq* polymerase errors (Acinas et al., 2005). The two diversity estimators used (Chao and Jackknife, see Table 25) and the rarefaction curves (Figure 24) indicated that while at 99% identity cutoff the clone libraries had not captured the full *amo*A diversity, most OTUs had been retrieved if the 95% identity cutoff was used. At 99% identity cutoff, none of the OTUs represented more than 10% of the sequences. At 95% cutoff, between 70% and 80% of the sequences were found in only three OTUs (Figure 25). The three clone libraries were examined using Venn diagrams and were found to be similar, sharing amongst each other between 49% and 58% of their sequences (Figure 26).

Despite the observed high diversity, it was possible to design a probe mix which covered all targets retrieved by clone libraries and which could be used as a single probe. The sequences from the three clone libraries were pooled and the probe design was done with PolyPro software, using a mismatch threshold 1 (Th1) of 5%. For a detailed description of the probe design see Appendix F, Figure 52 through Figure 57. The resulting probe mix (*amo*A-Nam, see Appendix F) contained 12 polynucleotides, with a ΔT_m of 1.7°C (Table 26).

From the total of 552 *amo*A target sequences, 77% were covered by two of the polynucleotides, while the remaining 23% were covered by the other 10 polynucleotides (Figure 27). From the 12 probes, three belonged and targeted cluster A1, seven cluster B, one cluster A2 and one the *Nitrosopumilus*-like cluster. At a Th1 of 5%, none of the probes had hits outside their own cluster (Figure 28).

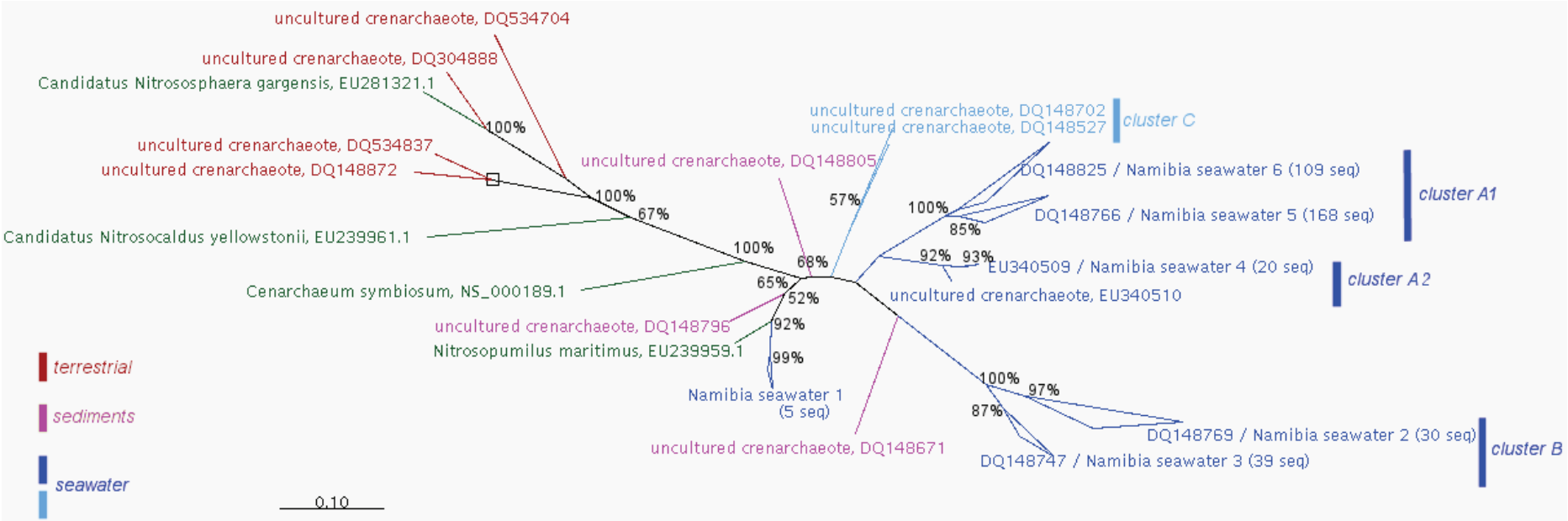


Figure 23: A maximum likelihood tree based on nucleotide sequences of crenarchaeotal *amo*A obtained from the three Namibian seawater samples. Bootstrap values (> 50% from 1000 bootstraps) are denoted at branch points. Sequences are color coded according to types of sample origins (terrestrial, marine sediments or seawater). The marine clusters A (here denoted A1), B and C, as defined by (Francis et al., 2005), plus a new cluster, A2, are also indicated. The majority of the Namibia seawater *amo*A group in marine clusters A1 and B. The Namibian *amo*A are represented as groups labeled 'Namibia seawater', with the number in brackets representing the number of unique *amo*A sequences from each group.

Table 25: Diversity of *amo*A gene (the 351 bp fragment used for probe design) in the three Namibian seawater samples – clone libraries results.

Clone library	*Total sequences*	*Diversity estimator*	*DNA OTUs*		
			unique	**99% identity**	**95% identity**
Station 249, depth 128 m	191	observed	126	58	14
		Chao	-*	103	17
		Jackknife	-	110	18
Station 213, depth 65 m	198	observed	135	52	13
		Chao	-	75	13
		Jackknife	-	77	15
Station 213, depth 76 m	163	observed	105	44	13
		Chao	-	65	13
		Jackknife	-	64	15
All samples	552	observed	286	96	19
		Chao	-	135	20
		Jackknife	-	136	22

* "-" = not determined

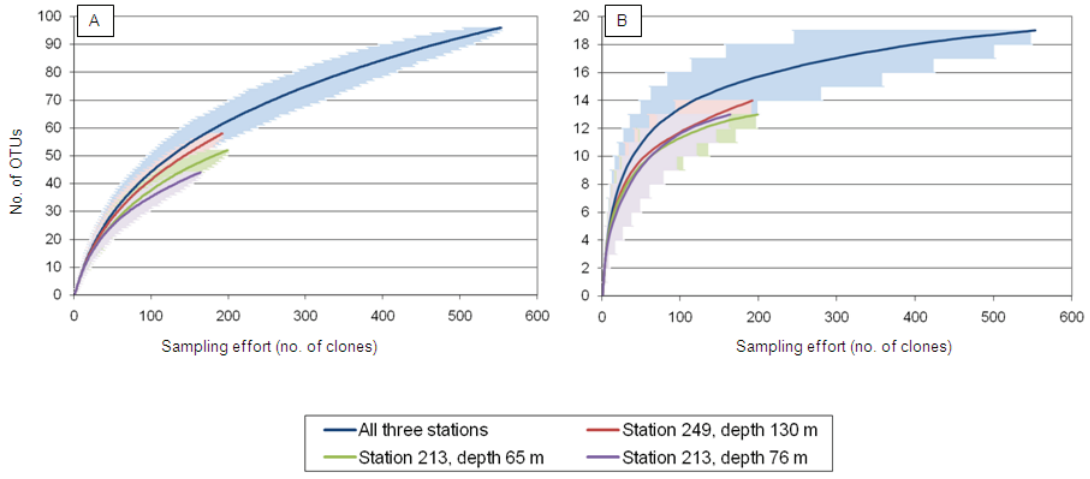


Figure 24: Namibian seawater samples - Rarefaction curves for the *amo*A gene (the 351 bp fragment used for polynucleotide probe design) at a 99% identity cutoff for OTU grouping (panel A) and a 95% identity cutoff for OTU grouping (panel B). Shaded areas represent the confidence intervals.

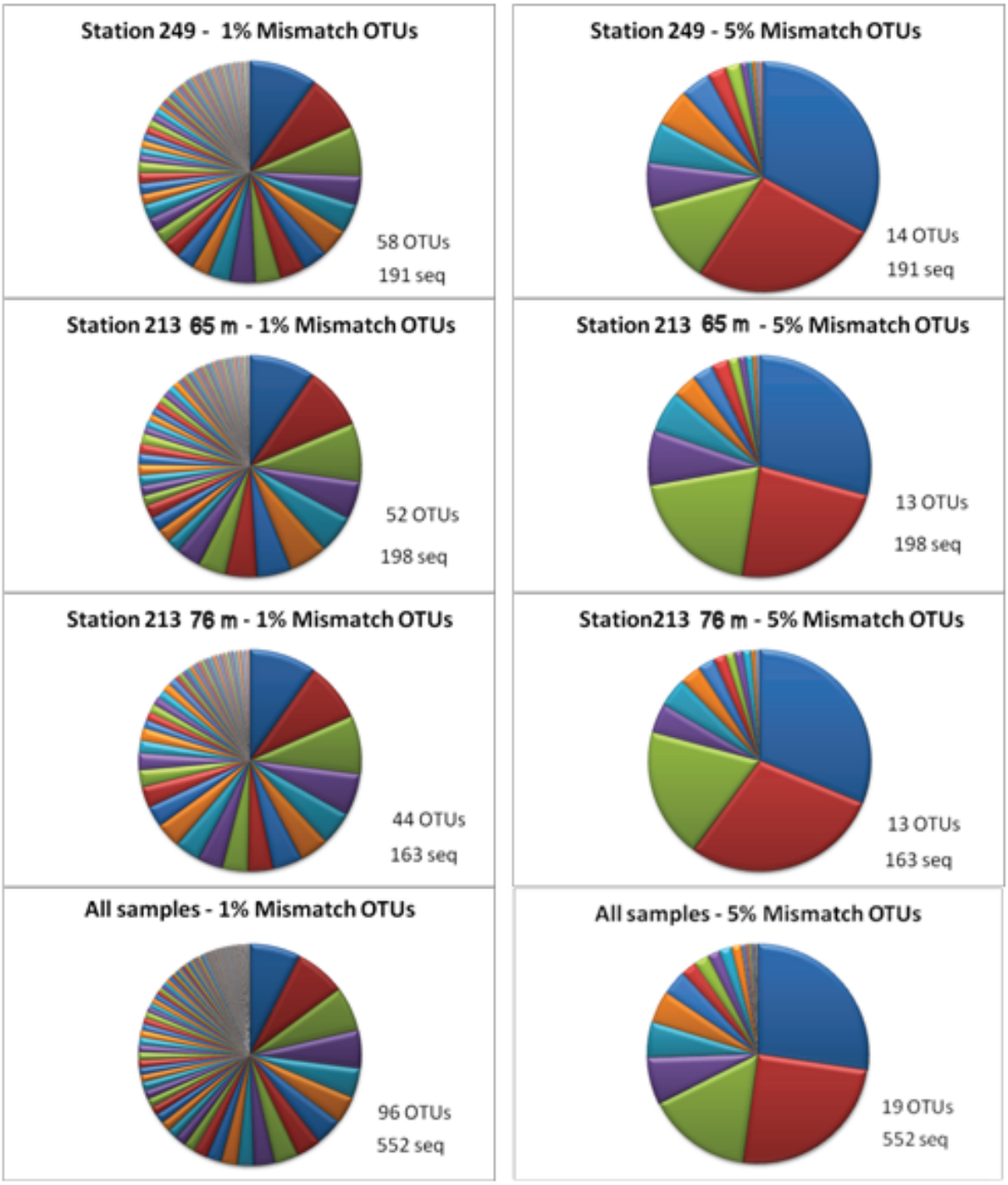


Figure 25: Namibian seawater samples - Distribution of the *amo*A gene (the 351 bp fragment used for polynucleotide probe design) in OTUs, at 99% and 95% identity cutoff. At 99% identity cutoff, none of the OTUs represents more than 10% of the sequences. The majority of the OTUs are represented by a very low number or individual sequences. Pooling the sequences from the three samples results in an almost doubling of the OTU number, as compared to individual clone libraries, and none of the individual OTUs are represented by more than 8% of the sequences. At a 95% identity cutoff, between 70% and 80% of the sequences are found in three OTUs. Pooling the sequences from the three clone libraries results in an increase of the number of OTUs with less than 50%, as compared to individual clone libraries, and 68% of the sequences are found in three OTUs.

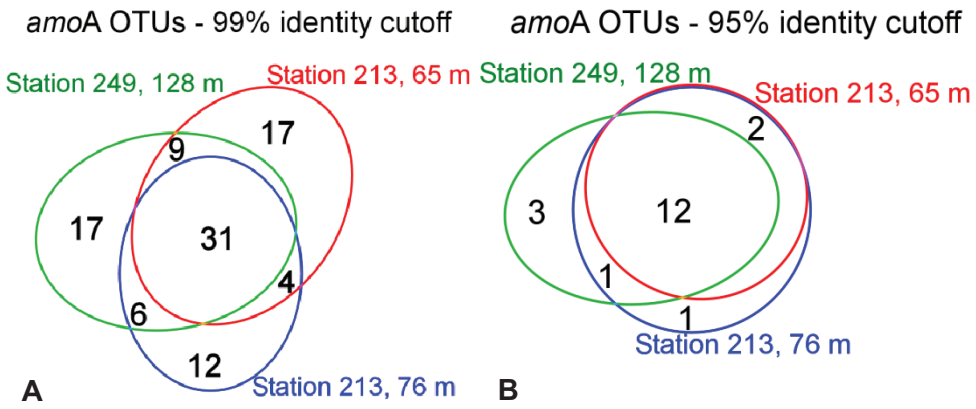


Figure 26: Distribution of the *amo*A gene (the 351 bp fragment used for polynucleotide probe design) across the three Namibian seawater samples. The OTUs were grouped according to a 99% identity cutoff (panel A) and a 95% identity cutoff (panel B). At a cutoff of 99% identity, the three clone libraries shared amongst them between 49% and 58% of their sequences, while between 21% and 28% were unique sequences and the remaining 21-24% sequences were shared between two clone libraries. When using 95% identity as a cutoff for OTU grouping, between 75% and 85% of the sequences were shared amongst the three clone libraries.

Table 26: The *amo*A-Nam probe mix used for geneFISH on Namibian seawater samples (targeting all retrieved crenarchaeotal *amo*A sequences). The parameters for this probe mix were: ΔT_m 1.7 °C, average %MM between probes: 18.77% ±6%. The probes had a length of 351 bp (position 95-445 *Cenarchaeum symbiosum* numbering).

Clone / Probe name	NCBI accession no.	Phylogenetic affiliation
amoA clib1-pl1 cl10	GQ500153	Marine Cluster A1, Namibia seawater 6
amoA clib1-pl1 cl34	GQ500177	Marine Cluster B, Namibia seawater 3
amoA clib1-pl1 cl53	GQ500196	Marine Cluster A1, Namibia seawater 5
amoA clib1-pl1 cl54	GQ500197	Marine Cluster B, Namibia seawater 2
amoA clib1-pl1 cl56	GQ500199	Marine Cluster A1, Namibia seawater 5
amoA clib1-pl1 cl59	GQ500201	Marine Cluster B, Namibia seawater 3
amoA clib1-pl1 cl75	GQ500213	Marine Cluster A2, Namibia seawater 4
amoA clib1-pl2 cl106	GQ500233	Marine Cluster B, Namibia seawater 2
amoA clib1-pl2 cl145	GQ500261	*Nitrosopumilus* –like, Namibia seawater 1
amoA clib2-pl2 cl66	GQ500359	Marine Cluster B, Namibia seawater 3
amoA clib2-pl3 cl105	GQ500387	Marine Cluster B, Namibia seawater 2
amoA clib3-pl1 cl74	GQ500555	Marine Cluster B, Namibia seawater 2

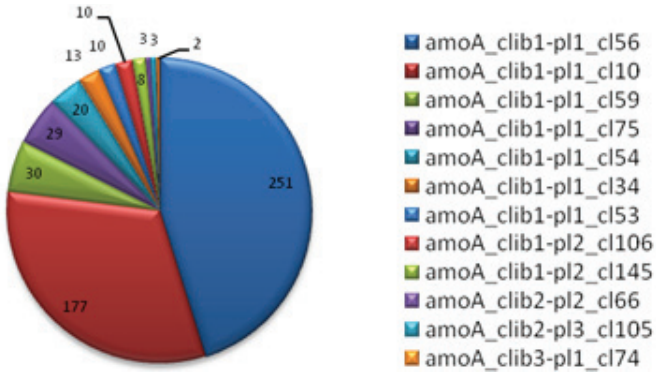


Figure 27: Namibian seawater samples - coverage of *amo*A targets by the 12 polynucleotides in the *amo*A-Nam probe (number of targets hit by each polynucleotide). A sequence is a target for a certain probe when the percentage mismatch between it and the probe is between 0% and 5%. The legend shows the probe names.

One question that might arise was whether the diversity had been sampled sufficiently by our cloning and sequencing efforts, such that the *amo*A-Nam probe mix truly covered the majority of the *amo*A alleles in these environmental samples. Since the mismatch threshold for probe design (Th1) was set to 5% mismatch, the coverage of the clone libraries should be estimated using the 95% identity cutoff for OTU grouping. Results (Table 25, Figure 24) indicated that the 19 OTUs retrieved represented >80% of the diversity. From the 12 polynucleotides of the *amo*A-Nam probe mix, two (cl56 and cl10) belonging to cluster A1 targeted 77% of all sequences. The other ten probes covered only between 0.36% and 5.43% of the sequences (Figure 27), but, on the other hand, they targeted Cluster B, A2 and *Nitrosopumilus*-like sequences (Figure 28). Although we took measures to reduce the PCR bias, the clone libraries cannot be used to infer cellular abundances of the different clades. Some sequences present in small numbers in the clone libraries could be numerically abundant in environmental samples. Therefore, all 12 polynucleotides were used together. Although a Th1 of 5% was used for designing the *amo*A-Nam probe mix, we expect that the probes will bind to targets with higher percent mismatches. However, as our results with clone 3G4 indicated (Table 31), the gene detection efficiency will decrease with increasing the percentage mismatch. Using only the two probes (cl56 and cl10, which cover 77% of the sequences) would have resulted in low or no detection of targets from clusters B, A2 and *Nitrosopumilus*-like *amo*A. Therefore, it can be safely assumed that the *amo*A-Nam probe mix binds the majority of *amo*A alleles in the three Namibian seawater samples.

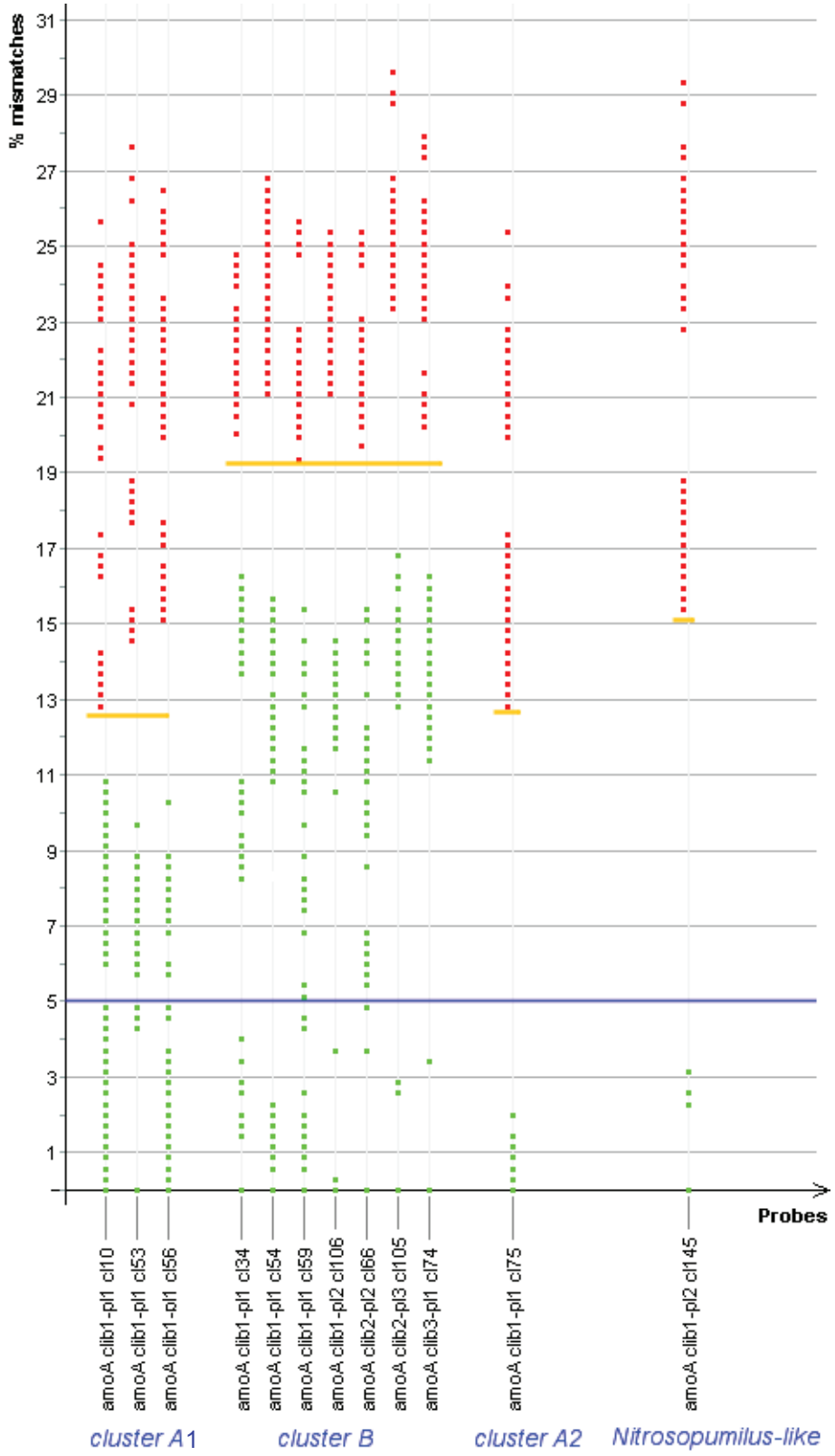


Figure 28: Hit map for the 12 polynucleotides in the *amo*A-Nam probe mix. For each probe the percentage mismatch with each sequence from the clone libraries is represented as a dot. The color of the dot is green when the respective sequence belongs to the same phylogenetic clade as the probe and red when it belongs to a different clade. The threshold for a probe to target a sequence was set to ≤ 5% mismatch (blue line in graph).

***Polynucleotide probe design for hyn*L and *apr*A genes**

Analysis of the *hyn*L clone libraries (a total of 61 clones) revealed the presence of one OTU group at 99% identity cutoff (to eliminate sequencing errors) in the two *Bathymodiolus puteoserpentis* individuals analyzed. To maximize the gene detection efficiency, we prepared four polynucleotide probes targeting four regions of the *hyn*L gene (see Table 24), about 350 bp each. The probes had between 0 and 1.1% mismatches with their targets and their mix was termed *hyn*L-Mix. The probe sequences were further searched using BLAST against the nucleotide database from NCBI. No identities above 74% were found with genes that were not *hyn*L. Therefore, it is expected that the four polynucleotide probes will not bind outside the *hyn*L gene when hybridized at -25°C below their T_m.

Analysis of the *apr*A clone library (a total of 48 clones) prepared from one representative of the *Bathymodiolus puteoserpentis* species, revealed the presence of one OTU group at 99% identity. Only one probe, *apr*A-Bath, of 352 bp length and perfect match with the target, was prepared. Using BLAST against the nucleotide database from NCBI, no identities above 73% were fund outside the *apr*A gene. Therefore, it is expected that the *apr*A-Bath probe will not bind outside the *apr*A gene when hybridized at -25°C below its T_m.

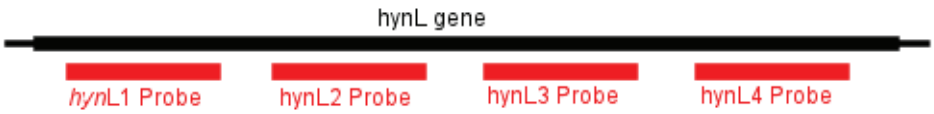


Figure 29: The *hyn*L probe mix.

3.2.2. Probe synthesis

The probes were synthesized by enzymatic incorporation of Dig labeled nucleotides either during *in vitro* transcription (in the case of ssRNA probes), or during probe synthesis PCR (in the case of dsDNA probes). The synthesis of ssRNA probes is depicted in Figure 30A, and had two steps: first, a PCR step, to obtain the T7 amplicon, which served as template in the next step – *in vitro* transcription. Before transcription, the template PCR products were purified by band extraction from agarose gels, to ensure probe specificity. To avoid binding to the corresponding mRNA, the ssRNA probes were synthesized as sense strands (the T7 promoter was at the forward end of the template). The synthesis of dsDNA probes had only a PCR step and is depicted in Figure 30B. The template for the PCR steps was represented either by *E. coli* cell lysates or by purified plasmid DNA. When synthesizing dsDNA probes, the best efficiency was obtained starting from purified plasmid DNA.

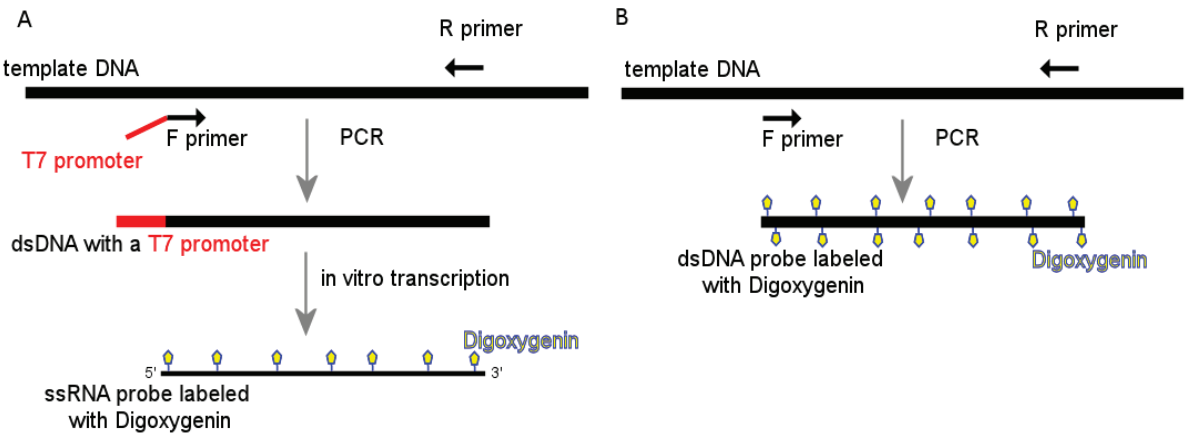


Figure 30: Polynucleotide probe synthesis. A. ssRNA probes were synthesized using as template a PCR amplicon which had a T7 promoter attached at the forward primer. *In vitro* transcription of this template resulted in the synthesis of the ssRNA (sense strand), with many Dig labels. **B. dsDNA probes** were synthesized by PCR, with Dig incorporation.

A very important step following the probe synthesis was probe purification with removal of labeled NTPs. Improper purification could lead to increased background during FISH procedures. Due to the high concentrations of NTPs used during the *in vitro* transcription step, the ssRNA probes were purified in two steps (see materials and methods), first by using of RNA-Easy spin columns (to remove the enzymes used during synthesis, digested DNA and NTPs) and then by using Sephadex columns (to remove any remaining NTPs). Because the concentrations of labeled NTPs used during synthesis of the dsDNA probes was much lower (~35 times lower) than during *in vitro* transcription, the probes were purified in one step only, using GeneClean spin columns (see materials and methods). After purification, the size and integrity of the probes was checked by gel electrophoresis (Figure 31 for ssRNA probes, Figure 32 and Figure 33 for dsDNA probes). Incorporation of Dig reduced the electrophoretic mobility of nucleic acids (Holtke and Kessler, 1990), so that they migrated slower than their unlabeled counterparts (see Figure 32 and Figure 33). This could be used as a confirmation for Dig incorporation.

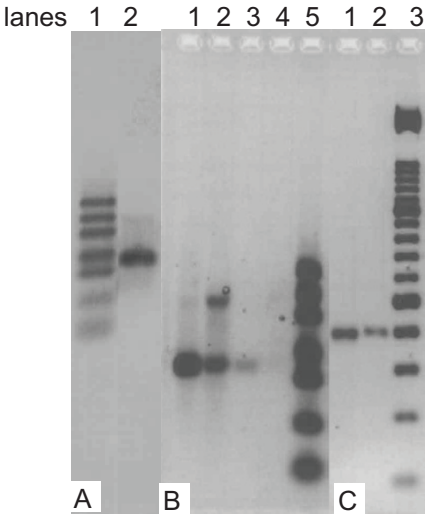


Figure 31: ssRNA probes (with Dig incorporation). **A.** lane 1 – ssRNA Ladder (Riboruler low range RNA ladder - Fermentas); lane 2 - *rpo*B ssRNA probe. **B.** lane 1, 3 - *dsrA* ssRNA probe; lane 2, 4 *hme*D ssRNA probe; lanes 1 and 2 were containing 10x more probe than lanes 3 and 4; lane 5 – ssRNA Ladder (Riboruler low range RNA ladder - Fermentas); **C.** lane 1 – T7-*dsrA* dsDNA (template for probe synthesis); lane 3 – T7-*hme*D dsDNA (template for probe synthesis); lane 3 – dsDNA ladder (100 bp DNA ladder, Roche).The Riboruler low range RNA ladder has the following bands: 100, 200, 300, 400, 600, 800 and 1000 nt. The 100 bp DNA ladder has bands once at each 100 bp and the bands for 500 bp and 1000 bp are thicker, for increased visibility.

