SYSTEMATIC OPTIMIZATION OF THE DETECTION OF SINGLE GENES IN MICROORGANISMS BY FLUORESCENCE *IN SITU* HYBRIDIZATION (GENEFISH)

DISSERTATION
ZUR ERLANGUNG DES GRADES EINES
DOKTORS DER NATURWISSENSCHAFTEN
– DR. RER. NAT. –

DEM FACHBEREICH BIOLOGIE / CHEMIE
DER UNIVERSITÄT BREMEN

VORGELEGT VON

JIMENA BARRERO CANOSA

Bremen, November 2015

Die vorliegende arbeit wurde vom Juni 2012 bis November 2015 am
Max-Planck-Institut für Marine Mikrobiologie in Bremen angefertigt

1. Gutachter: Prof. Dr. Rudolf Amann
2. Gutachter: Prof. Dr. Andreas Schramm

1. Prüfer: Prof. Dr. Michael Friedrich
2. Prüfer: Dr. Cristina Moraru

SUMMARY

Although microbes are the smallest living organisms on Earth, they are essential for shaping the global environment. Therefore, it is fundamental to understand the dynamics of microbial interactions with each other and their surroundings. The study of microorganisms is challenging because the majority is not cultivable. For this reason, culture-independent techniques have played an increasingly important role in microbial ecology. Fluorescence in situ Hybridization (FISH), in which a fluorescently labeled probe is used to detect target RNA or DNA molecules within an intact cell, is a well-established method for the identification and quantification of microorganisms. Over the years, this technique has contributed to a deeper understanding of the diversity and composition of microbial communities. Among the FISH-based techniques, geneFISH allows the linkage of particular genes to the microbial identity at the single cell level. The protocol involves ribosomal RNA-targeted Catalyzed Amplification Reported Deposition (CARD-FISH) for cell identification. For gene detection, digoxigenin-labeled polynucleotide probes are applied, to which specific antibodies conjugated with horseradish-peroxidases (HRP) can bind. The signal amplification is again achieved by CARD. So far, low gene detection efficiencies have prevented a quantitative determination of the cell fraction carrying the target gene. Additionally, the CARD step hinders the sub-cellular signal quantification, increases DNA degradation and cell damage and makes the protocol long and labor intensive. Therefore, the main aim of this thesis was to systematically improve the geneFISH protocol by (i) increasing the gene detection efficiency, (ii) simplifying the method and (iii) achieving a quantification of gene number per cell.

The first section of this thesis (**CHAPTER I**) deals with the improvement of the geneFISH protocol – especially, the efficiency of gene detection and the localization of the gene signal. To improve gene detection efficiencies, the number of probes used per target gene was increased. Originally, only a single digoxigenin-labeled polynucleotide probe was applied in geneFISH. During the course of this project, this number was sequentially increased to twelve. Our results revealed that for cells containing 3-8 copies of a target gene, four probes were sufficient to obtain a detection efficiency of

~93%. Secondly, to enable a more precise sub-cellular localization of the geneFISH signal, different dextran sulfate (DS) and tyramide concentrations were tested. These improvements allowed tracking the phage-host dynamics during an infection experiment with phage PSA-HP1 and *Pseudoalteromonas* H100. The new protocol was called phageFISH (**CHAPTER I: Manuscript I and Book chapter**). In a next step, the new protocol was applied to enrichment cultures to detect PSA-HP1 phages in natural seawater samples of the North Sea. The presence of the phage PSA-HP1 could be shown with PhageFISH after 75, 113 and 138 h in the later phases of the incubations (**CHAPTER I: Manuscript II**).

The second section (**CHAPTER II: Manuscript III**) focuses on the development of a simplified geneFISH protocol, named direct-geneFISH. Direct-geneFISH was first optimized in pure cultures of *Escherichia coli*. The rRNA and gene probes were hybridized simultaneously, reducing the overall handling time from 32 h to only 6 h. Furthermore, direct-geneFISH improved the original protocol by allowing a per cell quantification of the gene copy number and the sub-cellular localization of both the rRNA and the genes of interest. The newly developed protocol was compared to the CARD-based geneFISH, for the detection of the sulfate thiolhydrolase (soxB) gene in GSO-SUP05 cells in water samples from a sulfidic anoxic inversion in Rogoznica Lake, Croatia. The two protocols gave comparable results in their gene detection efficiency.

In the third section (**CHAPTER III: Manuscript IV: Preliminary results**), direct-geneFISH was applied to pure cultures of *Formosa* strain B cells isolated from Helgoland. Metagenomic data suggests that this Flavobacterium, *Formosa* strain B, is (i) a recurrent and clonal strain found in waters surrounding Helgoland and (ii) a specialist for mannan degradation. To test the latter hypothesis, pure cultures of *Formosa* strain B were subjected to probes, which were designed to target a gene encoding for a mannosyl hydrolase (GH92). Initial tests permitted the simultaneous visualization of both the GH92 gene and the rRNA in the strain. Future efforts will be focused on adapting the direct-geneFISH to simultaneously identify *Formosa* strain B harboring the GH92 gene in environmental samples.

ZUSAMMENFASSUNG

Obwohl Mikroorganismen die kleinsten lebenden Organismen der Erde sind, sind sie essenziell für die Gestaltung der globalen Umwelt. Deshalb ist es sehr wichtig, die Interaktionen der Mikroorganismen untereinander und mit ihrem Umfeld zu verstehen. Das Erforschen von Mikroorganismen ist schwierig, da die meisten von ihnen nicht kultiviert werden können. Daher spielen kultivierungsunabhängige Methoden eine immer wichtigere Rolle in der mikrobiellen Ökologie. Die Fluoreszenz-in-situ-Hybridisierung (FISH), bei der fluoreszierende Sonden zum Nachweis von RNA oder DNA Zielmolekülen in intakten Zellen verwendet werden, ist eine etablierte Methode zur Identifikation und Quantifizierung von Mikroorganismen. Diese Technik hat in den letzten Jahren maßgeblich zu einem tieferen Verständnis der Diversität und Zusammensetzung von mikrobiellen Gemeinschaften beigetragen. „GeneFISH“ ist eine der FISH-Techniken, welche erlaubt, bestimmte Gene zu einer einzelnen, identifizierbaren, mikrobiellen Zelle zuzuordnen. Das Protokoll beinhaltet zunächst die rRNA-basierte CARD-FISH (auf englisch: Catalyzed Amplification Reporter Deposition- FISH) für die Identifikation der einzelnen Zellen. Für den Nachweis der Gene werden Digoxigenin-markierte Sonden verwendet. An diese Sonden können im nächsten Schritt spezifische Antikörper, welche mit Meerrettichperoxidase (HRP; auf englisch „horseradish-peroxidase“) gebunden sind, binden. Die Signalverstärkung wird wieder durch CARD erreicht. Bis jetzt hat die geringe Effizienz des GeneFISH Protokolls die Quantifizierung der Zellen, welche das Gen enthalten, verhindert. Außerdem behindert der CARD-Schritt die subzelluläre Signalquantifizierung, erhöht die DNA-Degradierung und Zellschäden und macht das Protokoll lange und arbeitsintensiv. Deshalb ist das Hauptziel dieser Doktorarbeit die systematische Verbesserung des geneFISH Protokolls, indem (i) die Effizienz des Gennachweises erhöht wird, (ii) die Methoden vereinfacht werden und (iii) eine Quantifizierung der Gene pro Zelle erreicht wird.

Der erste Teil der Arbeit (**CHAPTER I**) handelt von der Verbesserung des geneFISH Protokolls– besonders von der Effizienz des Gennachweises und der Lokalisierung des

Gensignals. Um die Effizienz des Gennachweises zu erhöhen, wurde die Anzahl der Sonden, die pro Gen verwendet wurden, erhöht. Ursprünglich wurde nur eine Sonde pro Gen bei GeneFISH verwendet. Doch im Laufe dieses Projektes wurde diese Anzahl stufenweise auf zwölf Sonden pro Gen erhöht. Die Ergebnisse zeigten, dass für Zellen, die zwischen drei und acht Kopien des Zielgens enthielten, vier Sonden ausreichten, um eine Nachweiseffizienz von ~ 93 % zu erhalten. Um eine präzisere, subzelluläre Lokalisation des GeneFISH-Signals zu ermöglichen, wurden verschiedene Dextransulfat- und Tyramidkonzentrationen getestet. Die Verbesserungen erlaubten es, die Dynamiken einer Virus-Wirt-Interaktion während eines Infektionsexperiments mit dem Bakteriophagen PSA-HP1 und *Pseudoalteromonas* H100 zu verfolgen. Das neue Protokoll wurde „PhageFISH“ genannt (**CHAPTER I: Manuscript I and Book chapter**). Im nächsten Schritt wurde das neue Protokoll an Anreicherungskulturen angewendet, um PSA-HP1 Phagen in Meerwasserproben aus der Nordsee zu identifizieren. Nach Inkubationszeiten von 75, 113 und 138 h, konnte mit Hilfe von PhageFISH die Anwesenheit von PSA-HP1 Phagen nachgewiesen werden (**CHAPTER I: Manuscript II**).

Im Fokus des zweiten Teils (**CHAPTER II: Manuscript III**) steht die Entwicklung eines vereinfachten geneFISH Protokolls, das „direct-geneFISH“ genannt wurde. Direct-geneFISH wurde zunächst an Reinkulturen von *Escherichia coli* optimiert. Die rRNA- und Gensonden hybridisierten zeitgleich an ihre jeweiligen Zielmoleküle, was die Bearbeitungszeit von 32 auf nur 6 h verkürzte. Weitere Verbesserungen von direct-geneFISH gegenüber dem Originalprotokoll beinhalten die Quantifizierung der Genkopien pro Zelle und die subzelluläre Lokalisation sowohl von rRNA Molekülen als auch von Genen. Beide Protokolle (direct-geneFISH und das Originalprotokoll von geneFISH) wurden für den Nachweis des Sulfat-thiol-hydrolase Gens (*soxB*), in Wasserproben aus dem Drachenaugensee (Rogoznica, Kroatien) verwendet und lieferten vergleichbare Ergebnisse in der Detektionseffizienz.

Im dritten Kapitel (**CHAPTER III: Manuscript IV: Preliminary results**) wurde direct-geneFISH an Reinkulturen von *Formosa* Stamm B, deren Zellen vor Helgoland isoliert wurden, angewendet. Metagenomische Daten deuteten an, dass dieses

Flavobakterium, *Formosa* Stamm B, ein (i) wiederkehrender und klonaler Stamm aus den Gewässern um Helgoland und (ii) ein Spezialist für den Abbau von Mannan ist. Um letztere Hypothese zu testen wurde eine Gensonde entworfen, die spezifisch für eine Mannosyl-hydrolase (GH92) kodiert, und mit Reinkulturen von *Formosa* Stamm B hybridisiert. Vorläufige Tests erlaubten die gleichzeitige Visualisierung des GH92 Gens und der rRNA. Zukünftige Arbeiten werden sich darauf konzentrieren, das direct-geneFISH Protokoll für die Nutzung in Umweltproben anzupassen.

LIST OF ABBREVIATIONS

%GC	PERCENT MOLAR G + C
%MM	PERCENTAGE OF MISMATCHES
AFM	ATOMIC FORCE MICROSCOPY
BLAST	BASIC LOCAL ALIGNMENT SEARCH TOOL
bp	BASE PAIR
CARD	CATALYZED REPORTER DEPOSITION
CAZymes	CARBOHYDRATE-ACTIVE ENZYMES
CCD	CHARGED-COUPLED DEVICE
CLSM	CONFOCAL LASER SCANNING MICROSCOPY
DAPI	4',6-DIAMIDINO-2-PHENYLINDOLE
Dig	DIGOXIGENIN
dig-UTP	DIGOXIGENIN-LABELED URIDINE TRIPHOSPHATE
DNA	DEOXIRIBONUCLEIC ACID
DS	DEXTRAN SULFATE
dsDNA	DOUBLE STRANDED DNA
EDTA	ETHYLENEDIAMINETETRAACETIC ACID,
FISH	FLUORESCENCE *IN SITU* HYBRIDIZATION
FVIC	FREQUENCY OF VISIBLY INFECTED CELLS
GSO	GAMMA PROTEOBACTERIAL SULFUR OXIDIZERS
HRP	HORSERADISH PEROXIDASE
HCR	HYBRIDIZATION CHAIN REACTION
GHs	GLYCOSYL HYDROLASES
GH-92	GLYCOSYL HYDROLASE 92
kDa	KILODALTON
LSM	LASER SCANNING MICROSCOPY
MDA	MULTIPLE DISPLACEMENT AMPLIFICATION
MOI	MULTIPLICITY OF INFECTION
mRNA	MESSENGER RIBONUCLEIC ACID
PALM	PHOTOACTIVATED LOCALIZATION MICROSCOPY
PBS	PHOSPHATE-BUFFERED SALINE
PCR	POLYMERASE CHAIN REACTION
pDNA	PLASMIDIC DEOXYRIBONUCLEIC ACID
PFA	PARAFORMALDEHYDE
PFU	PLAQUE-FORMING UNIT
PNA	PEPTIDE NUCLEIC ACID
PT	POSITIONAL TREE
PUL	POLYSACCHARIDE UTILIZATION LOCI
QD	QUANTUM DOT
RCA	ROLLING CIRCLE AMPLIFICATION
RCB	ROSEOBACTER CLADE BACTERIA
RDP	RIBOSOMAL DATABASE PROJECT
RNA	RIBONUCLEIC ACID
rRNA	RIBOSOMAL RIBONUCLEIC ACID
RSOFT	REVERSIBLE SATURABLE OPTICAL FLUORESCENCE

	TRANSITIONS
RT	ROOM TEMPERATURE
SCG	SINGLE CELL GENOMICS
SDS	SODIUM DODECYL SULFATE
SIM	STRUCTURED ILLUMINATION MICROSCOPY
SNP	SINGLE NUCLEOTIDE POLYMORPHISM
STED	STIMULATED EMISSION DEPLETION
STORM	STOCHASTIC OPTICAL RECONSTRUCTION MICROCOPY
SR	SUPER RESOLUTION
SSC	SODIUM SALT CITRATE
TCC	TOTAL CELL COUNTS
TEM	TRANSMISSION ELECTRON MICROSCOPY
Tm	MELTING POINT TEMPERATURE
WBI	WASHING BUFFER I
WBII	WASHING BUFFER II

CONTENTS

INTRODUCTION

STUDY OF MICROORGANISMS IN MARINE ENVIRONMENTS

Marine microbes play a key role in the environment since they are involved in the global cycling of all biologically important elements. They have diverse life strategies and furthermore can interact among each other in many ways (Fuhrman *et al.*, 2015). The best way to understand microbial physiology is through culture-based techniques. However the enormous microbial diversity and the complex interactions have hindered its usage in microbial ecology, as less than 1% are cultivable (Amann *et al.*, 1995). Nowadays, molecular tools are frequently used for addressing fundamental ecological questions related to the diversity, composition and function of microbial communities.

The introduction of high throughput sequencing techniques enabled the discovery of novel genes and new metabolic pathways (Iverson *et al.*, 2012). Nonetheless, metagenomic predictions are not sufficient to address all ecologically relevant questions as in complex communities it is difficult to tease apart the genomes of the different microorganisms present and often the taxonomic affiliation of particular genes is impossible (Teeling and Glöckner, 2012).

Fluorescence *in situ* hybridization FISH is a molecular method routinely used for studying environmental microbes. In FISH, a fluorescently labeled nucleic acid probe is used to target complementary RNA or DNA molecules while preserving the cellular morphology and the *in situ* spatial organization. In microbial ecology, FISH is used to identify, quantify and infer relationships between microorganisms by targeting the ribosomal RNA (rRNA) (Amann *et al.*, 1990).

GENEFISH

GeneFISH is a FISH variant, which has the potential to link gene presence and cell identity of microbes *in situ*. The method combines the principles of catalyzed amplification reporter deposition CARD-FISH for cell identification, using horseradish

peroxidase (HRP)-labeled oligonucleotide probes targeting the rRNA (Pernthaler *et al*., 2002), and a polynucleotide probe, which is multiple-labeled with digoxigenin, followed by the binding of HRP-conjugated anti-DIG antibodies and CARD to amplify and visualize the signal for the gene detection (Figure 1).

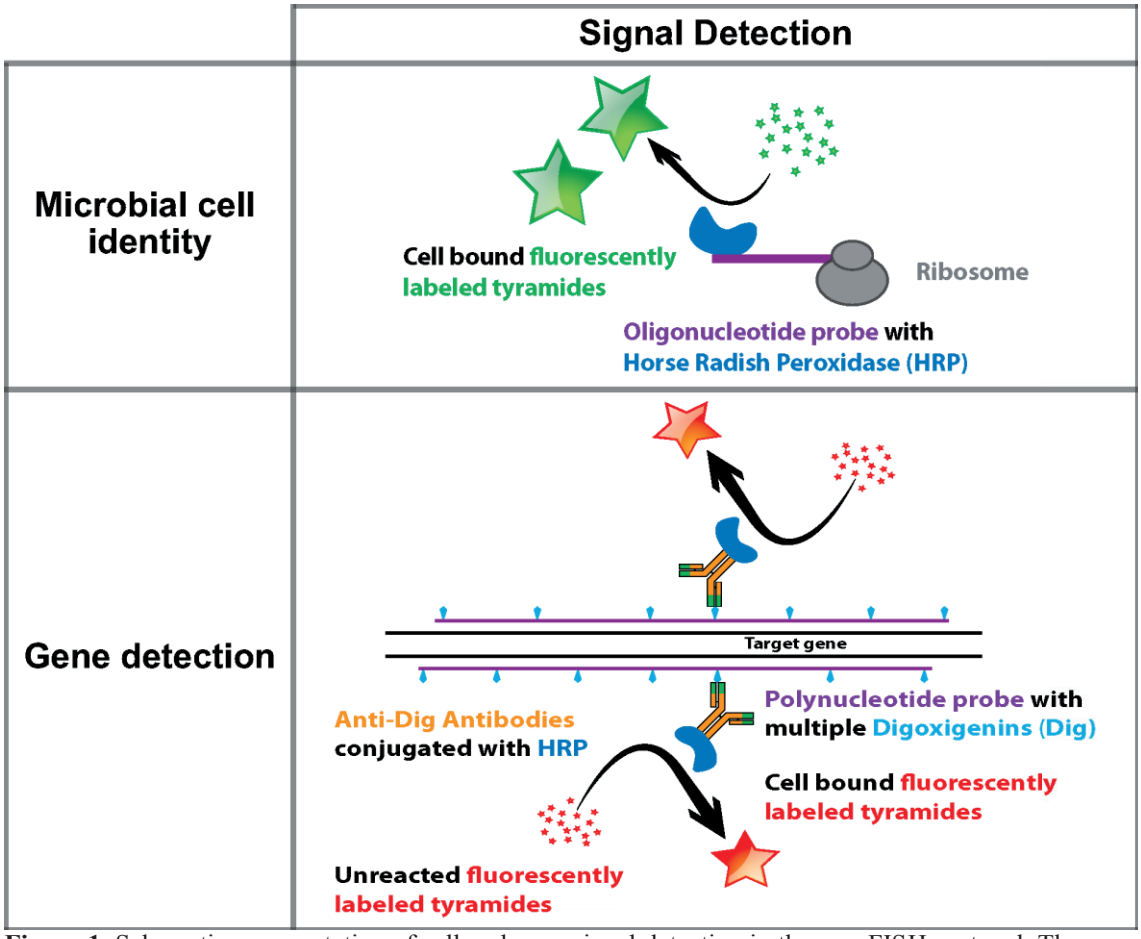

Figure 1. Schematic representation of cell and gene signal detection in the geneFISH protocol. The gene detection panel was modified from Moraru *et al*., 2010.

The success of geneFISH relies on many steps outlined in Figure 2. This includes also the definition of the ecological questions to be answered and the evaluation of the applicability of geneFISH, the selection and synthesis of the appropriate probes, the hybridization protocol, microscopy and data interpretation. In the next sections, the following steps are explained in depth: (I) Probe design and synthesis, (II) GeneFISH protocol, (III) geneFISH applications.

(I) PROBE DESIGN AND SYNTHESIS

GeneFISH can be used to qualitatively detect a specific gene of interest within either a pure culture or a mixed microbial community (Moraru *et al.*, 2010). Fundamental to this method is that a nucleic acid sequence is designed to complement the target gene of interest. If the designed probe successfully binds to the target sequence this is referred to as hybridization. To detect/visualize the hybridization the designed nucleic acid sequence is pre-labeled with a reporter molecule (O'connor, 2008).

The probe design, synthesis and labeling is thus crucial for the success of any FISH experiment. Probes can vary in their chemical backbone (e.g. RNA, DNA, PNA), length and the reporter label used (e.g. fluorochrome, enzyme or haptens). For the purpose of this thesis I discuss only the probes used for cell identification and gene detection in geneFISH (Figure 2).

Ribosomal RNA – targeted oligonucleotide probes can be taxonomically as broad as the taxonomic domain or can be specific down to the genus level. The DNA probes can vary in length between 15-30 bp (Amann *et al.*, 1990). In some cases they can be conveniently selected from a curated public database containing rRNA-based oligonucleotide probes already tested and optimized in previous studies such as ProbeBase. The access to this database is free, furthermore it contains the essential information of the probes specificity and coverage (Loy *et al.*, 2007). In other cases *de novo* design of a new probe is needed if the current database probes do not match the target rRNA. In this case new rRNA probes are designed with the ARB software package. The software features a “probe design” and “probe match” tool, which automatically search for suitable target sites. ARB probe design is a 3-step process: first the target group is selected, second “probe design” tool searches and ranks potential target sites in the positional tree (PT) server and third the “probe match” tool evaluates the potential probes against the whole database (Ludwig *et al.*, 2004). After an oligonucleotide probe is designed or selected (Figure 2), they are purchased from service partners specialized in nucleic acid synthesis (e.g. Biomers GmbH, Ulm). For CARD-FISH the probes carry a HRP at the 5’-end as the reporter molecule.

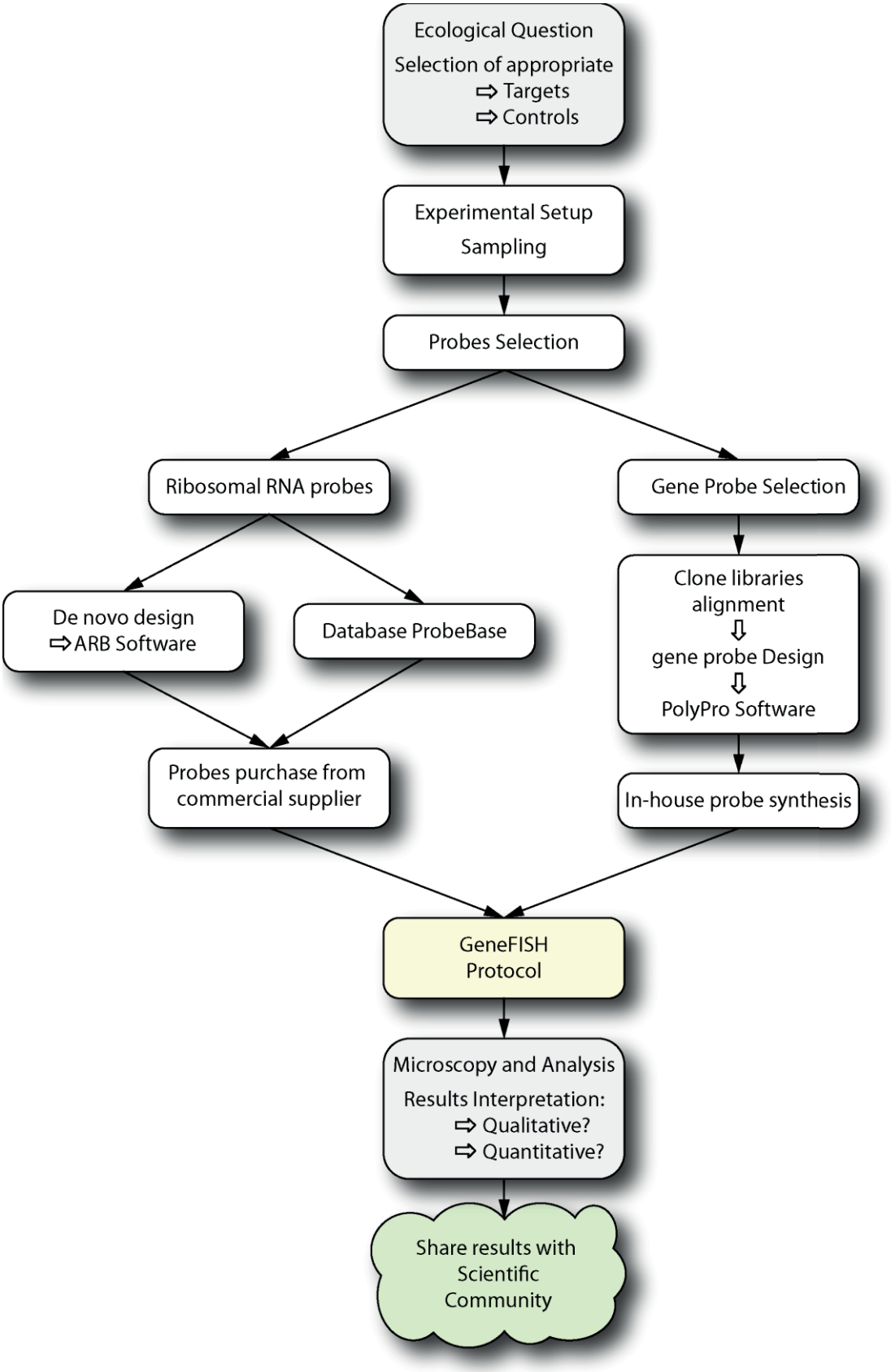

Figure 2. Workflow for geneFISH experiments. Details of the individual steps are explained below.

For gene detection, double stranded DNA (dsDNA) polynucleotide probes are preferentially used. The advantage of using polynucleotide probes in microbial ecology is that they are tolerant to a certain number of mismatches, which allows the detection of diverse genes. Moreover, a single probe carries many labels representing a simple way to amplify the signal (Moraru *et al.*, 2011). In contrast to the rRNA probes, a comprehensive gene probe database does not exist, yet. Therefore in most cases the design of gene probes is needed for every new geneFISH project (Figure 2). Like the 16S rRNA, the probe design process requires an initial alignment from gene sequences retrieved from the environment of interest. Based on the alignment, the target region is selected, which should belong to the translated region and ideally contains only few gaps and mismatches. The latter, if present, should be homogeneously distributed along the target regions (Moraru *et al.*, 2011). The probe length can vary from 100 to 960 bp (Moraru *et al.*, 2010; Kawakami *et al.*, 2012), nevertheless it is recommended to use shorter probes (100 to 450 bp) to minimize penetration and diffusion problems (Moraru *et al.*, 2010). When working with environmental samples the goal is to detect all or at least the majority of possible gene alleles. In cases where gene diversity is higher than 5-10% it is necessary to use multiple probes. Polypro software is useful in designing a mix of polynucleotide probes, which target all possible alleles of the target gene in the studied environment with a minimal amount of different probe variants. (Moraru *et al.*, 2011).

Probe synthesis and reporter labeling is done by the incorporation of modified nucleotides in the DNA synthesis by polymerase chain reaction (PCR) technology and chemistry. In this PCR reaction one fourth of the deoxy-thymine triphosphate (dTTP) is replaced with digoxigenin-label uridine triphosphate (dig-UTP). This ratio allows efficient labeling while yielding high double-stranded DNA. To achieve the highest yield, thermal cycling conditions used in the PCR reaction need to be optimized, first *in silico* by calculating thermal conditions and by verification with a gradient PCR. Afterwards synthesized probes need to be purified. In general, all PCR purification kits are adaptable for probe purification.

(II) GENEFISH PROTOCOL

Figure 3 illustrates the geneFISH protocol developed by Moraru and collaborators (2010). The protocol consists of an initial sample preparation and two consecutive hybridizations, the first for the cell identification and the second for gene detection. The total time required for carrying out the full protocol is ~32 h.

Like the widely used CARD-FISH technique, geneFISH requires careful consideration and optimization of the reagents used. This includes the selection of fixatives, solid supports for cell immobilization and optimal conditions for permeabilization. The purpose of the sample preparation step is preservation of the cell integrity while permeabilizing the membrane to facilitate the penetration of the amended probe into the cell (Amann and Fuchs, 2008). Good preparation of the sample is crucial for the success geneFISH. Specifically, permeabilization is needed to allow the penetration of high molecular mass molecules like the HRP (~44 kDa) (Bidnenko *et al*., 1998) and anti-dig fragments (240 kDa) (Wagner *et al*., 1998) into the cell, while also, the cell should remain robust enough to withstand repeated chemical and thermal treatments. Hence, to achieve the best results, the permeabilization needs to be both, effective and gentle. These conditions will vary depending on the target cell or group, for example gram negative and planktonic bacterial cells are effectively permeabilized with lysozyme (Pernthaler *et al*., 2002), for gram positive with lysins (Bidnenko *et al.*, 1998), lipases (Davenport *et al.*, 2000) and a combination of lysozyme followed by achromopeptidase treatments (Sekar *et al.*, 2003), and for the Archaea with proteinase K (Teira *et al.*, 2004; Lloyd *et al.*, 2013) or achromopeptidases (Kubo *et al.*, 2012) .

In the first hybridization for cell identification, the 16S rRNA targeted, (HRP)-labeled oligonucleotides probes hybridize with ribosomal RNAs. After a washing step un-hybridized probes are removed. This is followed by CARD amplification (Pernthaler *et al*., 2002). This requires that the samples are incubated with hydrogen peroxide and fluorescently labeled tyramides. In the presence of hydrogen peroxide, the HRP catalyzes the dimerization of phenolic-based compounds like the tyramides. In this process the tyramides are radicalized to highly reactive intermediates or free radicals. These free radicals can covalently bind to electron rich moieties of the proteins (e.g. tyrosine residues) amplifying the signal (Bobrow *et al.*, 1989). After the signal

amplification the introduced HRPs need to be deactivated to avoid false positives signals in the next hybridization.

In the second hybridization for the gene detection, both double stranded target and dig-probes need to be denatured. First, the samples are incubated at high temperature (85 ºC) for one hour and then immediately cooled down to between 20-30 ºC below the polynucleotide probes melting point (Tm). The Tm is calculated from the salt concentration in the hybridization buffer, the %GC-content and the probe length (Wetmur, 1991). After hybridization, the un-hybridized or non-specifically bound probes are removed in a stringent washing step. Thereafter the samples are incubated in blocking buffers to prevent the unspecific binding of the anti-dig antibodies in the next step. The anti-dig antibodies are conjugated to HRPs thus, after the immunochemical reaction between the dig molecules and the anti-dig antibodies the signal is amplified by CARD. Finally cells are counterstained with the DNA dye 4',6-diamidino-2-phenylindole, DAPI. The GeneFISH signals are detected by sensitive epi-fluorescence microscopy (Figure 2). Because of the high copy number the rRNA signal is spread more or less evenly through the cell cytoplasm, while the gene, usually localized on the DNA chromosome, is indicated by a single, dot-like signal.

Sample Preparation	Sample Fixation	Fixation conserves cell structure and morphology.	
	Sample Immobilization	Samples are immobilized on polycarbonate filters or glass slides.	
	Permeabilization	Permeabilization allows cell incorporation of: dig-labeled polynucleotide probe (390 Da per dig). anti-dig HRP-conjugated antibodies (260 kDa). HRP-rRNA HRP (40 kDa).	1 h
Cell identification	Inactivation of endogenous peroxidases	Cellular peroxidases have to be inactivated to avoid false positives. For inactivation incubation in 0.05M HCl.	10 min
	rRNA hybridization	HRP-conjugated oligonucleotide probes are used to identify microbial cells in environmental sample. (see Pernthaler *et al.,* 2004)	3 h 30 min
	CARD amplification	The HRP molecules amplify the rRNA signal by catalyzing the covalent binding of fluorescently labeled tyramides to the cellular proteins . This binding step is necessary to ensure that the signal is not wash off in the next steps.	30 min
	Inactivation of introduced peroxidases in rRNA-CARD	HRP molecules introduced in rRNA-CARD have to be inactivated to avoid false gene signals. For inactivation are incubated in 0.2 M HCl.	10 min
Gene detection	Gene hybridization	For gene hybridization: The samples have to be prehybridized in hybridization buffer that do not contain any probe. The step is necessary for decreasing the background. The gene probes and the target are denature in a 85 °C incubation. In the hybridization, the dig-labeled probes will bind to the targeted gene. Excess of probe is washed in a stringent washing buffer (low salt concentration 2° C above hybridization T).	1 h 1 h 17 h 1 h 30 min
	Antibody binding	The anti-dig-HRP will bind to Dig molecules on the hybridized probe. Blocking of unspecific antibody binding is crucial for avoiding high background.	2 h 45 min
	CARD amplification	The antibody bound will deposit between 460 to 2000 fluorochromes per probe. The gene amplification requires higher concentration of tyramide than CARD-rRNA, therefore unbound tyramide many washing steps are included after amplification to decrease the noise	1h 40 min
	Embedding and DNA staining	Embedding protects the fluorochromes against oxidative stress	
	Microscopy	Gene signals are dot size, for small cells the signal might be cell wide.	

Figure 3. GeneFISH protocol modified from Moraru *et al*., 2010.

(III) GENEFISH APPLICATIONS

GeneFISH can be applied to samples of diverse environments, for instance it has been used in planktonic microbial communities (Moraru *et al*., 2010), marine bacterial–eukaryotic symbiosis (Petersen *et al*., 2012; Bernhard *et al*., 2012), marine sediment enrichments (Lenk *et al*.,2012; Stagars *et al*., submitted) and groundwater samples (Matturro and Rossetti, 2015) (Table 1). GeneFISH was used to link the putative ammonia mono-oxigenase *amoA* gene to the Marine Group I Crenarchaeota in planktonic communities from the Benguela upwelling zone. The gene was targeted with a mix of 450 bp long probes that covered all *amoA* alleles present in the sample diversity. The single cells were identified with CREN554 probe specific for Marine Group I Crenarchaeota (Moraru *et al.*, 2010). The fluorescence images showed the overlapping gene and rRNA signal for 30% of the Crenarchaeota cells and furthermore the gene did not appear in other taxonomic groups, demonstrating for the first time the direct connection of ammonium oxidation and Crenarchaeota cells *in situ* (Moraru *et al.*, 2010).

GeneFISH also permitted the gene detection of bacteria living in symbiosis with eukaryotic organisms (Bernhard *et al*., 2012; Petersen *et al*., 2012). In the symbiosis of an allogromiid Foraminifera with Pseudomonas, the method was used to detect the presence of *nirK* gene that codes for dissimilatory nitrite reductase, an enzyme necessary in the reduction of nitrate to ammonium. FISH signals from *nirK* gene, were localized exclusively in the symbiont and not in the host genes. Furthermore, the pseudomonad cells were localized on the peripheries of nitrate containing vacuoles of the host suggesting that the endosymbiont is most likely responsible for denitrification in the symbioses (Bernhard *et al.*, 2012). Furthermore, geneFISH was also used to study a symbiosis of the deep-sea mussel *Bathymodiolus puteoserpentis* with its two bacterial endosymbionts, a methane-oxidizer (MOX) and a sulfur-oxidizer (SOX). The method resolved which of the 2 endosymbionts could potentially use hydrogen as an energy source as the gene that codes for the hydrogenase, *hupL,* was targeted. For gene detection, the use of 4 polynucleotides probes hybridizing consecutive regions of the gene helped increasing the sensitivity and, in combination with ribosomal probes specific for the identification of each symbiont, it could be shown that the SOX symbiont harbored the *hupL* gene (Petersen *et al.*, 2012).

The direct linking of gene presence and cell identity also can strengthen ecophysiology studies of uncultured microbial clades. For instance, geneFISH was applied to sulfide enrichments originating from coastal sediments in the Wadden Sea. The method allowed a direct link of the *dsrA* gene (dissimilatory sulfite reductase) to a phylotype of the Roseobacter clade bacteria (RCB), a clade with a high metabolic flexibility. The direct link would not have been possible by any other means because of the lack of a pure culture and because the probe targeting the *dsrA* gene was designed from a metagenomic fragment which lacked the 16S rRNA gene for phylogenetic identification (Lenk *et al*., 2012).

Table 1. GeneFISH reported in the literature, showing the origin of the samples evaluated, the source sample and the target for gene identity (rRNA) and gene detection, and the element cycle that the target gene is involved in. * Mixes that required multiple probes to cover allele diversity, ** Probes targeting consecutive regions of a gene to increase sensitivity. n.a not applicable (see text).

Environmental origin	Sample source	Cell Identity		Gene detection		Element cycle involved	Reference
		rRNA probe	Taxon target	Gene probe (length in bp)	Gene target		
Benguela upwelling zone - Namibian shelf	Planktonic	Cren554	marine group I Crenarchaeota	amoA 13 probes (450)*	Ammonia mono-oxigenase A (*amoA*)	Nitrogen cycle	Moraru *et al.*, 2010
Hydrothermal vent-*Bathymodiolus puteoserpentis*	Eukaryotic-prokaryotic symbioses	SOX-specific	Sulfur-oxidizer endosymbiont	** hupL-1 (350) hupL-2 (350) hupL-3 (350) hupL-4 (350)	Hydrogenase (*hupL*)	Hydrogen cycle	Petersen *et al.*, 2011
Santa Barbara basin, California, USA-allogromiid Foraminifera	Eukaryotic-prokaryotic symbioses	Pseudomonas-specific	Pseudomonas endosymbiont	nirK-probe (473)	dissimilatory nitrite reductase (*nirK*)	Nitrogen cycle	Bernhard *et al.*, 2012
Wadden Sea Sediment - Sylt, Germany	Sulfide enrichment	ROS537	Roseobacter clade	dsr285_RCB probe (359)	Reverse dissimilatory sulfite reductase (*dsrA*)	Sulfur cycle	Lenk *et al.*, 2012
Contaminated ground water with toxic chlorinated ethene	PCE-to-ethene dechlorinating enrichment culture contaminated ground-water	n.a	*Dehalococcoides mccartyi*	tceA-probe (558) vcrA-probe (223)	Reductive dehalogenase (*tceA* and *vcrA*)		Matturro & Rossetti., 2015
Guaymas Basin deep see sediments	Butane enrichment	DSS548	Deltaproteobacteria	masD 8 probes (300)*	(1-methylalkyl) succinate synthase (active subunit) (*masD*)	Carbon cycle (Hydrocarbon degradation)	Staggars *et al.*, submitted

CARD-FREE FISH FOR THE DETECTION OF LOW COPY TARGETS

Moraru and collaborators (2010) reported a detection efficiency of GeneFISH of only ~40%. Such a low efficiency currently limits the widespread usage of the protocol and therefore GeneFISH is often used only for qualitative analysis. Unlike the 16S rRNA-FISH, genes are typically present in one or few copies and therefore any gene hybridization signals need to be strongly amplified. The current protocol relies on signal amplification by CARD of fluorescent dyes. For CARD, cells need to be permeabilized to permit the penetration of the HRP-labeled oligonucleotide probes. Too harsh treatment leads to leakage of DNA out of the cells and consequently to loss of geneFISH signals. Similarly CARD includes incubations with hydrogen peroxide, which damages DNA as well. In summary, the hybridization efficiency is strongly depended on the quality of the target gene in the cell. Besides these drawbacks, CARD methods are laborious and time consuming. But there are alternatives and other, CARD-free approaches that have been developed to enhance FISH signals. These FISH techniques use dye-labeled probes to intra-cellularly localize nucleic acids (e.g., genes and mRNAs).

The simplest approach to amplify low copy signals is to use polynucleotide probes that carry multiple fluorescent labels. In microbes the approach is challenged by the small size of the genes (on average 1000 bp). The usage of fluorochrome-labeled probes requires a long target region to boost the sensitivity. For instance, the plasmid F partitioning and the intracellular localization of the origin of replication during *E.coli* cell division was shown with FISH targeting a region of 13.5 and 6.2 kb length respectively (Niki and Hiraga, 1997; Niki and Hiraga, 1998).

Alternative, an approach that is routinely used for the detection and quantification of low to moderate mRNAs copies is the usage of dye-labeled oligonucleotide probes. In microorganisms this has been used on *Escherichia coli* (Coleman *et al.*, 2007; Montero Llopis *et al.*, 2010; Skinner *et al.*, 2013) and on *Caulobacter crescentus* (Montero Llopis *et al.*, 2010). The high intensity signal is generated using a probe mix of 48-72 oligos and sensitive detection systems (i.e., high quantum efficiency CCD camera, a strong light source and high numerical aperture) (Coleman *et al.*, 2007; Skinner *et al.*, 2013). Alternative, the stochastic expression of a gene of *Bacillus subtilis*

was shown using six oligonucleotide probes multiple times labeled (Maamar *et al.*, 2007).