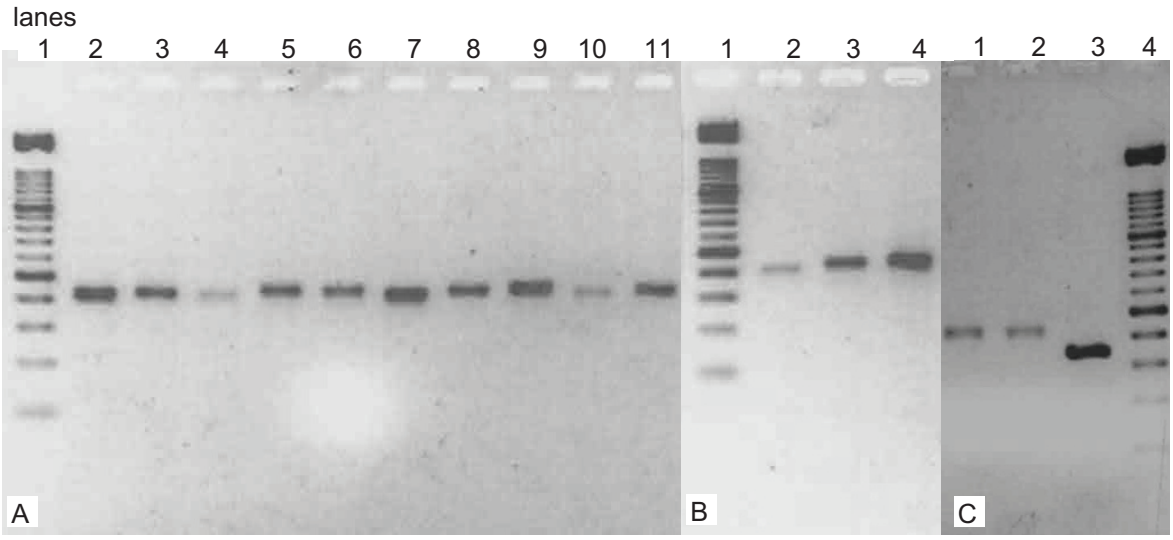


Figure 32: dsDNA probes - *amo*A-Nam probe mix (with Dig incorporation). A. lane 1 - dsDNA ladder (100 bp DNA ladder, Roche); lane 2-11 - *amo*A individual probes; **B.** lane 1 - dsDNA ladder (100 bp DNA ladder, Roche); lanes 2, 3 - *amo*A individual probes; lane 4 - NonPolyPr350 dsDNA probe. **C.** lanes 1, 2 - *amo*A individual probes; lane 3 – *amo*A target (without Dig incorporation); lane 4 - dsDNA ladder (100 bp DNA ladder, Roche). The 100 bp DNA ladder has bands once at each 100 bp and the bands for 500 bp and 1000 bp are thicker, for increased visibility.

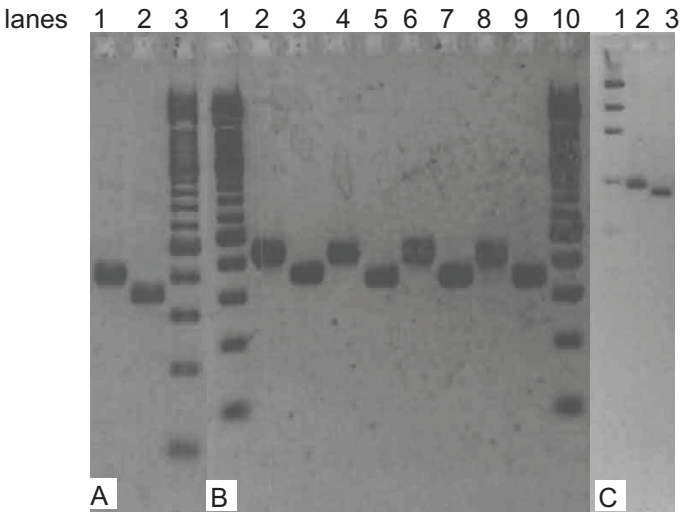


Figure 33: dsDNA probes (with Dig incorporation) and targets (without Dig incorporation). A. *amo*A-1E3. lane1 - dsDNA probe; lane2 - dsDNA target; lane 3 - dsDNA ladder (100 bp ladder, Roche). **B. *hyn*L.** lanes 1, 10 - dsDNA ladder (100 bp DNA ladder, Roche); lanes 2, 4, 6, 8 - dsDNA probes; lanes 3, 5, 7, 9 – dsDNA targets. **C. *FOS-rdsrA*.** lane 1 - dsDNA ladder (low mass DNA ladder, Invitrogen); lane 2 – dsDNA probe; lane 3 – dsDNA target. The 100 bp DNA ladder has bands once at each 100 bp and the bands for 500 bp and 1000 bp are thicker, for increased visibility. The low mass DNA ladder has the following bands: 100, 200, 400, 800, 1200 and 2000 bp.

3.2.3. Determination of the stringency parameters for hybridization with polynucleotide probes

The stringency parameters for hybridization included the composition of hybridization and washing buffers (Na^+ and formamide concentration), as well as hybridization and washing temperatures. The ideal parameters should allow for: (i) specific hybridization, (ii) optimum hybridization rate, and (iii) preservation of cellular morphology (low temperatures).

Determination of the stringency parameters required the knowledge of the melting temperature (T_m) of the hybrid formed between probe and target. Two methods were used to determine T_m: theoretical calculations and *in vitro* measurements. For the former, formulae implemented in PolyPro software were used to calculate the T_m. For *in vitro* measurements, a protocol for T_m determination with dsDNA binding dyes and real-time fluorescence detection (Monis et al., 2005; Gudnason et al., 2007) was modified and used.

In this *in vitro* method, the fluorescence conferred by the intercalating dye Syto9 is high when the DNA is double-stranded, while fluorescence levels drop when the DNA melts and becomes single-stranded. The T_m appears as a distinct peak when the negative of the first derivative of fluorescence is plotted versus temperature (see Figure 37). We modified

the original protocol to measure the T_m in hybridization-like and washing-like buffers. These buffers had the exact composition of the hybridization and washing buffers used for gene detection, except that they did not contain the blocking reagents (see materials and methods). The original protocols were performing the T_m measurements in PCR buffers, which contained only salts in low concentrations and a pH buffer. On the other hand, the hybridization-like and washing-like buffers contained chemicals which could interfere with the intercalating dye and consequently, with T_m detection. Therefore, the first thing that needed to be established was if T_m can be measured with Syto9 when chemicals like SDS and formamide or high salt concentrations are present in the buffer.

Initial method development

The first experiments were performed with an *apr*A dsDNA fragment, obtained by PCR from a fosmid (courtesy of Anke Meyerdierks). Addition of 0.1% SDS to a buffer containing 20 mM Tris-HCl, 75 mM NaCl produced no effect on the T_m values, the melting having two peaks, one at 87.7±0.1°C and the other at 88.9±0.1°C. However, it produced a significant decrease in the peak heights, of 5.6 and respectively, 5.0 folds. As expected, addition of 50% formamide decreased the T_m with about 0.6°C per 1% formamide (from 87.4±0.1°C and 88.7±0.1°C at 0%, to 55.4±0.3°C and 57.3±0.3°C at 50%), which is in the range for the formamide effect on the T_m of dsDNA (McConaughy et al., 1969; Blüthmann et al., 1973; Casey and Davidson, 1977; Hutton, 1977). As SDS, the formamide decreased the peak height, but the effect was less strong: 1 to 1.8 folds decrease. Both at 0% formamide and at 50% formamide, the melting showed 2 peaks (Figure 34A), which indicate the presence of multiple melting domains. *In silico* simulations of the melting using either MeltSIM (Blake et al., 1999) or Poland (Steger, 1994) software also revealed the presence of multiple melting domains - two and four, respectively (Figure 34 B and C). Therefore, these results confirmed the presence of the two melting peaks and made it unlikely that they were an artifact of the measurement method. As reported previously in studies measuring T_m with Syto9 in PCR buffers (Rasmussen et al., 2007), there was a difference between the measured T_m and the predicted T_m.

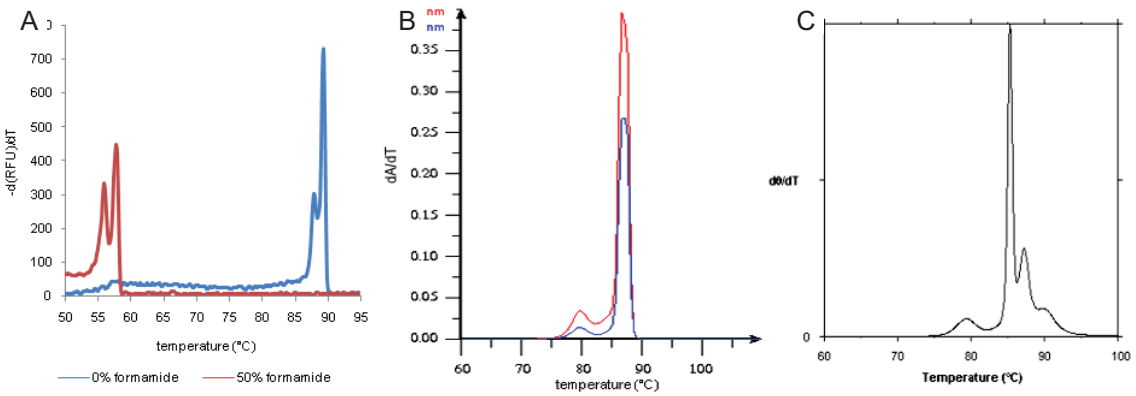


Figure 34: Melting of *apr*A dsDNA fragment (fosmid derived). A. Melting measurements with Syto9 dye in a buffer containing 75mM NaCl and no formamide (blue) or 50% formamide (red). B. *In silico m*elting simulation with Poland software (parameters: 75 mM Na^+, Blake and Delcourt). C. *In silico m*elting simulation with MeltSIM software (parameters: 75 mM Na^+). On the y axis is represented the change in dsDNA concentration with temperature, the dsDNA being expressed either in relative fluorescence units or in hypochromicity units.

The next test was to measure the T_m in a buffer with all chemicals together – SDS, formamide and high salt concentrations. This was the hybridization-like buffer (35% formamide, 1718 mM Na^+ and 0.1% SDS, see materials and methods) and the tests were done with *hyn*L1 probe and its target and probe-target hybrid (Table 28). Because these chemicals were producing a decrease in the peak height, various Syto9 concentrations were tested: 2 µM, 10 µM and 50 µM. At 2 µM the T_m peaks were very low or not detectable. The best defined peaks were obtained with 10 µM and 50 µM (Figure 35). Syto9 was the dye of choice for measuring T_m because, as compared with other dyes, it produces only a small increase in T_m with increasing dye concentration (Monis et al., 2005; Gudnason et al., 2007). Monis et al., 2005, for measurements performed in PCR buffer, reported a 2°C increase in T_m when the Syto9 concentration was increased from 2 µM to 33 µM. Our measurements in hybridization-like buffer showed no significant difference in T_m between 2 µM and 10 µM and a slight decrease, of less than 0.5°C, at 50 µM Syto9 (Figure 36). For further experiments, the 10 µM concentration was chosen.

The absence of T_m increase with increasing Syto9 concentration was probably due to the composition of the hybridization-like buffer. It is known that, at high salt concentrations, for e.g. above 0.7 M for buffers with 35% formamide (Hutton, 1977), the T_m is reaching a plateau and then starts to decrease (Hamaguchi and Geiduschek, 1962; Schildkraut and Lifson, 1965; Gruenwedel et al., 1971; Hutton, 1977). In the hybridization-like buffer the Na^+ concentration is ~1.7 M and the T_m is already in a plateau. Further addition of µM amounts of Syto9, which has a salt-like effect (Bjorndal and Fygenson, 2002), will not increase the T_m anymore. The slight decrease in T_m at 50 µM Syto9 suggests that addition of Syto9 will further destabilize the DNA duplex. On the other hand, Monis et al., 2005, have measured

the T_m in PCR buffers, where the salt concentration was low and small additions of Syto9 had an influence on T_m. The same effect is expected to happen in washing-like buffers, were the total Na^+ concentration is 23 mM, resulting in a T_m overestimation of about 3-5 °C (Rasmussen et al., 2007).

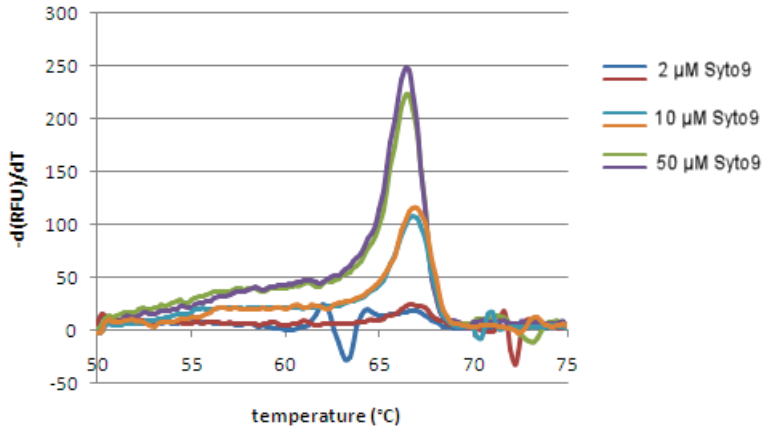


Figure 35: The effect of Syto9 dye concentration on the T_m peak height and shape. These T_m curves have been measured for the *hyn*L1 dsDNA probe. It can be noticed that an increase in Syto9 concentration produces an increase in peak height, without a significant variation in T_m.

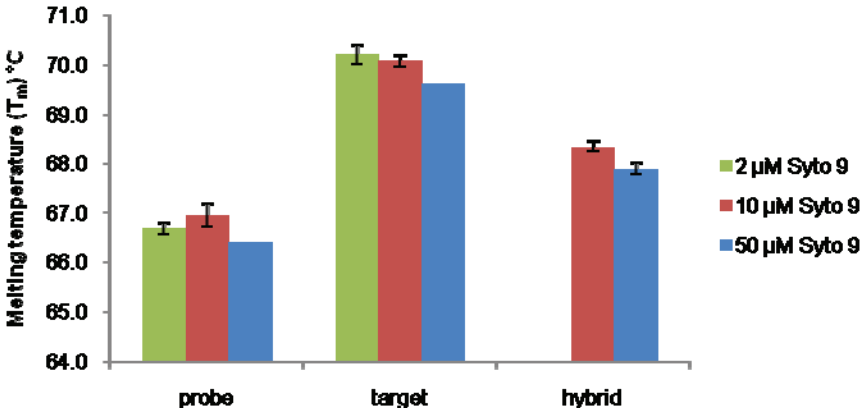


Figure 36: Variation of the melting temperature with the concentration of Syto9 dye. The T_m of *hyn*L1 dsDNA probe, target and hybrid were measured in hybridization-like buffer (35% formamide). The T_m did not vary much with the dye concentration, whose increase to 50 µM lead to a decrease in the T_m of ~ 0.5°C.

The initial tests showed that Syto9 could be used for measuring T_m in buffers containing formamide, SDS and high salt concentrations. The next step was to develop a strategy to measure the T_m of the hybrid formed between the probe and the target. Because the probe was Dig-labeled, while the target was not, the probe-target hybrid would have a strand with Dig and the other without. To measure the T_m of such a hybrid, the thermal protocol was modified to consist of denaturation, hybridization and subsequent melting. The

first two steps were necessary to promote the formation of hybrid molecules. The probe, target and probe-target hybrid were measured in individual reactions. As expected, they had different T_m, with that of the probe being the lowest and of the target the highest. The T_m of the hybrid was intermediary between the ones for probe and target, and its peak could be recognized by comparison with the peaks from only target or only probe reactions (Figure 37). This strategy for hybrid formation worked only in hybridization-like buffers. In the case of washing-like buffers, no hybrid peak was detected, due to the low Na^+ concentration, unfavorable for hybridization. Therefore, in the washing-like buffer only the T_m of the probe and the target was determined.

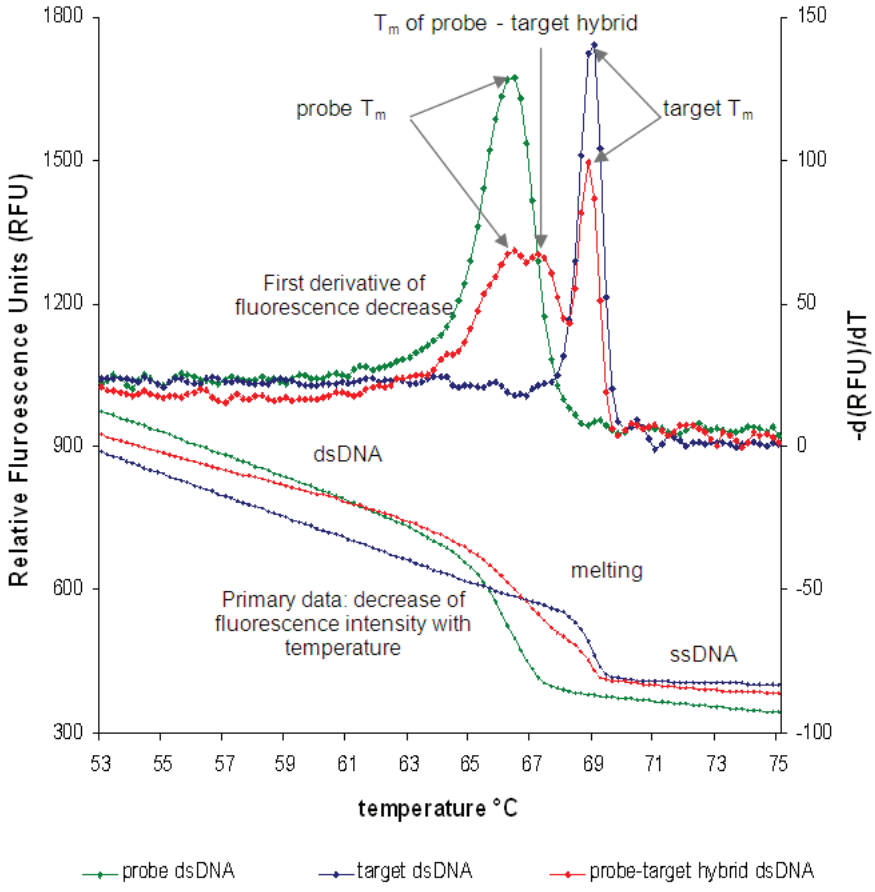


Figure 37: *In vitro* measurements of the melting temperature (T_m) for the *amo*A-1E3 probe (green) (both strands with Dig), target (blue) (neither of the strands with Dig) and hybrid (red) (one strand with Dig and one without). The measurements were performed in hybridization-like buffer, with 35% formamide and 1.718 M Na^+. The primary data represented the fluorescence (expressed as relative fluorescence units-RFU) decrease with increasing temperature. The first derivative showed the T_m as a peak. In the hybrid reaction, three T_m peaks appeared: one peak for the probe duplex, one peak for the hybrid duplex and one peak for the target. To identify the peaks, comparison with the individual reactions for probe and target was necessary.

***Determination of the stringency conditions for amo*A probes**

The crenarchaeotal *amo*A genes are so distant even from their closest bacterial homologue (Dunfield et al., 2007) that it is very difficult to obtain a good alignment at the nucleic acid level. The number of mismatches to our crenarchaeotal probes was well beyond those 20-30% mismatches reported by Wetmur (1991) to be the limit for hybridization. Hence, no binding was expected outside the crenarchaeotal *amo*A clade for both the *amo*A-1E3 and *amo*A-Nam probes. Therefore, the only criteria for choosing the hybridization temperature and composition of hybridization buffer were maintenance of cellular morphology and of an optimum hybridization rate.

To avoid degradation of the cellular morphology, the hybridization temperature had to be in the range 42-50°C. To have an optimum hybridization rate, the hybridization temperature was set to ~ 25°C below the T_m of the perfectly matched hybrid, according with previous findings for DNA:DNA hybrids (Wetmur, 1991). As a result, a hybridization buffer which was likely to give a T_m in the range 67-75°C was needed. To find such a hybridization buffer, initial calculations were performed by keeping the Na^+ concentration constant (at high levels, to promote hybridization kinetics) and varying the formamide concentration. A buffer with 35% formamide (see materials and methods) was found to give T_m in the desired temperature range, and therefore, it was used for further measurements.

For the *amo*A-1E3 probe, we compared the T_m of (i) probe DNA, labeled with Dig, (ii) target DNA, without Dig, and (iii) probe-target hybrid, which had one strand with Dig but the other strand without. In the hybridization-like buffer, the measured T_m were 66.2 °C for probe, 68.9 °C for target, and 67.2 °C for hybrid, respectively (Figure 37). The T_m difference between probe and target was about ~2.7°C. In washing-like buffer, the measured T_m were higher: 71.9 ± 0.4 °C and 75.2 ± 0.1 °C for probe and target, respectively.

In comparison, the T_m calculated for the target DNA were 73.4 °C in hybridization-like buffer and 69.9°C in washing like buffer. For *amo*A-Nam probe mix, the mean values for the calculated T_m of the perfectly matched targets were 73.4 °C in hybridization-like buffer and 69.9°C in washing like buffer (Table 27), similar to those calculated for *amo*A-1E3. To calculate the T_m of the mismatched hybrids, the same formula as for perfectly matched hybrids was used, with the addition of the term for mismatches. It was considered that 1% mismatch will give a 0.5-1.5°C decrease in the T_m, compared with the perfectly matched hybrid (Anderson and Young, 1985). Therefore, the calculated T_m for 5% mismatched hybrids were in the range 65.2 - 71.9 °C for hybridization-like buffer and 61.7 - 68.4 for washing like buffer.

The two methods used here to estimate T_m gave slightly different results, which was expected, as the theoretical calculations will only provide an estimation. Previous studies which measured the T_m in PCR buffer have found that the measured T_m is 3-5°C higher than

the estimated one (Rasmussen et al., 2007). For *amo*A-1E3 probe, in hybridization-like buffer the measured T_m was 3.4°C lower than the calculated one, while in washing-like buffer the measured T_m was 5.2°C higher than the calculated one. Due to the saturation effect that high salt concentrations have on T_m, it is expected that the *in vitro* measurements gave a good T_m in hybridization-like buffer. On the other hand, the low salt concentration in the washing-like buffer most probably led to a T_m overestimation due to the use of Syto9. While keeping in mind that both the measured and the calculated values did not give the exact T_m, this range of values can still be used to determine the parameters for hybridization.

Based on the above T_m values, the denaturation temperature, hybridization temperature and washing temperature of the *amo*A-1E3 probe were determined. The denaturation temperature was 75°C, chosen to be above the T_m of the probe and target duplexes in hybridization-like buffer. The hybridization temperature was calculated as follows: 67.2°C (the measured T_m of the hybrid in hybridization-like buffer) – 25°C (for optimum hybridization rate) = 42°C. The washing temperature was 42°C, chosen to be below the T_m in washing-like buffer.

The washing step after hybridization is important to melt short hybrids that might form during hybridization (Wetmur, 1991) between parts of the probe and random genomic DNA fragments. For this, a low salt buffer (washing buffer II) was used, as detailed in materials and methods. The choice for the washing temperature was made according to the T_m values for the washing-like buffer and it was lower than the T_m, to avoid melting of the *amo*A probe-target hybrids (this had to consider both perfectly matched and mismatched probe-target hybrids).

For the *amo*A-Nam probe mix the same parameters as for *amo*A-1E3 probe were used. Because the *amo*A-Nam probe mix was targeting alleles up to 5% mismatches, it was obvious that the choice of the hybridization temperature could not be the optimum value for all hybrids. Assuming a 0.5°C decrease in T_m for 1% mismatch, the deviation around the optimum would had been between -1.95°C and +1°C, while assuming a 1.5 °C decrease in T_m for 1% mismatch, the maximum deviation around the optimum would had been between -4.45°C and +1°C. This is placing many of the hybrids rather on the stringent side of the hybridization curve, but still close to their optimum hybridization rate. The washing temperature of 42°C was below the T_m of all probe-target hybrids.

Table 27: Calculated T_m for the polynucleotides forming the *amo*A probes. The T_m was calculated for the composition of the hybridization-like buffer (HB) (1718 mM Na^+ and 35% formamide) and the washing-like buffer II (WB) (23 mM Na^+).

Probe	%GC	Length	Perfectly matched targets		5% mismatched targets			
					T_m (°C) in HB		T_m (°C) in WB	
			T_m (°C) in HB	T_m (°C) in WB	0.5 °C per 1% mismatch	1.5 °C per 1% mismatch	0.5 °C per 1% mismatch	1.5 °C per 1% mismatch
***amo*A-1E3**	41.9	351	73.4	69.9	70.9	65.9	67.4	62.4
***amo*A clib1-pl1 cl10**	43.6	351	74.1	70.6	71.6	66.6	68.1	63.1
***amo*A clib1-pl1 cl34**	40.5	351	72.8	69.4	70.3	65.3	66.9	61.9
***amo*A clib1-pl1 cl53**	44.2	351	74.4	70.9	71.9	66.9	68.4	63.4
***amo*A clib1-pl1 cl54**	41.9	351	73.4	69.9	70.9	65.9	67.4	62.4
***amo*A clib1-pl1 cl56**	43.0	351	73.9	70.4	71.4	66.4	67.9	62.9
***amo*A clib1-pl1 cl59**	41.3	351	73.2	69.7	70.7	65.7	67.2	62.2
***amo*A clib1-pl1 cl75**	41.0	351	73.0	69.6	70.5	65.5	67.1	62.1
***amo*A clib1-pl2 cl106**	41.6	351	73.3	69.8	70.8	65.8	67.3	62.3
***amo*A clib1-pl2 cl145**	40.7	351	72.9	69.5	70.4	65.4	67.0	62.0
***amo*A clib2-pl2 cl66**	40.2	351	72.7	69.2	70.2	65.2	66.7	61.7
***amo*A clib2-pl3 cl105**	42.7	351	73.7	70.3	71.2	66.2	67.8	62.8
***amo*A clib3-pl1 cl74**	40.7	351	72.9	69.5	70.4	65.4	67.0	62.0
***amo*A-Nam mix**			ΔT_m =1.7	ΔT_m =1.7				
			Average = 73.4	Average = 69.9	Average = 70.9	Average = 65.9	Average = 67.4	Average = 62.4
					Min = 70.2	Min = 65.2	Min = 66.7	Min = 61.7
					Max = 71.9	Max = 66.9	Max = 68.4	Max = 63.4

***Determination of the stringency conditions for hyn*L**

To detect the *hyn*L gene, four polynucleotides targeting different regions of the gene were used in the same hybridization mixture. Therefore, the same rule applied as for all probe mixes: finding hybridization conditions that will be as close to optimum as possible for each of the polynucleotides. In this case, the problem was simplified by the fact that no mismatched alleles were targeted, and the only consideration in probe design was the ΔT_m of different probes.

For the *hyn*L-mix, the T_m was measured in hybridization-like buffer (35% formamide) for all four probes and their targets (Table 28). The T_m of the hybrid was determined only for *hyn*L1 probe. The measured T_m of the probes varied between 63.3°C and 66.3°C (ΔT_m=3°C), while that of the targets varied between 67.3°C and 70.2°C (ΔT_m=2.9°C). The *hyn*L1 probe-target hybrid had a T_m of 68.3°C. For the targets, the calculated T_m in hybridization-like buffer varied between 72.4°C and 74.9°C, and in washing-like buffer varied between 68.9°C and 71.4°C, with a ΔT_m of 2.6°C (Table 28). The denaturation temperature was chosen to be 75°C, above the T_m of all four *hyn*L probes and targets. The hybridization temperature was 42°C, which, considering the T_m of *hyn*L 1 as the highest from all four hybrids and a ΔT_m=3°C, was between 23.3 and 26.3°C below the T_m of the hybrids. This is placing the hybridization temperature in the optimum interval for kinetics. The washing temperature was also 42°C, much lower than the T_m of the hybrids. In these hybridization conditions, the *hyn*L-mix is expected to be specific for the *hyn*L gene.

Table 28: Calculated and measured T_m for the polynucleotides forming the *hyn*L probe mix. The T_m was calculated for the composition of the hybridization-like buffer (HB) -1.718 M Na^+ and 35% formamide, and of the washing-like buffer II (WB) - 0.023 M Na^+.

		Measured T_m (°C) in HB	calculated T_m (°C)		%GC	Length (bp)
			in HB	in WB		
***hyn*L1**	probe	66.3 ± 0.1			45.5	356
	target	70.2	74.9	71.4		
	hybrid	68.3				
***hyn*L2**	probe	64.2 ± 0.4			39.3	351
	target	68.1 ± 0.1	72.4	68.9		
***hyn*L3**	probe	63.3 ± 0.1			39.5	349
	target	67.3 ± 0.1	72.4	69		
***hyn*L4**	probe	65.1 ± 0.1			43.1	350
	target	69.2	73.9	70.4		
***hyn*L-mix**		Average probes = 64.7	Average = 73.4	Average = 69.9		
		Average targets = 68.7				
		Min probes = 63.3	Min = 72.4	Min = 68.9		
		Min targets = 67.3				
		Max probes = 66.3	Max = 74.9	Max = 71.4		
		Max targets = 70.2				
		ΔT_m probes = 3	ΔTm = 2.6	ΔTm = 2.6		
		ΔT_m targets = 2.9				
		Denaturation temperature = 75 °C				
		Hybridization temperature = 42 °C				
				Washing temperature = 42 °C		

***Determination of the stringency conditions for apr*A probe**

For *apr*A-Bath probe, the T_m was both measured and calculated in hybridization-like buffer (35% formamide) and washing-like buffer (Table 29). The measured T_m of the hybrid was 68°C and the hybridization temperature, 42°C, was in the optimum range for kinetics, about 26°C below T_m. In washing-like buffer, the measurements showed the existence of three melting domains for the probe and two for the target. The washing temperature (42°C) was below the T_m of all melting domains. In these conditions, the *apr*A-Bath probe should bind only to the *apr*A gene.

Table 29: Calculated and measured T_m for the *apr*A probe. The T_m was calculated for the composition of the hybridization-like buffer - HB (1.718 M Na^+ and 35% formamide) and of the washing-like buffer II - WB (0.023 M Na^+).

	%GC	Length (bp)	measured T_m (°C)				Calculated T_m (°C)	
			in HB	in WB			in HB	WB
				Peak 1	Peak 2	Peak 3		
probe			67±0.1	71.6±0.2	72.4±0.0	75±0.3		
target	43%	352	69.6±0.2	75±0.1	77.8±0.2		73.9	70.4
hybrid			68±0.1					
	Denaturation temperature = 75°C							
	Hybridization temperature = 42°C							
	Washing temperature = 42°C							

***Determination of the stringency conditions for FOS-rdsrA* probe**

The FOS-*rdsr*A probe was used to target alleles with 7.8% mismatches. Assuming a 1°C decrease in T_m for 1% mismatch, the calculated T_m was 68.8°C for a 45% formamide hybridization buffer and 71.1°C for washing buffer. The hybridization temperature was 42°C, about 26°C below the T_m of the mismatched hybrid. Two washing temperatures were tested: 42°C and 50°C.

***Determination of the stringency conditions for ssRNA* probe**

The stringency condition for ssRNA probes were estimated using only the T_m calculations. The calculated T_m in 50% formamide hybridization buffer were 81°C, 84.4°C and 85°C, while in the washing buffer were 80.5°C, 84°C and 84.6°C, for *rpo*B, *dsr*A and *hme*D probes, respectively. This places the optimum hybridization temperature above 60°C, since the for RNA:DNA hybrids the optimum is at 15-20°C below T_m (Birnstiel et al., 1972). The washing temperature was either 60°C or 70°C.

3.2.4. Gene detection using the geneFISH protocol

The main steps of the geneFISH protocol are illustrated in Figure 38. They consisted of two major parts: rRNA CARD FISH, followed by gene detection with ssRNA or dsDNA polynucleotide probes. The rRNA CARD-FISH part was performed as detailed elsewhere (Pernthaler et al., 2002a). This step allowed identification of single cells via hybridization with an rRNA targeted oligonucleotide probe. The protocol for gene detection was adapted from previously developed protocols for mRNA FISH (Wagner et al., 1998; Pernthaler and Amann, 2004). It was based on polynucleotide hybridization and included two steps of signal amplification (Figure 39). It started with the binding of the polynucleotide probe, which had multiple Dig labels. Antibodies conjugated with HRP were then added to bind to the Dig molecules. This was followed by a final amplification step, where the antibody bound HRP catalytically deposited many fluorescently labeled tyramides.

Initial calculations indicated that the signal amplification system used in geneFISH may deposit between 460 and 2000 fluorochromes per probe: a 300-400 nucleotide probe may have 6-16 Dig labels (Holtke and Kessler, 1990; Yu et al., 1994), each of which can bind an antibody conjugated with ~ 3 HRP molecules and each HRP molecule will in turn deposit between 26 to 41 fluorescently labeled tyramides (Hoshino et al., 2008). Previous studies using multiple Cy3 labeled probes (Niki and Hiraga, 1998; Viollier et al., 2004) indicated that around 200-400 fluorochromes in one localized spot were sufficient for the signal to be detected. Therefore, the signal amplification system used here should allow visualization of single targets.

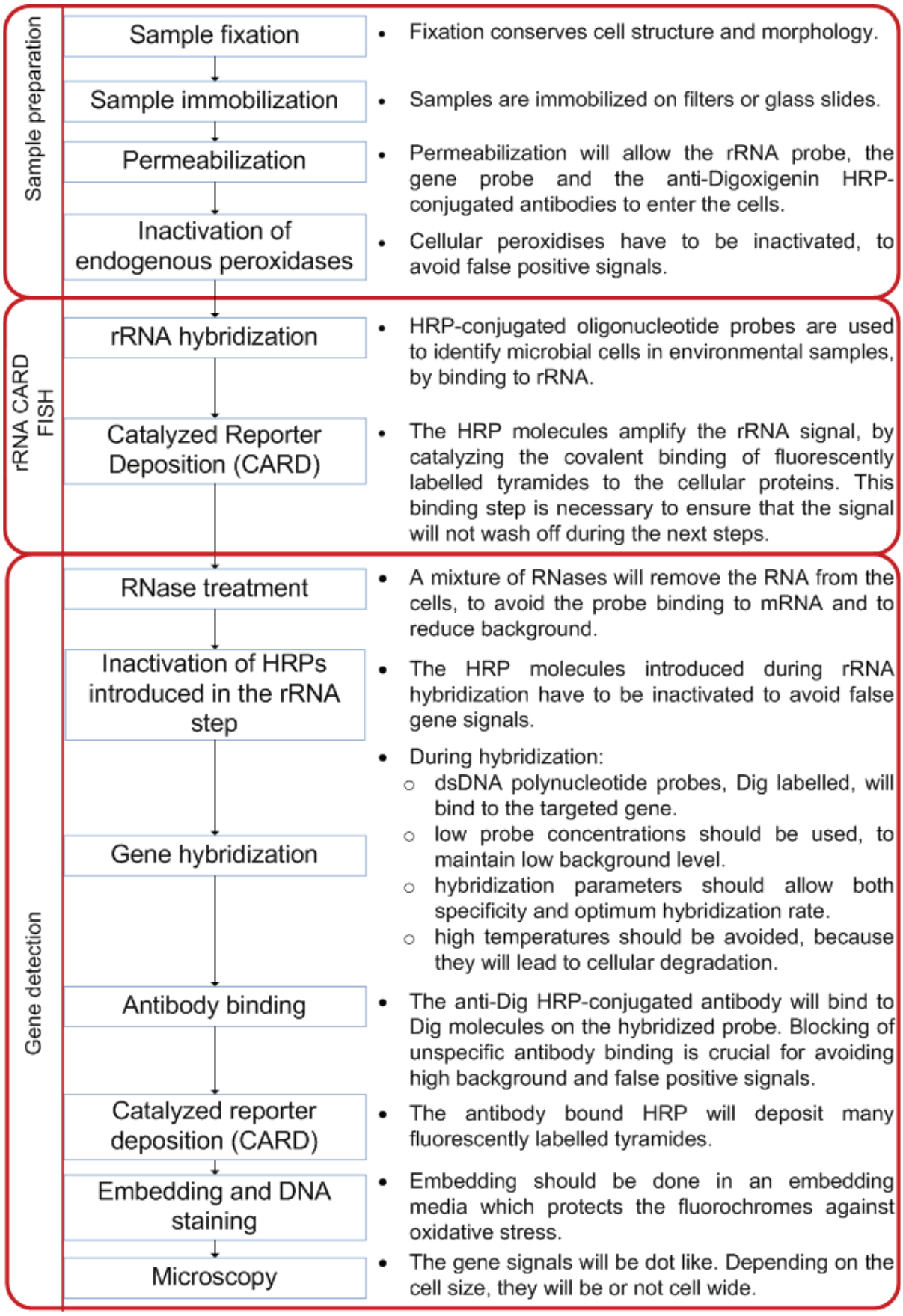


Figure 38: GeneFISH protocol

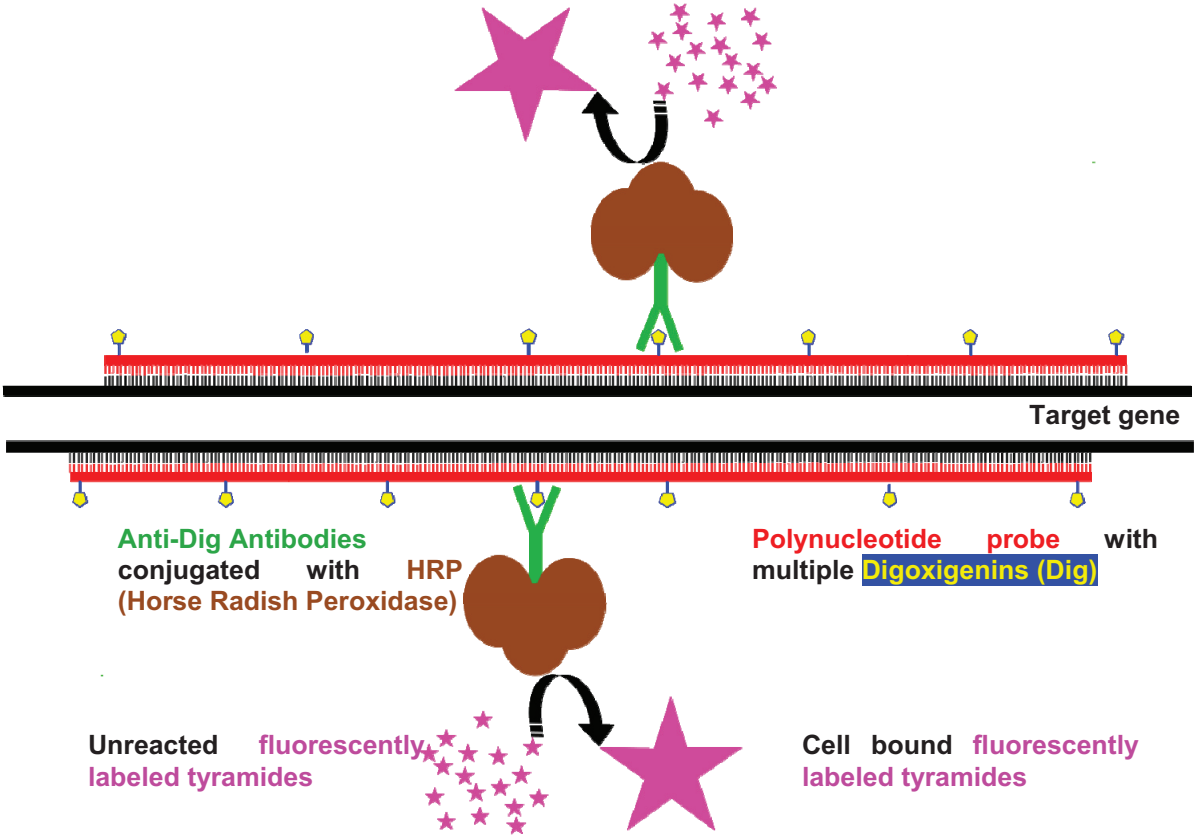


Figure 39: Schematic representation the geneFISH signal amplification.

Development of geneFISH protocol - Escherichia coli experiments

The initial geneFISH development was done with Dig labeled ssRNA probes. To avoid binding to mRNA and to ensure thus that the signal obtained was coming from the genes, the probe represented the sense strand. The first tests targeted the chromosome encoded *rpo*B gene in *E. coli* K12, with a single 359 nt polynucleotide probe (Table 30). The signals obtained were dot like, covering only part of the cells, as opposed to the ones usually obtained after rRNA FISH, which are cell wide. These signal characteristics made geneFISH very sensitive to background signals, which normally would be ignored in regular rRNA FISH. Introduction of an RNase digestion step previous to gene hybridization reduced the percentage of false positives from 10.5% to 3%, indicating that the ssRNA probe was giving background binding to rRNA. This phenomenon accompanying some ssRNA probes is known from Northern Blot hybridizations (Kulesza and Shenk, 2006; Flynt et al., 2009), and, given the amount of rRNA in the cells, compared with that of single genes, can have an important contribution to the background level, especially with high %GC probes. To avoid

probe degradation, RNAseI, a single strand endoribonuclease that can be inactivated by heat and SDS and cuts after each four bases (Meador et al., 1990) was used. The detection efficiency after the RNase step was 14%, with 3% false positives in the double digestion (RNase and DNase) control, when hybridized at 42°C in a 50% formamide hybridization buffer (see materials and methods) and washed stringently at 70°C.

These first experiments were important for establishing the main steps of the protocol. The initial steps were common for most of the FISH protocols: sample fixation and immobilization on solid support. Permeabilization was necessary to allow high molecular weight molecules (polynucleotides, HRP-conjugated antibodies, HRP-conjugated oligonucleotide probes) to diffuse into the cells. The rRNA CARD-FISH step had to be performed before the rRNA digestion step, and consequently, before gene hybridization. rRNA CARD-FISH was preferred over the standard rRNA FISH, because the CARD is fixing the signal inside the cells. Otherwise, the signal would had been lost during the more stringent gene hybridization step. On the other hand, this introduced an additional step in the protocol: inactivation of the HRP molecules introduced with the rRNA probes. This step was critical for avoiding false positive signals and had to be tested thoroughly before geneFISH. Different inactivation experiments indicated that the probe attached HRP needs a much stronger inactivation than is usually required for endogenous peroxidases. There are no standardized protocols for fixation, permeabilization and peroxidase inactivation. They differ with the sample type and with the microorganisms targeted. Different protocols for fixation, permeabilization and peroxidase inactivation have been previously described and discussed elsewhere (Pernthaler et al., 2002a; Pernthaler and Amann, 2004; Pernthaler et al., 2004; Pavlekovic et al., 2009).

Because single gene hybridizations are generally limited kinetically, different buffers known to increase the hybridization kinetics were tested. Two of them were commercially available – UltraHyb Ultrasensitive Hybridization Buffer and Northern Max Hybridization Buffer (see Appendix D). The composition of the commercial buffers was not available. There have been, however, reports about chemical agents, for e.g. CTAB, which can greatly increase the hybridization kinetics (Pontius and Berg, 1991). The physical appearance of the commercial buffers, especially that of UltraHyb, was an indication that the buffer was containing CTAB or a similar detergent. The third hybridization buffer was self made, with 1718 mM Na^+ and 50% formamide (see materials and methods). No signals were obtained with the Northern Max buffer, while the UltraHyb buffer gave a high background level. Moreover, the use of these two buffers resulted in a much degraded cellular morphology, most probably due to the high concentrations of detergents. In the end, the buffer of choice was the self made one, resulting in a lower background level and a less cell degradation.

To increase hybridization kinetics, the hybridization buffer contained a high salt concentration (1718 mM Na^+, see materials and methods) and 10% dextran sulfate, to promote high hybridization rates. The Na^+ concentration was higher than the one used in the mRNA FISH protocols: 1050 mM (Pernthaler and Amann, 2004) and 75 mM (Wagner et al., 1998). Blocking reagents (sheared salmon sperm DNA, yeast RNA, protein based blocking reagent) were included to minimize unspecific probe binding to different surfaces. Sodium dodecyl sulfate (SDS) was added as a denaturant, to help remove proteins from the chromosomal DNA and to facilitate probe diffusion by permeabilizing the cell walls. Ethylenediaminetetraacetic acid disodium dihydrate (EDTA) was added to chelate divalent cations, both to inactivate contaminant DNases and to control the stringency of hybridization. Formamide was used in order to decrease the temperature at which nucleic acids hybridize, since high temperatures are detrimental for cell integrity and morphology.

The posthybridization washes were similar to the ones used in Southern /Northern Blots (Sambrook, 2001) and included washing buffer I (WBI) with 2x SSC (390 mM Na^+) and 0.1% SDS, used to remove unbound probe molecules (Wetmur, 1991), and washing buffer II (WBII) with 0.1x SSC (19.5 mM Na^+) and 0.1% SDS, used as stringent washing. For probes with higher %GC (55-65 %GC) or for cases when discrimination between similar alleles is required, the stringency of the WBII can be increased by formamide addition. Alternatively, the washing temperature can be increased.

The antibody binding step is known to be the most responsible for high background levels. Different blocking reagents were tested: Max Block, casein and Western Blocking Reagents. The best were the last two, which gave similar results. However, the blocking can be sample depended, so, when necessary, other blocking reagents can be tested, either alone, or in combination, e.g. preimmune serum, bovine serum albumin and detergents.

To increase the efficiency of the CARD step, the concentration of the dextran sulfate in the amplification buffer was doubled for the gene amplification step, compared to the rRNA amplification step, as indicated by previous studies (Kubota et al., 2006). The tyramides used in the rRNA amplification step were labeled with $Alexa_{488}$, a green fluorochrome, while the ones used in the gene amplification step were labeled with $Alexa_{594}$ (see materials and methods). The use of these two dyes, together with the proper microscopy filter sets, avoids signal overlap in the two fluorescence channels. This is important, because leaking of the rRNA signal in the gene channel would increase the background fluorescence level, leading to decreased detection efficiency. Also important is the use of $Alexa_{594}$ for the gene signal, because it is a much stronger dye than $Alexa_{488}$, adding to the detection sensitivity of the method. For microscopes which offer the possibility of light intensity regulation, it is recommended to use low light levels, since it was noticed

that they perform better for gene signals detection, due to a decrease in the general fluorescence background and in the bleaching of the fluorochromes.

Following the experiment with the *rpo*B gene, a second experiment was performed, to test if the use of more than one polynucleotide probe is increasing the detection efficiency. The test system consisted of three *E. coli* clones, containing single copy fosmids (Table 30). The targeted genes were *dsrA* and *hme*D, the probes being perfect match with WS 39F7 clone (Mussmann et al., 2005) and mismatched (19% *dsr*A and 17.6% *hme*D) when compared with WS 7F8 clone (Mussmann et al., 2005). The third clone, NK54, contained no *dsrA* and *hme*D alleles and it was used as negative control. Two types of hybridizations were performed: using only *dsrA* probe (Figure 40), or, using both *dsrA* and *hme*D probes. As expected, the use of two polynucleotide probes having different target regions led to an increase in the detection efficiency (from ~18% when using only one probe, to ~28% when using two probes). On the other hand, the increase was not a doubling of the detection efficiency, probably because in many cases both probes bound in the same cell. This indicates that, to increase the hybridization efficiency to 100%, a large number of polynucleotide probes with different target regions would have to be used. The level of false positives was between 3-5%, without any significant differences between the negative control and the mismatched clone (WS 7F8), indicating that the percent mismatch between the probe and the target was too high for gene signals to be detected in these hybridization conditions.

Table 30: Overview of geneFISH experiments with ssRNA probes.

Sample	geneFISH						
	Target gene	Targets	Probes	%GC probe	Length probe (nt)	Probe-target identity	geneFISH detection
Chromosome localization							
Escherichia coli K12	*rpo*B	1	1	53%	359	100%	14%
DNase/RNase control							3%
Fosmid localization							
Escherichia coli clone WS 39F7	*dsr*A	1	1	*dsrA*: 57% *hmeD:* 58%	*dsrA*: 365 *hmeD:* 370	100%	18.3%
	*hme*D + *dsr*A	2	2			*hmeD:* 100% *dsrA*: 100%	27.8%
Escherichia coli clone WS 7F8	*dsr*A	1	1			81%	3.5%
	*hme*D + *dsr*A	2	2			*hmeD:* 82.4% *dsrA*: 81%	3.8%
Escherichia coli clone NK54 (negative control)	-	0	1			-	3.2%
	-	0	2			-	4.7%

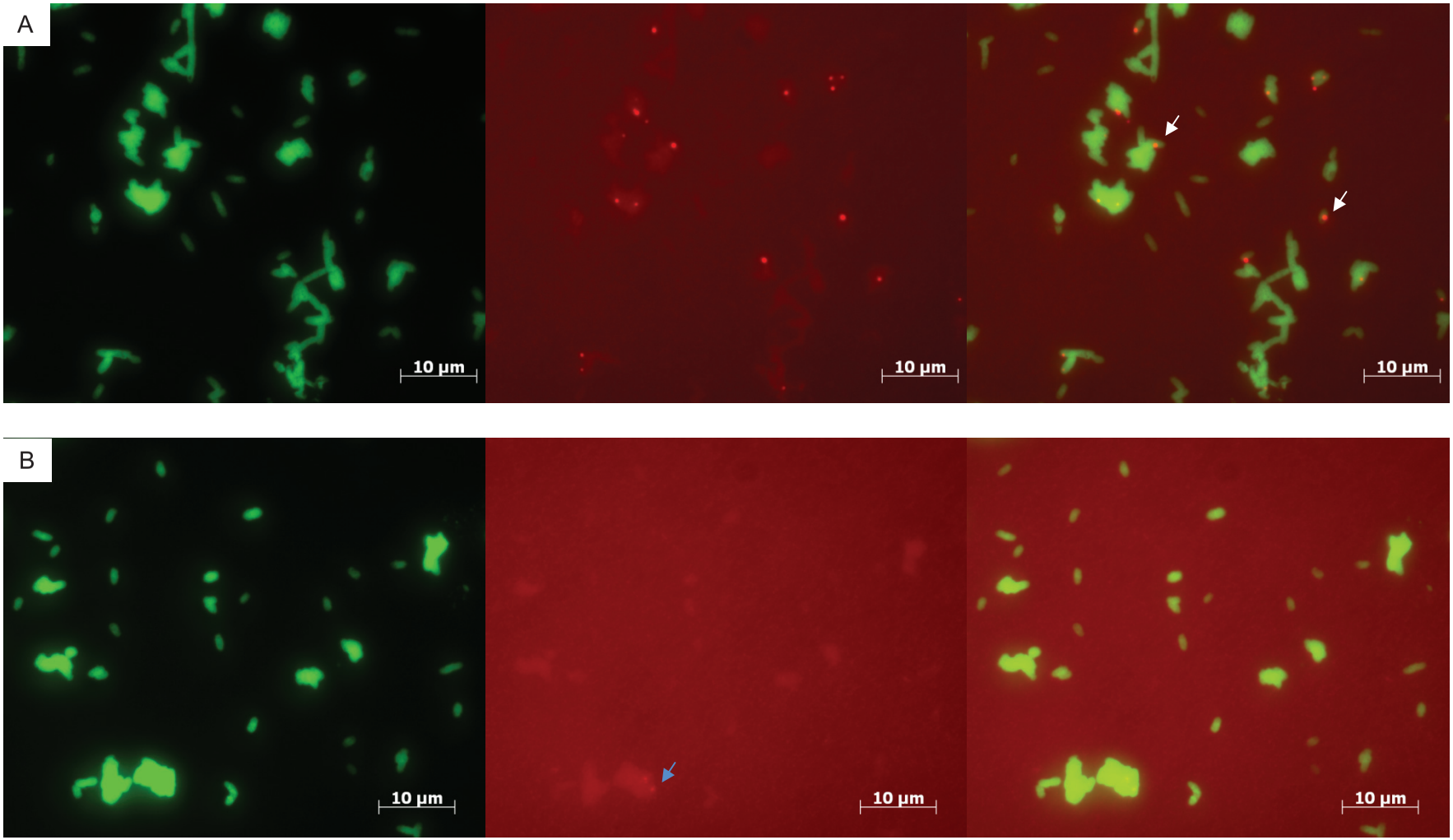


Figure 40: GeneFISH for simultaneous detection of *dsrA* gene with ssRNA probes (red) and of 16S rRNA with Eub338 probe (green) (via CARD-FISH) on *E. coli* clones. A. *E. coli* clone WS 39F7, positive control. **B.** *E. coli* clone NK54, no *dsrA* copy, negative control. White arrows indicate geneFISH signals in overlay (A). Blue arrow indicates false positive signals. It can be noticed that the cells had a degraded morphology. The hybridization temperature was 60°C, while the washing temperature was 70°C.

As indicated by the T_m of DNA:RNA hybrids, the optimum hybridization temperatures with ssRNA probes were relatively high (around 60°C). When hybridizing at these temperatures, the cell morphology had high levels of damage (Figure 40), raising concerns for the application of such a technique on environmental samples. Because most of the cases the DNA:DNA hybrids have lower T_m and lower temperature for optimum hybridization rate than the RNA:DNA hybrids (Wetmur and Davidson, 1968; Birnstiel et al., 1972; Hutton, 1977), the use of dsDNA probe reduces the hybridization temperature and the cellular morphology is better preserved. Moreover, DNA probes are less prone to degradation compared to RNA probes, which can be degraded by contaminating RNases, high temperatures or some chemicals (e.g. Mg_2^+) (Watson et al., 1984; Kierzek, 1992). Therefore, the next step in the development was the use of dsDNA probes. Although the dsDNA probes are not as prone as ssRNA probes to give background binding to rRNA, the RNase digestion step was kept, to ensure that the signal originated from the gene and not from the mRNA. This was necessary for method development. In regular use of the geneFISH protocol, this step can be omitted, as the presence of mRNA of the respective gene can increase the signal and the detection efficiency. Exceptions are the cases where discrimination of closely related alleles is required or when the presence of rRNA can give false positive signals (which can be checked with the DNase control).

The geneFISH protocol with dsDNA probes was first tested on pure cultures of *E. coli* EPI300 clones with crenarchaeotal *amo*A containing plasmids, including clones 1E3, 1E7 and 3G4. A culture of *E. coli* K12 strain, without *amo*A inserts, was included as a negative control. The copy number of *amo*A genes in 1E3 and 3G4 clones was between 1-2 per cell; the 1E7 clone was induced to a high copy number (10-200 per cell – according to kit manual and (Wild et al., 2002)).

A critical parameter for a good signal-to-noise ratio was the gene probe concentration. Different probe concentrations were tested - 0.25 pg μl^{-1}, 2.5 pg μl^{-1}, 25 pg μl^{-1} and 250 pg μl^{-1}. The hybridization time was 18-22 h or 41 h. The results are summarized in Table 31. The detection efficiency increased with probe concentration from 24% at 0.25 pg/µl to 44.3% at 250 pg/µl (clone 1E3, 18-22 h). On the other hand, the false positive level also increased, from 2.5±3.4% at 0.25 pg/µl to 19.0±2.4% at 250 pg/µl (*E. coli* K12, 18-22 h). The probe concentration which gave both a high detection efficiency and a low background was 2.5 pg/µl, with 42.8% detection efficiency and 1.4±1.1% false positives, followed close by 25 pg/µl, with 41% and 4.4±3.0%, respectively. Longer incubation times did not result in a significant increase in the detection efficiency, but rather in an increase in the background level. At optimum probe concentrations of 2.5 pg μl^{-1} and 18-22 h hybridization, the detection efficiency was 34% for the 3G4 clone and 92% for the 1E7 clone. The geneFISH signals obtained were dot-like (Figure 41), with most of the positive cells having one dot per cell. An

exception was the 1E7 clone, which possessed a higher number of gene copies per cell and where, consequently, multiple dots were observed (Figure 42A).

As opposed to the mRNA FISH protocols, where probe concentrations of 250 pg μl^{-1} (Pernthaler and Amann, 2004) or 25 ng μl^{-1} (Wagner et al., 1998) were used, it was found that the best signal to noise ratio was obtained at probe concentrations of 2.5 pg μl^{-1}.

In the case of the low copy number clones, not all the cells carrying the *amo*A fragment had a gene signal. This can be explained by the low number of targets per cell (1-2 copies). Hybridizations are equilibrium reactions in which far less than 100% of the targets are bound by the probe. When the target number is 1-2 per cell, this will result in part of the cells not having a gene signal at all. Moreover, different treatments necessary during the protocol are known to damage DNA: paraformaldehyde, acids, H_2O_2, high temperatures (Raap et al., 1986). DNA degradation by any of these treatments can lead to a reduction in the number of targets, including loss of target. This explanation for hybridization efficiency of 42% is supported by the experiments with the induced clone, where the increase in the number of targets per cell resulted in an increase of the hybridization efficiency to 90-100% (Table 31).

Table 31: GeneFISH on *Escherichia coli* clones with *amo*A-1E3 probe – gene detection efficiency (percentage of cells showing a gene signal).

Probe concentration (pg µl^{-1})	*E. coli* cl1E3 (100% identity), 1-2 *amo*A copies per cell		*E. coli* cl3G4 (95% identity), 1-2 *amo*A copies per cell		*E. coli* 1E7 (99% identity), 10-200 *amo*A copies per cell		*E. coli* K12 (target not present)	
	18-22 h	41 h	18-22 h	41 h	18-22 h	41 h	18-22 h	41 h
0.25	24.0±0.1%	-*	-	-	-	-	2.5±3.4%	-
2.5	42.8±2.4%	41.4±1.3%	34.2±1.6%	29.9±2.4%	92.0±1.7%	96.9±1.3%	1.4±1.1%	2.6±1.5%
25	41.0±6.2%	40.0±2.3%	34.0±3.6%		99.0±0.7%	98.1±1.2%	4.4±3.0%	6.0±0.8%
250	44.3±3.5%	-	-	-	-	-	19.0±2.4%	-

* "-" = not determined

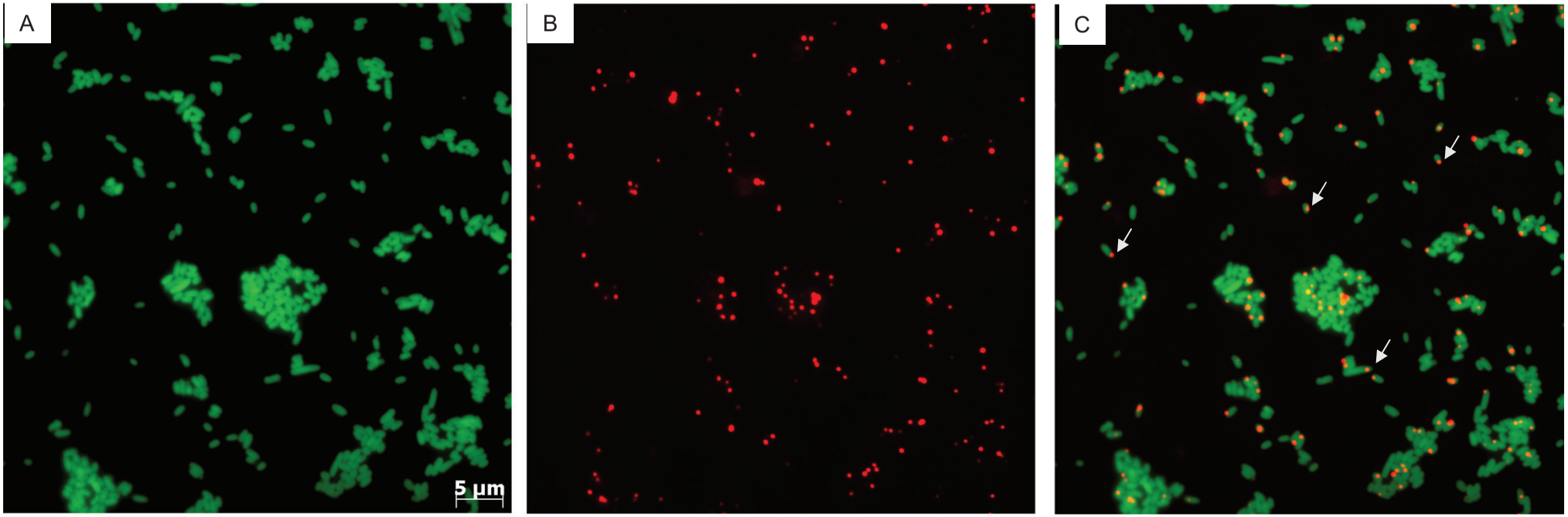


Figure 41: GeneFISH on *Escherichia coli* clone 1E3 (1-2 *amo*A copies per cell). Simultaneous 16S rRNA CARD-FISH with Eub338 (green) – a general bacterial probe - (Panel A) and gene detection with *amo*A-1E3 probe (red) targeting crenarchaeotal ammonia monooxygenase subunit A - (Panel B). Arrows indicate gene FISH signals in overlay (Panel C).