The probes can also be labeled with brighter molecules like quantum dots (QD). QD are spherical semiconductors crystallites of nanometer scale. Compared to the organic dyes QD are 10-20 times brighter, more photo-stable and have narrow emission spectra (Kwon, 2013). For *in situ* hybridizations, the QD have been used to simultaneously detect mRNA, mutations on human sperm (Pathak *et al.*, 2001), and to localized the plasmidic genes on *E.coli* (Wu *et al.*, 2010). However, the application of QD can be limited by their large size: their diameter varies from of 4 to 10 nm (Pawley, 2010) and their molecular weight ranges from 3.2 - 1000 kDa. (PlasmaChemGmbH©:www.plasmachem.com/download/Quantum_dots_ sizes_Plasma Chem _GmbH.pdf)

Another signal amplification technique is using so-called oligo-paints (Beliveau *et al*., 2012). The oligo-paints are single-stranded DNA fluorescently labeled probes used to visualize genomic regions of ~10 kb. The first step for oligo-paints synthesis is the construction of libraries of single stranded 72-mers; the libraries are designed with a bioinformatic pipeline (for details see Beliveau *et al*., 2012). Each 72-mer contains 32 bases homologous to the target (genome) and two 21 bases flanking regions for primer annealing. One primer, called MainStreet (Beliveau *et al.*, 2015), harbors a fluorochrome at the 5'-end and the other primer contains a recognition site for a nicking endonuclease, to make the oligopaint single stranded. The recognition site is oriented in a way that the nicking occurs at the 3'-end of the 32 complementary bases (Figure 4.A) (Beliveau *et al.*, 2012). Furthermore a sequence extension added to the MainStreet, permits the hybridization with a second labeled probe (Figure 4.B). In this way the signal is amplified directly in the place of hybridization. This improvement, in combination of stochastic optical reconstruction microscopy (STORM), permitted the detection with high-resolution (16 nm) of single nucleotide polymorphisms SNPs in eukaryotic cells. (Beliveau *et al.*, 2015).

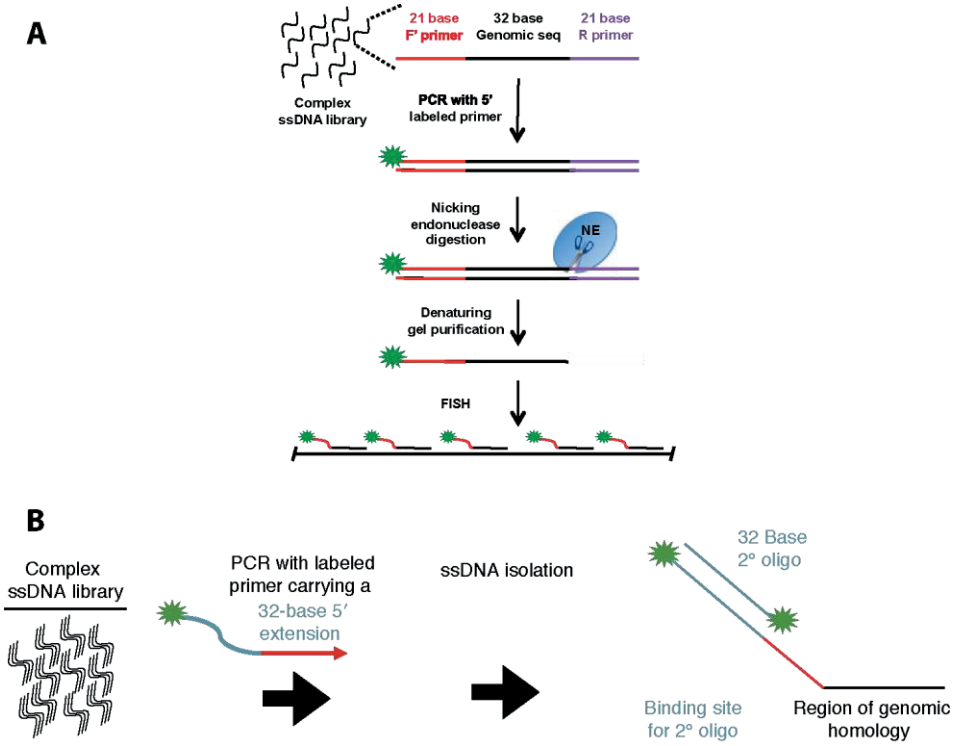

Figure 4. Oligo-paint synthesis strategy is based on the DNA amplification of complex ssDNA libraries. The libraries are composed of 72-mers. Each 72-mer consists of a genomic sequence (black lines) flanked by non-genomic regions for primer annealing (red and purple). One flanking region is annealed with a primer that carries a 5′ fluorophore, MainStreet (red), and the other primer contains a recognition site for endonuclease (purple). The probes are purified in denaturing gel electrophoresis. Image taken from (Beliveau *et al*., 2012). b. A secondary binding site for and oligo is introduced to the MainStreet (Red) by PCR amplification with a primer that carries the binding site. Image taken from (Beliveau *et al*., 2015)

A method that elegantly amplifies the FISH signals is the so-called hybridization chain reaction, HCR. In HCR the hybridization of an initiator probe activates a cascade of hybridization events between two DNA hairpins that have complementary tails (Figure 5) (Dirks and Pierce, 2004). HCR was adapted (Yamaguchi *et al*., 2015a; Yamaguchi *et al*., 2015b) and improved for the detection of microorganisms with low rRNA content, presenting results equivalent to CARD-FISH (Yamaguchi *et al*., 2015b). Additionally, HCR-FISH was also used for the detection of *luxA* gene transcripts of *Vibrio fischeri* living in symbiosis with a bobtail squid (Nikolakakis *et al.*, 2015).

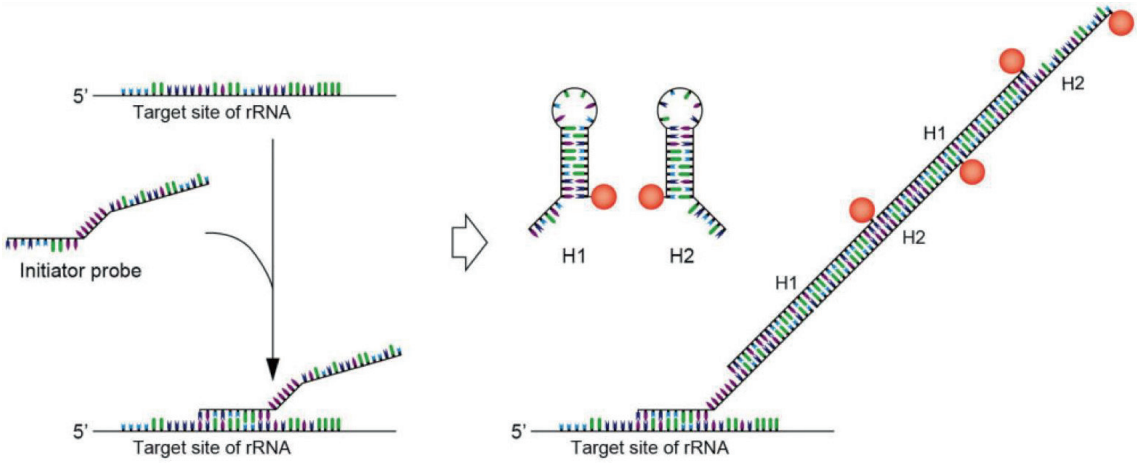

Figure 5. HCR-FISH on microbial cells. Image taken from (Yamaguchi *et al*., 2015a).

FLUORESCENCE MICROSCOPY

Fluorescence microscopy is a type of optical microscopy in which the magnified image is generated from the detection of molecules that are able to fluoresce. Fluorescence refers to the capacity of certain molecules, like fluorophores, to absorb light at a particular wavelength and release it with a different wavelength, that usually is longer than the original one (Lichtman and Conchello, 2005). The difference between the absorbed and emitted light wavelengths is known as the Stokes shift. This property is exploited in fluorescence microscopy. A fluorophore-labeled specimen is illuminated with light of a certain wavelength. The scattered light is then filtered out and only the light emitted by the fluorophore passes. The signal after the filtering component is detected. As a result, only the fluorophore-labeled molecules will be visible (Jost and Heintzmann, 2013).

This type of microscopy is the preferred tool for *in situ* visualization of environmental microbes because it is specific (i.e., only labeled molecules are visualized), versatile (i.e., samples can be labeled with multiple fluorophores for visualization and/or localization of different cellular components or cells) and sensitive. Furthermore, compared to other microscopy technologies (e.g., TEM and AFM), fluorescent microscopy is cheap and requires easy and less invasive sample preparation (Schermelleh *et al.*, 2010). Nevertheless, fluorescence microscopy is limited by resolution. Resolution is defined as the smallest distance between two points on a

specimen that can still be distinguished by the observer (Coltharp and Xiao, 2012). In the best case scenario, meaning that the sample preparation is excellent and the microscope is equipped with the finest optical lenses and detection systems, a light microscope can only resolve objects that are separated by 200 to 350 nm (Schermelleh *et al.*, 2010). Thereby, all valuable information below this limit is lost. The limit in resolution is due to intrinsic properties of the wavelengths, known as the diffraction limitation, and was mathematically explained by Ernst Abbe in 1873 (Fernández-Suárez and Ting, 2008). The diffraction limitation is approximately half the wavelength of detected light (Coltharp and Xiao, 2012).

ADVANCES IN RESOLUTION OF FLUORESCENCE MICROSCOPY: SUPER-RESOLUTION MICROSCOPY

Super-resolution techniques have been developed to overcome the diffraction limitation barrier (Figure 6). Super-resolution microscopy techniques are divided into two main categories: "patterned approaches" (Huang *et al.*, 2009) and single molecule localization (Gahlmann and Moerner, 2013). The first category uses spatial patterns for filtering the excited fluorescence. This category includes structure illumination microscopy (SIM), stimulated emission depletion (STED) and reversible saturable optical fluorescence transitions (RSOFT). The second category is based on the detection of single molecules that are photoswitchable or photoactivable, for example photoactivated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM). In the scope of this thesis, the three main super-resolutions techniques, PALM/STORM, SIM, STED, are briefly described.

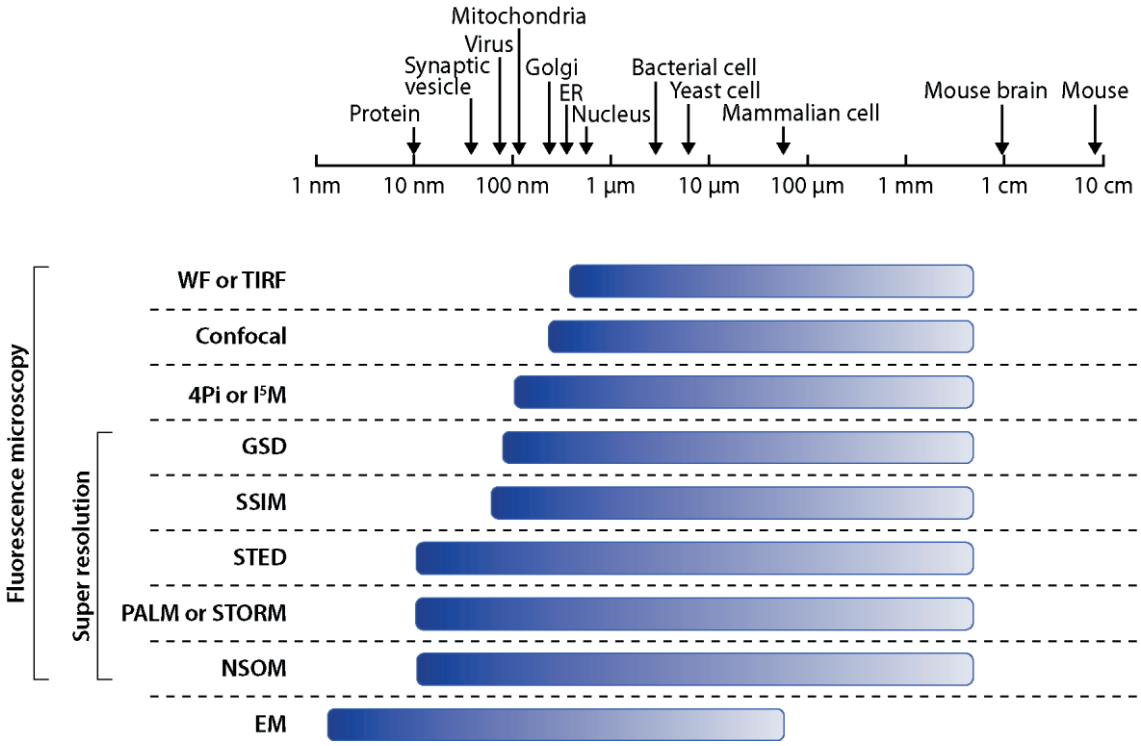

Figure 6. Resolution limits of several microscopy approaches. Image modified from Fernandez-Suarez and Ting, 2008.

PALM and STORM belong to pointillistic techniques, analog to a painting style in which discrete colored dots are applied in patterns to form an image. (Jost & Hentzmann, 2013). These microscopies are based on photoactivable (PALM) or photoswitchable (STORM) molecules that have the property of conversely move from a dark to an excited state (Huang *et al.*, 2009). This property allows the temporal separation of molecules enabling the accurate localization of single molecules (Schermelleh *et al.*, 2010). On a single cycle, most of molecules are off and only few are on an excited state; excited molecules are recorded. On next cycles, other molecules are stochastically on and localized (Figure 7.a). The full image, a pointillistic map, is reconstructed when this process is repeated thousands of times. The resolution limit is between 20 and 50 nm (Jost and Heintzmann, 2013).

In SIM microscopy, the laser light, used for fluorochromes excitation, passes through a movable optical grid generating an interference pattern. Multiple interference patterns are obtained with the angular rotation of the grid (Schermelleh *et al.*, 2010). Each interference pattern generates a sub-image and all sub-images acquired are processed by specialized software (e.g., Zen black edition) that generates the high-resolution image (See Jost and Heintzmann, 2013 for a detailed description of SIM

image reconstruction). The high-resolution image is based on moiré fringes created for superimposition of the interference patterns (Figure 7.b) (Coltharp and Xiao, 2012). SIM has a resolution of 50-130 nm on the xy-plane and 250-350 nm on the z-axis (Huang *et al.*, 2009).

In STED microscopy second laser (STED laser) is used to selectively deplete the fluorescence emitted out of the center of excitation (Hell and Wichmann, 1994) (Figure 7.c). The STED laser pattern is typically generated with the insertion of a phase mask into the light path that generates a donut-shaped pattern in the xy-plane (Huang *et al.*, 2009). The resolution of this SR microscopy is 30-100 nm on the xy-plane (Schermelleh *et al.*, 2010) and 30–40 nm on the z-axis (Huang *et al.*, 2009).

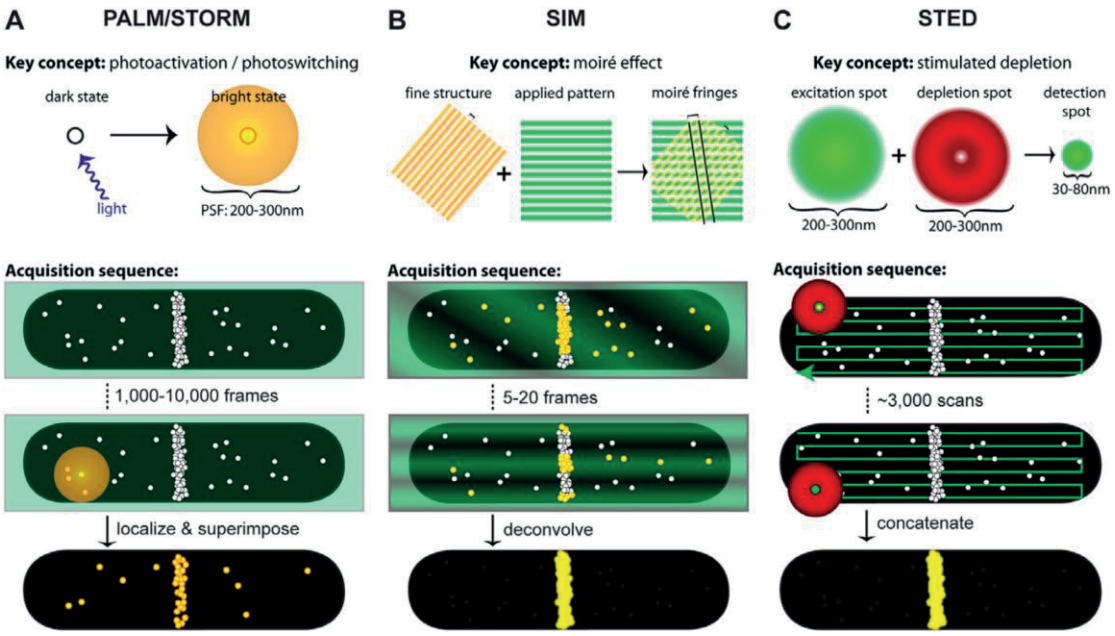

Figure 7. Main concepts and acquisition schematics for each super-resolution technique. White circles: molecule positions. Green: excitation light. Red: depletion light. Yellow: fluorescent proteins. Image taken from (Coltharp and Xiao, 2012).

To the date, super-resolution microscopy has been used in bacteria for understanding regulation processes, like transcription and translation coupling, chromosome partitioning and structural intracellular organization of nucleoid and cytoskeleton proteins. The studies reported so far, have focused on the localization of genetically modified proteins in model organisms (Reviewed in Colharp and Xiao, 2012; Gahlmann and Moerner, 2014).

THESIS OBJECTIVES

GeneFISH is a robust method for linking gene presence and cell identity in diverse environmental microbes. Nevertheless, the ~40% detection efficiency limited the method for quantitative applications. Furthermore, the protocol is long and labor intensive. On one hand, many steps decrease the quality of the sample. On the other, the complexity of the method does not allow the routinely usage.

Therefore the aim of this thesis was to systematically improve the geneFISH protocol; by increasing the sensitivity, reducing the overall handling time and to develop a simplified protocol making use of recent advances in fluorescence microscopy.

Objective I: To increase geneFISH detection efficiency and to shorten the protocol published by Moraru *et al*., 2010. The protocol optimizations allowed the development of a variation of geneFISH, PhageFISH. (**CHAPTER I: Optimization of CARD based geneFISH protocol**)

Objective II: To develop a simplified CARD-free geneFISH protocol in which the detection is based on fluorochromes carried by the probes. This will make the method faster and, coupled with super-resolution microscopy increase signal resolution and allow signal quantification. (**CHAPTER II: Development of direct-geneFISH**)

Objective III: To optimize direct-geneFISH on *Formosa* strain B cells to show the direct link of phylogenetic affiliation and mannan degradation capability as predicted from metagenomic analyses. Direct-geneFISH is optimized on pure cultures of *Formosa* strain B targeting a gene that codes for a glycosyl hydrolase involved in the hydrolysis of mannan, GH92. (**CHAPTER III: Direct-geneFISH on *Formosa* strain B**).

LIST OF MANUSCRIPTS

CHAPTER I

Manuscript I: Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses.

Elke Allers, Cristina L Moraru, Melissa B. Duhaime, Erica Beneze, Natalie Solonenko, Jimena Barrero-Canosa, Rudolf Amann and Matthew B. Sullivan.

Published in Environmental Microbiology 15 (8): 2306 -2318 (2013)

E.A., M.B.D., C.L.M., R.A. and M.B.S. conceived the study and wrote the manuscript. E.A., C.L.M., M.B.D., E.B., N.S., J.B.C., designed and performed the experiments. E.A., C.L.M., N.S., E.B. and J.B.C. evaluated the samples. E.A., C.L.M. and M.B.D. analyzed the data.

Book Chapter: PhageFISH for monitoring phage infections at single cell level.

Cristina L. Moraru and Jimena Barrero-Canosa.

Submitted as Chapter of the book Bacteriophages: methods and protocol. Editors: Martha Clokie, Rob Lavigne and Andrew Kropinski

C.L.M. wrote the chapter. J.B.C assisted in the writing. C.L.M., J.B.C., designed and performed the experiments.

Manuscript II: Detection of a virus infection in marine seawater samples by phageFISH

Jimena Barrero-Canosa, Cristina Moraru, Rudolf Amann and Bernhard M Fuchs.

Draft manuscript. Intended as brief communication in Environmental Microbiology Reports

B.M.F, R.A., J.B.C conceived the study. B.M.F., J.B.C., C.L.M and R.A designed the experiments. J.B.C performed experiments and evaluated the samples. J.B.C wrote the manuscript under editorial assistance of B.M.F.

CHAPTER II

Manuscript III: Direct geneFISH: A quantitative protocol for simultaneous Fluorescence *In Situ* Hybridization (FISH) of genes and rRNA in microorganisms at single cell level.

Jimena Barrero-Canosa, Cristina Moraru, Laura Zeugner, Bernhard M. Fuchs, Rudolf Amann.

Draft manuscript. Intended as paper in Environmental Microbiology

R.A., C.L.M., B.M.F. and J.B.C. conceived the idea. C.L.M., J.B.C., designed the experiments. J.B.C., L.Z. performed the experiments. J.B.C evaluated the samples. J.B.C and C.L.M. analyzed the data and wrote the manuscript under editorial assistance of R.A and B.M.F.

CHAPTER III

Manuscript IV: *In situ* detection of a glycosyl-hydrolase 92 in *Formosa* strain B - A case study

Jimena Barrero-Canosa, Karen Krueger, Tanja Fischer, Jens Harder, Hanno Teeling, Bernhard M Fuchs, Rudolf Amann.

Manuscript in preparation

B.MF. and R.A. conceived the idea. K.K., H.T. performed metagenomic analyses. J.H and T.F. Maintained and provided the pure cultures *Formosa* strain B and the strain closed genome. J.B.C. designed and performed geneFISH experiments. J.B.C. wrote the manuscript with the editorial assistance of B.M.F

OTHER CONTRIBUTIONS

Symbiont heterogeneity in the deep sea: Some sulfur-oxidizing symbiont strains can use methanol as energy source

Lizbeth Sayavedra, Miguel A. Gonzalez-Porras, Jimena Barrero-Canosa, Antony CP, Nicole Dubilier, Jillian M. Petersen.

Manuscript in preparation

J.B.C contribution: Direct-geneFISH experimental design

CHAPTER I:
OPTIMIZATION OF CARD BASED GENEFISH PROTOCOL

I.1 - MANUSCRIPT: SINGLE-CELL AND POPULATION LEVEL VIRAL INFECTION DYNAMICS REVEALED BY PHAGEFISH, A METHOD TO VISUALIZE INTRACELLULAR AND FREE VIRUSES.

ELKE ALLERS, CRISTINA L. MORARU, MELISSA B. DUHAIME, ERICA BENEZE, NATALIE SOLONENKO, **JIMENA BARRERO-CANOSA**, RUDOLF AMANN, MATTHEW B. SULLIVAN

ENVIRONMENTAL MICROBIOLOGY 2013 15(8): 2306 -2318

The pdf-document of this publication is not displayed due to copyright reasons. The publication can be accessed at: http://onlinelibrary.wiley.com/doi/10.1111/1462-2920.12100/pdf

I.2 - BOOK CHAPTER: PHAGEFISH FOR MONITORING PHAGE INFECTIONS AT SINGLE CELL LEVEL

CRISTINA MORARU & **JIMENA BARRERO-CANOSA**

SUBMITTED AS CHAPTER OF THE BOOK "BACTERIOPHAGES: METHODS AND PROTOCOLS"

EDITORS: MARTHA CLOKIE, ROB LAVIGNE AND ANDREW KROPINSKI

Title: PhageFISH for monitoring phage infections at single cell level

Cristina Moraru[1*], Jimena Barrero-Canosa[2]
[1] Institute for Chemistry and Biology of the Marine Environment, Oldenburg
[2] International Max Planck Research School of Marine Microbiology, Bremen
*Corresponding author, email address: CristinaMoraru@TheScientist.info

Summary/Abstract

PhageFISH uses the power of Fluorescence *in situ* Hybridization to monitor intracellular phage infections at single cell level. It combines host cell identification via rRNA probes and phage identification via phage specific gene probes, allowing not only the quantification of the infected cell fraction, but also the discrimination between infection stages. This book chapter is covering all the aspects of the procedure, from the phage probe design and synthesis, to the phageFISH protocol itself, to the microscopy and image analysis.

Key Words

phageFISH, virus, phage, microorganisms, fluorescence in situ hybridization, FISH, infection cycle, infection stages

1. Introduction

PhageFISH[1] is based on the detection of phage genes by Fluorescence *in situ* Hybridization (FISH) as a means to quantify viral infections in microorganisms, at single cell level. It has been applied in one-step growth experiments[1] to follow phage-host infection dynamics, where it provided two metrics: the fraction of infected cells (a quantitative metric), and the relative extent of per-cell phage infection (the phage signal area per cell, a semi-quantitative metric which allows discriminations between new and advanced infections). Using the two metrics, the method allowed modeling of the infection stages (adsorption, replication, assembly and lysis), and also, discrimination between subsequent waves of infection. Moreover, the protocol was able to detect free phage particles in the lysis stage of infection.

The protocol is based on the geneFISH protocol[2]. It combines host cell identification by rRNA-targeted oligonucleotides (Fig. 1, A, B and C) with phage gene detection (Fig. 1 D, E and F). For gene detection multiple ~300 bps long dsDNA polynucleotide probes (Fig. 1 D) are used which are labeled with digoxygenin (Dig). Subsequently, anti-Dig antibodies conjugated with horseradish peroxidase (HRP) are applied. The bound HRP enzymes catalyze the covalent binding of multiple fluorochrome-labeled tyramides to cellular proteins in a so-called Catalyzed Reporter Deposition step (CARD). This is resulting both in signal amplification and fixation of the signal inside the cells (Fig 1 E). By dual color epifluorescence microscopy host cells can be identified in one color and intra-cellular and extra-cellular phage particles in another color (Fig. 1F).

So far phageFISH has been applied to pure cultures, for modeling the infection dynamics of a lytic phage-host system. However, its use can be extended to the study of lysogenic systems and, because it allows both host and virus identification, also to the

study of more complex environmental systems. It should furthermore be possible to apply phageFISH not only to double stranded DNA viruses, but also to single stranded DNA viruses and RNA viruses.

2. Materials

Always use ultrapure, 0.22 µm filtered water for the preparation of solutions. Unless indicated otherwise, prepare and store the solutions at room temperature. Avoid exposing the fluorescent reagents to light, by storing them in non-transparent tubes/racks or wrapped in aluminum foil. Several of the chemicals used are toxic and/or volatile. Use appropriate protection measures, e.g. always work with formamide and paraformaldehyde in fume hood cabinets equipped with special waste disposal bins.

2.1. Stock solutions and chemicals

1. PCR Dig Probe Synthesis Kit (Roche, cat. no. 11636090910). Store at -20°C.
2. Alternative to the PCR Dig Probe Synthesis Kit: 1 mM Dig-dUTPs (Jena Biosciences, cat. no. NU-803-DIGXS), 5 Prime Master Taq Kit (5 Primer, cat. no. 2200230), 100 mM dNTP Set, PCR Grade (Invitrogen, cat. no. 10297-117). Store at -20°C.
3. Gene Clean Turbo kit (Q-Biogene, cat. no. 1102-600) or QIAquick PCR purification kit (Qiagen, cat. no. 28106).
4. 3-aminopropyl-triethoxysilane (TESPA), or poly-L-lysine
5. 1x PBS, pH 7.4: 1370 mM NaCl, 27 mMKCl, 80 mM Na_2HPO_4, and 20 mM KH_2PO_4, prepared from 10x PBS by mixing 1 part 10x PBS and 9 parts water.
6. 20% paraformaldehyde, electronic microscope grade (Electron Microscopy Sciences, cat. no. RT 15713).
7. 1 M Tris-HCl, pH 8.0
8. 0.5 M EDTA, pH 8.0
9. Buffer TE: 5 mM Tris, 1 mM EDTA, pH 8.0.
10. 20% SDS
11. 20x SSC
12. 5 M NaCl.
13. 37% HCl (~ 10 M HCl).
14. 96% Ethanol.
15. Lysozyme, powder (AppliChem, cat. no. A4972.0010).
16. 10 mg ml^{-1} sheared salmon sperm DNA (Ambion, cat. no. AM 9680).
17. 10 mg ml^{-1} yeast RNA (Ambion, cat. no. AM 7118).
18. Dextran sulfate, sodium salt.
19. Formamide, 100%, molecular grade (Sigma).
20. Nucleic acid blocking reagent (Roche, cat. no. 11096176001).
21. Maleic acid buffer: 100 mM Maleic acid, 150 mM NaCl, pH 7.5, adjusted with concentrated or solid NaOH, sterile.
22. Alexa488 and Alexa594 labeled tyramides (see **Note 1**).
23. RNase I (Ambion, cat. no. AM 2295).
24. RNase A (Sigma, cat. no. R4642-10).

25. Anti-Dig-POD Fab fragments (Roche, cat. no. 11207733910): prepare stock solution as indicated by the manufacturer, store at +4°C. Do not vortex antibody containing solutions!
26. Western Blocking Reagent, Solution (Roche, cat. no. 11921673001).
27. HRP-labeled 16S rRNA-targeted oligonucleotide probes (Biomers, Germany): prepare stock solutions as indicated by the manufacturers. From the stock solutions, prepare working solution to a concentration of 50 ng μl^{-1}. Never freeze HRP stocks once thawed. Do not vortex HRP probes.
28. SlowFadeGold (Invitrogen, cat. no. S36936) or ProLong Gold antifade reagent (Invitrogen, cat. no. P36930).
29. 4′,6-diamidino-2-phenylindole (DAPI), 1 mg ml^{-1}(Sigma-Aldrich).

2.2. Glassware and plastic ware

1. Thin forceps, from acid, bases, organic solvents and temperature resistant materials (for example from Electron Microscopy Sciences, cat. no. 72692-F).
2. Petri dishes, various sizes, sterile, DNase free.
3. 15 and 50 ml Falcon tubes, sterile, DNase free.
4. Scalpels: sterile, disposable.
5. Hybridization chambers: any tightly closing, temperature resistant container that seals with a silicone O-ring, e.g. food containers used in the kitchen.
6. Secure-Seal hybridization chamber gasket (Molecular probes, cat.no. S24732).
7. 0.22 µm sterile syringe filters.
8. 0.2 µm polycarbonate membrane filters (GTTP, Millipore, cat. no. GTTP02500).
9. Diamond Retractable Tip Scriber: for writing on glass, metal and plastic (Electron Microscopy Sciences, cat. no. 70036).
10. Glassslides, frosted.
11. Poly-L-Lysine Coated Slides (Electron Microscopy Sciences, 63410-02)
12. Coverslips, # 1.5, high precision (Marienfeld, cat. no. MARI0107052v).

2.3. Laboratory equipment

1. Incubators: at different temperatures, at. 37°C, 42°C, 46°C, 85-90°C.
2. Slide Denaturation/Hybridization system machine (alternative to incubators).
3. Water baths: at different temperatures, at 37°C, 42°C, 48°C.
4. Table top centrifuges (14000 rpm max)
5. Agarose gel electrophoresis devices.
6. Spectrophotometer or other device for measuring nucleic acid concentration (e.g Nanodrop)
7. Filtration device (e.g. Millipore) and vacuum pump.
8. Epifluorescence microscope: equipped with a 63x or 100x objective, with fluorescent filter sets suitable to discriminate between Alexa488 and Alexa594 fluorescence, with a black and white CCD camera and with software for image capture. Example of filter sets: for Alexa488 - 472/30 excitation, 520/35

emission and 495 Beam Splitter; for $Alexa_{594}$- 562/40 excitation, 624/40 emission and 593 Beam Splitter.

2.4. Software

1. PolyPro[3]: for polynucleotide probe design (http://www.mpi-bremen.de/genefish.html)
2. DAIME[4]: for microscopy image analysis (http://www.microbial-ecology.net/daime/).

2.5. Working solutions

1. Permeabilization solution: 0.5 mg ml^{-1} lysozyme (AppliChem, cat. no. A4972.0010), 1x PBS pH 7.4, 0.1 M Tris-HCl pH 8.0 and 0.05 M EDTA. First, prepare 50 ml of permeabilization buffer, by mixing 5 ml 10x PBS, 5 ml 1 M Tris-HCl pH 8.0, 5 ml 0.5 M EDTA pH 8.0 and 35 ml water. Second, prepare a 5 mg ml^{-1} lysozyme solution, e.g. by dissolving 50 mg lysozyme in 10 ml permeabilization buffer. Warm up at 37°C to dissolve, if necessary. Finally, mix 1 part of 5 mg ml^{-1} lysozyme solution with 9 parts permeabilization buffer, to get a final concentration of 0.5 mg ml^{-1} lysozyme. Always use a freshly made permeabilization solution.
2. 0.01 M HCl: to 50 ml of water add 50 µl 37% HCl, mix.
3. 0.2 M HCl: to 49 ml of water add 1 ml HCl, mix.
4. rRNA hybridization buffer: 35% formamide (see **Note 2**), 10% dextran sulfate (DS), 0.9 M NaCl, 20 mM Tris-HCl pH 8.0, 1% nucleic acid blocking reagent, 0.25 mg ml^{-1} sheared salmon sperm DNA, 0.25 mg ml^{-1} yeast RNA and 0.02% SDS. For preparation of 40 ml hybridization buffer proceed as follows. In a 50 ml tube add 4 g DS, 7.2 ml 5 M NaCl, 0.8 ml 1 M Tris-HCl pH 8.0 and 4 ml water, close the tube and shake/vortex strongly to disperse the DS in solution. Incubate on water bath at 37-48°C to dissolve the DS, vortex and shake from time to time. After the DS has dissolved completely, allow the solution to cool down to RT, and then add 4 ml 10% nucleic acid blocking reagent, 1 ml 10 mg ml^{-1} sheared salmon sperm DNA, 1 ml 10 mg ml^{-1} yeast RNA, 17.5 ml 100% formamide and 40 µl 20% SDS (see **Note 3**). If necessary, adjust with water until the mark of 40 ml. Vortex to mix components, followed by quick spin. Optionally, the buffer can be filtered through 0.22 µm sterile syringe filters. Aliquot and store at -20°C. Before use warm up at 37°C, to re-dissolve any precipitates.
5. rRNA hybridization buffer-probe mix: to 900 µl rRNA hybridization buffer add 3 µl of 50 ng $µl^{-1}$ HRP labeled rRNA probe, to a final concetration of 0.17 ng $µl^{-1}$ (see **Note 4**). Mix by gentle shaking or up and down pipeting. Do not vortex, otherwise the HRP probes could be damaged. Prepare freshly, just before use, keep at room temperature until hybridization.

6. rRNA hybridization - washing buffer: 70 mM NaCl (see **Note 5**), 5 mM EDTA pH 8, 20 mM Tris-HCl and 0.01% SDS. In a 50 ml tube, add 700 µl 5 M NaCl, 500 µl 0.5 M EDTA, 1 ml 1 M Tris-HCl, fill with water up the 50 ml mark and then add 25 µl 20% SDS.
7. rRNA CARD buffer: 1x PBS, 10% DS, 0.1% nucleic acid blocking reagent and 2 M NaCl. For preparation of 40 ml rRNA CARD buffer proceed as follows. In a 50 ml tube add 4 g DS, 4 ml 10X PBS pH 7.4, 16 ml NaCl and water to a final volume of 40 ml, close the tube and shake/vortex strongly to disperse the DS in solution. Incubate on water bath at 37-48°C to dissolve the DS, vortex and shake from time to time. After the DS has dissolved completely, allow the solution to cool down to RT, and then add 400 µl 10% nucleic acid blocking reagent, vortex, quick spin. Filter through 0.22 µm sterile syringe filters. Aliquot and store at +4°C. Before use warm up at 37°C, to re-dissolve any precipitates.
8. rRNA CARD buffer-tyramide mix: to the rRNA CARD buffer, add H_2O_2 to a final concentration of 0.0015% and Alexa488 tyramides to a concentration of 0.33 µg ml^{-1} (see **Note 6**). For example, to 3 ml of buffer add 30 µl of 100x H_2O_2 (always prepare freshly by mixing 1 µl 30% H_2O_2 and 200 µl 1x PBS) and 1 µl of 1 mg ml^{-1} Alexa488 tyramides, vortex, quick spin. Prepare freshly, just before use, and keep a room temperature, in dark, until rRNA CARD.
9. RNase solution: 0.1 U $µl^{-1}$ RNase I, 75 µg ml^{-1} RNase A, 0.1 M Tris-HCl, pH 8.0. For preparation of 12 ml RNase solution, in a 15 ml tube add 10.8 ml water, 1.2 ml 1 M Tris-HCl pH 8, 15 µl 100 U $µl^{-1}$ RNase I and 30 µl 10 mg ml^{-1} RNase A.
10. Gene hybridization buffer: 35% formamide, 5x SSC, 10% DS, 0.1% SDS, 20 mM EDTA, 1% nucleic acid blocking reagent, 0.25 mg ml^{-1} sheared salmon sperm DNA and 0.25 mg ml^{-1} yeast RNA. For preparation of 40 ml gene hybridization buffer proceed as follows. In a 50 ml tube add 4 g DS, 10 ml 20x SSC, 1.6 ml 5 mM EDTA pH 8 and 4.4 ml water. Close the tube and shake/vortex strongly to disperse the DS in solution. Incubate on water bath at 37-48°C to dissolve the DS, vortex and shake from time to time. After the DS has dissolved completely, allow the solution to cool down to RT, and then add 4 ml 10% nucleic acid blocking reagent, 1 ml sheared salmon sperm DNA, 1 ml yeast RNA, 14 ml formamide and 200 µl 20% SDS. Vortex to mix components, followed by quick spin. Optionally, the buffer can be filtered through 0.22 µm sterile syringe filters. Aliquot and store at -20°C. Before use warm up at 42°C, to re-dissolve any precipitates.
11. Gene hybridization buffer-probe mix: to the gene hybridization buffer add each probe to a final concentration of 5 pg $µl^{-1}$. For example, to 1 ml gene hybridization buffer add for each probe 1 µl of 5 ng $µl^{-1}$ probe stock. Vortex to mix, followed by quick spin. Prepare freshly, in the same day as hybridization, and store at room temperature before use. See **Note 7** for the concentration of the negative control probe.

12. Gene hybridization washing buffer I: 2x SSC and 0.1% SDS. In a 50 ml Falcon tube, add 5 ml 20x SSC, water to a final volume of 50 ml and 250 µl 20% SDS, vortex. Prepare for use within 1-2 days; keep at 42°C (see **Note 8**).
13. Gene hybridization washing buffer II: 0.1x SSC and 0.1% SDS. In a 50 ml tube, add 250 µl 20x SSC, water up to 50 ml and 250 µl 20% SDS, vortex. Prepare for use within 1-2 days; keep at 42°C (see **Note 8**).
14. Antibody blocking/washing solution: 1x PBS and 1% Western Blocking Reagent (WBR) (see **Note 9**). To a 50 ml tube, add 5 ml 10x PBS, 5 ml 10% WBR and water to a final volume 50 ml, vortex. Prepare in the same day as the antibody step, keep at room temperature until use.
15. Antibody binding solution: 1x PBS, 1% Western Blocking Reagent (see **Note 9**) and 0.3 U ml^{-1} (500x dilution of the 150 U/ml stock) anti-Dig HRP-conjugated antibody. To a 15 ml Falcon tube, add 1 ml 10x PBS, 1 ml 10% WBR, 8 ml water and 20 µl antibody (just before use, centrifuge the antibody stock solution for 10 min at 10 000x g and +4°C, to deposit any precipitates). Mix gently. Do not vortex antibody containing solutions. Prepare freshly, keep at room temperature until use.
16. Gene CARD amplification buffer: 1x PBS, 20% DS, 0.1% blocking reagent, and 2 M NaCl. For preparation of 40 ml gene CARD buffer proceed as follows. In a 50 ml Falcon tube add 8 g DS, 4 ml 10x PBS pH 7.4, 16 ml NaCl and 15.6 ml water, close the tube and shake/vortex strongly to disperse the DS in solution. Incubate on water bath at 37- 48°C to dissolve the DS, vortex and shake from time to time. After the DS has dissolved completely, allow the solution to cool down to RT, and then add 400 µl 10% nucleic acid blocking reagent, vortex, and quick spin. Filter through 0.22 µm sterile syringe filters. Aliquot and store at +4°C. Before use warm up at 37°C, to re-dissolve any precipitates.
17. Gene CARD amplification buffer-tyramide mix: to the rRNA CARD buffer, add H_2O_2 to a final concentration of 0.0015% and Alexa594 tyramides to a concentration of 2 µg ml^{-1}. For example, to 1 ml of buffer add 10 µl of 100x H_2O_2 (always prepare freshly by mixing 1 µl 30% H_2O_2 and 200 µl 1x PBS) and 2 µl of 1 mg ml^{-1} Alexa594 tyramides, vortex, quick spin. Prepare freshly, just before use, and keep a room temperature, in dark, until rRNA CARD.
18. Antifade reagent – 2 µg ml^{-1} DAPI mix: to 1 ml antifade reagent (SlowFade Gold or ProLong Gold) add 2 µl of 1 mg ml^{-1}, vortex, quick spin. Store at -20°C.