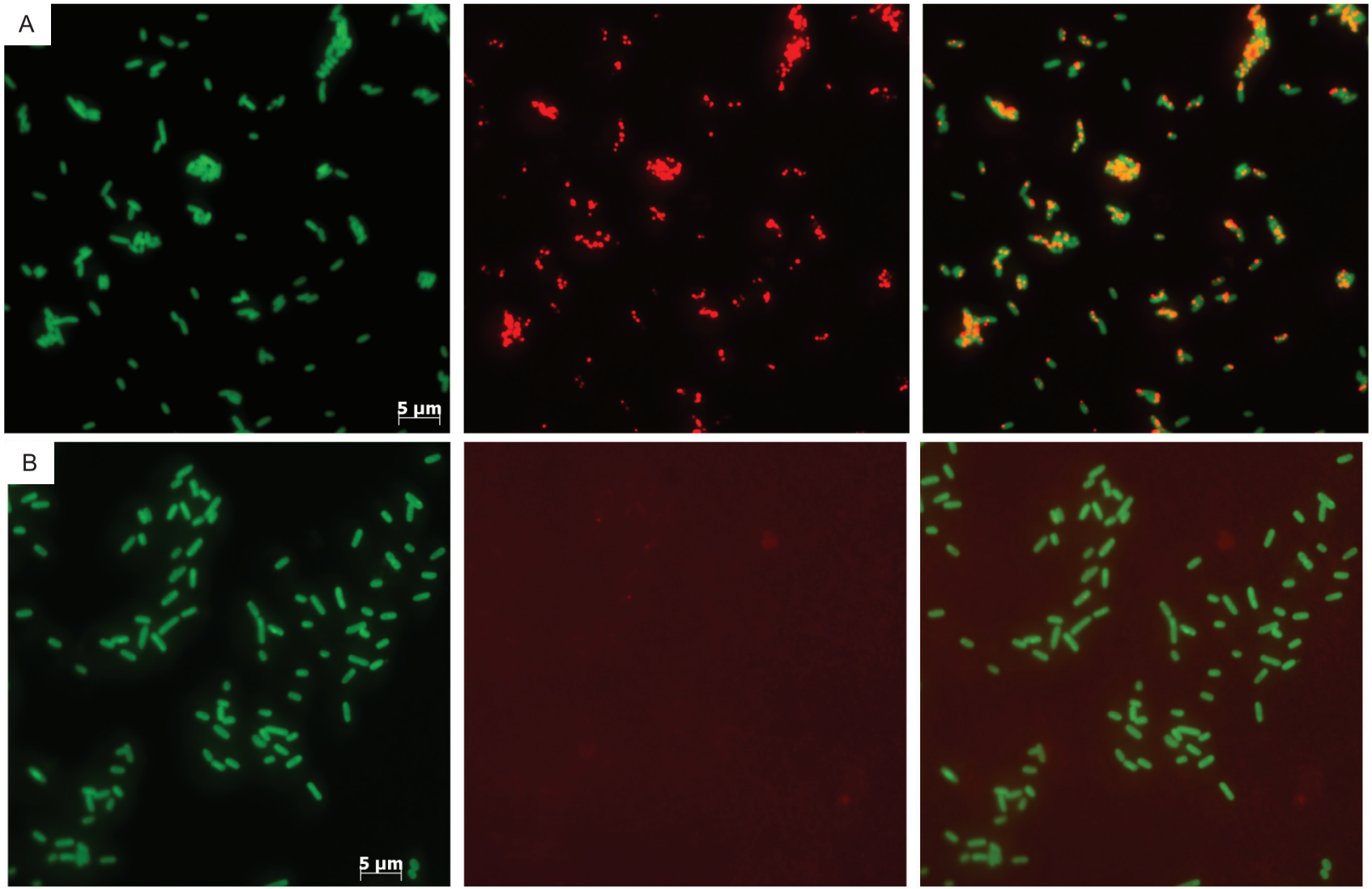


Figure 42: GeneFISH for simultaneous detection of *amo*A gene with *amo*A-1E3 probe (red) and 16S rRNA CARD-FISH with EUB338 - general bacterial probe, (green) on *Escherichia coli* clones. A. *E. coli* clone 1E7, induced to contain 10-200 copies of the *amo*A gene fragment per cell, positive control. B. *E. coli* K12, no *amo*A copy, negative control.

Detection of crenarchaeotal amoA gene in Namibian seawater samples

One question that could be addressed with such a FISH technique linking gene presence with cell identity would be whether the many putative ammonia monooxygenase sequences recently retrieved from various marine environments truly belong to *Crenarchaeota*. *Crenarchaeota*, especially those in Marine Group I (MGI), are widespread in marine environments, and they dominate microbial communities in the mesopelagic ocean with relative abundances of up to 40% of total microbial communities (Karner et al., 2001; Teira et al., 2006; Schattenhofer et al., 2009). Recent studies suggest that they might be capable of ammonia oxidation, a trait that has been shown previously exclusive to two specialized groups of *Proteobacteria* (Prosser and Nicol, 2008). Ammonia oxidation is mediated by the enzyme ammonia monooxygenase (AMO), and *amo*A, the gene encoding the alpha subunit of AMO, is often used as a functional gene marker. It shows congruent phylogeny to that of 16S rRNA (Purkhold et al., 2000). A first indication that *Crenarchaeota* possess ammonia oxidation potential came from the assembly of *amo*-like genes on archaeal scaffolds from the Sargasso Sea metagenomic dataset (Venter et al., 2004; Schleper et al., 2005). Later, both the *amo* genes and the crenarchaeotal 16S rRNA were found in the same metagenomic fragment from a soil clone library (Treusch et al., 2005). However, the first direct evidence came from the isolation of the crenarchaeon *Candidatus* "Nitrosopumilus maritimus" from a marine aquarium. This crenarchaeon has been shown to oxidize ammonia to nitrite, via hydroxylamine, and it carries the archaeal type of putative *amo*A gene (Könneke et al., 2005). Based on the similarity between the 16S rRNA and *amo*A genes from *Candidatus* "Nitrosopumilus maritimus" with sequences retrieved from different seawater samples, it has been suggested that Marine Group I Crenarchaeota present in marine waters also possess *amo*A genes and thus might play an important role in the marine nitrogen cycle (Prosser and Nicol, 2008). Additional evidence for this is the resemblance of the kinetics of ammonia oxidation by *Candidatus* "Nitrosopumilus maritimus" to that of *in situ* nitrification in marine systems (Martens-Habbena et al., 2009). A number of studies have quantified both the gene copy number of archaeal 16S rRNA and putative *amo*A genes, although the relationships between these two in environmental samples were only inferred indirectly (Lam et al., 2007; Mincer et al., 2007; Agogue et al., 2008; Park et al., 2008). To date, there has been no *in situ* localization of these environmental *amo*A sequences in single crenarchaeotal cells.

The geneFISH was applied for the Namibian seawater samples, with modifications of the permeabilization, peroxidase inactivation, rRNA hybridization and rRNA CARD steps, as have previously been found necessary for hybridizing *Crenarchaeota* (see materials and methods). The gene probe used was the *amo*A-Nam probe mix and the two negative controls were DNA digestion and NonPolyPr350 probe. The NonPolyPr350 probe was

synthesized from an artificial sequence, which gave no significant resemblance via BlastN searches (Johnson et al., 2008) with bacterial / archaeal sequences deposited in the public database Gen Bank (see materials and methods). For the three samples hybridized, *amo*A gene signals which were strictly colocalized with crenarchaeotal rRNA probe signals (Figure 43) were obtained. rRNA CARD-FISH with Cren554 probe showed abundances of *Crenarchaeota* between 10% and 21%. Out of these crenarchaeotal cells, 30% - 34% showed positive *amo*A gene signals with the *amo*A-Nam probe, compared to ~0.7% for both DNA digestion with *amo*A-Nam probe and NonPolyPr350 probe without DNA digestion (Table 32, Figure 44).

Therefore, in the Namibian seawater samples, ~30% of the crenarchaeotal cells identified by 16S rRNA-targeted CARD-FISH were detected to carry an *amoA* gene signal. This provided the first direct visual link between single celled *Crenarchaeota* in these seawater samples and the presence of putative *amoA* genes. The vast majority of *amo*A signals originated from cells identified as *Crenarchaeota*. Compared with the detection efficiency in *E. coli* which had 1-2 target copies per cell (~ 42% for perfectly matched hybrids), the percentage of *amoA* signals (~30%) in these Namibia seawater samples strongly suggest that most if not all crenarchaeotal cells in these seawater samples possess the *amoA* gene. For studies on other microorganisms, this type of assumption might not always be valid, because the detection efficiency of geneFISH depends on the number of targets per cell, which in turn depends on the cell cycle (Kubitschek and Freedman, 1971; Breuert et al., 2006; Lundgren et al., 2008).

Two negative controls were used for the environmental samples: DNA digestion and NonPolyPr350. The negative controls were necessary to confirm that the signals represent hybridization events and did not come from endogenous or introduced peroxidases, from unspecific probe or antibody binding to cellular matrixes, cell walls, etc. Both controls gave similar results, with the false positive signals being less than 1%. This confirmed that the NonPolyPr350 is a good negative control. On routine geneFISH applications, only one control is necessary. The NonPolyPr350 is preferred, because it does not introduce an extra step in the protocol, which increases the hands on time as well as cell damage. However, if the template for NonPolyPr350 is not available, then DNase digestion can also be used.

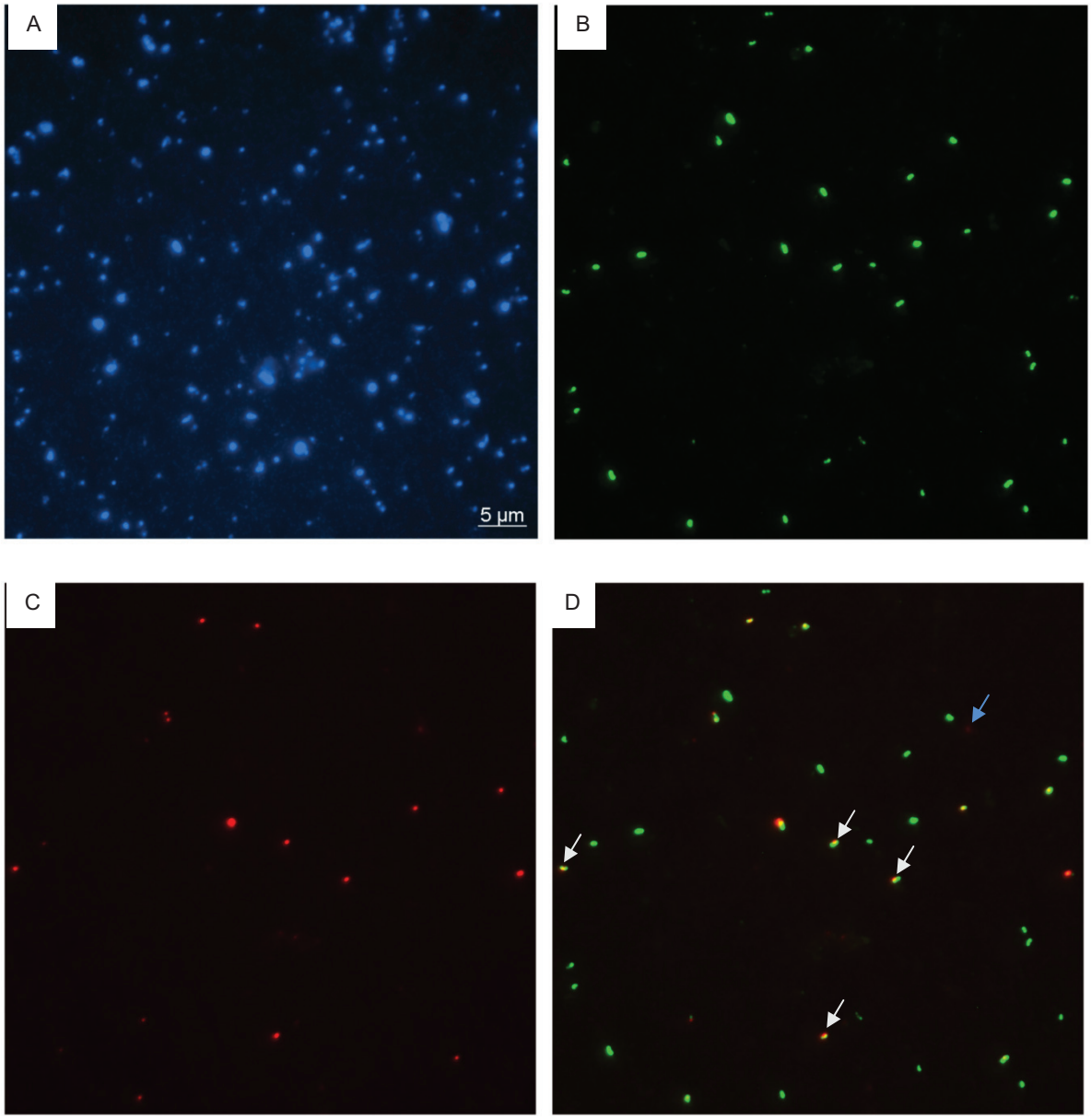


Figure 43: Gene FISH on the Namibian seawater samples, station 249. All cells stained with DAPI (blue) (Panel A). Simultaneous 16S rRNA CARD FISH with Cren554 probe (green) for marine *Crenarchaeota* (Panel B) and gene detection with *amo*A-Nam probe mix (red) for ammonia monooxygenase subunit A gene (Panel C). White arrows indicate geneFISH signals in overlay and the light-blue arrow indicates non-specific signal (Panel D).

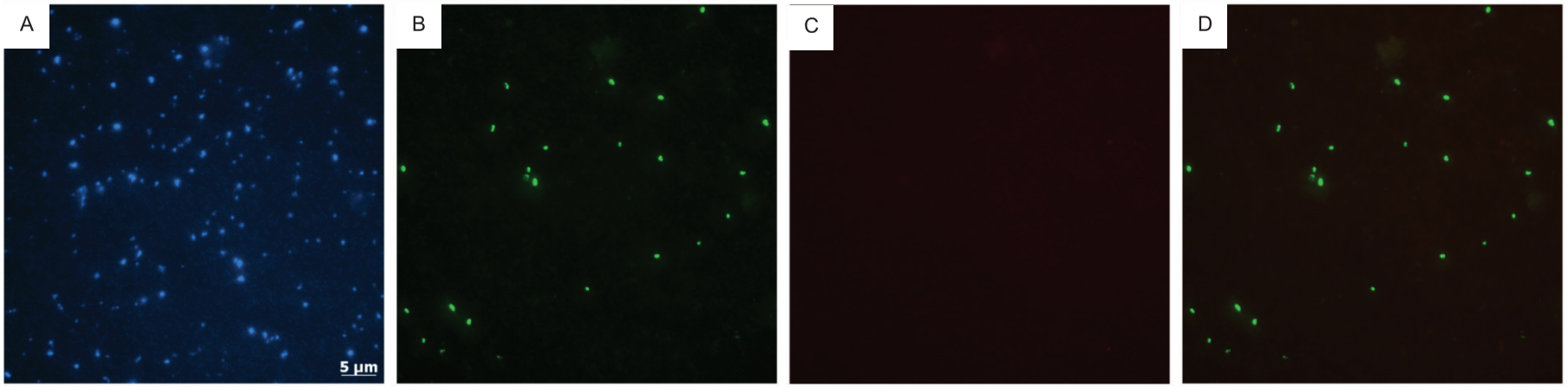


Figure 44: Gene FISH with the negative control probe for gene detection (NonPolyPp350 - red, Panel C) and 16S rRNA CARD-FISH with Cren554 probe for marine *Crenarchaea* (green, Panel B) on water column samples from Namibia, station 249. Panel D represents the overlay of the 16S and gene channels. All cells stained with DAPI: blue (Panel A).

Table 32: GeneFISH on the three Namibian seawater samples – percentage of crenarchaeotal cells showing a gene signal.

Station	1 day of hybridization		2 days of hybridization		
	*amo*A-Nam	NonPolyPr350	*amo*A-Nam	NonPolyPr350	DNase I +
249	30.6±9.2%	0.5±1.1%	29.1±7.4%	0.4±0.9%	-
213, 65 m	-*	-	34.0±7.0%	0.7±1.5%	0.7±1.14%
213, 76 m	-	-	30.5±6.1%	0.6±1.4%	0.7±1.1%

* "-" = not determined

Detection of genes in thiotrophic symbionts of Bathymodiolus puteoserpentis

The geneFISH protocol was applied also for the detection of genes in another environmental sample – that is, the symbiotic system that *Bathymodiolus puteoserpentis* and its bacterial partners are forming (Duperron et al., 2006). The gills of *Bathymodiolus puteoserpentis* mussel have bacteriocytes - specialized cells that harbor two types of symbiotic bacteria, a methane-oxidizer and a sulfur-oxidizer.

Two genes were targeted: *apr*A and *hyn*L. A single polynucleotide probe was used for *apr*A, while four polynucleotide probes targeting different regions were used for *hyn*L. As negative control, the NonPolyPro350 probe was used. The same general protocol for geneFISH with dsDNA probes was applied, with modifications of the sample preparation, rRNA CARD-FISH and peroxidase inactivation steps, as required for these samples (see materials and methods). Both the *apr*A and *hyn*L probes showed colocalization with the sulfur-oxidizing bacteria (Figure 45 and Figure 46), confirming findings from earlier experiments (Pernthaler et al., in prep; Zielinski et al., in prep). The signals were again dot like and their density was much higher for *hyn*L than for *apr*A, most probably due to the use of four target regions for *hyn*L. The eukaryotic tissue had a high autofluorescence level, which probably obscured some of the gene signals. Background signals were also an issue, but they were mostly localized outside the bacteriocyte area, therefore not interfering with gene detection. They were most probably caused by unspecific binding of the antibodies to the mussel tissue.

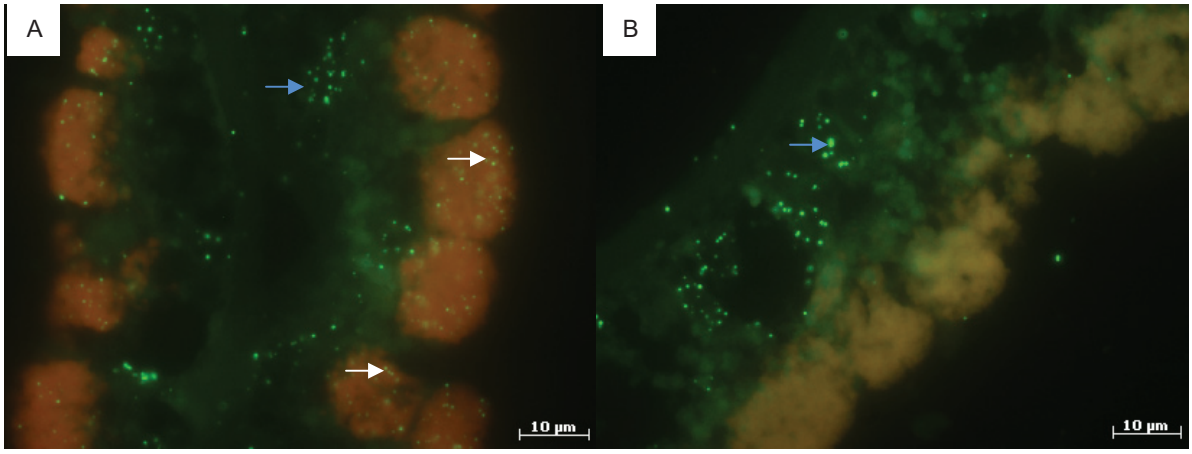


Figure 45: GeneFISH for simultaneous detection of *apr*A gene (green) and identification of bacterial cells with rRNA probe (red) (probe BMARt-193 for Thiotrophic symbionts) on bacterial gill symbionts from *Bathymodiolus puteoserpentis*. A. Overlay *apr*A probe and BMARt-193 probe. B. Overlay NonPolyPr350 probe (green) and BMARt-193 probe (red). White arrows indicate geneFISH signals, blue arrows indicate background signals.

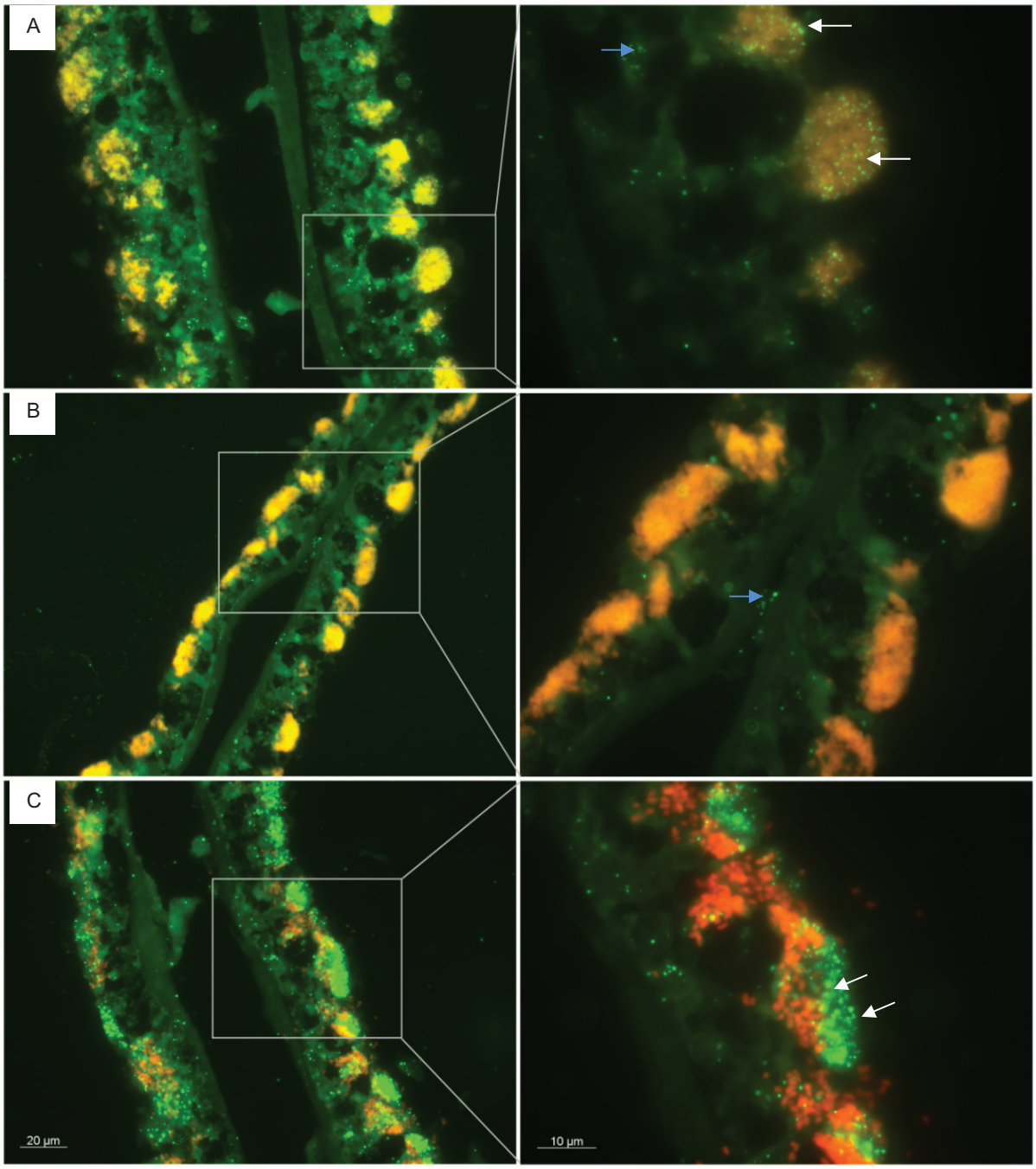


Figure 46: Gene FISH for simultaneous detection of *hyn*L gene (green) and identification of bacterial cells with rRNA probes (red) (probe BMARt-193 for Thiotrophic symbionts, probe BangM-138 for Methanotrophic symbionts) on bacterial gill symbionts from *Bathymodiolus puteoserpentis*. . A. Overlay *hyn*L probe and BMARt-193 probe. B. Overlay NonPolyPr350 (green) probe and BMARt-193 probe. C. Overlay *hyn*L probe and BangM-138 probe. White arrows indicate geneFISH signals, blue arrows indicate background signals.

Detection of rdsrA gene in enrichment culture AK199

The geneFISH protocol was further applied on an enrichment sample (AK199). The enrichment was obtained using anoxic marine sediments as inoculum and it was composed mostly from *Alphaproteobacteria* (*Roseobacter* clade) and *Gammaproteobacteria*. The target gene was *rdsrA*, for which earlier experiments suggested that it might be also found in *Roseobacter* (Lenk et al., in prep). Phylogenetic analysis of the *rdsrA* sequences cloned from this enrichment showed the presence of two clades, one grouping with sequences from *Gammaproteobacteria*, and the other grouping close with sequences from *Alphaproteobacteria* (Lenk et al., in prep). The probe used (FOS-*rdsr*A) was already available from a separate study, being perfect match with a *rdsrA* sequence present on a metagenomic fragment (Lenk et al., in prep). Taking into consideration only the target region for the FOS-*rdsr*A probe, the sequences from the clone libraries were represented by one *Alphaproteobacteria* allele, with 7.8% mismatches with the probe and two *Gammaproteobacteria* alleles, with 33.2% and 39% mismatches with the probe. Therefore, it was expected that the FOS-*rdsr*A probe will target the *Alphaproteobacteria* and not the *Gammaproteobacteria* alleles. Microscopic examination indicated colocalization of the FOS-*rdsr*A probe with the *Roseobacter* cells, and not with the *Gammaproteobacteria* cells (Figure 47). The lack of gene signals with the *Gammaproteobacteria* cells are a clear indication of specific binding, despite the high %GC of the probe used.

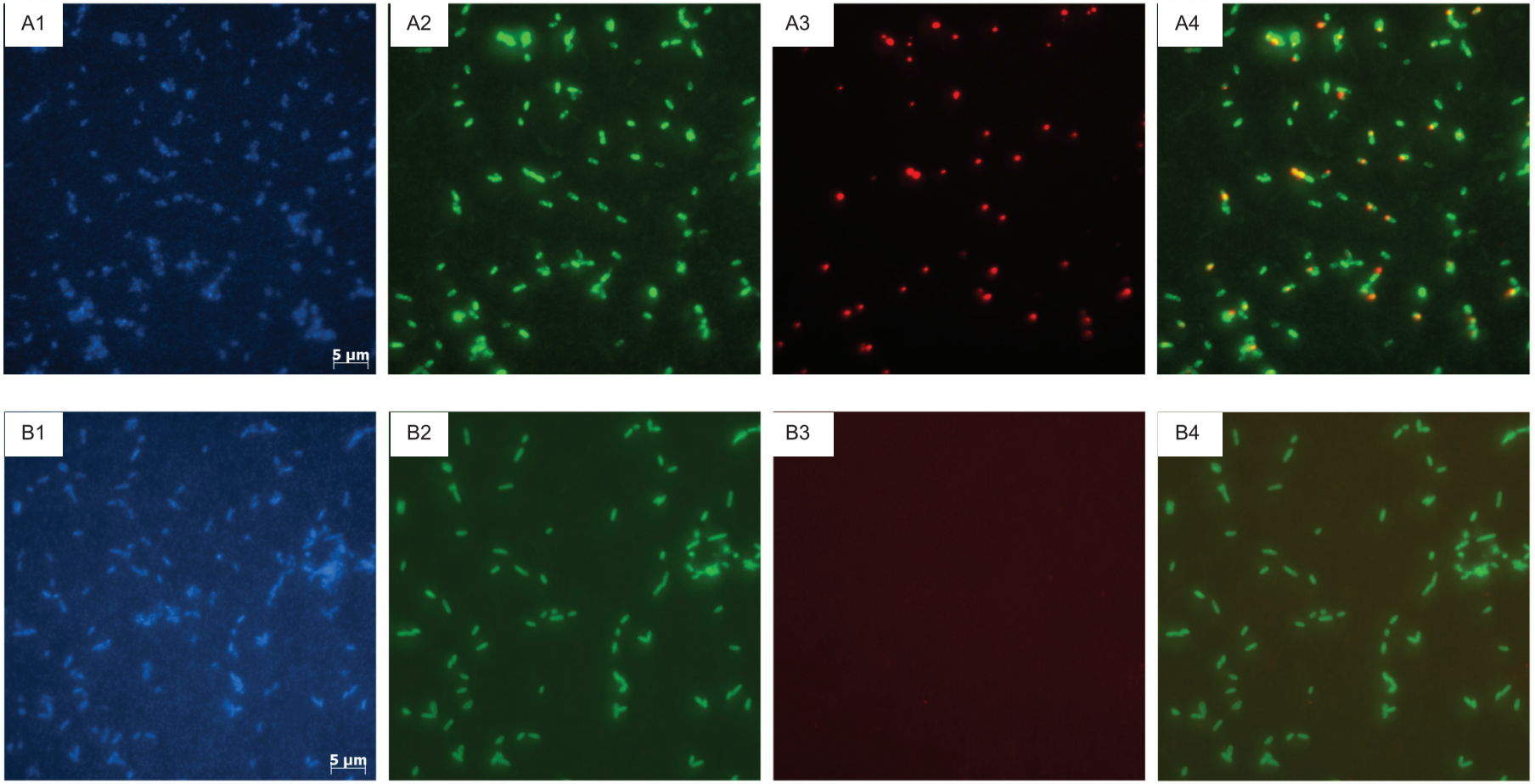


Figure 47: Gene FISH on AK199 enrichment sample. All cells stained with DAPI (Panels A1 and B1). (i**)** Simultaneous 16S rRNA CARD-FISH with Ros537 probe for marine *Roseobacter* (Panel A2) and gene detection of *rdsrA* (Panel A3), with an overlay in panel A4. (ii**)** Simultaneous 16S rRNA CARD-FISH with Gam42a probe for *Gammaproteobacteria* (Panel B2) and hybridization with NonPolyPr350 (Panel B3), with an overlay in panel B4.

3.3. Development of *in situ* mRNA and gene detection protocols with NanoSIMS – preliminary results

NanoSIMS is a nano-scale secondary-ion mass spectrometer which can determine the elemental composition of single cells. To be detected with NanoSIMS, *in situ* hybridization events have to be accompanied by introduction into the cells of chemical elements less likely to be found in environmental samples. Such examples are isotopes of different elements (e.g. ^{14}C, ^{33}P), or halogens like fluorine, iodine or bromine. These halogens have been recently used for phylogenetic identification of microorganisms based on *in situ* rRNA hybridization (Behrens et al., 2008; Li et al., 2008; Musat et al., 2008), either by direct binding of halogenated compounds to rRNA probes, or by deposition of halogen containing tyramides. By extension, it might be possible for the catalyzed reporter deposition of halogen compounds to be used for detection of geneFISH or mRNA FISH signals with NanoSIMS. However, methods based on CARD do not allow the quantification of target molecules, and the antibody step used in these protocols might give a background too high for NanoSIMS analysis.