3. Methods

3.1. Phage probe design

1. Use the bioinformatics tools you have in your laboratory to select a phage specific genomic region (not more than 70% identity with other sequences, see **Note 10**), which is not found in the other microbial members of the sample being studied (e.g. the host, or co-infecting viruses, other bacteria, archaea, viruses, etc., which might be present in the sample). This requires having a

priori genomic / metagenomic information for the sample of interest. If such information is not available, than compare your phage genome against the NCBI database.

2. Next, in the selected region identify at least 6 probes (see **Note 11**), each ~300 bps, with similar %GC. In a first step, calculate the variation of the %GC along the selected DNA region, using the bioinformatics tools in your laboratory. One such tool is DAN (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::dan), which can calculate the denaturation profile, including the %GC. When using DAN, chose as parameters: "window size" = 100, "shift increment" = 1, "DNA concentration" = 1 nm, "salt concentration" = 1000 mM and "output format" = excel. In Excel, plot the base position (for DAN, that would be column "Start" in the output file) versus the %GC (see **Note 12**). To select the probes, avoid sequence stretches with high variations (e.g. more than 10 units) in the %GC. Furthermore, if possible, chose sequence stretches with low %GC (see **Note 13**). Perform a Blast search with the individual probes against the database relevant for your samples, to identify potential non-specific binding sites (regions with more than 80% identity on higher than 20-30 base stretches). If necessary, discard probes with potential for non-specific binding. The last step is to design primers for the ends of each 300 bps probe.

3.2. Phage probe synthesis

The phageFISH probes are dsDNA molecules labeled with Dig. They are produced by incorporating Dig into dsDNA during a PCR reaction (see Fig. 1 D).

1. Optimize the PCR conditions (to avoid expenses you can leave out Dig) for the template and primers of interest. Use either viral DNA or plasmid DNA as PCR template.
2. Proceed to the probe synthesis PCR. A labeling kit is available from Roche - the PCR Dig Probe Synthesis Kit. Follow manufacturer's instructions for probe synthesis. Because Dig incorporation results in lower product yields, you can use two kit reactions per probe and pool them for purification. For an alternative to the kit, see **Note 14**.
3. Purify the PCR products using PCR purification kits, as for example the Gene Clean Turbo kit or QIAquick PCR purification kit. At the end of the purification, elute the probes in TE buffer, pH 8.0.
4. Check the probes on 2.5-3% agarose gels. Due to Dig incorporation, the probes will migrate slower than the unlabeled counterparts.
5. Measure the probe concentration using a spectrophotometer.
6. Prepare stocks of 5 ng μl^{-1} each probe, by diluting in TE buffer, pH 8.0.
7. Store at -20°C.

3.3. Determination of the stringency parameters for gene hybridization

The hybridization and washing stringency will influence not only the specificity, but also the detection efficiency of the hybridization. The stringency refers to how close to the melting temperature of the probe-target hybrids the hybridization or washing takes place. It can be modulated by modifying the formamide concentration in the hybridization buffer and by changing the hybridization and the washing temperatures.

1. Calculate the formamide concentration which will allow a hybridization temperature in the range 42-50°C (see **Note 15**). For this, use the HPC module of the PolyPro software to calculate for each probe/target pair the graph of the melting temperature variation function of formamide concentration. As input parameters use DNA:DNA hybridization, Na+ concentration of 975 mM, formamide in between 1 and 100 % and criterion of 25 (distance from Tm, see **Note 16**) . As output option, chose "Hybridization temperature function of % formamide". Using the graphs for all probe/target pairs, select a formamide concentration that will give a hybridization temperature in the range 42-50°C for every probe/target pair.
2. Use the chosen formamide concentration to test the detection efficiency at different gene hybridization temperatures (in the range 42-50°C), by performing the whole phageFISH protocol. The optimum hybridization temperature will be the one giving the highest detection rate (after correction for the false positives in the negative control – see **Note 17**).

3.4. phageFISH protocol

During the FISH procedure the cells are immobilized on solid support, most on 0.22 μm pore size polycarbonate filters or by glass slides. See **Note 18** for general instructions for handling polycarbonate filters and **Note 19** for glass slides. Avoid excessive light exposure during the procedure. Avoid sample drying, unless specifically instructed in the protocol (see **Note 20**). Unless stated otherwise, perform all incubations at room temperature (RT). Pre-warm or pre-cool the reagents before use, to bring them at the required incubation temperature. Whenever working with toxic substances – paraformaldehyde during fixation, formamide during hybridization, etc. – use a chemical fume hood.

1. Sample fixation and immobilization. Add 20% PFA directly to the sample of interest, to result in a final concentration of 1-4% PFA. Incubate for 1 hour at RT or overnight at +4°C (see **Notes 21 and 22**). To remove the PFA and bring the cells on solid support, filter the fixed culture on 0.22 μm polycarbonate filters (see **Notes 23** and **24** for other methods to remove PFA and immobilize the cells). Apply a vacuum pressure as low as possible as and not higher than 0.2 mBar. Initially, test different culture volumes to see which one gives a uniform distribution of cells on the filter, avoiding too little or too many cells. Consider using a volume that will give a denser cell distribution because during the

phageFISH protocol some cells will detach from the solid support (see **Note 25**). After filtering the fixed culture, filter through 10-15 ml 1xPBS for washing and then 10-15 ml water. Allow filters to air-dry and store at -20°C, or directly proceed further.

2. Permeabilization. Overlay the filters with permeabilization solution, incubate on ice for one hour, and then wash 5 min with 1x PBS and 1 min with water. Permeabilization can be sample specific. See **Note 26** for more details.
3. Inactivation of endogenous peroxidases. Immerse the filters in 0.01M HCl for 10 min, and then wash 1x PBS for 5 min, water for 1 min and 96% ethanol for 1 min. Allow filters to air dry and store at -20°C, or directly proceed further. Inactivation of peroxidases can be sample specific. See **Note 27** for more details.
4. rRNA hybridization. Place the filters face up on Petri Dishes and cover the samples with the hybridization mix. Transfer them to a humidity chamber (see **Note 20**). Incubate at 46°C for 1.5-3 h. For washing, quickly rinse the samples in rRNA washing buffer, then transfer them in fresh rRNA washing buffer and incubate for 15 min at 48°C.
5. CARD for rRNA detection. Incubate the samples for 10-15 min in 1x PBS. Further, transfer the samples in rRNA-CARD buffer-Alexa488 tyramide mix and incubate for 10 min at 37 °C (see **Note 28**). For washing, quickly rinse the filters in 1x PBS, then transfer in 1x PBS, 10 min at 46 °C, followed by 1 min with water and 1 min with 96% ethanol. Allow filters to air-dry and store at -20°C, or directly proceed further.
6. RNase treatment (see **Note 29**). Cover the samples with RNase solution and incubate overnight at 37 °C. Wash twice for 5 min in 1x PBS and once for 1 min with water.
7. Inactivation of HRP introduced with the rRNA probe. Immerse the samples in 0.2 M HCl for 10 min. Wash with 1x PBS for 1 and 5 min, then 1 min with water, 1 min with 96% ethanol. Allow filters to air-dry and store at -20°C, or directly proceed further.
8. Gene hybridization - prehybridization. Place the filters face up on Petri Dishes and overlay them with 30-100 µl of hybridization buffer. Introduce the Petri Dish in a humidity chamber (see **Note 20**) and incubate for 0.5-1 h at the hybridization temperature (see "Determination of stringency parameters for gene hybridization" section).
9. Gene hybridization – denaturation and hybridization. On a Petri dish, place as many 30-100 µl droplets of gene hybridization buffer-probe mix as the number of filters. Gently, remove the filters from the prehybridization buffer and place them face down into the droplets of gene hybridization buffer-probe mix. See **Note 30** for working on glass slides.

   Place the samples back in the humidity chamber and denature for 1 h at 85°C-90°C (see **Note 31**). Further, quickly move the humidity chambers in an oven set at the hybridization temperature and hybridize for 2 h or overnight (see **Note 32**). For washing, immerse samples in gene washing buffer I, 3x for 1 min at RT

and 30 min at 42 °C, followed by gene washing buffer II, 3x for 1 min at RT and 1.5 h at 42 °C, in a slow shaking water bath, and finally 1 min in 1x PBS at RT.

10. Antibody binding. Incubate the filters in antibody blocking solution for 30 min. Transfer samples in antibody binding solution and incubate for 1.5 h. For washing, immerse samples in antibody washing solution for 1 min and then 3x 10 min. Slow shaking (e.g. 20 rpm) during these steps could improve results, but could also result in cell loss.
11. CARD for gene detection. Cover samples with gene CARD buffer-Alexa594 tyramide mix and incubate for 45 min at 37 °C. Quick wash for 1 min in 1x PBS at RT, and then for 5 min and 2x 10 min with 1x PBS in a 46 °C oven, slow shaking, then 1 min with water, 1 min with 96% ethanol. Allow filters to air-dry and store at -20°C, or directly proceed further.
12. Embedding and counterstaining. Place the filters face up on a microscopy glass slide, in a droplet of embedding reagent. On top of the sample add sufficient embedding reagent (usually 2-5 µl per filter piece), so that when placing next a coverslip on top, the whole sample surface will be covered by embedding reagent. If a non-hardening media is used, e.g. SlowFade Gold, then samples can be imaged immediately. If a hardening media is used, e.g. ProLong Gold, then allow the samples to cure for 24 h at RT.

3.5. Microscopy and data analysis

1. Image acquisition. Use the Alexa488 filter set to image the 16S rRNA signals and the Alexa594 filters set to image the phage signals. Because the phage signals can vary greatly with respect to intensity, take a series of images with increasing exposure times. Short exposure times will capture the advanced infections, which give strong signals, while overlooking the weak signals. The long exposure times will result in overly saturated signals from the advanced infections, while capturing the weak signals from early infections / free phage particles. Adjust the lowest exposure time such that the advanced infections are visible, but their signal is not over saturated. Adjust the highest exposure time such that the early infections / free phage particles are visible, but not over saturated. In between the lowest and highest exposure time, set several time exposures. These will help with image analysis later. For each sample acquire several images representing different fields of view. Each field of view will have images corresponding to different acquisition channels – the Alexa488 channel, and the many Alexa594 channels, each at a different exposure time.
2. Image processing with DAIME (see **Note 33**).

   - From the time exposure series, chose the exposure time where the strong signals from advanced infections are not overexposed, but also, where the weak signals from early infections / free phage particles are also present, albeit almost invisible (see **Note 34**). This exposure time will be used for phage signal quantification, and thus, termed further the “phage quantification exposure time”.

- Import images in DAIME grouped in "stacks". One stack will contain all images (all fields of view) from one replicate corresponding to an acquisition channel. Each replicate sample per time point will have three stacks: the rRNA signals, the phage signals from the phage quantification exposure time and the phage signals from the highest exposure time. To ensure correct correlation between the three channels of one field of view, each field of view should receive a number and this number should be present at the end of all images from that field (see **Note 35**).

- Segment objects in each stack. This will identify each individual cell or phage signal and will transform them in "objects". Test different segmentation algorithms, choosing the one which best identifies individual signals (see **Notes 33 and 36**).

- Create a 4th stack by replicating the phage quantification stack and removing the mask. Transfer the object layer from the phage highest exposure time stack to the newly created phage quantification stack – further named "single phage quantification stack".

- In "Visualizer" module, create a new Session by adding the following segmented stacks: the 16S rRNA stack, the phage quantification stack and the single phage quantification stack (from the above step). For each field of view, compare the objects in the two phage stacks, removing any duplicated object (if the same object appears in both stacks, remove one of them).

- For each stack, measure the segmented objects ("Analysis menu") and export the data as csv files. The most important parameters are: total area, mean intensity, pixels, centroid X (pix) and centroid Y (pix).

3. Data analysis.

- Import the data in Excel. Calculate the phage signal intensity of each phage signal by multiplying the mean intensity column with the pixels column.

- For all time points in the experiment, plot the phage signal intensity versus the phage signal area (total area parameter) in the same graph (data grouped by time points and replicates).

- Use the graph generated above to define three phage signal size classes. Since the time 0 infections should represent new infections, use their phage signal area range as the definition for the first size class. Further, identify the first time point where both the signal area and the signal intensity are highest. This time point is most likely represented by advanced infections (late replication and encapsidation). Use the phage signal area range in this time point as the definition for the third size class. All the signals in between the 1st and the 3rd size class can be considered as the second size class (ongoing replication).

- To calculate the cell fraction infected by phages per each time point, count all the cells displaying phage signals and calculate their percentage relative to the total host cells (the number of objects in the corresponding 16S rRNA stack). Most of the times, to one cell corresponds one phage signal. In the cases where multiple phage signals are presented per one cell, they have to be counted as one

(this information can be obtained by comparing the pixel coordinates of the cells with that of phage signals – the centroid X (pix) and centroid Y (pix) columns for each stack).

4. Image processing for "display". Because the difference in signal strength prevents acquisition of images where both early and advanced infections are visible in a meaningful way, image processing algorithms have to be applied on the time exposure series, to merge the information from the different exposures. Such algorithms are called High Dynamic Range algorithms and they often are included in software packages for image processing.

4. Notes

1. Preparation of fluorescently labeled tyramides has been described by Pernthaler and Pernthaler (2005)[5].
2. To ensure specific hybridization, the formamide concentration in the rRNA hybridization buffer is probe specific. Thus, when preparing the rRNA hybridization buffer, both the water and 100% formamide volumes have to be adjusted accordingly. For more information about the formamide concentration of established probes, you can consult the online database http://www.microbial-ecology.net/probebase/
3. To avoid precipitation, always add SDS as the last component.
4. A higher concentration of 16S rRNA probe can be used if the target cells are low in ribosomal content. Do not exceed a probe concentration of 0.5 ng μl^{-1} as this can result in strong background signals.
5. The NaCl concentration in the washing buffer varies with the probe, more exactly, with the formamide concentration used in the hybridization buffer for each probe. The following is listing the correspondence between the formamide (FA) concentration and NaCl concentration, when washing at 48°C: 0% FA – 900 mM NaCl, 5% FA – 636 mM NaCl, 10% FA – 450 mM NaCl, 15% FA – 318 mM NaCl, 20% FA – 225 mM NaCl, 25% FA – 159 mM NaCl, 30% FA – 112 mM NaCl, 35% FA – 80 mM NaCl, 40% FA – 56 mM NaCl, 45% FA – 40 mM NaCl, 50% FA – 28 mM NaCl,, 55% FA – 20 mM Nacl, 60% FA – 14 mM NaCl.
6. The concentration of the Alexa488-tyramides can be adjusted depending on the requirements of the microbial cells. For fast growing cells, with high number of ribosomes, a lower concentration of tyramides can be used (e.g. 0.33 $\mu g\ ml^{-1}$), while more tyramides are necessary for slow growing cells, with low number of ribosomes. Generally, do not exceed a concentration of 2 $\mu g\ ml^{-1}$, as it can result in background formation. If you want to preserve sub-cellular localization of the rRNA signal, use less Alexa488 tyramides. Furthermore, too much Alexa488 tyramides in the rRNA step can saturate the tyramide binding sites in the cell, leaving too little binding sites for the Alexa594 tyramides in the gene CARD step, and thus, decreasing phage detection.

7. When using the Non-Poly350Pr as negative control, add it to a concentration equivalent to that of all the probes in the positive control probe mix. For example, if the positive control probe mix contains 6 probes, the final concentration of the Non-Poly350Pr should be 30 pg μl^{-1}.
8. Store gene hybridization washing buffers at the washing temperature, if different from 42°C.
9. The antibody step can result in false positives signals. To reduce the background formation, western blocking reagent is used before and during the antibody binding step. If increased background is a problem for your samples, you can test stronger blocking mixes, e.g. by adding bovine serum albumin or sheep serum to the antibody blocking and binding solutions. On the other hand, a too strong blocking can reduce the detection efficiency, as it competes with the specific antibody binding sites.
10. When working with a group of similar phages, keep in mind that polynucleotide probes are not able to discriminate between closely related sequences. On the other hand, this enables the use of polynucleotide probes for targeting mismatched sequences[2,3]. One polynucleotide probe can be used to detect targets with 5% mismatches, with a slight decrease in the hybridization rate (and thus, detection efficiency). Increasing the mismatches will lead to a further decrease in the hybridization rate, until, at around 20% mismatches, no hybridization will take place. The PolyPro software[3] can be used to design polynucleotide probes for multiple alleles. From a multiple alignment of the allelic regions, select 300 bps stretches which show minimum sequence variation. For each of the 300 bps stretches, the PolyPro software will search for the probe/probe mixes which will enable binding to all targeted alleles.
11. The detection efficiency, i.e. the percentage of the cells which show a positive gene signal from all cells having the respective target gene, depends both on the number of polynucleotide probes used and on the number of targets per cells. Allers, Moraru et al. 2013[1] show that, for 3-8 gene copies per cell, while 1 probe gave 70% detection efficiency, increasing the number of polynucleotides increased the detection efficiency to >90% with 4 polynucleotides and 98% with 12 polynucleotides. For cells with a high number of target copies, 1 probe was sufficient for 90% detection efficiency, while 3 probes gave 100%. Calculations indicate that, for cells with one gene copy, at least 12 polynucleotides will be necessary for 100% detection[a].

    On the other hand, using similar approach Matturro et al. [6] reached 100% detection efficiency with a single polynucleotide probe. The factors that could contributed to the increased on the detection efficiency is the higher denaturation

[a] Detailed information can be found on the Supplementary data from the manuscript: "Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses" Allers *et al*., 2013.

temperature (from 85°C for 1 h in a humidity chamber, to 90°C for 20 min in PCR tubes), a better accessibility to chromosomal regions that are actively transcribed because they tend to stand in a decondensate state[7], and a higher number of targets per cell often found in dividing bacteria.

12. In the plot, the %GC at a certain base position will correspond to the sequence stretch from the respective base position plus 99 bases (this for cases when the "window size" parameter for calculating the %GC has been set to 100).
13. Target sequences with lower %GC will denature more easily and allow access of the probes to the individual DNA strands.
14. Alternatively, probes can be synthesized by using Digoxigenin-11-dUTP nucleotides, any *Taq* polymerase kit (e.g. 5 Prime MasterTaq kit) and unlabeled nucleotides. High Fidelity polymerases shoul not be used because they do not incorporate modified nucleotides. For synthesis, start preparing a mix of unlabeled nucleotides with a concentration of 2 mM per dATP, dCTP and dGTP and 1.3 mM dTTP. For example, add 2 µl of 100 mM dATP, 2 µl 100 mM dCTP, 2 µl of 100 mM GTP and 1.3 µl 100 mM dTTP to 92.7 µl water. Mix by vortexing and store at -20°C. Next, prepare a 100 µl PCR reaction with the following concentrations: 1x *Taq* buffer, 200 µM of each dATP, dCTP and dGTP, 130 µM dNTP, 70 µM Dig-11-dUTP, 1 µM of each primer, 0.04 U of *Taq* polymerase and 50 ng of DNA template. The concentration of the DNA template can affect the success of obtaining a high probe concentration, it is recommended to do serial dilutions of 1:10, 1:100 and 1:1000 of the template and run PCR reactions in parallel to find the optimal DNA concentration for the PCR mix.
    For the PCR thermocycling, start with an initial denaturation step of 5 min at 95°C, followed by 30 cycles of: denaturation (1 min at 95°C), annealing (1 min at 55-60°C) and elongation (1-3 min at 72°C). Finalize the reaction with an elongation step at 72°C for 10 min. The annealing temperature and the elongation time have to be optimized for the primers and template of interest [b].
15. Because high temperatures are damaging the cells, formamide is used in hybridization buffers to lower the melting temperatures and to allow stringent hybridization at relatively low temperatures (42-50°C). On the downside, high concentrations of formamide are decreasing the hybridization rate.
16. Theoretical considerations indicate that the highest hybridization rate for polynucleotides is at 25°C below their Tm. Getting closer to the Tm will decrease the hybridization rate (and thus, detection efficiency), while getting away from the Tm will not only decrease the hybridization rate, but also favor the formation of short mismatched hybrids.
17. There are two controls which can be used for phageFISH – i) the same sample, but not infected, and/or ii) a negative control gene probe, e.g. NonPolyPro350[2].

[b] Experiments carried out with an alternative kit for labeling polynucleotide probes via PCR are detailed in the Appendix I.2.1.1

18. Working with filters
    - To label the filters, use a carbon pencil and write preferably on the edge of the filter, where there are no cells
    - To treat the filters with different reagents:
        - If the reagents used are rather expensive and economical use is preferred (e.g. when working with enzymes or with hybridization buffers): place the filters face up on a Petri dish plate, add the reagents on top of the filters while making sure that the filters are completely covered.
        - If the reagents are relatively inexpensive (e.g. when inactivating the endogenous peroxidases with 0.1 M HCl): fill a 25/50 ml Petri dish (depending on the number of filters, crowding should be avoided) with the reagent of interest and immerse the filters in it. Make sure the filters are completely immersed and they are not floating on top of the solution.
    - To wash the filters:
        - If the incubation is performed at room temperature or in the oven, then place the filters in a 50 ml Petri Dishes filled with the washing solution of interest
        - If the incubation is performed in the water bath, then place the filters in 50 ml tubes filled with the washing solution of interest. To remove them from the tube, either pour the solution in a Petri Dish and remove the filters from there, or pour the solution through a ceramic sieve which will catch the filters.
    - To cut the filters in smaller pieces, use sterile scalpels and clean them with ethanol for reuse.
    - To dry the filters, first blot them on chromatography paper to eliminate most of the liquid, and then leave them to air dry on fresh chromatography paper. An alternative to chromatography paper is the "Kimwipes" paper, from Kimtech Science.
19. Working with glass slides
    - Buy slides already coated, e.g. with TESPA or L-polylysine. Make sure that the slides are clean before using them, to avoid background problems.
    - Alternatively, coat your own slides, e.g. with L-polylysine or with TESPA, following manufacturer's instructions. Before coating, clean the slides to remove any particles or grease, e.g. by incubating 20 min in a mixture of 10% HCl and 70% ethanol, rinsing with miliQ water followed by 95% ethanol and air-drying at 60°C.
    - To label the slides, use a Diamond Pen to make marks in the glass.
    - To mark the sample area on the glass slide, use a Diamond Pen to draw a circle or rectangle on the back of the slide. Do not draw on the sample face of the slide, since the scratched glass could result into an uneven surface and interfere with microscopy latter on.

- To treat the sample on glass slides with different reagents:
    - If economical use is preferred: place the glass slides face up, add the reagents on top of the sample are while making sure that the sample area is completely covered.
    - If the reagents are relatively inexpensive: fill a slide staining jar with the reagent of interest and immerse the slides in it. Make sure that the sample area or any other slide surfaces previously coming into contact with other reagents are completely immersed.
- To wash the samples on glass slides:
    - If the incubation is performed at room temperature or in the oven, then place the glass slides in staining jars filled with the washing solution of interest.
    - If the incubation is performed in the water bath, then place the glass slides (up to 2 slides per Falcon tube, back to back) in 50 ml Falcon tubes filled with the washing solution of interest.

20. Avoiding sample drying
    - Do not allow samples to dry unless indicated in the protocol, otherwise background formation will occur, especially in the steps using DS.
    - Avoid drying during incubations and washing steps by completely immersing the filters in the respective buffers.
    - Drying can especially occur when smaller volumes of buffers are used and/or the incubations are performed at higher temperatures. To avoid drying in these cases, the samples have to be placed in humidity chambers. A humidity chamber is represented by any tightly closing container that seals with a silicone O-ring, e.g. food containers used in the kitchen. Polypropylene containers work well for low temperature incubations (e.g. 46 °C). However, they will get deformed at high temperatures (e.g. 85°C), and in this case containers with a glass bottom part should to be used. To create and maintain humidity in the chamber, line the bottom of the container with tissue paper and soak it with liquid. The soaking liquid is either water (when the reagents are not volatile), or a water-volatile reagent (e.g. formamide, paraformaldehyde) mixture (when reagents used are volatile; in this case, the volatile reagents should have the same concentration in the soaking mixture as in the buffers themselves). For samples immobilized on filters, the filters can be placed face-up in Petri dishes, covered with buffer, and then the Petri dishes placed in humidity chambers. For samples immobilized on slides, place a PCR tube rack in the humidity chamber and then place the slides on top. Care should be taken for the humid paper not to come into contact with the sample area.
21. The fixation procedure can be sample specific. However, most often fixation is performed by using PFA. Generally, avoid long incubation times with PFA, as it can result in increased cell autofluorescence. Another fixative is ethanol, mostly used for fixation of gram-positive bacteria [8].

22. More details about variations in protocols for fixation, permeabilization, endogenous peroxidase inactivation can be found in the literature [9,10,11,12].
23. For cell concentration and PFA removal, three procedures can be used: centrifugation, filtration in Swinnex filter holders (Millipore) containing 0.22 μm polycarbonate filters or direct filtration on 0.2 μm polycarbonate membrane filters, in filter towers (described in the methods section). In all cases, harsh conditions which would promote cell lysis should be avoided.

    During centrifugation, use the minimum centrifugal force required to pellet the cells, preferably in a swinging bucket rotor. After centrifugation, remove the supernatant and suspend the cells in 1xPBS. Repeat the wash step, to ensure that all PFA has been removed. You can proceed immediately to the sample immobilization step. Alternatively, for long term storage, add 96% ethanol in a 1:1 ratio with the 1xPBS. Store the cells at -20°C.

    During filtration in Swinnex filter holders, gently push the liquids through the device using a syringe, first the fixed cells, then at least 20 ml of 1x PBS. This will remove the culture media and the PFA, while concentrating the cells in the liquid left on top of the membrane filter. To recover the cells, invert the Swinnex filter holder, position its cell end in a 2 ml tube and unscrew it. The 1xPBS containing the cells will go into the 2 ml tube. Remove the filter from the holder and place it in the 2 ml tube, then gently wash its surface by pipetting up and down before removing it. You can proceed immediately to the sample immobilization step. Alternatively, for long term storage, add ethanol in a 1:1 ratio with the cell suspension, and store at -20°C.
24. For the cells which in the fixation step were either centrifuged or processed through a filter holder (see **Note 23**), a separate step for immobilization on solid support is necessary. The solid support is represented either by 0.2 μm polycarbonate filters, or by coated glass slides (see **Note 19**). To immobilize on filters, mix different volumes of cell suspension with 10 ml 1xPBS and filter on 0.22 polycarbonate filters, followed by washing with 10-15 ml 1xPBS and 10-15 ml water. Use DAPI staining and microscopy to check which cell suspension volume gives the best distribution of cells on filters. To immobilize on glass slides, spot 10-100 μl of cell suspension in the marked area (see **Note 19**) and air-dry at 37°C. To remove precipitated salts, wash once in water and then let dry.
25. To mitigate cell loss, agarose embedding is routinely used in rRNA CARD-FISH protocols[11]. However, we have found that it significantly decreases detection efficiency of the phage genes.
26. Permeabilization is necessary to allow intracellular diffusion of the high molecular reagents used during phageFISH, e.g. HRP-oligonucleotide probes, polynucleotide probes, HRP-antibody conjugates. Depending on the cell type, different permeabilization reagents can be used, e.g. achromopeptidase for gram-positive bacteria, lysozyme for gram-negative bacteria, detergents, acids and/or proteinases for archaea. See **Note 22** for literature references.

27. For most samples, 0.1 M HCl will be sufficient to inactivate the endogenous peroxidases. However, there are samples for which a stronger inactivation is necessary (e.g. by using H_2O_2 and/or methanol). See **Note 22** for literature references. The efficiency of the inactivation can be tested by performing only the rRNA CARD-FISH part of the phageFISH protocol, both with a 16S probe (as positive control) and without any probe (as negative control).The rRNA CARD-FISH part consists from sample fixation, immobilization, permeabilization, inactivation of endogenous peroxidases, rRNA hybridization, CARD for rRNA detection and embedding and counterstaining. If microscopic evaluation reveals no signals in the negative control and signals in the positive control, then the inactivation procedure has worked. If not, further optimizations have to be performed.
28. For cells with low ribosomes content, the signal intensity of the rRNA can be enhanced by increasing Alexa488-tyramide concentration (see **Note 6**) and also, by performing the incubation with Alexa488 tyramides for longer times (e.g. 20-30 min), and/or, at 46°C.
29. RNase treatment is necessary when the discrimination of the different infection stages based on the per cell phage signal area is desired. Otherwise, binding of the phage probes to the corresponding mRNA would result in overestimation of the phage signal area.
30. If the samples are on glass slides, add the same volume of hybridization mix as for prehybridization, however, with double the probe concentration than the intended one.
31. The temperature control during the denaturation/hybridization steps is very important for the success of the method, as it influences both detection efficiency (to access the target gene, high temperature for longer times are necessary) and the cell morphology and loss (most often, high temperatures for prolonged times are damaging the cells, especially for fragile virus infected cells). When working with humidity chambers and ovens, decreasing the denaturation time from 1 h at 85°C to 30 min at 85°C lead to a decrease in the detection efficiency for early infection stages (low numbers of virus genomes per cell), while it had no influence on the late infection stages (high numbers of virus genomes per cell). On the other hand, the cells looked more damaged after 1 h denaturation. A more precise control of the temperature can be obtained in a PCR thermocycler, as Matturro et al.[6] have shown, by immersing the filters in PCR tubes with hybridization buffer. In this case, only 20 min at 90°C where sufficient to obtain 100% detection efficiency (see **Note 11**). To obtain a similar control of the temperature when working with slides, a slide thermocycling machine can be used, as exemplified below.

    Note that the slide thermocycling machines vary in their operation mode. For illustration, we are describing the conditions used with the TDH-500 Slide denaturation/hybridization system from Hangzhou Allsheng Instruments Co.

- Load the 4 liquid tanks (two on each side of the slide heating plate) with the water-formamide solution corresponding to the formamide concentration in the hybridization buffer (e.g. 35% formamide solution). In each tank, introduce a strip of a KimWipe, in such a way that one end is in the tank and the other end is overlaying the slide heating plate. This will allow the formation of a humid atmosphere during slide incubation. Prepare the glass slides with the samples as described in **Note 19**.
- To avoid evaporation of the hybridization mix, it is recommended to use secure-seal hybridization chambers of a 9 mm diameter and 0.8 mm depth. Put the chamber on top of the sample region on the glass slide (see **Note 19**). Apply pressure to seal the secure chamber to the slide.
- After the secure-seal chamber is properly bound, put the glass slides on the heating plate. Add the hybridization mix, making sure it covers the sample entirely. A total of 50 µl is needed to cover a chamber of 9 mm diameter and 0.8 mm deep.
- Close the slide thermocycling machine and verify that it is tightly sealed. If not put some weight on top of the machine (e.g. 2 L Duran bottles filled with water).
- Incubate for 5- 15 min at denaturation temperature (85 to 95°C) and then for 2 hours or overnight at hybridization temperature. The optimal denaturation conditions will vary depending on the sample and therefore they needed to be optimized for each specific viral-host system.

32. A significant decrease the gene hybridization time is obtained by increasing the DS concentration in the hybridization buffer from 10% to 20%. Prepare the gene hybridization buffer as follows: in a 50 ml Falcon tube add 8 g DS and 10 ml of 20 SSC, vortex vigorously. Add 1.6 ml 5mM EDTA pH 8 and 4 ml water. Close the tube and shake/vortex strongly to disperse the DS in solution. Incubate on water bath at 46-48°C to dissolve the DS, vortex and shake from time to time. It might take a long time (e.g. overnight) to dissolve all the DS. After the DS has dissolved completely, allow the solution to cool down to RT, and then add 4 ml 10% nucleic acid blocking reagent, 1 ml 10 mg ml^{-1} sheared salmon sperm DNA, 1 ml 10 mg ml^{-1} yeast RNA, 14 ml 100% formamide and 200 µl 20% SDS. Vortex to mix components. The buffer needs to be filtered through 0.22 µm sterile syringe filters. Aliquot and store at -20°C[c].
33. DAIME will be efficient for image analysis every time when the cells are sufficiently separated to be recognized the software (through the use of segmentation algorithms) as separate objects. When the segmentation algorithm cannot recognize individual cells, then the image analysis has to be performed in

[c] The effect of the hybridization time and the dextran sulfate concentration the detection efficiency are detailed in the chapter appendix I.2.1.2

a semi-manual way. For example, counting can be performed using the AxioVision software (from Zeiss), by manually marking the cells in the $Alexa_{488}$ channel and the corresponding gene signals in the $Alexa_{594}$ channel with the "Events" tool from the "Measure" menu. The number of events can be determined using the "measure events" function. Similarly, when the phage signals cannot be segmented by DAIME, their signal intensity has to be measured again semi-manually, with the help of image analysis software. For example, the free hand tool in Zen Lite 2011 (Blue Edition; Carl Zeiss, Germany) software can be used to mark the phage signals and measure their signal intensity.

34. Most of the times, the early infections will also be recorded in the low exposure images, but at a very low intensity and they will become visible to the eye only after image enhancement (e.g. increased brightness and magnification) and comparison with the same field of view, bur from a higher exposure times (which indicate where to look).
35. The images could be labeled for example using the following pattern: replicate_timepoint_channel__XX (where XX is the number of the field of view). See DAIME manual for further information regarding naming of files for import into stacks.
36. In our hands, the edge thresholding algorithm with worked the best.

5. Acknowledgments

The authors would like to thank Rudi Amann and Bernhard Fuchs (Max Planck Institute for Marine Microbiology, Bremen) for critical reading of the manuscript and helpful discussions, to Elke Allers (Accelerate Diagnostics, Tucson), Mellisa Duhaime (University of Michigan) and *Matthew* B. *Sullivan (University of Arizona)* for the close partnership which lead to the development of the phageFISH method, and finally, to Max Planck Society, Germany, for funding.

6. References

1. Allers, E. Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R., and Sullivan, M.D. Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15,** 2306–18 (2013).
2. Moraru, C., Lam, P., Fuchs, B., Kuypers, M. & Amann, R. GeneFISH – an *in situ* technique for linking gene presence and cell identity in environmental microorganisms. *Environ Microbiol* **11,** 3057-73 (2010).
3. Moraru, C., Moraru, G., Fuchs, B. & Amann, R. Concepts and software for a rational design of polynucleotide probes. *Environ Microbiol Rep* **3,** 69–78 (2011).
4. Daims, H., Lücker, S. & Wagner, M. Daime, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8***,* 200-213 (2006).
5. Pernthaler, A. & Pernthaler, J. Simultaneous fluorescence *in situ* hybridization of mRNA and rRNA for the detection of gene expression in environmental microbes. *Meth. Enzymol* **397,** 352–371 (2005).
6. Matturro, B. & Rossetti, S. GeneCARD-FISH: Detection of tceA and vcrA reductive dehalogenase genes in *Dehalococcoides mccartyi* by fluorescence *in situ* hybridization. *J Microbiol Methods* **110**, 27-32 (2015).
7. Wang, Y., Maharana, S., Wang, M. D. & Shivashankar, GV. Super-resolution microscopy reveals decondensed chromatin structure at transcription sites. *Sci rep* **4,** 4477 (2014).
8. Roller, C., Wagner, M., Amann, R., Ludwig, W. & Schleifer, K. H. In situ probing of gram-positive bacteria with high DNA G + C content using 23S rRNA-targeted oligonucleotides. *Microbiol* **140,** 2849–58 (1994).
9. Manz, W., Amann, R., Ludwig, W., Wagner, M. & Schleifer, K.-H. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: problems and solutions. *Syst appl microbiol* **15,** 593–600 (1992).
10. Pernthaler, A. *et al.* Sensitive multi-color fluorescence in situ hybridization for the identification of environmental microorganisms. *Mol Microbiol Ecol Manual.* **1-2,** 711–725 (2004).
11. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68,** 3094–3011 (2002).
12. Pavlekovic, M, Schmid, MC & Schmider-Poignee, N. Optimization of three FISH procedures for in situ detection of anaerobic ammonium oxidizing bacteria in biological wastewater treatment. *J Microbiol Methods* **78,** 119–126 (2009).

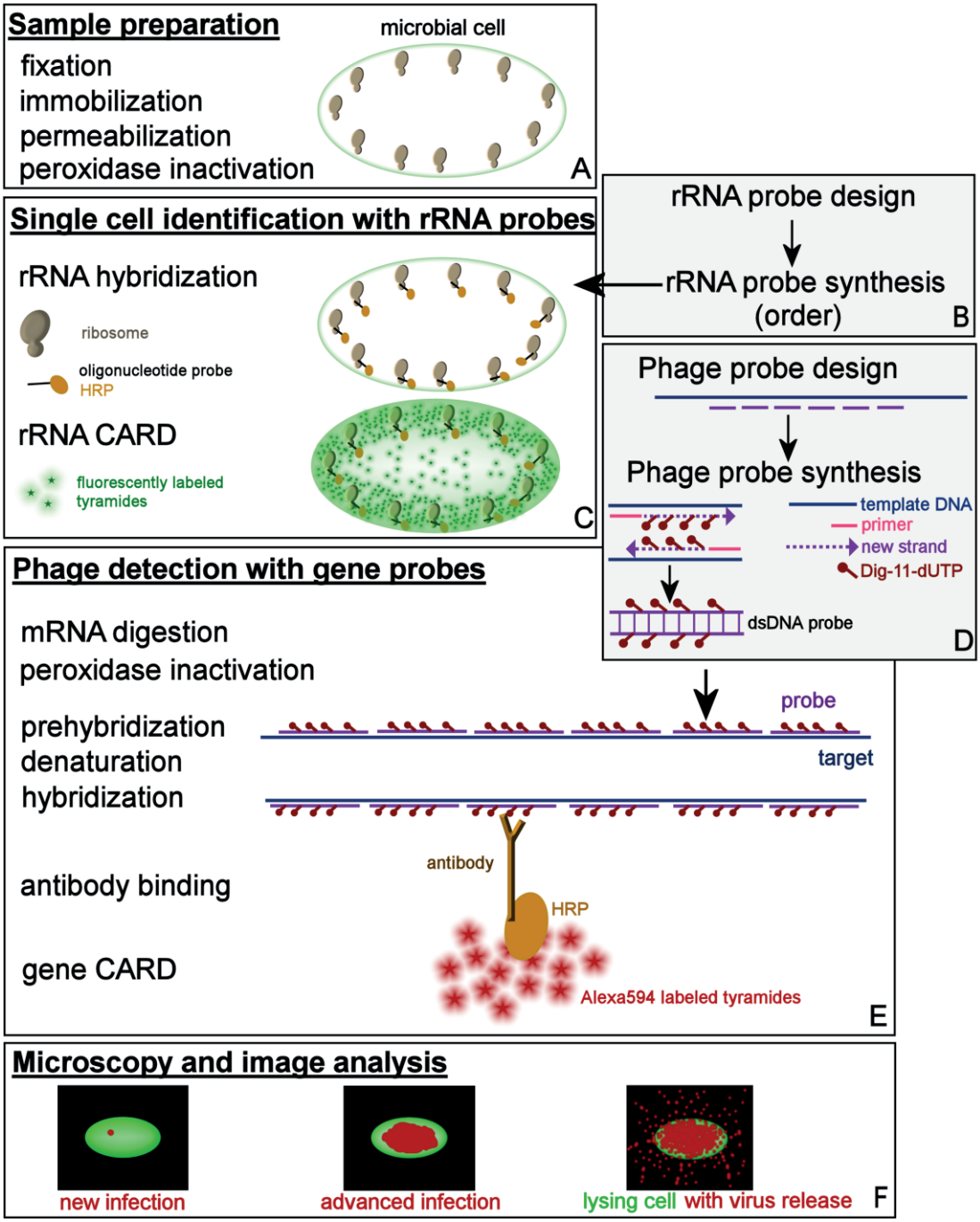

Figure 1. Workflow of the phageFISH protocol.