On the other hand, the use of polynucleotide probes directly labeled with halogens might give enough sensitivity for detection with NanoSISM. Such polynucleotide probes can be prepared by incorporation during *in vitro* transcription of fluorine labeled UTP and CTP (Figure 48). The labeling density is high, the U and C being totally replaced by their fluorinated analogs. We have used in this study such fluorinated probes for detection of *nif*H mRNA in *Crocosphaera watsonii* cells.

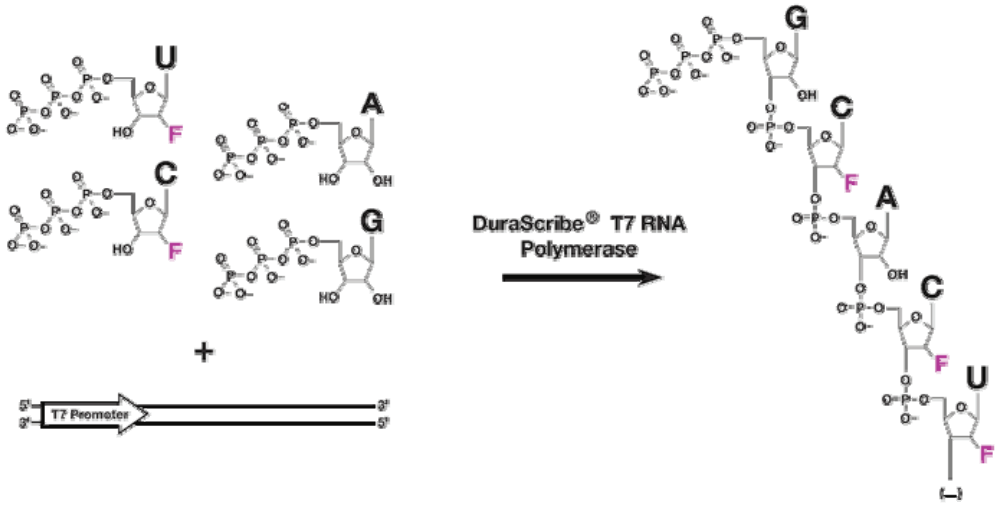


Figure 48: Incorporation of fluorinated CTP and UTP in ssRNA probes during *in vitro* transcription using DuraScribe T7 RNA Polymerase (figure adapted from http://www.epibio.com – Epicentre being the producer of DuraScribe transcription kits).

3.3.1. Probe design and synthesis

The probe synthesis process was similar to that of ssRNA probes described in the geneFISH section (3.2.2. Probe synthesis), consisting in a template generating step by PCR and a probe synthesis step, by using a special kit for *in vitro* transcription – DuraScribe from Epicentre. Two types of probes were synthesized: the antisense strand, which is the reverse complement for the mRNA, therefore being the positive probe, and the sense strand, which is in the same direction as the mRNA, therefore being the negative probe. The probes were perfect match with *Crocosphaera watsonii nif*H gene (position 127-447) and each had 321 nt. The positive probe had 167 fluorine atoms per molecule, while the negative probe had 154 fluorine atoms (Table 33), without any significant differences in the fluorine content.

Table 33: Labeling density for F-ssRNA probes.

	% A	% U*	%C*	%G	length	Fluorine atoms per probe molecule	
						Sense strand (negative control probe)	Antisense strand (positive control probe)
nifH *Crocosphaera*	26.5	28.7	19.3	25.2	321	154	167
nifH *cloned*	25.5	29.3	21.2	24	321	162	159

* **fluorine labeled**

3.3.2. Determination of the stringency conditions for hybridization with F-RNA probes

The fluorine atoms in the nucleotide analogs are actually replacing the 2' OH group on the ribose ring. As a result, the fluorinated nucleotides are neither deoxyribonucleotides nor ribonucleotides. The molecule formed by transcription (F-RNA) is a mix between ribonucleotides and these analogs, and thus, it is not a true ssRNA molecule. Therefore, the hybrids of such probe with a true RNA target (F-RNA:RNA) might have different stabilities than true RNA:RNA hybrids. Earlier studies (Cerqueira et al., 2008) have indicated that the introduction of such fluorinated analogs into oligonucleotide probes leads to an increase in T_m. Also, our initial experiments with F-RNA probes indicated that the T_m is indeed higher than that of true RNA:RNA hybrids.

As a result, the composition of the hybridization buffer was a compromise between increasing hybridization kinetics and lowering the T_m. To increase the kinetics, the buffer contained 1.43 M Na^+ and 10% dextran sulfate. To decrease the T_m, 50% or 70% formamide was added. The calculated T_m for RNA:RNA hybrids was 87.6°C for 50% formamide and 80.6°C for 70% formamide. To avoid damage to cell morphology (Raap et al., 1986), the hybridization temperature was set to 55°C. Our experiments indicate that the T_m for the F-RNA:RNA hybrids is at least 6°C higher than the calculated one for RNA:RNA hybrids (see

below). Consequently, the difference between T_m and hybridization temperature was about 38.6°C for the 50% formamide buffer and about 31.6°C for the 70% formamide buffer. Because both values were higher than the optimum temperatures for hybridization of the standard hybrid types (Wetmur and Davidson, 1968; Birnstiel et al., 1972; Hutton, 1977), we assumed the same is true for F-RNA:RNA hybrids. Also, both values were in the range where unspecific binding can occur. However, higher formamide concentrations would have decreased too much the hybridization kinetics, while Na^+ concentrations lower than 660 mM could be obtained only if dextran sulfate was eliminated from the buffer.

To ensure the hybridization specificity, we designed the washing as the stringency determining step, because during washing both low Na^+ and high formamide concentrations can be used. The T_m was measured with the method described in the geneFISH section 3.2.3.. Shortly, F-RNA:RNA hybrids (for the positive probe) were prepared, separated from single-stranded molecules by agarose electrophoresis and then purified from the gel. The T_m of these hybrids was then measured in a buffer with similar composition with the washing buffer II (23 mM Na^+, 0 to 60% formamide), plus the Syto9 dye. For all formamide concentration tested (with the exception of 0% formamide, where some of the peaks were below the detection limit for the method) the hybrids showed three melting peaks, corresponding to three melting domains (Table 34). For all three peaks, there was a linear correlation between T_m and formamide concentration (Figure 49), although with different slopes (-0.44°C, -0.49°C or -0.52°C per 1% formamide). The measured T_m were all higher than the calculated ones with the formulas derived from Wetmur, both for RNA:RNA or RNA:DNA hybrids. The difference was higher for the lower formamide concentrations, than for higher ones, indicating that, not only the T_m of F-RNA:RNA hybrids is higher than that of RNA:RNA, but also, the destabilizing influence of formamide is much stronger for the F-RNA:RNA hybrids. For 50% formamide, the difference between the calculated temperature and the highest peak was about 11°C. If we correct this value for the Syto9 influence on T_m measurements, by subtracting 5°C (Rasmussen et al., 2007), than, the difference between the real and calculated T_m is about 6°C.

Two additional software, Meltsim (Blake et al., 1999) and Poland (Steger, 1994) were used to simulate the melting for this *nif*H region. None of them showed the presence of three melting domains, regardless if the parameters used were for dsDNA, dsRNA or RNA:DNA. Since it has been previously shown that the T_m readings with Syto9 and these two software give similar results concerning the number of melting domains (Rasmussen et al., 2007), it can be concluded that the presence of the three melting domains is due to the special nature of the F-RNA:RNA hybrids. An alternative explanation would be the presence of formamide, which is not modeled in either of the software.

The 60% formamide buffer was chosen for stringent washing. The washing temperature was set to 57°C, which, if the Syto9 influence on the measured T_m value is considered (62°C – 5°C), it is equal with the lowest T_m peak (Table 34). However, as long as the other two T_m peaks are higher than the washing temperature, the specific hybrids should not melt. The measurements were performed on a *nif*H probe derived from a cloned fragment, which has ~18% mismatches with the probe for *Crocosphaera watsonii*. Because the %GC is similar between the *nif*H probes for *Crocosphaera* (%GC = 44.5) and for this clone (%GC = 45.2), we used the same stringency conditions also for the *Crocosphaera* probe.

Table 34: Melting temperatures of *nif*H hybrids in washing buffer II (23 mM Na^+, different formamide concentrations): calculated and measured values.

Formamide (% vol)	*nif*H cloned					*nif*H Crocosphaera watsonii	
	Measured T_m (F-RNA:RNA)			Calculated T_m		Calculated T_m	
	Peak 1	Peak 2	Peak 3	RNA:RNA	RNA:DNA	RNA:RNA	RNA:DNA
0	93.9 ± 0.1			80.8	74.3	80.2	73.7
10	88.1 ± 0.1	90.6 ± 0.0	93.5 ± 0.1	77.3	69.3	76.7	68.7
20	81.2 ± 1.7	83.9 ± 1.8	87.5 ± 1.6	73.8	64.3	73.2	63.7
30	76.6 ± 0.8	79.4 ± 0.8	83.4 ± 0.8	70.3	59.3	69.7	58.7
40	71.9 ± 0.1	74.9 ± 0.1	79.7 ± 0.4	66.8	54.3	66.2	53.7
50	67.1 ± 0.4	70.2 ± 0.6	75.2 ± 0.6	63.3	49.3	62.7	48.7
60	62.0 ± 0.3	65.3 ± 0.1	71.0 ± 0.0	59.8	44.3	59.2	43.7

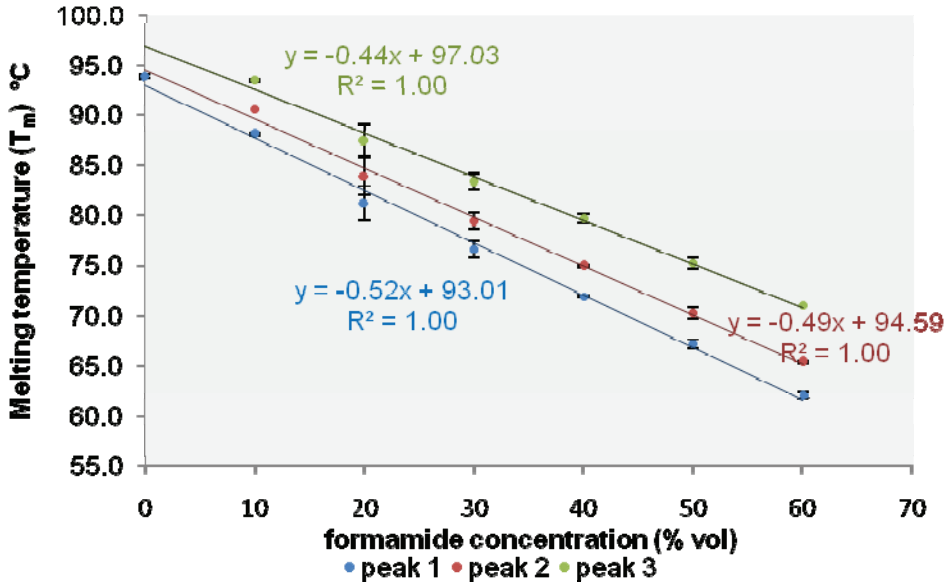


Figure 49: Variation of the melting temperature with formamide concentration. The T_m of *nif*H probe-target hybrid were measured in washing-like buffer II (23 mM Na^+, different formamide concentrations). For each formamide concentration, 3 melting peaks were visible, indicating the existence of 3 melting domains for the probe-target hybrid.

3.3.3. mRNA- HISH SIMS protocol development

The *Crocosphaera watsonii* cultures used in this experiment have been grown in conditions which favored expression of *nif*H mRNA (see materials and methods). The mRNA HISH-SIMS protocol used was a simple one, including only a few steps (Figure 50). The first step was formaldehyde fixation, which is not only preserving cell structure and morphology, but is also inactivating RNases, followed by sample immobilization on Pd/Au coated filters and permeabilization, to allow the polynucleotide probe to enter the cells. For hybridization, two buffers were tested, with 50% and 70% formamide (see materials and methods). The probe concentration was 5 ng/µl, as previously used for hybridization with polynucleotide probes (Pernthaler et al., 2002b). After five hours of hybridization, two washings were performed: the first with a high salt buffer, to remove unbound probe, and the second with the 60% formamide buffer described above, to melt unspecific hybrids. Two probes were used, the antisense strand, as positive probe, and the sense strand, as negative probe.

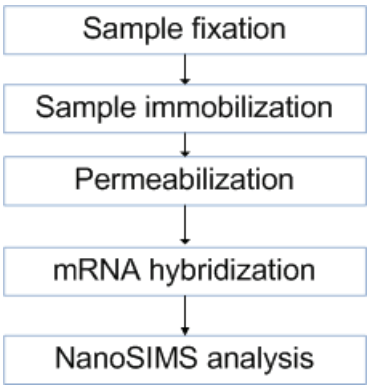


Figure 50: mRNA HISH-SIMS protocol.

Although massive cell loss from the filters prevented a comprehensive statistical analysis of the hybridized samples, the initial results were encouraging. For the positive probe (Figure 51, A1 and A2), the fluorine signal was present in seven out of nine analyzed cells, while for the negative control probe (Figure 51, B1 and B2), no fluorine signal was detected in the three cells analyzed.

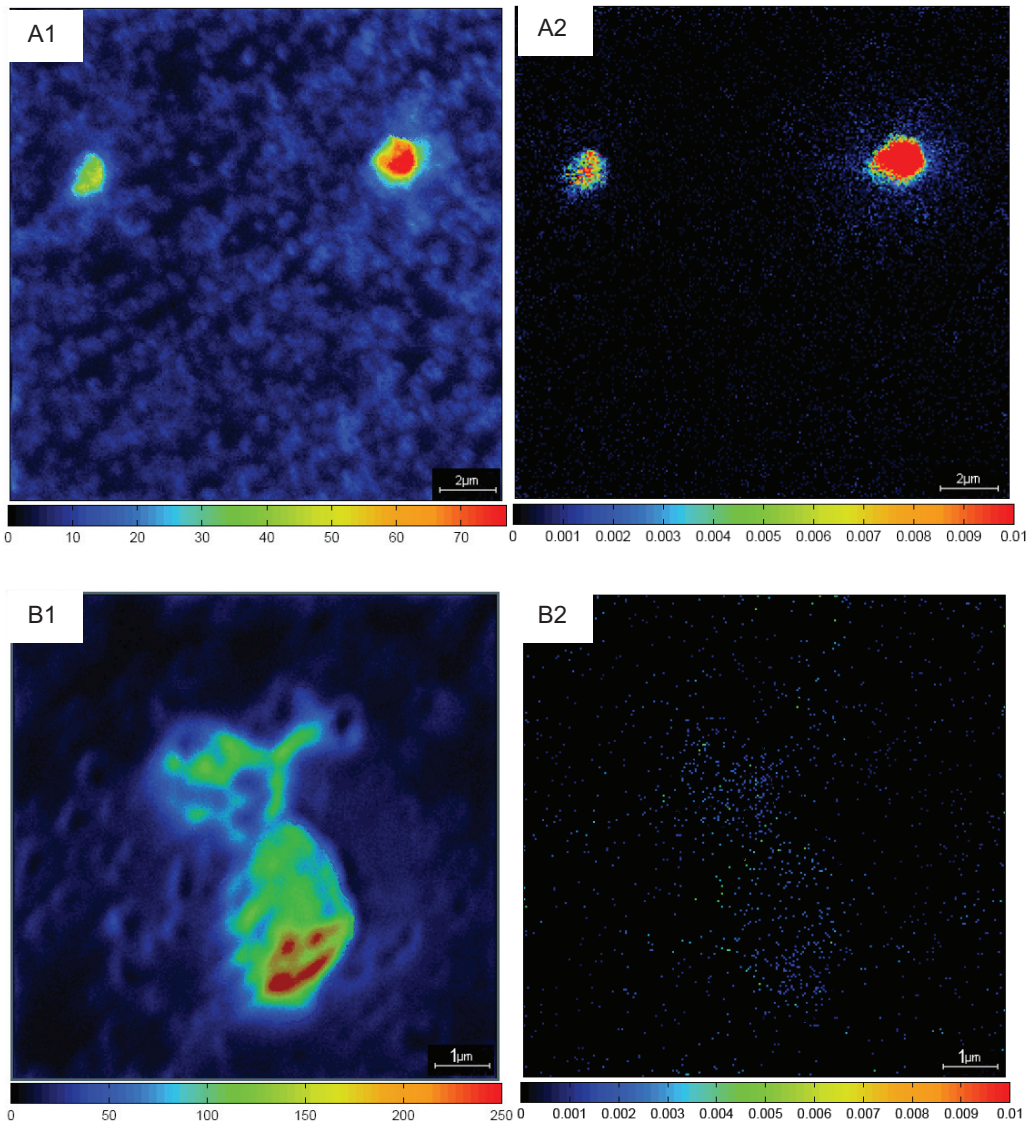


Figure 51: mRNA HISH-SIMS for the *nif*H gene on *Crocosphaera watsonii* cells. Parallel secondary ion images for $^{12}C^{14}N^-$ (A1 and B1), to indicate the cell biomass, and $^{19}F/^{12}C$ (A2 and B2) to indicate the mRNA signal. A1 and A2 – the positive probe, B1 and B2 – the negative probe.

There is still a lot of work to be done for the development of mRNA HISH-SIMS. The results obtained in the first hybridization experiment need to be confirmed by further experiments. The hybridization conditions need to be improved, so that they get close to the optimum hybridization temperature. Moreover, the probes tested here were having a relatively low %GC. For probes with higher %GC, the T_m would be so high that it could create problems with finding suitable hybridization and washing conditions. A possible solution could be the use of tetraethyl ammonium chloride (TEACl), which has been shown to reduce the effect of the %GC on T_m (Melchior and Hippel, 1973; Chang et al., 1974).

The detection limit of the method should be established by correlations with quantitative RT-PCR for cultures with different expression levels. mRNA HISH-SIMS, provided the proper standards are used, is also offering the possibility for *in situ* target quantification. That would allow researchers to quantify at single cell level the expression of different RNA molecules. To be useful in environmental samples, the mRNA HISH-SISM technique should be combined with rRNA HISH-SIMS, for phylogenetic identification of the analyzed cells.

3.4. Concluding remarks and outlook

The newly developed geneFISH protocol was successfully applied on samples with different degrees of complexity and difficulty, from pure cultures, to enrichments, to a eukaryote-bacteria symbiotic system and to seawater samples (Table 35). It allowed linking the gene presence with cell identity.

The geneFISH protocol is characterized by less than 50% gene detection efficiency, which limits its use in quantitative studies. To determine how many cells of the population have the target gene, a detection efficiency of ~100% is needed. The detection efficiency increased both when multiple probe-target regions were used (double detection of *dsr*A and *hme*D in *E. coli* clone WS 39F7, detection of four probe-target regions of the *hyn*L gene in *Bathymodiolus puteoserpentis*) and when the target copy number was increased (detection of *amo*A gene in induced copy control clone 1E7). These two strategies could be used for achieving 100% detection efficiency. The first one, multiple probe-target regions, could be used when large genes are targeted. However, most of the times the gene fragments are retrieved by PCR from environmental samples and are rather short, allowing only for one probe-target region.

Therefore, the most useful strategy would be to add a target amplification step at the beginning of the geneFISH protocol. Such an amplification step should be isothermal, should avoid any denaturations (either thermal or chemical) which might lead to strand breakage and subsequent target loss, should use primers with wide coverage and should result in long amplicons, which will not diffuse out of the cells. A Multiple Displacement Amplification (Lovmar and Syvänen, 2006; Lasken, 2007) variant might qualify for this. Such a quantitative geneFISH protocol would be an even more useful tool to address certain ecological questions. For instance, a previous published study using quantitative PCR reported a discrepancy between the abundances of crenarchaeotal 16S rRNA and *amoA* genes in different marine environments (Agogue et al., 2008). While these results suggested that not

all crenarchaeotal cells carried the *amoA* gene, the fact that these were gene fragments amplified in separate PCRs precludes a direct inference. Hence, the use of geneFISH may help to test such observations. In addition, the application of the geneFISH protocol may be further extended to examine the relative distribution of various sub-populations of functional gene variants. For instance, more than two major clusters of crenarchaeotal *amoA* genes have been identified in this study (Figure 23). An interesting follow-up study would then be to use polynucleotide probes to discriminate between these clusters identified thus far. A plot of the percent mismatches of each of the 12 probes with all *amoA* sequences (Figure 28) shows that the *amoA*-Nam mix could be split in four probe mixes. By modifying the hybridization stringency, these four probe mixes could be used to discriminate between Clusters A1, A2, B, and *Nitrosopumilus*-like *amoA*.

A crucial role in gene detection is played by the probe design process. A concept and software for rational polynucleotide probe design were developed in the first part of this thesis. Applying this concept to three metabolic marker genes revealed the following about the use of polynucleotide probes in FISH: (i) a single probe is not sufficient to detect all alleles of a gene; (ii) single probes can be used mostly at the genus level; (iii) probe mixes cannot be used to detect all alleles of a gene, because of differences in the melting temperature; (iv) probe mixes can be used for identifying a gene mostly at the genus and family level. The best strategy is to design sample specific probes based on a comprehensive knowledge of the diversity of the target gene in the respective sample. The choice for the Th1 value is very important and can influence both probe coverage and detection efficiency. The experiment with the FOS-*rdsr*A probe showed that a Th1 of 7.8% MM is still giving a geneFISH signal. Quite probably, an even higher value for the Th1 would give geneFISH signals. On the other hand, the experiments with the *amo*A-1E3 probe showed a ~20% decrease in the detection efficiency when the Th1 was 5% MM. This suggests that, with increasing the Th1 value the detection efficiency will decrease.

For answering qualitative questions, the decrease in the detection efficiency is not restrictive as long as it is still above the background level which accompanies the gene detection. This background varies with sample type, and it was 1.5% for *E. coli* clones and 0.7% for Namibian seawater samples, while with the *Bathymodiolus puteoserpentis* it appeared mostly on the eukaryotic tissue. It is most probably due to the use of antibodies and to the dot like signals characteristic for geneFISH. Such a small dot would not be considered a false positive for techniques were the whole cell is stained, like rRNA FISH. However, this background level, combined with lower detection efficiency, will restrict the application of geneFISH for detection of genes present in rare populations. Consequently, the Th1 used for designing probes for such rare populations should be as low as possible.

Since the hybridization rate is influenced by temperature in relation to T_m, the detection rate is dependent on the hybridization parameters. Ideally, the hybridization should take place at ~25°C below the T_m of the hybrids. Since the T_m equations do not always reflect the real values, a second method for T_m estimation – *in vitro* measurements – was developed in this thesis. Based on direct measurements of the T_m of the probe-target hybrids, the optimum hybridization parameters can be determined. The use of the T_m measurements can be extended further to determination of the hybridization parameters for discrimination of two closely related alleles, by measuring the T_m of the target hybrid and non-target hybrid and selecting a hybridization temperature in between. Another advantage of the *in vitro* T_m measurements is their ability to reveal the existence multiple melting domains and their T_m, while the equations will give only one average T_m.

A further improvement in the geneFISH protocol, besides adding a target amplification step, would be the use of TEACl as component of the hybridization buffer. The main effect of TEACl is the reduction, even elimination, of the %GC influence on T_m. As a result, the ΔT_m of the probe mixes would be zero or close to zero, which would make possible the use of probe mixes for higher taxonomic levels, where otherwise the T_m difference would be prohibitory. Even general probe mixes for genes could be designed and used, provided the existence of comprehensive sequence data bases. Moreover, the use of TEACl would boost hybridization kinetics and narrow the temperature range for hybrid melting, improving discrimination between closely related hybrids. Other advantages of using the TEACl would be the lowering of the melting temperature, which in turn leads to a low hybridization temperature. Only a single hybridization buffer would be needed for all types of hybrids, regardless of their %GC.

Another line of development is the mRNA and gene detection with NanoSIMS. The main advantage of NanoSIMS in microbial ecology is that it measures substrate incorporation rates at the single cell level. Combined with mRNA HISH-SIMS, this would be a powerful tool for studying the physiology and cell biology of yet uncultivated microorganisms. The initial results with the mRNA HISH-SIMS using fluorinated ssRNA probes are encouraging. However, further work is necessary for the development of such a protocol and for establishing the detection sensitivity, i.e. if single hybridization events can be detected. An alternative would be the deposition of halogen-containing tyramides in the last amplification step of the geneFISH protocol.

Table 35: Overview of geneFISH experiments with dsDNA probes.

Sample	GeneFISH							rRNA probes
	gene	Allele no.	Probe no.	Probe – target identity	Probe Length (bp)	Probe %GC	Gene detection	
pure cultures – plasmid localization								
Escherichia coli clone 1E3 (1-5 copies/cell)	*amo*A	1	1	100%	351	41%	42%	Eub338
Escherichia coli clone 3G4 (1-5 copies/cell)		1		95%			34%	
Escherichia coli clone 1E7 (25-50 copies/cell)		1		100%			97%	
enrichment culture – chromosome localization								
AK199	*rdsrA*	1	1	7.8%	359	65%	n.d.*	Ros537
environmental samples – chromosome localization								
Bathymodiolus puteoserpentis (symbionts)	*hyn*L	1	4 , for different target regions	100%	349-356	39 - 45%	n.d.	BMARt-193 / BangM-138
	*apr*A	1	1	100%	352	43%	n.d.	BMARt-193
Namibian upwelling system station 213, 65m	*amo*A	96	12, for one target region	95-100%	351	40 - 44%	30%	Cren554
station 213, 76m							34%	
station 249 (130m)							30%	

*n.d. = not determined

Appendix

Appendix A: primers

Table 36: List of primers used in this study.

Primer name	Primer sequence	Annealing temperature (°C)	Reference
	***amo*A amplicons**		
Arch-amoAF	STAATGGTCTGGCTTAGACG	53	(Francis et al., 2005)
Arch-amoAR	GCGGCCATCCATCTGTATGT		
1f	CAGGAGACTACATCTTCTA	57	
2f	CGGGAGACTATATCTTCTA	57	
3f	CAGGAGACTATATCTTCTA	57 / 48	
4f	CAGGGGACTACATCTTCTA	57	
5f	CTGGTGACTATATCTTCTA	57 / 48	
6f	CAGGAGACTACATCTTTTA	57	
7f	CAGGAGATTATATCTTCTA	48	
8f	CTGGAGACTATATCTTCTA	57	
9f	CTGGAGATTATATTTTCTA	48	
3r	TCAAATTGACCATGTTGAA	57	
4r	TCAGCTGTACCATATTGAA	57	
5r	TCAGGTTTACCATATTGAA	48	
6r	TCAGGTTTACCATGTTGAA	57	
7r	TCAACTGTACCATATTGAA	57 / 48	
8r	TCAACTGTACCATGTTGAA	57 / 48	
9r	TCAAGTTGACCATGTTGAA	57	
10r	TCAATTGTACTATATTGAA	48	
	***nif*H amplicons**		
nifH 1	TGYGAYCCNAARGCNGA	57	(Zehr and McReynolds, 1989)
nifH 2	ADNGCCATCATYTCNCC		
nifH_B-F	TCTACTCGTTTGATGCTTCAC	55	
T7_nifH_B-F	gcc agt gaa ttg taa tac gac tca cta tag gg TCTACTCGTTTGATGCTTCAC		
nifH_D-R	TGTAACGATGTAGATTTCTTG		
T7_nifH_D-R	gcc agt gaa ttg taa tac gac tca cta tag gg TGTAACGATGTAGATTTCTTG		
nifHCr-F	TCTACCCGTTTAATCCTCAAC	48	
T7-nifHCr-F	gcc agt gaa ttg taa tac gac tca cta tag gg TCTACCCGTTTAATCCTCAAC		
nifHCr-R	GGTAACGATGTAGATTTCTTG		
T7-nifHCr-R	gcc agt gaa ttg taa tac gac tca cta tag gg GGTAACGATGTAGATTTCTTG		

Primer name	Primer sequence	Annealing temperature (°C)	Reference
	***hynL* amplicons**		
hynl1-F	GTGATTGTTAATGCGGTTTC	57	
hynl1-R	CTAGATTTGGCATGCTTAGG		
hynl2-F	GGGCAATTAGGACCCTTTA		
hynl2-R	ACCCCATAGTTGTCCCTTAT		
hynl3-F	AAAAGTGTCATGGATTACGG		
hynl3-R	AACGAGATAAAGGACCCACT		
hynl4-F	CTTTAGCTGGCACTGAACTT		
hynl4-R	CTCAAAGGCACCAATATTTC		
	***apr*A amplicons**		
aprA-Bath F	ACCGTCTTTGAATCGAGC	48	
aprA-Bath R	TGAATCTTATAAGCCTATCG		
AM55 21/C12	TGGCAGCTCATGATCAATGG		
AM56 21/C12	GCACCCACGGGACCGTA		
	***rpo*B amplicons**		
rpoB R	CGTTGCATGTTCGCACCCAT	61	
T7-rpoB F	gcc agt gaa ttg taa tac gac tca cta tag gg AACATCGGTCTGATCAAC		
	***rdsr*A amplicons**		
rdsr1-R	GCGCCGGGCGGTGCATCTC	56	
rdsr1-F	GAAGTATCCCGAGTCGAAGG		
	***hme*D amplicons**		
T7 HmeD_384 for	gcc agt gaa ttg taa tac gac tca cta tag gg GAYCCCAARAAYATGCC	61	
HmeD_704 rev	GATCCTTGAGGTCATCCA		
	***dsr*A amplicons**		
DsrA_520R	CACTCGCACCTSGMCAT	61	
T7 MM Dsr1F deg	gcc agt gaa ttg taa tac gac tca cta tag gg GGCCAYTGGAARCAYG		
	NonPolyPr350		
NonPolyPr350-F	ACAGTCGAATGTCTACCTAC	58	
NonPolyPr350-R	AATATTGTGCAGTCGGATC		
	Screening PCR		
M13F	GTAAAACGACGGCCAG	50	
M13R	CAGGAAACAGCTATGAC		
	Sequencing primers		
T7	TAATACGACTCACTATAGGG	50	
T3	AATTAACCCTCACTAAAGGG	55	
SP6	ATTTAGGTGACACTATAG	50	
HUPL 540	GAAGCAGATTTAATGGCGGTAGC	50	

Appendix B: Media and buffers

Luria Bertani (LB)

Table 37: Luria Bertani (LB) medium

component	amount
tryptone	10 g
Yeast extract	5 g
NaCl	10 g
water	Up to 1 L

Preparation*, **:

- add chemicals, add water up to 970 ml
- adjust pH to 7.0
- add water up to 1 L
- autoclave

* for preparing solid LB medium, add 1.5% agar before autoclavation.