I.2.1 - APPENDIX BOOK CHAPTER: PHAGEFISH FOR MONITORING PHAGE INFECTIONS AT SINGLE CELL LEVEL

I.2.1.1 - ALTERNATIVE PCR- DIG LABELING PROBE KIT

The more straightforward way to synthesize and label dig-DNA polynucleotide probes is through the incorporation of modified nucleotides in a PCR reaction (McCreery, 1997). The only commercial supplier of a full component kit is Roche (PCR-DIG Probe Synthesis Kit, cat no. 11636090910). Alternatively, it is possible to create a "self-made" kit acquiring each component separately considering:

(i) Nature of the DNA polymerase: high fidelity polymerases are inefficient in the incorporation of modified nucleotides. Taq polymerase allows the dig-dUTPs incorporation (Anderson *et al*., 2005). The Taq fidelity has been extensively studied, and depending on the assays used, the error rate values ranges from 10^{-5} to 2 x 10^{-4} (McInerney *et al*., 2014). For a PCR reaction, the average mutation frequency *f* can be calculated based on the number of PCR cycles *n* and the mutation rate *p* according the formula $f = \frac{np}{2}$ (Pfeifer, 2006). For gene probe synthesis with 30 PCR cycles and a mutation rate of 2 x 10^{-4}, the mutation frequency *f* is 3 x 10^{-3}. This means a mutation every 333 bp ($\frac{1}{f}$). A mutation of 1 bp in a 350 bp probe corresponds to 0.3% of mismatch ($\frac{1\ bp\ (mismatch)}{350\ bp\ (probe\ length)}\ x\ 100\%$). For geneFISH this error can be negligible.

(ii) Selection of the modified nucleotides and their concentration in the PCR reaction: a wide range of modified nucleotides is commercially available. The analog Digoxigenin-11-dUTP nucleotide can be bought individually from several commercial distributors. The 11 refer to the length of the carbon linker arm between the digoxigenin and the nucleotide (Figure I.1). A longer linker arm allows a better accessibility for the antibody (conjugated with the HRP). Nevertheless, longer arms decrease the yield of the PCR (Jena Biosciences; Yu *et al*., 1994).

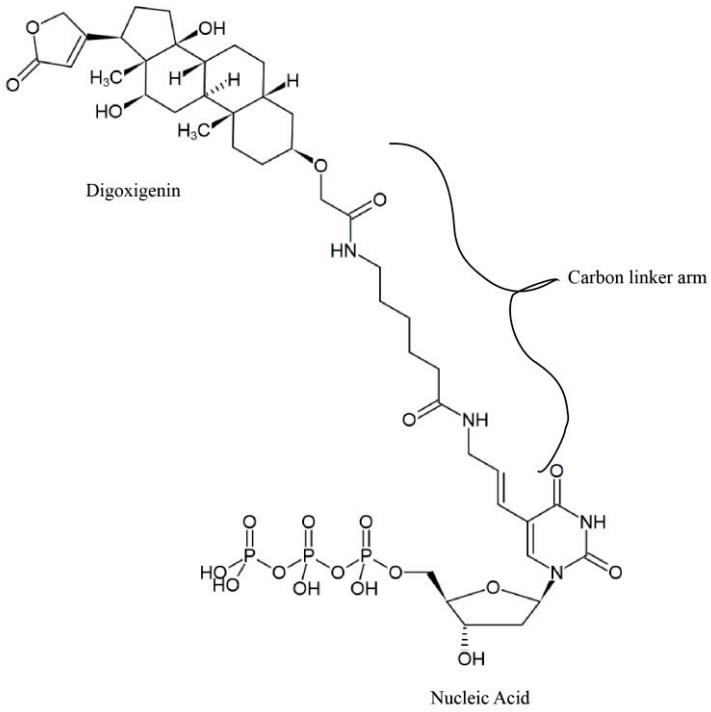

Figure I.1. Structural formula of dig-11-UTP. Image modified from Jena Bioscience dig-UTP datasheet.

The ROCHE kit and a self-made kit labeling methods were tested for geneFISH on 2% PFA fixed *E.coli* B/r cells using gene probes synthetized with both methods. Per kit, 2 polynucleotide probes (300 and 350 bp) were synthesized. The 300 bp polynucleotide probe (rpoB-PR12) targeted rpoB gene (β subunit of the polymerase) of *E.coli* and the 350 bp long Non-poly350 probe was used for negative control (Moraru *et al*., 2010). The probes labeled with Roche kit were synthesized according to the manufacturer's instructions and the self-made synthesis details are described in Note 14 of the book chapter. The geneFISH protocol was performed as described by Moraru *et al.*, 2010. Three technical replicates (~1000 cells) were counted per hybridization. The gene detection efficiency with the probe synthesized with Roche kit was on average 27% and for the self-made kit was 30%. The negative controls were 2% and 3% respectively (Figure I.2). Hence, the two labeling methods give comparable results for geneFISH hybridizations.

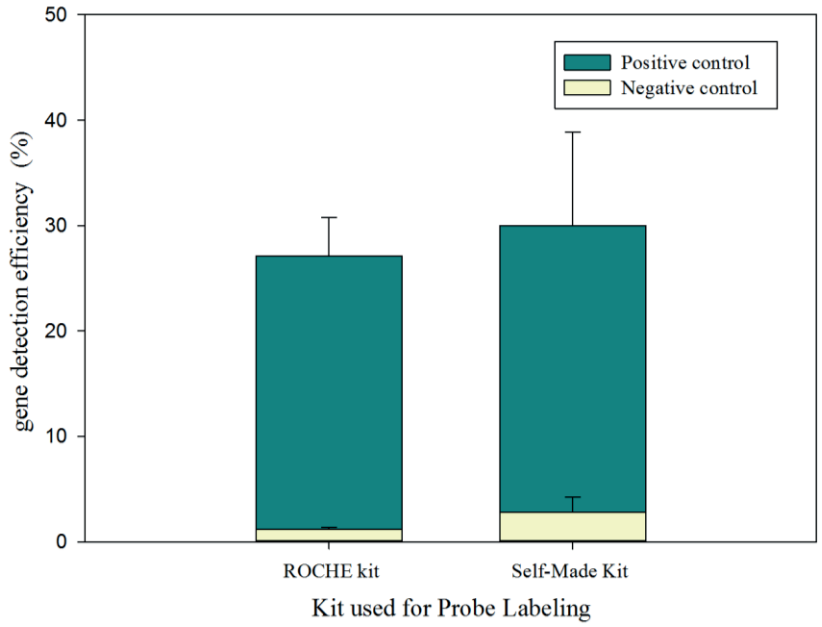

Figure I.2. GeneFISH detection efficiency for a polynucleotide dig-probe (rpoB-PR12) labeled with two approaches: using a commercial kit (ROCHE) and a self-made kit. Positive control: *E.coli* B/r and rpoB-PR12. Negative control *E.coli* B/r + Non-poly350Pr

I.2.1.2 - SHORTENING GENEFISH HYBRIDIZATION TIME AND EFFECT OF DEXTRAN SULFATE IN THE GENE HYBRIDIZATION BUFFER.

A drawback of geneFISH is that the protocol is laborious and time consuming. The time needed for executing a whole experiment is approximately 32 h (Moraru et al., 2010). One aim is to shorten the total time necessary to carry out the protocol. The long incubation time required for gene hybridization (17 h) highly contributes for the long span of the protocol. Therefore, the effect of shortening the gene hybridization time from 17 to 2 h was evaluated. For this, the protocol was carried out on *E.coli* clones harboring the *unk*-gene on copy-control plasmids (3 - 8 copies per cell). Another *E. coli* strain that did not harbor *unk* was used as negative control. A total of 6 polynucleotide probes (300 bp each) targeting consecutive regions of *unk* were used. The only modified step was the reduction in incubation time from 17 to 2 h. The remaining steps were followed as previously described (Allers *et al.*, 2013). Three replicates (≈1000 cells) were counted per hybridization. The detection efficiencies were 95% for 17 h and 92% for 2 h hybridization. Furthermore, shorter hybridization resulted in less noise 3% for 17 h to 1% 2 h (Figure I.3.a). Consequently, the gene hybridization can be decreased to 2 h without jeopardizing the efficiency of gene signal detection. Moreover decreasing

the gene hybridization time represents a reduction to half the time (from 32 to 16 h) needed for doing the whole geneFISH protocol.

The hybridization rates can also be increased with the addition of inert polymers. For nucleic acid hybridizations, inert polymers create a crowding effect so that the apparent concentration of the probes is augmented (Wetmur, 1991). Dextran sulfate (DS) has been commonly used for this purpose (Wetmur, 1975; Wahl *et al*., 1979; Yamaguchi *et al*., 2015). Hence, the effect of increasing the dextran sulfate on the gene hybridization buffer for was also investigated. For this, gene hybridization buffers containing 3 different DS concentrations were tested: 10% (standard concentration in gene hybridization buffer), 20% and 25% (w/v) DS. The hybridizations were done on the *E.coli* cells described in the previous paragraph.

For gene detection, 1 to 3 polynucleotide probes (300 bp each) targeting consecutive regions of the *unk* gene were hybridized for 2 h. Per hybridization, 3 replicates (~1000 cells) were counted. The detection efficiencies for hybridizations carried out in a buffer with 10% (w/v) were 58 %, 85% and 87% when using 1, 2 and 3 polynucleotide probes respectively. For 20% (w/v) DS the detection efficiencies were 65 %, 91 % and 97% respectively. For 25% (w/v) DS the detection efficiencies were 71%, 87% and 95% respectively. For all the experiments, the negative controls were below 3% (Figure I.3.b). The results shows that the best hybridization buffer for short gene hybridizations was containing 20% (w/v) DS since it resulted in a ~10% increase in the detection efficiency. Nevertheless, it is important to emphasize that this DS concentration should only be used for 2 h incubations because in preliminary for longer than 4 h gene hybridizations the noise was above 10%.

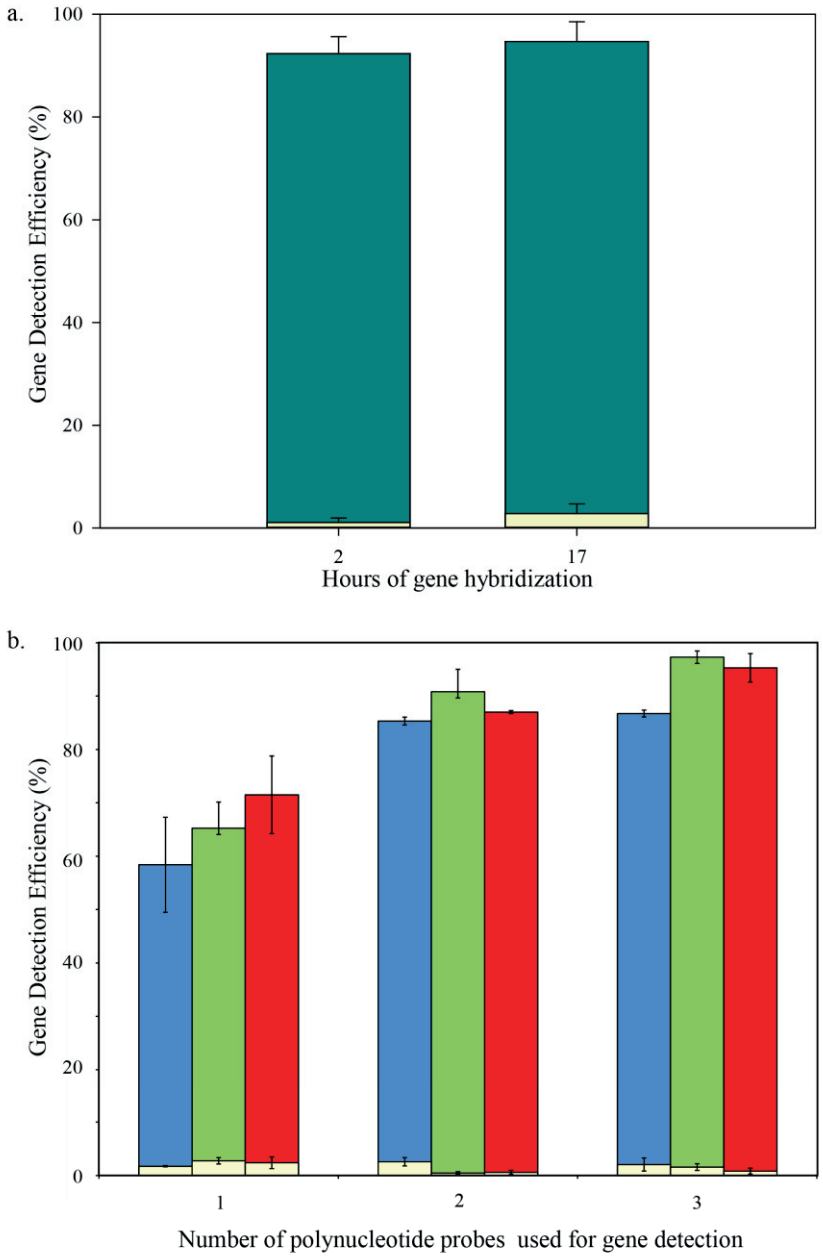

Figure I.3. Results from geneFISH experiments carried out on *E.coli* clones that harbor the target gene (*unk*) on copy-control plasmids (3 - 8 copies per cell). Another *E.coli* clone that did not carry *unk* was used for negative control. Gene detection efficiency defined by $= \frac{cells\ presenting\ \ gene\ signal}{Total\ cells} \times 100\%$. Figure I.3.a. Detection efficiency when 2 and 17 h incubation time. All the hybridizations were completed with 6 dig-labeled probes (300 bp each) targeting consecutive regions of the *unk* gene. Cyan: positive controls. Beige: Negative controls. Figure I.3.b. The effect of dextran sulfate in the gene detection efficiency for short hybridizations (2 h) and when the target gene is hybridized with 1 to 3 dig-labeled probes. DS concentrations in % (w/v) tested: 10 – blue, 20 - green and 25- red. Beige bars: negative controls.

I.2.1.3 - REFERENCES

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R., and Sullivan, M.D. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15**: 2306-2318.

Anderson, J.P., Angerer, B., and Loeb, L.A. (2005) Incorporation of reporter-labeled nucleotides by DNA polymerases. *Biotech* **38**: 257-264.

McCreery, T. (1997) Digoxigenin labeling. *Mol Biotechnol* **7**: 121-124.

McInerney, P., Adams, P., and Hadi, M.Z. (2014) Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol Biol Int* **2014**: 287430.

Moraru, C., Lam, P., Fuchs, B.M., Kuypers, M.M., and Amann, R. (2010) GeneFISH--an in situ technique for linking gene presence and cell identity in environmental microorganisms. *Environ Microbiol* **12**: 3057-3073.

Pfeifer, J.D. (2006) Polymerase chain reation In Pfeifer, J.D., ed. *Molecular genetic testing in surgical pathology*: Lippincott, Williams & Wilkins, Philadelphia, PA: 86-110

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 U S A* **76**: 3683-3687.

Wetmur, J.G. (1975) Acceleration of DNA renaturation rates. *Biopol* **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.

Yamaguchi, T., Fuchs, B.M., Amann, R., Kawakami, S., Kubota, K., Hatamoto, M., and Yamaguchi, T. (2015) Rapid and sensitive identification of marine bacteria by an improved in situ DNA hybridization chain reaction (quickHCR-FISH). *Syst Appl Microbiol*. **38**: 400–405.

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.

I.3 - DETECTION OF A VIRUS INFECTION IN MARINE SEAWATER SAMPLES BY PHAGEFISH

JIMENA BARRERO-CANOSA, CRISTINA MORARU, RUDOLF AMANN,

BERNHARD M. FUCHS

Detection of a Virus Infection in Marine Seawater Samples by PhageFISH

Jimena Barrero-Canosa[a], Cristina Moraru[b], Rudolf Amann[a], Bernhard M. Fuchs[a*]

[a] Department of Molecular Ecology Max Planck Institute for Marine Microbiology, Celsiusstr. 1, 28359 Bremen, Germany

[b] Institute for Chemistry and Biology of the Marine environment (ICBM), Carl-von-Ossietzky-Straße 9-11 D-26111 Oldenburg, Germany

*to whom correspondence should be addressed:
Max Planck Institute for Marine Microbiology
Department of Molecular Ecology
Celsiusstr. 1, 28359 Bremen, Germany
Office: +49 421 2028935, email: bfuchs@mpi-bremen.de

Keywords: Fluorescence *in situ* hybridization, phage detection, *Pseudoalteromonas*, PSA-HP1, Helgoland.

Running title: Chasing the PSA-HP1 phage in Helgoland seawater.

Abstract

In the ocean phages are ecologically important and abundant entities. Nevertheless, understanding their ecological role and host-interactions is method-limited. In recent years diverse methods to study environmental phages have been developed. One example is phageFISH, a method based on fluorescence *in situ* hybridization that allows the simultaneous visualization of phage and host with high sensitivity. PhageFISH was demonstrated first on a phage - bacterial host model system. PSA-HP1 is a strain-specific phage for *Pseudoalteromonas* strain H100 isolated from the North Sea. Our aim was to detect the phage in natural seawater samples taken from the North Sea. To increase the abundance of the host, we took advantage of the fact, that *Pseudoalteromonas* often is enriched in dilution cultures with nutrient amendments. We prepared two cultures, one with pure seawater and the other enriched with Zobell's media. Incubation was at room temperature for up to four days. Subsamples were taken every 12 - 16 h. Cell numbers of *Pseudoalteromonas* were quantified by fluorescence *in situ* hybridization (FISH). Indeed after a few days of incubation rapid growth of *Pseudoalteromonas* cells could be detected. As expected, PhageFISH could confirm the presence of the phage PSA-HP1 after 75, 113 and 138 h in the later phase of the incubations. The detection of this specific virus in natural seawater samples is remarkable because it allowed for the first time the identification of a bacterial strain at high taxonomical resolution, and simultaneously the detection of its specific phage. Our protocol showcases the potential of phageFISH for microbial ecology and opens a window to follow specific viruses in environmental samples.

Introduction

Phages are ecologically important components of the marine environments, primarily because they can lyse bacteria thus affecting the structure of microbial communities and transferring biomass back into the dissolved organic matter pool (Bratbak *et al*., 1994; Middelboe and Lyck, 2002). Phages can also accelerate the export of carbon and other organic compounds below the photic zone by increasing the sinking rates of infected cells (Buesseler *et al.*, 2007). Additionally, they can increase genetic diversity through viral-mediated gene transfer (Breitbart *et al*., 2007) and are able to manipulate bacterial metabolism, altering host gene expression or the expression of phage-encoded auxiliary metabolic genes (Breitbart, 2012). Despite their importance, virus ecology remains poorly understood due to the lack of appropriate methods for their study. In the last 20 years, the attention to marine phages has notably increased. Consequently, diverse methods have recently been developed that include approaches to assess single cell host-phage interactions (Dang and Sullivan, 2014), for example, phageFISH (Allers *et al*., 2013).

PhageFISH is a method based on fluorescence *in situ* hybridization that allows the simultaneous identification of phage and host with high sensitivity, infection tracking and detection of replicated, encapsidated and silent phages (Allers *et al*., 2013). For the host identification it uses rRNA CARD-FISH (Pernthaler *et al*., 2002). For phage detection, one viral gene is targeted with digoxigenin-labeled polynucleotide probes and the signal is amplified with a combined antibody-CARD system (Moraru *et al*., 2010). Potentially, phageFISH could be used for the *in situ* identification of phages in environmental samples. To the date PhageFISH studies have only documented phage-host dynamics of model systems under controlled infections in pure cultures (Allers *et al*., 2013; Holmfeldt *et al*., 2014).

One of the model systems in which phageFISH has been applied is PSA-HP1 – *Pseudoalteromonas* (Allers *et al*., 2013). PSA-HP1 phage is a marine podovirus isolated from Helgoland waters in the early 1990s and it infects *Pseudoalteromonas* strain H100 with high specificity (Wichels *et al*., 1998). *Pseudoalteromonas* are fast growing heterotrophic Gammaproteobacteria, frequently isolated from North Sea samples despite its low representation in the bacterioplankton community of only a few percent during spring and summer (Eilers *et al*., 2000).

Our aim was to detect a PSA-HP1 infection in natural seawater samples from Helgoland by PhageFISH. We were focusing on the PSA-HP1 system because the phage-host system is well characterized (Wichels *et al*., 2002), the isolation site of the phage and host easily accessible and the host a fast growing copiotroph (Deng *et al.*, 2013). We could successfully demonstrate the simultaneous identification of bacterial host and phage infection in natural seawater samples and will discuss strengths and limitations of the phageFISH method.