** for preparing solid LB medium with Antibiotica, let cool after autoclavation and than add the antibiotica to the desired concentration.

Ampicillin stock solution

Dissolve ampicillin Na salt in water to a concentration of 100 mg/ml. Filter sterilize (0.2 µm) and store at -20°C.

Chloramphenicol stock solution

Dissolve chloramphenicol in absolute ethanol to a concentration of 23.5 mg/ml. Filter sterilize (0.2 µm) and store at -20°C.

Appendix C: Nucleic acid purification protocols

Amplicon purification with Gene Clean Turbo kit

- to 1 part PCR product add 5 parts salt solution
- mix gently
- transfer DNA solution to a cartridge assembled in a 2 ml cap-less catch tube
- 5 sec at 14 000 × g, empty catch tube
- repeat the last 2 steps until all gel solution has passed through the cartridge
- + 500 µl wash solution
- 5 sec at 14 000 × g, empty catch tube
- + 500 µl wash solution
- 5 sec at 14 000 × g, empty catch tube
- 4 min at 14 000 × g
- transfer cartridge into a new catch tube (with cap)
- + 30 µl elution solution (= water, RNase free)
- 5 min at RT
- 1 min at 14 000 × g
- transfer DNA solution into new tubes
- Store at 4°C (amplicons for cloning) or at -20°C (probes, targets)

Purification of nucleic acids from agarose gels with Pure Link Quick Gel extraction kit

- weigh each gel slice and add 6 µl of GS1 buffer for each mg of gel
- incubate at 50°C until the gel has melted
- load 850 µl gel solution on a purification column
- 1 min at 12 000 × g, empty flow through
- repeat the last 2 steps until all gel solution has passed through the column
- + 500 µl GS1 buffer
- 1 min at RT
- 1 min at 12 000 × g, empty flow through
- + 700 µl W9 buffer
- 5 min at RT
- 1 min at 12 000 × g, empty flow through
- 1 min at 12 000 × g
- place purification column in a 1.5 ml recovery tube
- + 50 µl TE (prewarmed at 70°C), directly in the centre of the column
- 10 min at RT
- 2 min at 12 000 × g
- transfer DNA solution into new tubes

- Store -20°C

Sephadex purification

- take multiscreen column loader 45 µl
- place some Sephadex G-50 Superfine on the plate
- fill all the wells carefully with Sephadex using the spatule
- put the rest into the with box containing the spatule
- put MultiScreen-HV plate on top of the multiscreen column loader, placing the Multiscreen-HV plate directly next to the spike
- rotate the assembly (180°) and tip on the top the column loader that the Sephadex falls into the new plate
- add 300 µl autoclaved MilliQ water to as many wells as needed for purification
- store at least for 3 h at RT or overnight at 4°C
- assemble MultiScreen-HV, blue frame, and 96 well wash plate
- centrifuge for 5 min at 910 x g in a swing out plate rotor
- empty wash plate
- add 150 µl autoclaved MilliQ water to the MultiScreen-HV
- centrifuge vor 5 min at 910 xg in a swing out plate rotor
- add 15 µl “Lichrosolv” ultrapure water to each sequencing reaction or 5 µl “Lichrosolv” ultrapure water to each screening PCR reaction
- put MultiScreen-HV on a new 96 well round bottom plate, with blue frame
- transfer each sequencing reaction to one well of the MultiScreen-HV
- centrifuge vor 5 min at 910 xg in a swing out plate rotor
- seal 96 well plate with plastic foil and store at 4°C

ssRNA purification with RNeasy Min Elute Cleanup kit

- to 20 µl sample, add 80 µl of water
- + 350 µl RLT buffer, mix
- + 250 µl 96% ethanol, mix
- transfer 700 µl of sample to a spin column
- 15 sec at 8000 × g at RT, empty flow through
- transfer spin column in a new 2 ml collection tube
- + 500 µl RPE Buffer
- 15 sec at 8000 × g at RT, empty flow through
- + 500 µl 80% ethanol
- 2 min at 8000 × g at RT, empty flow through
- transfer spin column in a new 2 ml collection tube

- 5 min at 16000 × g at RT
- transfer spin column in a new 1.5 ml collection tube
- + 14 µl water
- 1 min at 16000 × g at RT
- transfer RNA solution into new tubes
- store -20°C

ssRNA purification with Nuc Away Spin columns

- tap the column to settle the dry gel in the bottom of the spin column.
- hydrate the column with 650 µL of RNase-free water. Cap, vortex, tap out air bubbles, and hydrate at room temperature 5–15 min.
- place the NucAway spin column in a 2 ml collection tube and spin the column at 750 x g for 2 min to remove excess interstitial fluid, keeping track of the orientation of the column in the rotor.
- discard the collection tube and immediately apply the sample (20–100 µl) to the center of the gel bed at the top of the column.
- ***Note***: Do not disturb the gel surface or contact the sides of the column with the pipette tip or reaction mixture.
- place the NucAway spin column in the 1.5 ml elution tube and place in the rotor, maintaining orientation.
- spin the NucAway spin column in the tube at 750 x g for 2 min. The sample will be in the elution tube.
- discard the NucAway spin column and continue with your procedure.

Appendix D: Materials

Product	Company	Cat. No.
1 M Tris, pH 8.0	Ambion	AM9856
20% SDS	Ambion	AM9820
20x SSC	Ambion	AM9765
5 M NaCl	Ambion	AM9759
AmpliScribe T7 High Yield Transcription kit	Epicentre	AS3107
Anti-Dig-POD Fab fragments	Roche	11 207 733 910
Blocking Reagent for nucleic acid hybridizations	Roche	11096176001
Copy Control cDNA, Gene and PCR Cloning Kit	Epicentre	CCPCR1CC
DuraScribe T7 Transcription kit	Epicentre	DS010925
Gene Clean Turbo kit	Q-Biogene	1102-600
Lysozyme	AppliChem	A4972.0010
Microcon YM100	Millipore	42413
Northern Max Hybridization Buffer	Ambion	AM8677
NucAway Spin Columns	Ambion	AM10070
PBS 10x, pH 7.4	Ambion	AM9625
PCR Dig Probe Synthesis Kit	Roche	11636090910
PCR Water	Sigma	W4502
Plasmid Maxi Kit	Qiagen	12163
Plasmid MiniPrep Kit	Millipore	LSKP09623
ProLongGold	Invitrogen	P36930
Quick Gel extraction kit	Invitrogen	K2100-12
RNase free DNaseI	Epicentre	D9902K
RNase I (Cloned) 100 U/µl	Ambion	AM2294 / AM2295
RNaseA	Sigma	R4642-10
RNeasy Min Elute Cleanup kit	Invitrogen	74204
Sheared Salmon Sperm DNA (sheared, 10 mg/ml)	Ambion	AM9680
SlowFadeGold	Invitrogen	S36936
Syto9	Invitrogen	S-34854
TopoTA Cloning Kit for Sequencing	Invitrogen	K4575-40
UltraHyb Ultrasensitive Hybridization Buffer	Ambion	AM8669
Western Blocking Reagent, Solution	Roche	11921673001
Yeast RNA (10 mg/ml)	Ambion	AM7118

Appendix E: Polynucleotide probe design – results table

Table 38: Probe mixes formed by alleles of the *amo*A and *pmo*A genes. The probe mixes marked in blue color were calculated by putting together probe mixes from lower taxonomic levels.

Taxonomic level	Taxonomic Division	No. of Targets	No. of Probes	ΔT_m (°C)	Th1 (% mismatches)	Th2 (% mismatches)	Th2-Th1 (ΔT_m corrected) (°C)
***amo*A**							
Phylum	Proteobacteria	169	35	7.9	10	28	
Class	GammaProteobacteria	14	3	1.1	10	29	
Genus	Nitrosococcus	14	3	1.1	10	29	
Class	BetaProteobacteria	155	32	7.9	10	28	
Family	Nitrosomonadaceae	155	32	7.9	10	28	
Genus	Nitrosomonas	70	19	4	10	19	13.8
	Nitrosospira	85	13	2.9	10	20	9.5
***pmo*A**							
	All	191	31	12.4	10	28	
Phylum	Verrucomicrobia	12	6	5.6	10	33	
	Proteobacteria	179	25	10.7	10	28	
Class	GammaProteobacteria	56	16	8.7	10	27	
Order	Methylococcales	56	16	8.7	10	27	
Family	Crenotrichaceae	6	1	-	10	31	
Genus	Crenothrix	6	1	-	10	31	
Family	Methylococcaceae	50	15	8.7	10	27	
Genus	Methylothermus	1	1	-	0	15	
	Clonothrix	1	1	-	0	18	
	Methylosoma	1	1	-	0	13	
	Methylococcus	3	1	-	10	14	8.5
	Methylohalobius	2	2	0.5	10	14	14.4
	Methylomonas	11	1	-	10	17	9.2
	Methylocaldum	13	3	1.5	10	15	4.05
	Methylomicrobium	4	none	-	-	-	
	Methylosarcina	3	3	0.8	6	8	
	Methylobacter	7	none	-	-	-	
	Methylobacter-Methylomicrobium – Methylosarcina Group	14	7	1.8	10	13	2.6
Class	AlphaProteobacteria	123	9	3.3	10	26	
Order	Rhizobiales						

Family	Beijerinckiaceae	2	1	-	10	28	
Genus	Methylocapsa	2	1	-	10	28	
Family	Methylocystaceae	121	8	1.7	10	26	
Genus	Methylocystis - Methylosinus Group, pmoA1	107	3	1.1	10	24	13.8
	Methylocystis - Methylosinus Group, pmoA2	14	5	1. 7	10	20	14.6

Table 39: Probe mixes formed by alleles of the *apr*A gene. The probe mixes marked in blue color were calculated by putting together probe mixes from lower taxonomic levels.

SOP/ SRP	Taxonomic level	Taxonomic Division	No. of Targets	No. of Probes	ΔT_m (°C)	Th1 (% mismatches)	Th2 (% mismatches)	Th2-Th1 (ΔT_m corrected) (°C)
*apr*A								
all			208	115	10.4	10	n.a.	
		SOP	58	26	9.4	10	18	
		SRP	147	85	9.8	10	19	
	Domain	Archaea	9	8	6.3	10	33	
	Phylum	Crenarchaeota	5	4	6.3	10	34	
	Family	Thermoproteaceae	5	4	6.3	10	34	
SRP	**Genus**	Caldivirga	2	1	-	10	36	
SRP/ SOP *		Pyrobaculum*	3	3	1.7	10	34	
SRP	**Phylum**	Euryarchaeota	4	4	1.7	10	33	
	Class	Archaeoglobi	4	4	1.7	10	33	
	Genus	Archaeoglobus	4	4	1.7	10	33	
	Domain	Bacteria	198	107	10.4	10	33	
SRP	**Phylum**	Nitrospirae	2	1	-	10	13	12
	Genus	Thermodesulfovibrio	2	1	-	10	13	12
SOP	**Phylum**	Chlorobi	10	4	7.4	10	29	
	Family	Chlorobiaceae	10	4	7.4	10	29	
	Genus	Chlorobaculum	3	none	-	5	5	
		Chlorobium	7	none	-	5	5	
SRP	**Phylum**	Firmicutes	27	13	8.0	10	12	0.7
	Class	Clostridia	27	13	8.0	10	12	0.7
	Genus	Thermacetogenium	1	1	-	0	13	
	Family	Peptococcaceae	26	12	7.0	10	19	6.3
	Genus	Desulfotomaculum	26	12	7.0	10	19	6.3
SRP	**Phylum**	Thermodesulfobacteria	4	2	2.1	10	30	
	Family	Thermodesulfobacteriaceae	4	2	2.1	10	30	
	Genus	Thermodesulfatator	1	1		0	24	
		Thermodesulfobacterium	3	1	-	10	24	17.1

	Phylum	Proteobacteria	154	86	10.4	10	20	
SOP	**Class**	Betaproteobacteria	12	6	2.25	10	10	0
	Genus	Thiobacillus	12	6	2.25	10	10	0
		Thiobacillus group I	5	3	0.8	10	14	4.6
		Thiobacillus group II	7	3	1.4	10	10	0
	Class	Gammaproteobacteria	36	17	7.8	10	11	0.5
	Order	Thiotrichales	3	2	1.3	10	20	19.6
	Family	Thiotrichaceae	3	2	1.3	10	20	19.6
	Genus	Beggiatoa	1	1	-	0	23	
		Thiothrix	2	1	-	10	20	20.8
	Order	Chromatiales	33	15	4.9	10	11	1.1
	Genus	Thioalkalivibrio	1	1	-	0	13	
	Family	Chromatiaceae	32	14	4.9	10	11	1.1
	Genus	Thiorhodovibrio	1	1	-	0	11	
		Chromatium	1	1	-	0	9	
		Thermochromatium	1	1	-	0	6	
		Thiohalocapsa	1	1	-	0	9	
		Rhabdochromatium	1	1	-	0	11	
		Thiolamprovum	1	1	-	0	9	
		Thiorhodococcus*	2	2	1.7	10	12	10.9
		Thiodictyon	2	none				
		Lamprocystis	2	none				
		Halochromatium	2	2	1.4	5	9	8.1
		Thiococcus	3	2	0.6	5	9	8.6
		Allochromatium	3	none				
		Thiocapsa**	6	none				
		Thiocystis*	6	5	1.2	5	7	6.8
SRP	**Class**	Deltaproteobacteria	106	63	9.8	10	19	
	Order	Desulfarculales	1	1	-	0	20	
	Genus	Desulfarculus	1	1	-	0	20	
	Order	Syntrophobacterales	12	8	3.6	10	20	11.5
	Family	Syntrophobacteraceae	10	6	3.6	10	20	11.5
	Genus	Thermodesulforhabdus	1	1	-	0	19	
	Family	Desulfacinum	2	1	-	10	14	6.6
		Desulforhabdus	3	1	-	10	19	9.4
		Syntrophobacter	4	3	1.7	10	11	11.5
		Syntrophaceae	2	2	0.5	10	22	22.6
	Genus	Desulfomonile	1	1	-	0	25	
		Desulfobacca	1	1	-	0	23	
	Order	Desulfovibrionales	41	20	9.3	10	13	1.9
	Family	Desulfonatronumaceae	1	1	-	0	14	
	Genus	Desulfonatronum	1	1	-	0	14	
	Family	Desulfomicrobiaceae	2	1	-	10	16	11.9
	Genus	Desulfomicrobium	2	1	-	10	16	11.9
	Family	Desulfohalobiaceae*	4	4	8.0	10	11	3.0
	Genus	Desulfonatronovibrio	1	1	-	0	18	

	Family	Desulfonauticus	1	1	-	0	27	
		Desulfothermus	1	1	-	0	20	
		Desulfohalobium	1	1	-	0	11	
		Desulfovibrionaceae	34	14	6.1	10	12	0.4
	Genus	Desulfovibrio	34	14	6.1	10	12	0.4
	Order	Desulfobacterales	51	33	7.3	10	17	
	Family	Desulfobulbaceae	15	11	5.0	10	14	3.9
	Genus	Desulfofustis	1	1	-	0	16	
	Family	Desulfotalea	1	1	-	0	16	
		Desulfocapsa	2	2	1.7	10	14	13.8
		Desulforhopalus	3	2	1.6	10	13	12.25
		Desulfobulbus	6	4	3.2	10	14	6.25
		Desulfobacteraceae	36	22	7.3	10	13	11.4
	Genus	Desulforegula	1	1	-	0	14	
		Desulfosarcina	1	1	-	0	12	
		Desulfotignum	1	1	-	0	8	
		Desulfospira	1	1	-	0	8	
		Desulfobotulus	1	1	-	0	15	
		Desulfobacula	2	1	-	10	14	13.2
		Desulfocella	2	1	-	10	15	8.5
		Desulfofrigus	2	1	-	10	16	11.0
		Desulfatibacillum	2	1	-	10	14	14.0
		Desulfofaba	2	2	0.5	10	15	15.5
		Desulfonema	3	3	5.0	10	11	7.7
		Desulfobacter	4	3	0.7	5	8	5.4
		Desulfococcus	5	2	0.35	10	14	5.6
		Desulfobacterium	9	9	5.1	5	7	5.0
Unclassified deltaprot		Desulfocaldus	1	1	-	0	12	
no affiliation in a Phylum	**Genus**	Thermodesulfobium	1	1	-	0	31	

Apendix F: Probe design of *amo*A-Nam probe mix

Figure 52 through Figure 57 - a step by step description of the probe design process for the *amo*A-Nam probe mix. At the end, the sequence of each *amo*A probe is given.

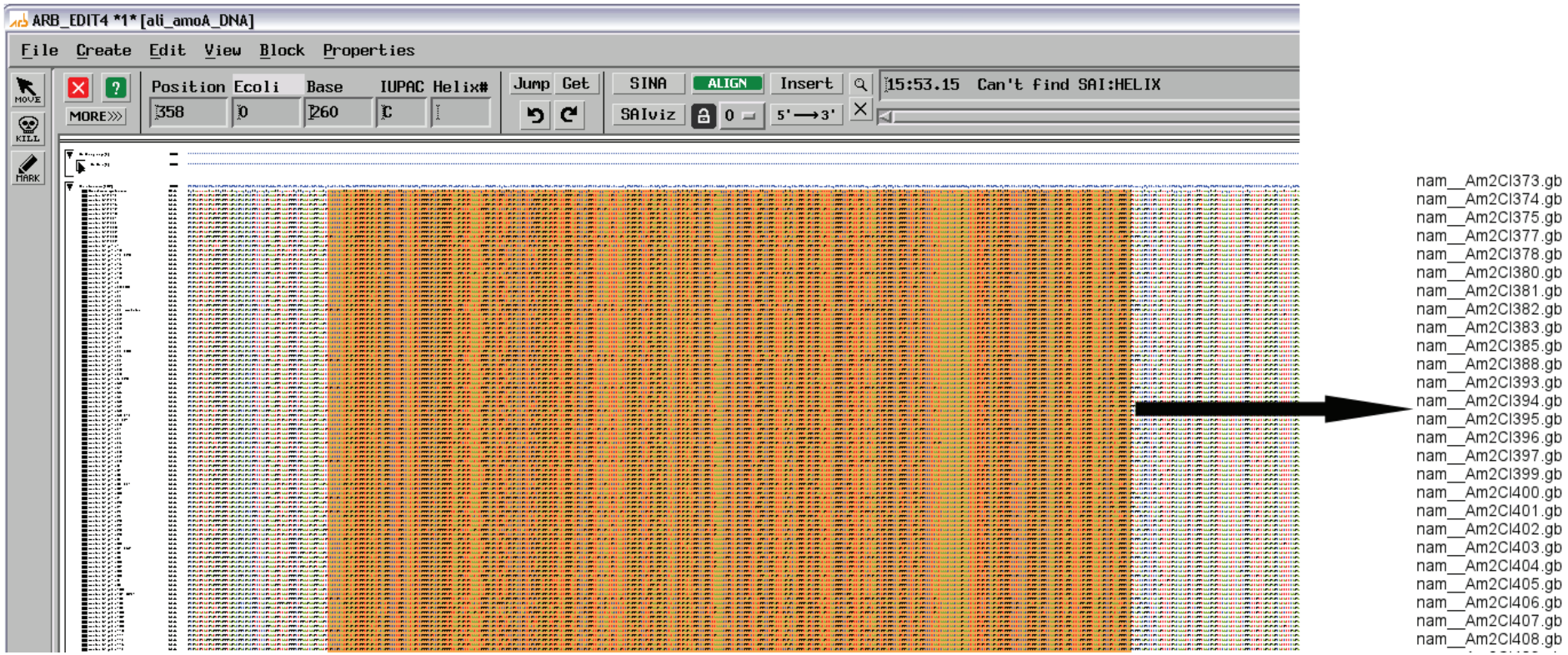
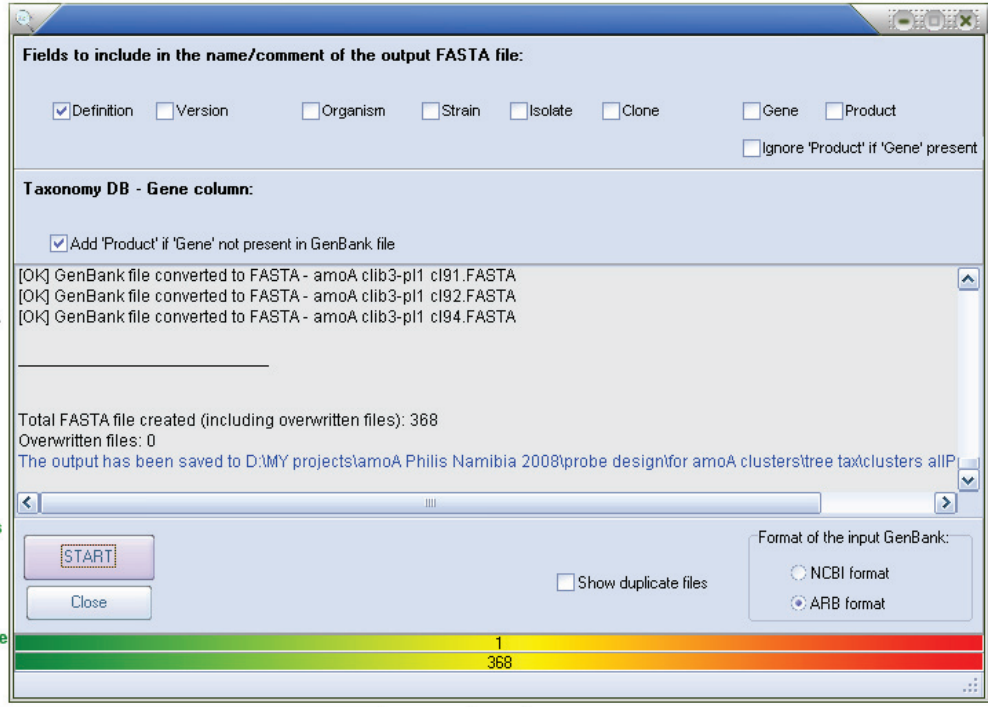


Figure 52: A nucleotide multiple alignment of *amo*A from Namibia seawater samples, performed with the integrated aligner from ARB. The target region for probe design is marked in orange (position 95-445 *Cenarchaeum symbiosum* numbering). The sequences are trimmed according to the marked region and exported in GenBank format, with phylogenetic information as inferred from the maximum likelihood tree (see **Figure 23**).

A: GenBank files with phylogenetic information in the SOURCE field.

```
LOCUS       Am2Cl499     351 bp ss-DNA          DNA
DEFINITION  amoA clib1-pl2 cl123
ACCESSION   GQ500246
SOURCE      Archaea;Crenarchaea;Marine;Cluster B; Namibia seawater 2;
REFERENCE   1
ORIGIN
1     CTGGAGATTA TATTTTCTAT ACTGACTGGG CCTGGACATC ATTTGTAGTA
51    TTCTCTATAG CTAACACGTT GATGACGGTA GTTGGAGCTG TATATTACAT
151   CACATTTACT GGTGTTCCTG GTACAGGTGC ATATTATGGC TTAATAATGC
201   AAGTCTATAC ATGGGTTGCAAAGGTTGCAT GGATGGCACT AGGTTATCCA
251   GTTGATTTCA TTGTTCATCC TATGTGGATA CCATCTTGTA TGCTATTGGA
301   CTTGGCATAT TGGGCGACAAAGAAGAATAA GCACTCGCTG ATATTCTTTG
351   GTGGCGTTTT AGTCGGAATG TCAATGCCGC TGTTTAACAT GGTACAATTG A
//
```

B: GTE module from PolyPro software transforms the GenBank files in FASTA files and creates the Taxonomy database using phylogenetic information from the SOURCE field of the GenBank files.



C: In the Taxonomy database the columns from Domain to Genus can be filled with information from the SOURCE field of the GenBank files. Since there is no information available for the amoA sequences regarding the taxonomic levels from Class to Genus, these columns are filled with tree derived information (See Figure 23). **The Strain column contains the name of the corresponding FASTA file (which is the same as the comment).**

Mismatch table | Taxonomy DB | Log

D:\MY projects\amoA Philis Namibia 2008\probe design\for amoA clusters\tree tax\clusters allPr\Taxonomy\Taxonomy.TAX

Gene	Domain	Phylum	Class	Order	Family	Genus	Species	Strain
	Archaea	Crenarchaea	Marine	Cluster B	Namibia seawater 3			amoA clib1-pl2 cl110
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl111
	Archaea	Crenarchaea	Marine	Cluster B	Namibia seawater 3			amoA clib1-pl2 cl113
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl114
	Archaea	Crenarchaea	Marine	Cluster B	Namibia seawater 3			amoA clib1-pl2 cl115
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl116
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl117
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl118
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl119
	Archaea	Crenarchaea	Marine	Cluster B	Namibia seawater 2			amoA clib1-pl2 cl123
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl125
	Archaea	Crenarchaea	Marine	Cluster A	Namibia seawater 5 + 6			amoA clib1-pl2 cl126

D: The output FASTA files have as comment the DEFINITION field from the GenBank files, as chosen in the GTE module.

```
> amoA clib1-pl2 cl123
CTGGAGATTATATTTTCTATACTGACTGGGCCTGGACATCATTTGTAGTA
TTCTCTATAGCTAACACGTTGATGACGGTAGTTGGAGCTGTATATTACAT
CACATTTACTGGTGTTCCTGGTACAGGTGCATATTATGGCTTAATAATGC
AAGTCTATACATGGGTTGCAAAGGTTGCATGGATGGCACTAGGTTATCC
AGTTGATTTCATTGTTCATCCTATGTGGATACCATCTTGTATGCTATTGG
ACTTGGCATATTGGGCGACAAAGAAGAATAAGCACTCGCTGATATTCTTT
GGTGGCGTTTTAGTCGGAATGTCAATGCCGCTGTTTAACATGGTACAATT
GA
```

Figure 53: The GenBank files (A) containing the trimmed sequences are transformed into FASTA files (D) using the GTE module (B) from the PolyPro software. The phylogenetic information from the SOURCE field of the GenBank files is used to create the Taxonomy database (C).

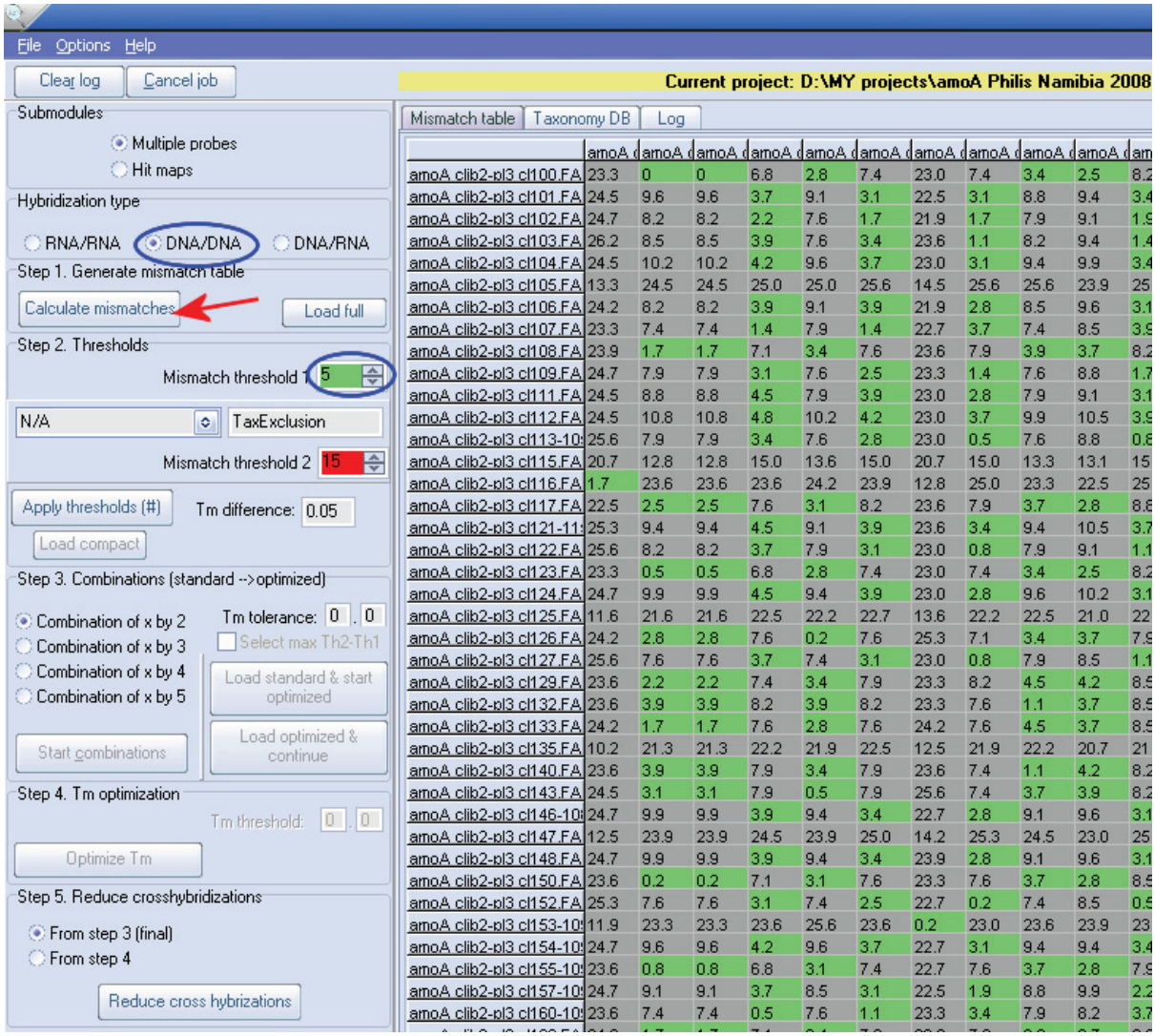


Figure 54: The PPD module from the PolyPro software receives the FASTA files generated by the GTE as input. All the *amo*A sequences are given both as probes and as targets. The hybridization type is set to DNA:DNA and a mismatch table is calculated between probes (horizontal header) and targets (vertical header). In the mismatch table, the cells with percentage mismatch (%MM) lower than mismatch threshold 1 (Th1) are marked in green.

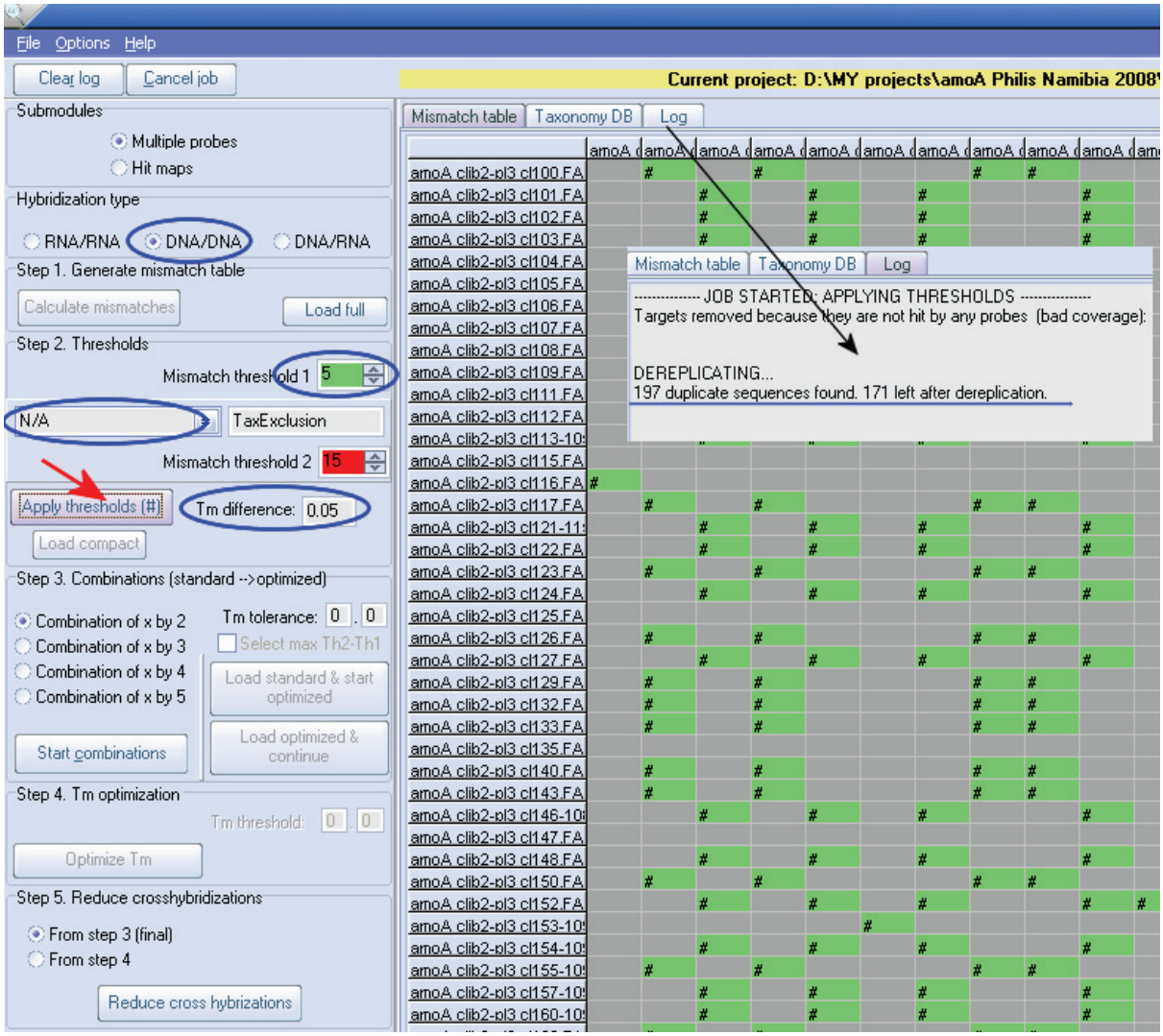


Figure 55: The threshold for a probe to hit a target (Th1) is set to ≤ 5% mismatch. No phylogenetic clade is selected. Therefore, the probes will be designed for all crenarchaeal *amo*A targets retrieved from the Namibian seawater samples. The probes with identical target groups and a T_m difference lower than 0.05 °C are considered replicates and only one of them will be kept. As a consequence, the number of probes decreases, as reported in the LOG. The mismatch table is transformed into a hit matrix, which will be used in the next step to calculate the probe mixes.

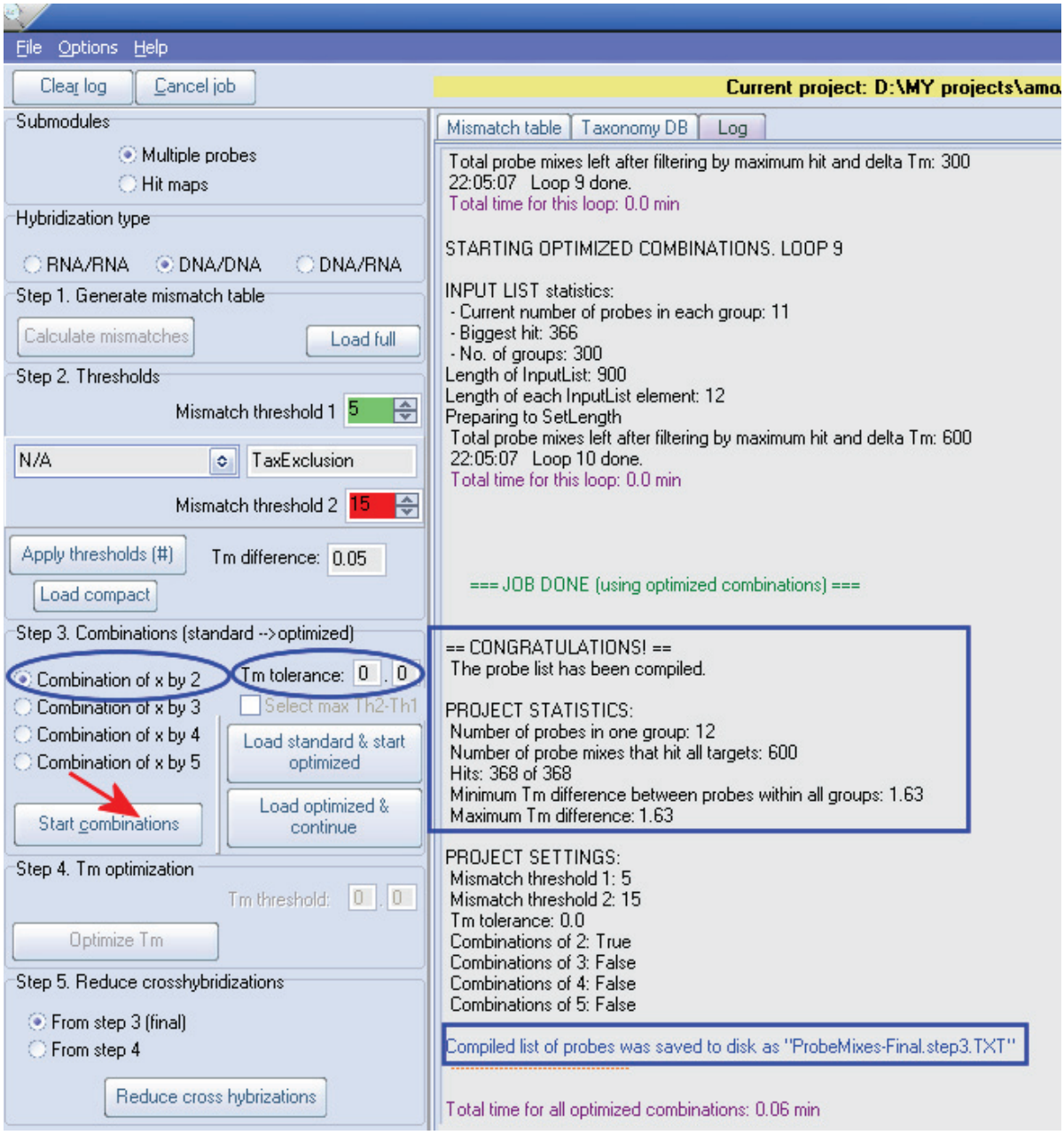


Figure 56: The probe mixes which hit all the targets are calculated. The combination algorithm is initiated by combinations of 2, followed by optimized combinations. Since the T_m tolerance parameter was set to 0, only the probe mixes with the lowest ΔT_m were selected during the combination algorithm. As a result, all the 600 probe mixes have the same ΔT_m (1.63 °C), as reported in the LOG.

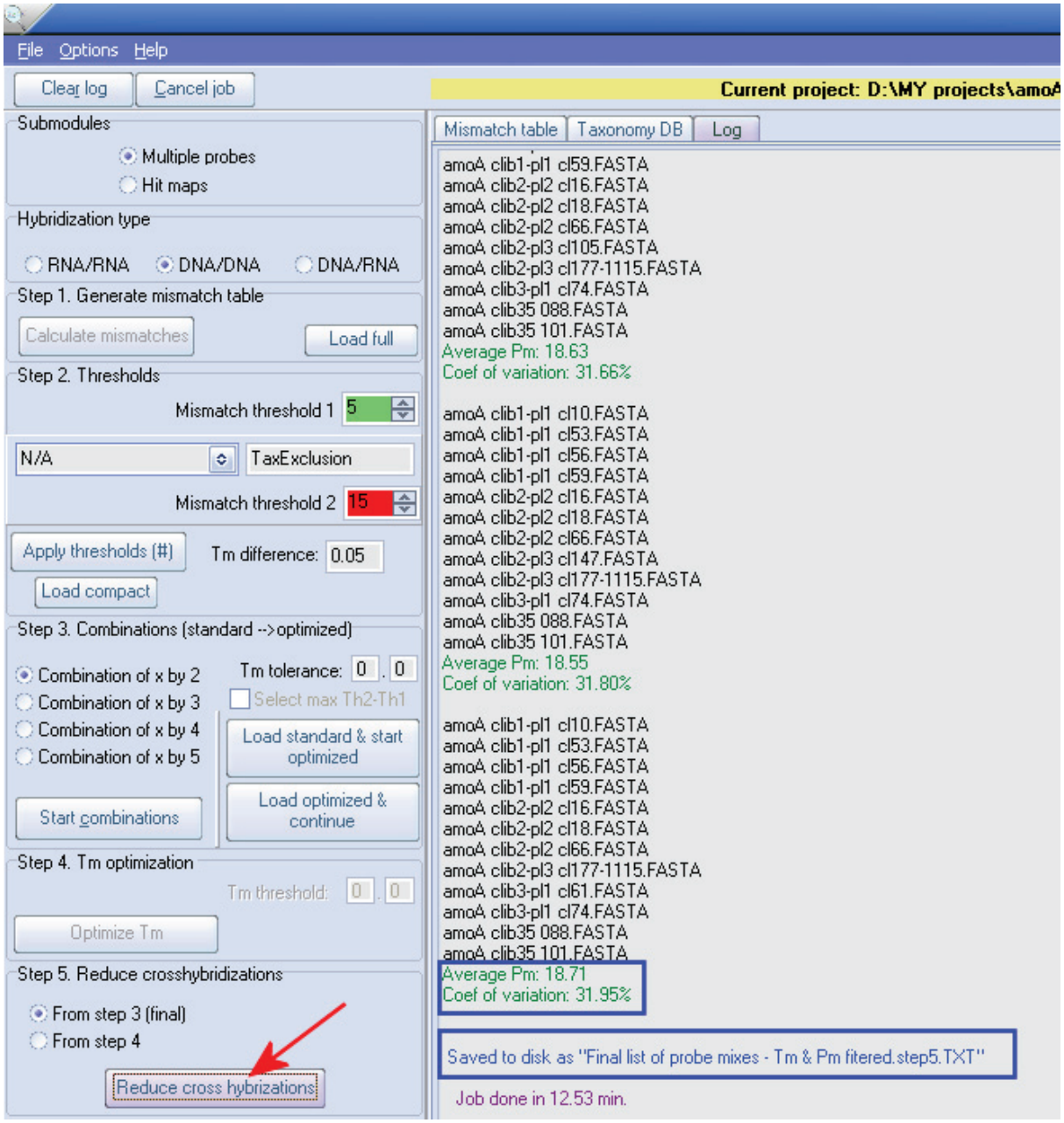


Figure 57: Step 5 was used to select probe mixes with dsDNA probes less likely to crosshybridize. From all probe mixes resulting in step 3, only the ones with the lowest similarity between the probes were selected.

amoA polynucleotide probes:

>amoA-1E3 probe

CTGGTGACTATATCTTCTATACTGATTGGGCCTGGACATCATTTGTAGTATTTTCCATTG
CCAACACTTTGATGACTGTTGTTGGTGCTGTGTATTATCTTACATTTACGGGTGTACCTG
GTACTGCATCGTATTACGGTCTGATTATGCAGGTCTATACATGGGTTGCAAAAGTTGCA
TGGTTTGCACTTGGTTACCCGGTGGATTTCATCGTTCATCCGATGTGGATTCCATCTTG
TATGTTATTGGACTTGGCATATTGGGCGACGAAGAAGAATAAGCACTCGCTGATATTCT
TTGGCGGAGTTTTAGTTGGAATGTCAATGCCACTGTTCAATATGGTACAGTTGA

***amo*A-Nam probe mix**

>amoA-Nam mix, probe amoA clib1-pl1 cl10

CAGGAGACTACATCTTCTATACTGATTGGGCCTGGACATCGTTTGTAGTATTCTCAATAT
CTCAAACATTGATGCTTGCGGTAGGTGCAGCATATTATCTGACATTTACTGGAGTTCCA
GGAACCGCAACGTATTATGCGCTGATTATGACAGTCTATACTTGGATTGCTAAAGGTGC
ATGGTTCGCATTAGGTTACCCATACGACTTCATCGTTACACCAGTTTGGCTTCCGTCAG
CAATGCTGTTGGATCTGGCTTACTGGGCGACAAAGAAGAATAAGCACTCCTTGATACTG
TTCGGCGGAGTCTTAGTTGGAATGTCATTACCACTATTCAACATGGTCAACTTGA

>amoA-Nam mix, probe amoA clib1-pl1 cl34

CGGGAGACTATATCTTCTACACTGATTGGGCTTGGACATCATTTGTCGTATTCTCAATTG
CCAACACTTTGATGACTATTGTTGGTGCTGTGTATTATCTTACATTTACAGGCGTACCTG
GTACTGCAACGTACTATGGTCTGATTATGCAAGTCTATACTTGGGTTGCAAAAGTTACAT
GGTTTGCACTTAACTATCCAGTAGATTTCATCGTTCATCCAATGTGGATTCCATCGTGTA
TGCTGTTGGACTTGGCATATTGGGCGACAAAAAAGAATAAACACTCGCTGATATTCTTT
GGTGGGGTTTTAGTTGGAATGTCAATGCCACTGTTCAACATGGTACAGTTGA

>amoA-Nam mix, probe amoA clib1-pl1 cl53

CAGGAGACTATATCTTCTACACTGATTGGGCCTGGACATCGTTTGTAGTATTTTCAATAT
CTCAAACATTGATGCTCGCTGTAGGTGCAGCATATTATCTGACCTTCACTGGAGTTCCA
GGAACCGCAACGTATTATGCGCTTATCATGACGGTCTATACGTGGATTGCCAAAGGTGC
ATGGTTCGCATTAGGTTACCCATACGACTTTATTGTTACACCAGTTTGGCTACCGTCAG
CAATGCTGTTGGACTTAGCGTACTGGGCAACAAAGAAGAATAAGCACTCTCTGATACTG
TTCGGCGGAGTCTTAGTTGGAATGTCGTTGCCACTATTCAACATGGTCAATTTGA

>amoA-Nam mix, probe amoA clib1-pl1 cl54

CAGGAGACTATATCTTCTATACTGACTGGGCTTGGACATCATTCGTAGTTTTCTCTATCG
CCAACACTTTGATGACGGTAGTCGGAGCTGTGTATTACATTACATTTACTGGTGTTCCT
GGTACAGGCGCGTATTACGGCTTGATAATGCAGGTCTATACATGGGTCGCTAAAGTTG
CATGGTTTGCACTTGGTTATCCGGTGGATTTCATTGTTCATCCAATGTGGATTCCATCAT
GCATGTTATTGGATTTGGCATATTGGGCTATGAAGAAGAATAAGCACTCGCTGATATTCT
TTGGTGGAGTTTTAGTGGGAATGTCAATGCCGCTCTTCAATATAGTACAATTGA

>amoA-Nam mix, probe amoA clib1-pl1 cl56

CAGGGGACTACATCTTCTATACTGATTGGGCCTGGACATCGTTTGTAGTATTCTCAATAT
CTCAAACATTAATGCTCACTGTAGGTGCGTGTTACTATCTCACCTTCACTGGAGTTCCA
GGAACCGCAACGTATTATGCACTAATTATGACAGTCTACACTTGGATTGCAAAAGGTGC
ATGGTTCGCATTAGGTTACCCATATGACTTCATCGTTACACCAGTTTGGCTACCATCAG
CAATGCTGTTGGACTTGGCGTACTGGGCAACAAAGAAGAATAAGCACTCTCTGATACTG
TTCGGCGGAGTCTTAGTTGGAATGTCATTACCACTATTCAACATGGTCAATTTGA

>amoA-Nam mix, probe amoA clib1-pl1 cl59

CTGGTGACTATATCTTCTATACTGATTGGGCCTGGACATCATTTGTAGTATTTTCCATTG
CCAACACTTTGATGACTGTTGTTGGTGCTGTGTATTATCTTACATTTACTGGTGTACCTG
GTACTGCATCGTATTACGGTCTGATTATGCAGGTCTATACATGGGTTGCAAAAGTTGCA
TGGTTTGCACTTGGTTACCCAGTGGATTTCATCGTTCATCCGATGTGGATTCCATCTTGT
ATGTTATTGGACTTGGCATATTGGGCGACGAAGAAGAATAAGCACTCGCTGATATTCTT
TGGCGGAGTTTTAGTTGGAATGTCAATGCCACTGTTCAATATGGTACAGTTGA

>amoA-Nam mix, probe amoA clib1-pl1 cl75

CAGGAGATTATATCTTCTATACTGACTGGGCTTGGACATCGTTTGTAGTATTCTCAATAT
CTCAGACGTTGATGCTTGCAGTAGGTGCAGCATACTATCTCACATTTACTGGAGTTCCA
GGTACGGCGACATATTATGCGCTAATTATGACAGTATATACTTGGATTGCAAAAGGTGC
ATGGTTTGCACTTGGTTATCCATATGACTTTATTGTCACACCAGTCTGGTTGCCATCAGC
AATGCTATTAGATTTGGCATACTGGGCGACGAAGAAGAATAAGCACTCGCTGATACTGT
TTGGTGGAGTACTAATTGGAATGTCACTGCCATTATTCAATATGGTAAACCTGA

>amoA-Nam mix, probe amoA clib1-pl2 cl106

CAGGAGACTATATCTTCTATACAGACTGGGCTTGGACATCATTTGTGGTATTTTCCATAG
CTAATACTCTGATGACAGTAGTTGGAGCTGTTTACTATATCACGTTTACAGGTGTTCCAG
GCACAGGCGCATATTATGGCCTGATTATGCAAGTCTATACTTGGGTCGCTAAAGTTGCA

TGGTTTGCACTTGGCTATCCTGTGGATTTCATTGTTCATCCAATGTGGATTCCATCTTGT
ATGCTATTGGACTTGGCGTATTGGGCTACAAAGAAGAATAAACACTCGCTGATATTCTTT
GGCGGAGTTTTAGTCGGAATGTCGATGCCATTATTCAACATGGTACAGTTGA

>amoA-Nam mix, probe amoA clib1-pl2 cl145

CAGGAGACTACATCTTTTACACTGACTGGGCTTGGACATCGTATACGGTATTTTCAATAT
CGCAAACTTTGATGCTTATTGTAGGAGCAACATATTATCTTACATTTACTGGCGTTCCAG
GCACAGCAACGTACTACGCTCTAATTATGACAGTATACACATGGATAGCAAAAGGTGCA
TGGTTTGCACTCGGATATCCATATGACTTCATTGTAACTCCAGTTTGGTTACCATCAGCA
ATGCTTTTGGATTTAGTCTACTGGGCAACAAAGAAGAACAAGCACTCCTTGATACTGTTT
GGCGGCGTACTGGTAGGAATGTCTTTACCATTATTCAACATGGTAAACCTGA

>amoA-Nam mix, probe amoA clib2-pl2 cl66

CTGGTGACTATATCTTCTATACTGATTGGGCCTGGACGTCTTTCGTAGTATTTTCTATAG
CTAACACTTTGATGACAGTTGTCGGTGCTGTGTATTATCTTACATTTACTGGAGTACCGG
GTACTGCATCGTATTACGGTTTAATTATGCAAGTCTATACATGGGTTGCAAAAGTTGCAT
GGTTTGCACTTGGTTACCCAGTGGATTTCATTGTTCATCCAATGTGGATTCCATCGTGTA
TGTTATTGGACTTGGCATATTGGGCGACAAAGAAGAATAAACACTCGCTGATATTCTTTG
GCGGAGTTTTAGTAGGAATGTCAATGCCACTGTTCAATATGGTACAGTTGA

>amoA-Nam mix, probe amoA clib2-pl3 cl105

CTGGAGACTATATCTTCTATACAGATTGGGCATGGACTTCCTTTGTAGTTTTCTCCATTG
CCAATACTCTCATGACAGTCGTCGGAGCAGTATATTATATCACATTTACAGGCGTACCT
GGAACGGGCGCGTATTATGGTTTAATTATGCAAGTCTATACTTGGGTTGCCAAAGTGGC
ATGGTATGCACTTGGCTATCCTGCAGATTTCATCGTTCATCCAATGTGGATTCCATCATG
CATGTTGTTGGATTTGGCATATTGGGCTACGAAGAAGAATAAGCATTCGCTGATATTCTT
TGGAGGAGTTTTGGTCGGTATGTCAATGCCGCTGTTCAATATGGTACAGCTGA

>amoA-Nam mix, probe amoA clib3-pl1 cl74

CTGGAGATTATATTTTCTATACTGACTGGGCCTGGACATCATTTGTAATATTCTCTATAG
CTAACACGTTGATGACGGTAGTTGGAGCTGTATATTACATCACATTTACTGGTGTTCCTG
GTACAGGTGCATATTATGGCTTAATAATGCAAGTCTATACGTGGGTTGCAAAGGTTGCA
TGGATGGCACTGGGCTATCCAGTTGATTTCATTGTTCATCCTATGTGGATACCATCGTG
TATGCTATTGGATTTGGCATATTGGGCGACAAAGAAGAATAAGCATTCGCTGATATTCTT
TGGTGGAGTTTTAGTCGGAATGTCGATGCCGCTATTCAACATGGTACAGTTGA

Publication list

The work presented in this thesis is part of several manuscripts. Two of the manuscripts are submitted, while the others are in preparation.

1. **Moraru, C.**, Moraru, G., Fuchs, B.M., and Amann, R. (submitted) Concepts and software for a rational design of polynucleotide probes. *Environ. Microbiol.*
   C.M. contribution: polynucleotide probe design concept, software and algorithm design, validation and testing, case studies, manuscript writing.

2. **Moraru, C.**, Lam, P., Fuchs, B.M., Kuypers, M.M.M. and Amann, R. (submitted) GeneFISH – an *in situ* technique for linking gene presence and cell identity in environmental microorganisms. *Environ. Microbiol.*
   C.M. contribution: amoA clone libraries preparation and sequencing, sequence analysis, probe design and synthesis, T_m measurements and calculation of hybridization specificity conditions, rRNA FISH, geneFISH protocol concept, development and application, manuscript writing.

3. **Moraru, C.**, Musat, N., Polareky, L., Mohr, W., LaRoche J., Wagner, T., Kuypers, M.M.M. and Amann, R. (in prep) mRNA-HISH SIMS – an *in situ* technique for detection and quantification of mRNA in environmental microorganisms, at single cell level.
   C. Moraru and N. Musat have contributed equally both to the concept development and to the practical work. C.M. contribution: sample collection and fixation, nifH clone library, probe design and synthesis, T_m measurements and calculation of hybridization specificity conditions, mRNA-HISH SIMS protocol concept, data analysis and manuscript writing.

4. Lenk, S., **Moraru, C.**, Krupke, A., Harder, J., Kube, M., Amann, R., and Mussmann, M. (in prep) dsrAB carrying genome fragment reveals a contribution to sulfur compound oxidation by marine sedimentary *Roseobacter* - Insights from metagenomics, enrichments and FISH.
   *C.M. contribution: gene probe design and synthesis, geneFISH for the rdsr*A *gene.*

5. Zielinski, F. U., Petersen, J. M., **Moraru, C.**, Wetzel, S., Pape, T., Seifert, R., Wenzhöfer, F., Amann, R. and Dubilier, N. (in prep) The sulfur-oxidizing endosymbiont of the hydrothermal vent mussel *Bathymodiolus puteoserpentis* (*Bivalvia: Mytilidae*) uses hydrogen as an energy source.
   C.M. contribution: analysis of hynL clone libraries, gene probe design and synthesis, geneFISH for the hynL gene.

Other contributions

1. Ploug, H., Musat, N., Adam, B., **Moraru, C.**, Lavik, G., Bergman, B., and Kuypers, M.M.M.. (submitted) Carbon and nitrogen fluxes associated with *Aphanizomenon* sp. in the Baltic Sea. *ISME-J.*
   C.M. contribution: sample collection, fixation and preparation, incubation experiments.

Acknowledgements

To my academic supervisor, Prof. Dr. Rudolf Amann, my special thanks for entrusting me with this work. Thanks for continuous support during the years, for invaluable guidance and discussions.

To my second academic supervisor, Dr. Bernhard Fuchs, my thanks for guiding me through my PhD work and bearing with me through its final stages.

There are many people which have contributed to my work presented in this thesis. Thanks to Gabriel Moraru, for programming the PolyPro software. Thanks to Phyllis Lam, for preparation of the Namibian seawater samples and for her numerous advises during my work here. Thanks to Daniela Franzke for performing DNA extractions from seawater samples. My HIWI, Cristiana Lungu, thanks for her help with sequencing the clone libraries. To Jillian Petersen and Silke Wetzel, thanks for sharing their samples of *Bathymodiolus puteoserpentis*, for performing the DNA extractions and preparation of *hyn*L and *apr*A clone libraries, for performing the rRNA CARD-FISH step on *Bathymodiolus puteoserpentis* and preparing the DNA fosmid extracts used for probe synthesis. To Sabine Lenk and Marc Mussmann, thanks for sharing and preparing the AK199 enrichment samples and the fosmid samples, for performing the rRNA CARD-FISH step on AK199 and for preparing the DNA fosmid extracts used for probe synthesis. To Gordana Ivcic, thanks for help with preparation of the *dsr*A and *hme*D probes. To Niculina Musat, thanks for working close with me on the concept rRNA HISH-SIMS method, for performing the mRNA HISH-SIMS and the NanoSISM analysis. To Wiebke Mohr, thanks for preparing and sharing the *Crocosphaera watsonii* cultures. To Helle Ploug and Birgit Adam, thanks for the help with Baltic seawater sample collection. To Jörg Wulf and to all the members of the molecular ecology department for continuous support in the lab.

I would like to give thanks the person who has first opened the life of research for me - Ileana Stoica, my preferred undergrad professor. Thank you for being my role model and inspiration, not only in science, but also in life.

Special thanks to my parents for their support and for the courage with which they endured my time away from the home land. Thanks to Gabriel Moraru, my husband and best friend, for his continuous support and invaluable help.

References

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**: 8966-8969.

Agogue, H., Brink, M., Dinasquet, J., and Herndl, G.J. (2008) Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. *Nature* **456**: 788-791.

Amann, R., and Fuchs, B.M. (2008) Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol* **6**: 339-348.

Amann, R.I., Zarda, B., Stahl, D.A., and Schleifer, K.H. (1992) Identification of individual prokaryotic cells by using enzyme-labeled, rRNA-targeted oligonucleotide probes. *Appl Environ Microbio* **58**: 3007-3011.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**: 1919-1925.

Amasino, R.M. (1986) Acceleration of nucleic acid hybridization rate by polyethylene glycol. *Anal Biochem* **152**: 304-307.

Anderson, M.L.M., and Young, B.D. (1985) Quantitative filter hybridisation. In *Nucleic acid hybridisation - a practical approach*. Hames, B.D., and Higgins, S.J. (eds). Oxford: Oxford University Press, pp. 73-110.

Banér, J., Nilsson, M., Mendel-Hartvig, M., and Landegren, U. (1998) Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Research* **26**: 5073-5078.

Behrens, S., Losekann, T., Pett-Ridge, J., Weber, P.K., Ng, W.-O., Stevenson, B.S. et al. (2008) Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence *in situ* hybridization (EL-FISH) and NanoSIMS. *Appl Environ Microbio* **74**: 3143-3150.

Birnstiel, M.L., Sells, B.H., and Purdom, I.F. (1972) Kinetic complexity of RNA molecules. *J Mol Biol* **63**: 21-39.

Bishop, J. (1972) Molecular hybridization of ribonucleic acid with a large excess of deoxyribonucleic acid. *Biochem J* **126**: 171-185.

Bjorndal, M.T., and Fygenson, D.K. (2002) DNA melting in the presence of fluorescent intercalating oxazole yellow dyes measured with a gel-based assay. *Biopolymers* **65**: 40-44.

Blake, R., Bizzaro, J., Blake, J., Day, G., Delcourt, S., Knowles, J. et al. (1999) Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics* **15**: 370-375.

Blüthmann, H., Brück, D., Hübner, L., and Schöffski, A. (1973) Reassociation of nucleic acids in solutions containing formamide. *Biochem Biophys Res Commun* **50**: 91-97.

Bodkin, D.K., and Knudson, D.L. (1985) Assessment of sequence relatedness of double-stranded RNA genes by RNA-RNA blot hybridization. *J Virol Methods* **10**: 45-52.

Bonner, J., Kung, G., and Bekhor, I. (1967) A method for the hybridization of nucleic acid molecules at low temperature. *Biochemistry* **6**: 3650-3653.

Bonner, T.I., Brenner, D.J., Neufeld, B.R., and Britten, R.J. (1973) Reduction in the rate of DNA reassociation by sequence divergence. *J Mol Biol* **81**: 123-135.

Bothe, H., Jost, G., Schloter, M., Ward, B.B., and Witzel, K.-P. (2000) Molecular analysis of ammonia oxidation and denitrification in natural environments. *FEMS Microbiol Rev* **24**: 673-690.

Brahic, M., and Haase, A.T. (1978) Detection of viral sequences of low reiteration frequency by *in situ* hybridization. *Proc Natl Acad Sci USA* **75**: 6125-6129.

Breuert, S., Allers, T., Spohn, G., and Soppa, J. (2006) Regulated polyploidy in halophilic *Archaea*. *PLoS One* **1**: e92.

Britten, R.J., and Davidson, E.H. (1985) Hybridisation strategy. In *Nucleic acid hybridisation - a practical approach*. Hames, B.D., and Higgins, S.J. (eds). Oxford: Oxford University Press, pp. 3-16.

Caccone, A., DeSalle, R., and Powell, J. (1988) Calibration of the change in thermal stability of DNA duplexes and degree of base pair mismatch. *J Mol Evol* **27**: 212-216.

Casey, J., and Davidson, N. (1977) Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res* **4**: 1539-1552.

Cerqueira, L., Azevedo, N.F., Almeida, C., Jardim, T., Keevil, C.W., and Vieira, M.J. (2008) DNA mimics for the rapid identification of microorganisms by fluorescence *in situ* hybridization (FISH). *Int J Mol Sci* **9**.

Chang, C.-T., Hain, T.C., Hutton, J.R., and Wetmur, J.G. (1974) Effects of microscopic and macroscopic viscosity on the rate of renaturation of DNA. *Biopolymers* **13**: 1847-1858.

Chen, Y.-B., Zehr, J.P., and Mellon, M. (1996) Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. IMS 101 in defined media: evidence for a circadian rhythm. *Journal of Phycology* **32**: 916-923.

Chien, Y.H., and Davidson, N. (1978) RNA:DNA hybrids are more stable than DNA:DNA duplexes in concentrated perchlorate and trichloroacetate solutions. *Nucleic Acids Res* **5**: 1627-1637.

Collins, M.L., Irvine, B., Tyner, D., Fine, E., Zayati, C., Chang, C. et al. (1997) A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res* **25**: 2979-2984.

Condon, C. (2003) RNA processing and degradation in *Bacillus subtilis*. *Microbiol Mol Biol Rev* **67**: 157-174.

Csákia, R., Hanczára, T., Bodrossya, L., Murrellc, J.C., and Kovács, K.L. (2001) Molecular characterization of structural genes coding for a membrane bound hydrogenase in *Methylococcus capsulatus* (Bath) *FEMS Micobiology Letters* **205**: 203-207.

Daum, M., Zimmer, W., Papen, H., Kloos, K., Nawrath, K., and Bothe, H. (1998) Physiological and molecular biological characterization of ammonia oxidation of the heterotrophic nitrifier *Pseudomonas putida*. *Curr Microbiol* **37**: 281-288.

DeLong, E.F., Taylor, L.T., Marsh, T.L., and Preston, C.M. (1999) Visualization and enumeration of marine planktonic *Archaea* and *Bacteria* by using polyribonucleotide probes and fluorescent *in situ* hybridization. *Appl Environ Microbiol* **65**: 5554-5563.

Deutscher, M.P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res* **34**: 659-666.

Dunfield, P.F., Yuryev, A., Senin, P., Smirnova, A.V., Stott, M.B., Hou, S. et al. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**: 879-882.

Duperron, S., Nadalig, T., Caprais, J.-C., Sibuet, M., Fiala-Medioni, A., Amann, R., and Dubilier, N. (2005) Dual symbiosis in a *Bathymodiolus sp.* mussel from a methane seep on the Gabon continental margin (Southeast Atlantic): 16S rRNA phylogeny and distribution of the symbionts in gills. *Appl Environ Microbio* **71**: 1694-1700.

Duperron, S., Bergin, C., Zielinski, F., Blazejak, A., Pernthaler, A., McKiness, Z.P. et al. (2006) A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ Microbiol* **8**: 1441-1447.

Ehrhardt, C.J., Haymon, R.M., Sievert, S.M., and Holden, P.A. (2009) An improved method for nanogold *in situ* hybridization visualized with environmental scanning electron microscopy. *J Microsc* **236**: 5-10.

Eilers, H., Pernthaler, J., Peplies, J., Glöckner, F.O., Gerdts, G., and Amann, R. (2001) Isolation of novel pelagic Bacteria from the German Bight and their seasonal contributions to surface picoplankton. *Appl Environ Microbio* **67**: 5134-5142.

Flynt, A., Liu, N., Martin, R., and Lai, E.C. (2009) Dicing of viral replication intermediates during silencing of latent Drosophila viruses. *Proc Natl Acad Sci USA* **106**: 5270-5275.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity and diversity of ammonia-oxidizing Archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* **102**: 14683-14688.

Friedrich, M.W. (2002) Phylogenetic analysis reveals multiple lateral transfers of adenosine-5'-phosphosulfate reductase genes among sulfate-reducing microorganisms. *J Bacteriol* **184**: 278-289.

Friedrich, R., and Feix, G. (1972) RNA-RNA hybridization in aqueous solutions containing formamide. *Anal Biochem* **50**: 467-476.

Fritz, G., Buchert, T., Huber, H., Stetter, K.O., and Kroneck, P.M.H. (2000) Adenylylsulfate reductases from Archaea and Bacteria are 1:1 alphabeta-heterodimeric iron-sulfur flavoenzymes-high similarity of molecular properties emphasizes their central role in sulfur metabolism. *FEBS Lett* **473**: 63-66.

Galau, G.A., Britten, R.J., and Davidson, E.H. (1977a) Studies on nucleic acid reassociation kinetics: Rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation. *Proc Natl Acad Sci USA* **74**: 1020-1023.

Galau, G.A., Smith, M.J., Britten, R.J., and Davidson, E.H. (1977b) Studies on nucleic acid reassociation kinetics: Retarded rate of hybridization of RNA with excess DNA. *Proc Natl Acad Sci USA* **74**: 2306-2310.

Gérard, E., Guyot, F., Philippot, P., and López-García, P. (2005) Fluorescence *in situ* hybridisation coupled to ultra small immunogold detection to identify prokaryotic cells using transmission and scanning electron microscopy. *J Microbiol Methods* **63**: 20-28.

Gillespie, S., and Gillespie, D. (1971) Ribonucleic acid–deoxyribonucleic acid hybridization in aqueous solutions and in solutions containing formamide. *Biochem Journal* **125**: 481-487.

Gordon, J.A. (1965) Stabilizing effectiveness of alkali metals toward ribonucleic acid from the bacterial virus R17. *Biopolymers* **3**: 5-14.

Gruenwedel, D.W., Hsu, C.-H., and Lu, D.S. (1971) The effects of aqueous neutral-salt solution on the melting temperatures of deoxyribonucleic acids. *Biopolymers* **10**: 47-68.

Gudnason, H., Dufva, M., Bang, D.D., and Wolff, A. (2007) Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic Acids Res* **35**: e127.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95-98.

Hamaguchi, K., and Geiduschek, E.P. (1962) The effect of electrolytes on the stability of the deoxyribonucleate helix. *J Am Chem Soc* **84**: 1329-1338.

Hanaki, H., Odawara, T., Nakajima, N., Shimizu, Y.K., Nozaki, C., Mizuno, K. et al. (1998) Two different reactions involved in the primer/template-independent polymerization of dATP and dTTP by Taq DNA polymerase. *Biochem Biophys Res Commun* **244**: 210-219.

Hanaki, K., Odawara, T., Muramatsu, T., Kuchino, Y., Mizuno, K., and Yoshikura, H. (1997) Primer/template-independent synthesis of poly d(A-t) by Taq polymerase. *Biochem Biophys Res Commun* **238**: 113-118.

Hodson, R.E., Dustman, W.A., Garg, R.P., and Moran, M.A. (1995) *In situ* PCR for visualization of microscale distribution of specific genes and gene products in prokaryotic communities. *Appl Environ Microbiol* **61**: 4074-4082.

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

Holtke, H.-J., and Kessler, C. (1990) Non-radioactive labeling of RNA transcripts in vitro with the hapten digoxigenin (DIG); hybridization and ELISA-based detection. *Nucleic Acids Res* **18**: 5843-5851.

Hopman, A.H.N., Ramaekers, F.C.S., and Speel, E.J.M. (1998) Rapid synthesis of biotin-, digoxigenin-, trinitrophenyl-, and fluorochrome-labeled tyramides and their application for *in situ* hybridization using CARD amplification. *J Histochem Cytochem* **46**: 771-778.

Hoshino, T., Noda, N., Tsuneda, S., Hirata, A., and Inamori, Y. (2001) Direct detection by *in situ* PCR of the amoA gene in biofilm resulting from a nitrogen removal process. *Appl Environ Microbio* **67**: 5261-5266.

Hoshino, T., Yilmaz, L.S., Noguera, D.R., Daims, H., and Wagner, M. (2008) Quantification of target molecules needed to detect microorganisms by fluorescence *in situ* hybridization (FISH) and catalyzed reporter deposition - FISH. *Appl Environ Microbiol* **74**: 5068-5077.

Hutton, J.R. (1977) Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic Acids Research* **4**: 3538-3555.

Hutton, J.R., and Wetmur, J.G. (1973a) Renaturation of bacteriophage φX174 DNA-RNA hybrid: RNA length effect and nucleation rate constant. *J Mol Biol* **77**: 495-500.

Hutton, J.R., and Wetmur, J.G. (1973b) Effect of chemical modification on the rate of renaturation of deoxyribonucleic acid. Deaminated and glyoxalated deoxyribonucleic acid. *Biochemistry* **12**: 558-563.

Jacobs, D., Angles, M.L., Goodman, A.E., and Neilan, B.A. (1997) Improved methods for in situ enzymatic amplification and detection of low copy number genes in bacteria. *FEMS Microbiology Ecology* **152**: 65-73.

Jaeger, J.A., Zuker, M., and Turner, D.H. (1990) Melting and chemical modification of a cyclized self-splicing group I intron: similarity of structures in 1 M Na^+, in 10 mM Mg^{2+}, and in the presence of substrate. *Biochemistry* **29**: 10147-10158.

Jain, C. (2002) Degradation of mRNA in *Escherichia coli*. *IUBMB Life* **54**: 315-321.

Jensen, R.B., and Shapiro, L. (1999) The *Caulobacter crescentus smc* gene is required for cell cycle progression and chromosome segregation. *Proc Natl Acad Sci USA* **96**: 10661-10666.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**: W5-W9.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507-510.

Kenzaka, T., Tamaki, S., Yamaguchi, N., Tani, K., and Nasu, M. (2005a) Recognition of individual genes in diverse microorganisms by cycling primed *in situ* amplification. *Appl Environ Microbiol* **71**: 7236–7244.

Kenzaka, T., Ishidoshiro, A., Yamaguchi, N., Tani, K., and Nasu, M. (2005b) rRNA sequence-based scanning electron microscopicdetection of Bacteria. *Appl Environ Microbio* **71**: 5523-5531.

Kenzaka, T., Tani, K., Sakotani, A., Yamaguchi, N., and Nasu, M. (2007) High-frequency phage-mediated gene transfer among *Escherichia coli* cells, determined at the single-cell level. *Appl Environ Microbiol* **73**: 3291-3299.

Kierzek, R. (1992) Nonenzymatic hydrolysis of oligoribonucleotides. *Nucleic Acids Res* **20**: 5079-5084.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543-546.

Kubitschek, H.E., and Freedman, M.L. (1971) Chromosome replication and the division cycle of *Escherichia coli J Bacteriol* **107**: 95-99.

Kubota, K., Ohashi, A., Imachi, H., and Harada, H. (2006) Visualization of *mcr* mRNA in a methanogen by fluorescence *in situ* hybridization with an oligonucleotide probe and two-pass tyramide signal amplification (two-pass TSA-FISH). *J Microbiol Methods* **66**: 521-528.

Kulesza, C.A., and Shenk, T. (2006) Murine cytomegalovirus encodes a stable intron that facilitates persistent replication in the mouse. *Proc Natl Acad Sci USA* **103**: 18302-18307.

Lam, P., Jensen, M.M., Lavik, G., McGinnis, D.F., Müller, B., Schubert, C.J. et al. (2007) Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. *Proc Natl Acad Sci USA* **104**: 7104-7109.

Lasken, R.S. (2007) Single-cell genomic sequencing using Multiple Displacement Amplification. *Current Opinion in Microbiology* **10**: 510-516.

Lenk, S., Moraru, C., Krupke, A., Harder, J., Kube, M., Amann, R., and Mussmann, M. (in prep) *dsr*AB carrying genome fragment reveals a contribution to sulfur compound oxidation by marine sedimentary *Roseobacter* - Insights from metagenomics, enrichments and FISH.

Lesnik, E.A., and Freier, S.M. (1995) Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**: 10807-10815.

Li, T., Wu, T.-D., Mazéas, L., Toffin, L., Guerquin-Kern, J.-L., Leblon, G., and Bouchez, T. (2008) Simultaneous analysis of microbial identity and function using NanoSIMS. *Environ Microbiol* **10**: 580-588.

Lichtman, J.W., and Conchello, J.-A. (2005) Fluorescence microscopy. *Nature Methods* **2**: 910-919.

Lovmar, L., and Syvänen, A.-C. (2006) Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. *Human mutation* **27**: 603-614.

Ludwig, W., Dorn, S., Springer, N., Kirchhof, G., and Schleifer, K.H. (1994) PCR-based preparation of 23S rRNA-targeted group-specific polynucleotide probes. *Applied and Environmental Microbiology* **69**: 3236-3244.

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

Lundgren, M., Malandrin, L., Eriksson, S., Huber, H., and Bernander, R. (2008) Cell cycle characteristics of Crenarchaeota: unity among diversity. *The Journal of Bacteriology* **190**: 5362-5367.

Manz, W., Amann, R., Ludwig, W., Wagner, M. & Scheifer, K. H. (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of Proteobacteria: problems and solutions. Syst. Appl. Microbiol. 15, 93-600.

Marmur, J., and Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J Mol Biol* **5**: 109-118.

Martens-Habbena, W., Berube, P.M., Urakawa, H., de la Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**: 976-979.

Maruyama, F., Kenzaka, T., Yamaguchi, N., Tani, K., and Nasu, M. (2003) Detection of bacteria carrying the stx2 gene by *in situ* loop-mediated isothermal amplification. *Appl Environ Microbiol* **69**: 5023-5028.

Maruyama, F., Kenzaka, T., Yamaguchi, N., Tani, K., and Nasu, M. (2005) Visualization and enumeration of bacteria carrying a specific gene sequence by *in situ* rolling-circle amplification. *Appl Environ Microbiol* **71**: 7933-7940.

Maruyama, F., Tani, K., Kenzaka, T., Yamaguchi, N., and Nasu, M. (2006) Quantitative determination of free-DNA uptake in river bacteria at the single-cell level by *in situ* rolling-circle amplification. *Appl Environ Microbio* **72**: 6248-6256.

Massana, R., Murray, A.E., Preston, C.M., and DeLong, E.F. (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl Environ Microbiol* **63**: 50-56.

Mayer, G., and Bendayan, M. (1999) Immunogold signal amplification: application of the CARD approach to electron microscopy. *J Histochem Cytochem* **47**: 421-430.

Mayer, G., Leone, R.D., Hainfeld, J.F., and Bendayan, M. (2000) Introduction of a novel HRP substrate – nanogold probe for signal amplification in immunocytochemistry. *The J Histochem Cytochem* **48**: 461-469.

McConaughy, B.L., Laird, C.D., and McCarthy, B.J. (1969) Nucleic acid reassociation in formamide. *Biochemistry* **8**: 3289-3295.

Meador, J., Cannon, B., Cannistraro, V.J., and Kennell, D. (1990) Purification and characterization of *Escherichia coli* RNase I. *Eur J Biochem* **187**: 549-553.

Melchior, W.B., and Hippel, P.H.V. (1973) Alteration of the Relative Stability of dA - dT and dG* dC Base Pairs in DNA. *Proc Natl Acad Sci USA* **70**: 298-302.

Meyer, B., and Kuever, J. (2007a) Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-5'-phosphosulfate reductase-encoding genes (*apr*BA) among sulfur-oxidizing prokaryotes. *Microbiology* **153**: 3478-3498.

Meyer, B., and Kuever, J. (2007b) Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase from sulfate-reducing prokaryotes - origin and evolution of the dissimilatory sulfate-reduction pathway. *Microbiology* **153**: 2026-2044.

Meyer, B., and Kuever, J. (2008) Homology modeling of dissimilatory APS reductases (AprBA) of sulfur-oxidizing and sulfate-reducing prokaryotes. *PLoS One* **3**: e1514.

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007) Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9**: 1162-1175.

Moir, J.W.B., Crossman, L.C., Spiro, S., and Richardson, D.J. (1996) The purification of ammonia monooxygenase from *Paracoccus denitrificans*. *FEBS Lett* **387**: 71-74.

Monis, P.T., Giglio, S., and Saint, C.P. (2005) Comparison of SYT09 and SYBR Green I for real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis. *Anal Biochem* **340**: 24-34.

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F. et al. (2008) A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc Natl Acad Sci USA* **105**: 17861-17866.

Mussmann, M., Richter, M., Lombardot, T., Meyerdierks, A., Kuever, J., Kube, M. et al. (2005) Clustered genes related to sulfate respiration in uncultured prokaryotes support the theory of their concomitant horizontal transfer. *J Bacteriol* **187**: 7126-7137.

Mussmann, M., Hu, F.Z., Richter, M., de Beer, D., Preisler, A., Jørgensen, B.B. et al. (2007) Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol* **5**: e230.

Nakajima, N., Sata, T., Hanaki, K., Kurata, T., and Yoshikura, H. (1999) Application of the hybridization AT-tailing method for detection of human immunodeficiency virus RNA in cells

and simian immunodeficiency virus RNA in formalin-fixed and paraffin-embedded tissues. *J Virol Methods* **81**: 169-177.

Nakajima, N., Ionescu, P., Sato, Y., Hashimoto, M., Kuroita, T., Takahashi, H. et al. (2003) *In situ* hybridization AT-tailing with catalyzed signal amplification for sensitive and specific *in situ* detection of human immunodeficiency virus-1 mRNA in formalin-fixed and paraffin-embedded tissues. *Am J Pathol* **162**: 381-389.

Nakajima, N., Hanaki, K., Shimizu, Y.K., Ohnishi, S., Gunji, T., Nakajima, A. et al. (1998) Hybridization-AT-tailing (HybrAT) method for sensitive and strand-specific detection of DNA and RNA. *Biochem Biophys Res Commun* **248**: 613-620.

Nakatsu, C.H., and Forney, L.J. (2004) Parameters of nucleic acid hybridization experiments. In *Molecular Microbial Ecology Manual*. G.A Kowalchuk, Bruijn, F.J.d., I.M Head, A.D. Akkermans, and Elsas, J.D.v. (eds): Springer, pp. 321-332.

Ness, J.V., and Hahn, W.E. (1982) Physical parameters affecting the rate and completion of RNA driven hybridization of DNA: new measurements relevant to quantitation based on kinetics. *Nucleic Acids Res* **10**: 8061-8077.

Niki, H., and Hiraga, S. (1997) Subcellular distribution of actively partitioning F plasmid during the cell division cycle in *E. coli*. *Cell* **90** 951-957.

Niki, H., and Hiraga, S. (1998) Polar localization of the replication origin and terminus in *Escherichia coli* nucleoids during chromosome partitioning. *Genes Dev* **12**: 1036-1045.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., and Hase, T. (2000) Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res* **28**: e63.

Orosz, J.M., and Wetmur, J.G. (1977) DNA melting temperatures and renaturation rates in concentrated alkylammonium salt solutions. *Biopolymers* **16**: 1183-1199.

Park, S.-J., Park, B.-J., and Rhee, S.-K. (2008) Comparative analysis of archaeal 16S rRNA and *amo*A genes to estimate the abundance and diversity of ammonia-oxidizing archaea in marine sediments. *Extremophiles* **12**: 605-615.

Pavlekovic, M., Schmid, M.C., Schmider-Poignee, N., Spring, S., Pilhofer, M., Gaul, T. et al. (2009) Optimization of three FISH procedures for *in situ* detection of anaerobic ammonium oxidizing bacteria in biological wastewater treatment. *J Microbiol Methods* **78**: 119-126.

Pernthaler, A., and Amann, R. (2004) Simultaneous fluorescence i*n situ* hybridization of mRNA and rRNA in environmental bacteria. *Appl Environ Microbiol* **70**: 5426-5433.

Pernthaler, A., and Pernthaler, J. (2005) Simultaneous fluorescence *in situ* hybridization of mRNA and rRNA for the detection of gene expression in environmental microbes. *Methods Enzymol* **397**: 351-371.

Pernthaler, A., Pernthaler, J., and Amann, R. (2002a) Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68**: 3094–3101.

Pernthaler, A., Pernthaler, J., and Amann, R. (2004) Sensitive multi-color fluorescence in situ hybridization for the identification of environmental microorganisms. In *Molecular Microbial Ecology Manual*. Kowalchuk, G.A., de Bruijn, F.J., Head, I.M., Akkermans, A.D., and van Elsas, J.D. (eds): Springer Netherlands.

Pernthaler, A., Zielinski, F.U., Borowski, C., and Dubilier, N. (in prep) Expression patterns of mRNAs for methanotrophy and thiotrophy in symbionts of the hydrothermal vent mussel Bathymodiolus puteoserpentis.

Pernthaler, A., Preston, C.M., Pernthaler, J., DeLong, E.F., and Amann, R. (2002b) Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine Bacteria and Archaea. *Appl Environ Microbio* **68**: 661-667.

Pilhofer, M., Pavlekovic, M., Lee, N.M., Ludwig, W., and Schleifer, K.-H. (2009) Fluorescence *in situ* hybridization for intracellular localization of *nif*H mRNA. *Syst Appl Microbiol* **32**: 186-192.

Player, A.N., Shen, L.-P., Kenny, D., Antao, V.P., and Kolberg, J.A. (2001) Single-copy gene detection using branched DNA (bDNA) *in situ* hybridization. *J Histochem Cytochem* **49**: 603-611.

Pogliano, J., Ho, T.Q., Zhong, Z., and Helinski, D.R. (2001) Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc Natl Acad Sci USA* **98**: 4486-4491.

Pol, A., Heijmans, K., Harhangi, H.R., Tedesco, D., Jetten, M.S.M., and Op den Camp, H.J.M. (2007) Methanotrophy below pH1 by a new *Verrucomicrobia* species. *Nature* **450**: 874-878.

Pontius, B.W., and Berg, P. (1991) Rapid renaturation of complementary DNA strands mediated by cationic detergents: A role for high probability binding domains in enhancing the kinetics of molecular assembly processes. *Proc Natl Acad Sci USA* **88**: 8237-8241.

Prosser, J.I., and Nicol, G.W. (2008) Relative contributions of Archaea and Bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**: 2931-2941.

Purkhold, U., Pommerening-Röser, 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 *amo*A sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* **66**: 5368–5382.

Raap, A.K., Marijnen, J.G.J., Vrolijk, J., and Ploeg, M.v.d. (1986) Denaturation, renaturation, and loss of DNA during in situ hybridization procedures. *Cytometry* **7**: 235-242.

Rasmussen, J., Saint, C., and Monis, P. (2007) Use of DNA melting simulation software for *in silico* diagnostic assay design: targeting regions with complex melting curves and confirmation by real-time PCR using intercalating dyes. *BMC Bioinformatics* **8**: 107.

Rees, W., Yager, T., Korte, J., and Hippel, P.v. (1993) Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**: 137-144.

Sambrook, J. (2001) *Molecular cloning manual*. New York: Cold Spring Harbor.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078-2093.

Schildkraut, C., and Lifson, S. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers* **3**: 195-208.

Schleper, C., Jürgens, G., and Jonuscheit, M. (2005) Genomic studies of uncultivated Archaea. *Nat Rev Microbiol* **3**: 479-488.

Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501-1506.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing MOTHUR: open source, platform-independent, community-supported

software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Schmeckpeperl, B.J., and Smith, K.D. (1972) Use of formamide in nucleic acid reassociation. *Biochemistry* **11**: 1319-1326.

Schwartz, D.E., and Wash, R. (1989) Use of volume exclusion agents for the enhancement of *in situ* hybridization. In. USA: Microprobe Corporation, Bothel, Wash.

Smolina, I., Lee, C., and Frank-Kamenetskii, M. (2007) Detection of low-copy-number genomic DNA sequences in individual bacterial cells by using peptide nucleic acid-assisted rolling-circle amplification and fluorescence *in situ* hybridization. *Appl Environ Microbiol* **73**: 2324-2328.

Somerville, C.C., Knight, I.T., Straube, W.L., and Colwell, R.R. (1989) Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Appl Environ Microbio* **55**: 548-554.

Speel, E.J.M., Hopman, A.H.N., and Komminoth, P. (1999) Amplification methods to increase the sensitivity of *in situ* hybridization: play CARD(s). *J Histochem Cytochem* **47**: 281-288.

Springer, M.S., Davidson, E.H., and Britten, R.J. (1992) Calculation of sequence divergence from the thermal stability of DNA heteroduplexes. *J Mol Evol* **34**: 379-382.

Steger, G. (1994) Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic Acids Res* **22**: 2760-2768.

Tani, K., Kurokawa, K., and Nasu, M. (1998) Development of a direct *in situ* PCR method for detection of specific Bacteria in natural environments. *Appl Environ Microbiol* **64**: 1536-1540.

Teira, E., Lebaron, P., Aken, H.v., and Herndl, G.J. (2006) Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic. *Limnol Oceanogr* **51**: 2131-2144.

Teira, E., Reinthaler, T., Pernthaler, A., Pernthaler, J., and Herndl, G.J. (2004) Combining catalyzed reporter deposition-fluorescence *in situ* hybridization and microautoradiography to detect substrate utilization by Bacteria and Archaea in the deep ocean. *Appl Environ Microbiol* **70**: 4411-4414.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.

Torre, J.R.d.l., Walker, C.B., Ingalls, A.E., Könneke, M., and Stahl, D.A. (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* **10**: 810-818.

Trebesius, K., Amann, R., Ludwig, W., Muhlegger, K., and Schleifer, K.-H. (1994) Identification of whole fixed bacterial cells with nonradioactive 23S rRNA-targeted polynucleotide probes. *Appl Environ Microbiol* **60**: 3228-3235.

Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.-P., and Schleper, C. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic Crenarchaeota in nitrogen cycling. *Environ Microbiol* **7**: 1985-1995.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.

Viollier, P.H., Thanbichler, M., McGrath, P.T., West, L., Meewan, M., McAdams, H.H., and Shapiro, L. (2004) Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc Natl Acad Sci USA* **101**: 9257-9262.

Wagner, M., Schmid, M., Juretschko, S., Trebesius, K.-H., Bubert, A., Goebel, W., and Schleifer, K.-H. (1998) *In situ* detection of a virulence factor mRNA and 16S rRNA in *Listeria monocytogenes*. *FEMS Microbiol Lett* **160**: 159-168.

Wahl, G.M., Stern, M., and Stark, G.R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc Natl Acad Sci USA* **76**: 3683-3687.

Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H., and Itakura, K. (1981) The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit ß-globin DNA. *Nucleic Acids Res* **9**: 879-894.

Wartell, R.M., and Benight, A.S. (1985) Thermal denaturation of DNA molecules: A comparison of theory with experiment. *Physics Reports* **126**: 67-107.

Watson, N., Gurevitz, M., Ford, J., and Apirion, D. (1984) Self cleavage of a precursor RNA from bacteriophage T4. *J Mol Biol* **172**: 301-323.

Wetmur, J.G. (1975) Acceleration of DNA renaturation rates. *Biopolymers* **14**: 2517-2524.

Wetmur, J.G. (1991) DNA probes: applications of the principles of nucleic acid hybridization. *Crit Rev Biochem Mol Biol* **26**: 227-259.

Wetmur, J.G., and Davidson, N. (1968) Kinetics of renaturation of DNA. *J Mol Biol* **31**: 349-370.

Wilcox, J. (1993) Fundamental principles of *in situ* hybridization. *J Histochem Cytochem* **41**: 1725-1733.

Wild, J., Hradecna, Z., and Szybalski, W. (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res* **12**: 1434-1444.

Yu, H., Chao, J., Patek, D., Mujumdar, R., Mujumdar, S., and Waggoner, A.S. (1994) Cyanine dye dUTP analogs for enzymatic labeling of DNA probes. *Nucleic Acids Res* **22**: 3226-3232.

Zehr, J.P., and McReynolds, L.A. (1989) Use of degenerate oligonucleotides for amplification of the *nif*H gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl Environ Microbiol* **55**: 2522-2526.

Zhou, J., Bruns, M., and Tiedje, J. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbio* **62**: 316-322.

Zhou, Y., Calciano, M., Hamann, S., Leamon, J.H., Strugnell, T., Christian, M.W., and Lizardi, P.M. (2001) *In situ* detection of messenger RNA using digoxigenin-labeled oligonucleotides and rolling circle amplification. *Exp Mol Pathol* **70**: 281-288.

Zielinski, F.U., Petersen, J.M., Moraru, C., Wetzel, S., Pape, T., Seifert, R. et al. (in prep) The sulfur-oxidizing endosymbiont of the hydrothermal vent mussel *Bathymodiolus puteoserpentis (Bivalvia: Mytilidae)* uses hydrogen as an energy source.

Zimmermann, J., Ludwig, W., and Schleifer, K.H. (2001) DNA polynucleotide probes generated from representatives of the genus *Acinetobacter* and their application in fluorescence *in situ* hybridization of environmental samples. *Syst Appl Microbiol* **24**: 238-244.

Zwirglmaier, K. (2005) Fluorescence *in situ* hybridisation (FISH): the next generation. *FEMS Microbiol Lett* **246**: 151-158.

Zwirglmaier, K., Ludwig, W., and Schleifer, K.-H. (2003) Improved fluorescence in situ hybridization of individual microbial cells using polynucleotide probes: the network hypothesis. *System Appl Microbiol* **26**: 327-337.

Zwirglmaier, K., Ludwig, W., and Schleifer, K.H. (2004) Recognition of individual genes in a single bacterial cell by fluorescence *in situ* hybridization - RING-FISH. *Mol Microbiol* **51**: 89-96.

Anlage zur Dissertation

Name: Cristina Liliana Moraru Ort, Datum: Bremen, 10-03.2010

Anschrift: Celsiusstrasse 1, 28359 Bremen, Germany

Erklärung

Gem. § 6 (5) Nr. 1-3 PromO

Ich erkläre, dass ich

die Arbeit ohne unerlaubte fremde Hilfe angefertigt habe,

keine anderen, als die von mir angegebenen Quellen und Hilfsmittel benutzt habe und

die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Cristina Moraru

(Unterschirift)