Methods

Surface seawater was collected in Helgoland (54°10'58.3''N, 7°53'19.9''E) in April 2013. Seawater was equally transferred (1 L each) into autoclaved and acid washed (1 M HCl) bottles. In one bottle, the water remained untreated whereas the other was enriched 1:100 with Zobell's marine media (0.5 % (w/v) peptone, 0.1% (w/v) yeast extract, 0.66 mM $FePO_4$). The bottles were incubated at room temperature in the dark and subsamples were fixed every 12 to 16 h. For fixation, two duplicates of 10 ml were collected and fixed by adding paraformaldehyde (PFA) to a final concentration of 1%, followed by 1 h incubation at room temperature. The fixation was stopped by filtration on polycarbonate filters (diameter, 47 mm; pore size, 0.2 μm; type GTTP; Millipore, Germany, cat. no. GTTP02500) and kept at -20 ºC until analysis. The filters were used for total cell counts (TCC), FISH, phageFISH, and PCR-amplification.

Total cell counts (TCC)

Filter pieces were stained for 5 min with 1 μg ml^{-1} 4', 6-diamidino-2-phenylindole (DAPI) at room temperature. The excess of dye was washed with ultrapure water (MilliQ; Millipore). The counts were determined by epi-fluorescence microcopy. On average ~1000 cells were counted.

Filter PCR and clone libraries

A filter piece was used for analyzing microbial composition of the incubations when they present high cell density (untreated: 90 h; enriched 75 h). The analysis was carried out via 16S rRNA gene amplification, cloning and Sanger sequencing. For initial 16S rRNA gene amplification, a portion of the filters were cut in small pieces and placed in PCR tubes with 100 μl ultrapure water each. The tubes were kept shaking overnight at maximum speed. The water was used as template for the 16S rRNA gene amplification. The gene was amplified with general bacterial primers GM3F and GM4R (Muyzer *et al*., 1995). The PCR reactions mix was composed of 0.5 μM GM3F, 0.5 μM GM4R, 1X PCR buffer, 250 μM dNTPs, 0.02 U μl^{-1} Polymerase (5Prime, Mastermix). The thermo-cycling conditions were an initial denaturation at 94 °C for 5 min followed for 30 cycles with 94 °C 1 min, 48 °C 1 min 72 °C for 1 min with a final elongation at 72 °C for 10 min. The DNA concentration was determined spectrophotometrically with

NanoDrop 1000 (Fisher Thermo Scientific). A 1% agarose electrophoresis was also run to confirm that the PCR product corresponded to the 16S rRNA gene size. The PCR products were purified with gene clean turbo kit (MP Biomedicals) and eluted with 30 µl water. The cloning was done with TOPO-TA cloning kit (Invitrogen) following manufacturer's instructions. The clone libraries were PCR screened with the primers M13 R and M13 F (Messing, 1983) and checked on 1% agarose gel electrophoresis. PCR products were purified with Sephadex columns (Amersham Bioscience) in a Duopore® VDF multiscreen 96-well plate (Millipore). The purified PCR products were used for Sanger sequencing. The sequencing PCR was done with Big Dye Sequencing Terminator Kit (Applied Biosystems) following manufacturer's instructions. The sequencing PCR products were purified on Sephadex (Amersham Bioscience) columns and the capillary electrophoresis was run on an ABI 3700 instrument. Nucleotide sequences were cleaned edited with Geneious (Biomatters LTd). The identity of the sequences was verified with BlastN query from NCBI (Johnson et al., 2008) and on Ribosomal Database Project (RDP).

Catalyzed Amplification Reporter Deposition-Fluorescence *in situ* hybridization CARD-FISH

CARD-FISH was used to monitor the changes of bacterial abundance over time. Since PSA-HP1 phage host is one strain of *Pseudoalteromonas* spp, the screening was focused to this genus. Therefore, the hybridizations were carried out with PSA184 probe (CCCCTTTGGTCCGTAGAC) (Eilers *et al.*, 2000). The samples were also hybridized with EUB338 I-III (Amann *et al*., 1990; Daims *et al*., 1999) and NON338 (Wallner *et al*., 1993) probes as positive and negative controls. All hybridizations were done as previously described (Pernthaler *et al*., 2002). The filters were incubated 1 h in 0.5 mg $m1^{-1}$ lysozyme at 4 °C for permeabilization. The hybridization was carried out for 3 h at 46 °C using the published formamide concentration assuring stringent hybridization (35% for EUB I-III and 30% for NON338 and PSA184). For CARD, the samples were incubated for 45 min at 37 °C in a buffer containing 1 µg $µl^{-1}$ Alexa488 tyramide and 0.0015% H_2O_2. Cells were counterstained with 1 µg ml^{-1} DAPI and mounted with a mix 3:1 Citifluor:Vectashield.

PhageFISH

PhageFISH was followed according to Allers and coworkers (2013) with the following modifications. The gene hybridization buffer had double the concentration of dextran sulfate (final concentration: 20% (w/v)) and the gene probes were hybridized for 2 h. The host cell were identified with PSA184 probe. PSA-HP1 phage was detected and identified with a mix of 4 dig-labeled polynucleotides probes (300 bp each) targeting consecutive regions of the phage *unk*-gene (Allers *et al*., 2013). For negative control, the samples were hybridized with Nonpoly350 probe (Moraru *et al*., 2010).

Microscopy

For TCC, FISH and phageFISH the filters were visualized using epi-fluorescence microscope Axioscop II motplus coupled with a Axio-MR m CCD camera (Carl Zeiss) with the following fluorescence filters: DAPI (365/10 nm excitation, 420 LP emission, FT 395 Beam Splitter), Alexa488 (472/30 excitation, 520/35 emission, 495 Beam Splitter) and Alexa594 (462/40 excitation, 624/40 emission, 593 Beam Splitter).

Results and discussion

Our aim was to visualize an infection of PSA-HP1 phage in "environmental-like" conditions *in situ*. Since the relative abundance of *Pseudoalteromonas* (PSA-HP1 host) in the environment was low, it was unlikely that we would have picked up a virus infection by microscopy of a few milliliter of seawater. Hence we incubated seawater from Helgoland Roads (the phage isolation site) to promote the growth of the host. Additionally, to increase the probability of phage detection, we prepared two incubations: one consisting of untreated seawater and another containing seawater enriched with Zobell media (1:100). The reasons to target PSA-HP1 instead of other phages were that (i) the phage-host interactions have been well characterized on pure cultures (Wichels *et al*., 2002) as well as visualized with phageFISH (Allers *et al*., 2013), (ii) the accessibility to the phage isolation site is easy, and (iii) the host is a fast growing copiotroph (Deng *et al.*, 2012).

The microbial growth was monitored every 12 to 16 h. In the untreated seawater incubation, the cell abundance slightly fluctuated during the first 27 h. The cell density reached the highest numbers at 75 h (2.0×10^6 cell ml^{-1}) and the death phase began at 89 h (Figure 1.a). In Zobell supplemented seawater, the population increased more than two orders of magnitude within the first 51 h. The highest density was observed at 113 h (9×10^7 cells ml^{-1}) after which the population started to decrease. The decrease became more pronounced from 122 to 138 h (Figure 1.b).

The 16S rRNA gene diversity was determined by clone libraries, which showed a dominance of *Pseudoalteromonas* especially in the enriched seawater incubation (Figure 1). Next, the abundance of *Pseudoalteromonas* spp. of all the sampling points was assessed with CARD-FISH using a genus-specific rRNA probe, PSA184. The abundance of *Pseudoalteromonas* spp, first increased, reached a plateau and entered a phase of cell decrease. Our FISH counts showed that *Pseudoalteromonas* was one of the dominant genera in our incubations (Figure 1) and accounted for up to 89% of the microbial population in the media-enriched incubation and about 40% in the natural seawater incubation. The cell abundance curves allowed the identification of the time points with a high probability for the detection of the PSA-HP1 phage. These were the points in which the incubations presented high *Pseudoalteromonas* spp density followed by a sharp decrease. This sharp decrease may be explained by the fact that in dense

microbial populations, cell lysis mediated by phages is most likely to happen (Suttle, 2007). In our enrichments time points where virus – mediated mortality could have happened were from 65 to 122 h in the untreated seawater and from 113 to 138 h in the Zobell supplemented seawater incubations (Figure 1).

Next, we used PhageFISH for the *in situ* detection of a PSA-HP1 phage infection. The bacterial host was identified with the PSA184 probe and the phage was identified with a mix of 4 polynucleotide probes (300 bp each) targeting a marker gene of unknown function (*unk*) of the PSA-HP1 virus (Allers *et al*., 2013). We were able to visualize PSA-HP1 infections in both of the incubations (Figure 1). In the untreated seawater, the phage was observed on ~2% of *Pseudoalteromonas* cells at time point 75 h of the incubation. In the enriched seawater, we were able to detect the phage in ~3% at 113 h and ~1% at 138 h of the *Pseudoalteromonas* cells present. Interestingly, the infection was always observed on neighboring cells. We noticed variations both on the phage signal size and in the cell morphology as shown in Figure 2. This may indicate different stages of phage infection (Allers *et al.*, 2013). It was reported for the PSA-HP1 phage that in pure cultures the full infection cycle takes only 1-1.5 h (Allers *et al*., 2013). Since our sampling intervals were 12-16 h we were lucky to catch the infection in both incubations. Taking into account the fast infection cycle the second infection observed on the enriched seawater (Figure 1.b) could be a second infection wave due to higher strain abundance as consequence of the enrichment. If sampling would have been denser (e.g. every 20 min), we might have observed the infection on more time points. However, due to the limited volume of the incubations and lack of knowledge of the microbial dynamics in our incubations, we decided to keep the interval longer to increase our chances to detect the phage at least in one point.

Based on viral tagging experiments it was suggested that the adsorption of PSA-HP1 in the ocean always leads to an infection (Deng *et al*., 2013). In the context of our experiment, this means that all the cells that presented phage signal were definitely infected. Moreover, we are confident that the size and differences on the phage signal revealed a lytic infection (Allers *et al*., 2013; Dang and Sullivan, 2014). Therefore we estimated the percentage of mortality mediated by PSA-HP1 (See Table 1). We approximated that PSA-HP1 was responsible of at least 7.6% and 6.8% of *Pseudoalteromonas* lysis from 75 to 89 h in the untreated seawater and from 113 to 122 h in the enriched seawater, respectively. It is unclear if these percentages represent all

cells of *Pseudoalteromonas* strain H100 because the rRNA probe allows the identification only down to genus level. Considering the total microbial population (TCC), PSA-HP1 infection explained the decrease to 1% of the population from 75 to 89 h incubation in the untreated water, and 3% from 113 to 122 h in the enriched incubation. However, our approximation might be underestimated due to a lower sensitivity of detection of newly infected cells (Allers *et al*., 2013).

Although the abundance was high, the percentage of infected *Pseudoalteromonas* was low. This was somewhat expected for the following reasons: i) - the *Pseudoalteromonas* strain needs to be the specific PSA-HP1 host strain, *Pseudoalteromonas* sp. H100, otherwise no infection will take place given the high specificity of the virus (Wichels *et al.*, 2002; Deng *et al.*, 2013); and ii) – the overall likelihood to be infected is low. Our study shows that with a limited knowledge of the host, it is possible to tackle viruses from natural seawater. Even more interesting, the virus was first isolated in the 1990's (Wichels *et al*., 1998). Twenty years later we were able to detect, within the identification range of our probe, the very same virus in the same environment, which gives a strong hint of recurrence of this phage in Helgoland waters. Since the phage is strain specific this is also direct evidence of the host recurrence. This is consistent with the annual reappearance other main taxa found in Helgoland Roads (Lucas *et al*., 2015; Teeling et al., submitted).

Figures and Tables

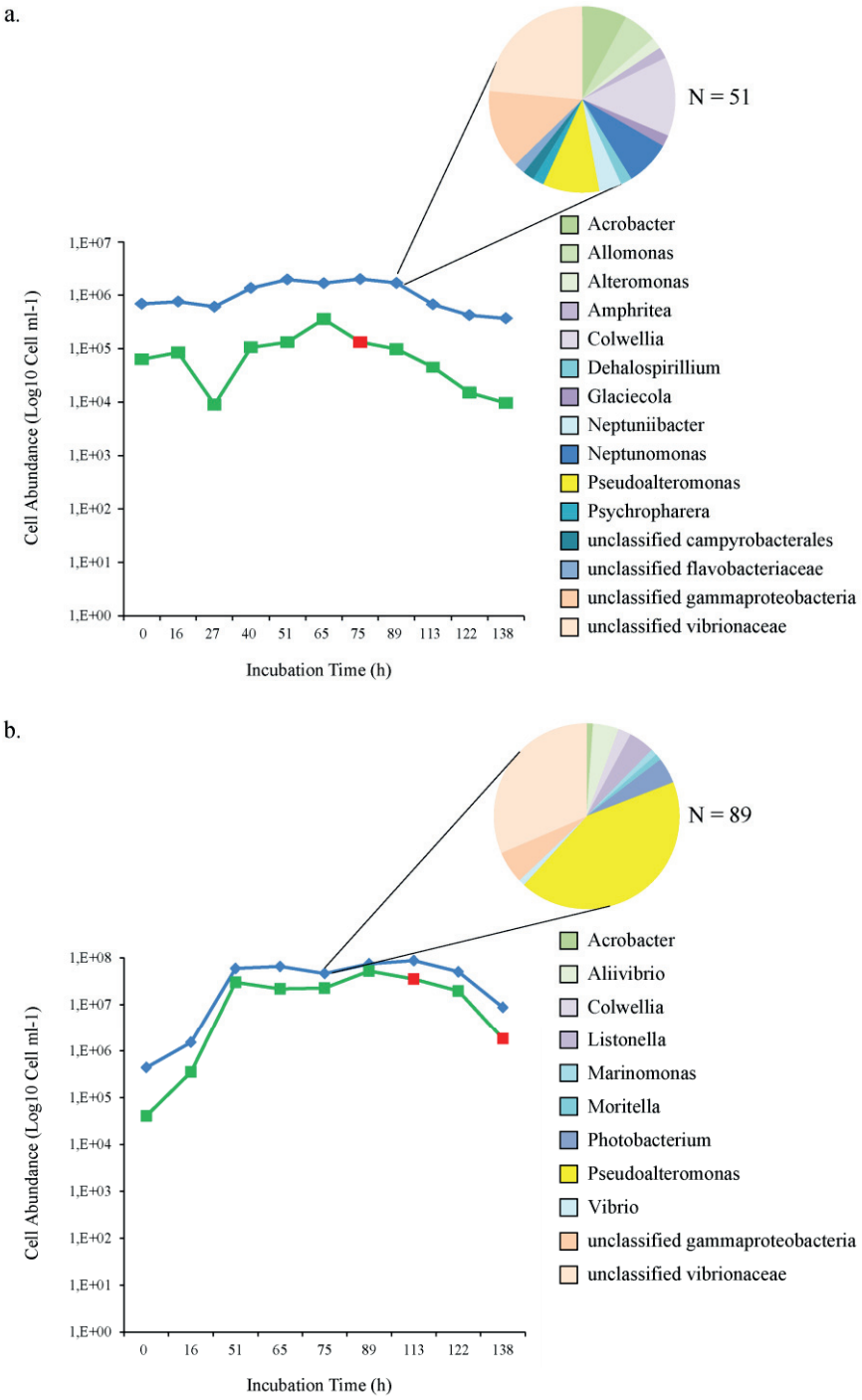

Figure 1. Cell abundance monitored in incubations with seawater recollected in Helgoland-April 2013. Two incubations were prepared: a) corresponds to untreated sea water incubated for 138 h at room temperature and b) to seawater enriched 1:100 with Zobell media. Bacterial abundances are presented in logarithmic scale. Blue line: Total cell abundance after DAPI staining. Gren line: *Pseudoalteromonas* spp. abundance determined by CARD-FISH. Red dots show the points were PSA-PH1 was observed. Pie chart showing the 16S rRNA gene diversity from the clones libraries prepared at the point time indicated in the graph, in yellow *Pseudoalteromonas* spp.

Figure 2. PSA-HP1 infection with PhageFISH from untreated seawater at 75 h incubation**.** In green *Pseudoalteromonas* sp. and red PSA-HP. Scale bar 2 µm arrows showing cells at different infection stage. The SR-SIM images were generated using ELYRA PS.I microscope (Carl Zeiss Germany) from Z-stacks of 0.87 µm thin containing 5 phase shifts and 3 rotations using a 63X Plan-Apochromatic oil immersion objective. The pictures were taken with a PCO edge sCMOS camera. For the excitation and emission of the dyes we used the following settings: 488 nm laser (100 mW) and beam splitters 495-575 for $Alexa_{488}$; 561 nm laser (100 mW) and beam splitters BP 570-650 for Alexa594; 405 nm laser (50 mW) and beam splitter 420-480 for DAPI. The images were reconstructed using ZEN software (black edition, 2011, ZEISS) based on the structured illumination algorithm

Table 2. *Pseudoalteromonas* mortality mediated by PSA-PH1 phage from time point 75 to 89 h in untreated incubation and from 113 h to 122 h in enriched incubation.

	PSA-HP1 phage observed				***Pseudoalteromonas* spp. mortality after phage observed**			
	Incubation time (h)	Abundance (cell ml^{-1})	Cells infected (%)	Cells infected (cell ml^{-1})	Incubation time (h)	Abundance (cell ml^{-1})	Cell(ti)-Cell(ti+1)	Mortality mediated by PSA-HP1
Seawater	75	1E+05	2%	3E+03	89	9E+04	4E+04	7,6%
Seawater	113	4E+07	3%	1E+06	122	2E+07	1E+07	6,8%
+Zobell	138	2E+06	1%	2E+04	n.a.			

References

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R., and Sullivan, M.B. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15**: 2306–2318.

Amann, R., 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 Microbio* **56**: 1919–1925.

Bratbak, G., Thingstad, F., and Heldal, M. (1994) Viruses and the microbial loop. *Microb Ecol* **28**: 209–221.

Breitbart, M. (2012) Marine viruses: truth or dare. *Ann Rev Mar Sci* **4**: 425–448.

Breitbart, M., Thompson, L., Suttle, C.A., and Sullivan, M.B. (2007) Exploring the vast diversity of marine Viruses. *Oceanogr* **20**: 135–139.

Buesseler, K.O., Lamborg, C.H., Boyd, P.W., Lam, P.J., Trull, T.W., Bidigare, R.R., Bishop, J., Casciotti. K., Dehairs,F., Elskens, M., Honda, M., Karl, D.M., Siegel, D.A., Silver M.W., Steinberg, D.K., Valdes, J., Van Mooy, B., and Wilson, S. (2007) Revisiting carbon flux through the ocean's twilight zone. *Science* **316**: 567–570.

Daims, H., Brühl, A., Amann, R., Schleifer, K.-H., and Wagner, M. (1999) The Domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434–444.

Dang, V.T. and Sullivan, M.B. (2014) Emerging methods to study bacteriophage infection at the single-cell level. *Front Microbiol* **5**: 724.

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2013) Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *Bio* **4**:e00516-12.

Eilers, H., Pernthaler, J., Glöckner, F.O., and Amann, R. (2000) Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl Environ Microbio* **66**: 3044–3051.

Holmfeldt, K., Howard Varona, C., Solonenko, N., and Sullivan, M.B. (2014) Contrasting genomic patterns and infection strategies of two co-existing Bacteroidetes podovirus genera. *Environ Microbiol* **16**: 2501–2513.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucl. Acids Res* **36**: W5–9.

Lucas, J., Wichels, A., Teeling, H., Chafee, M., Scharfe, M., and Gerdts, G. (2015) Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. *FEMS Microbiol. Ecol* **91**: fiv099.

Messing, J. (1983) New M13 vectors for cloning. In, *Recombinant DNA Part C*, *Method Enzymol* Elsevier: 20–78.

Middelboe, M. and Lyck, P.G. (2002) Regeneration of dissolved organic matter by viral lysis in marine microbial communities. *Aquat. Microb. Ecol* **27**: 187–194.

Moraru, C., Lam, P., Fuchs, B.M., Kuypers, M.M.M., and Amann, R. (2010) GeneFISH - an *in situ* technique for linking gene presence and cell identity in environmental microorganisms. *Environ Microbiol* **12**: 3057–3073.

Muyzer, G., Teske, A., Wirsen, C.O., and Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea

hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**: 165–172.

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

Suttle, C.A. (2007) Marine viruses-major players in the global ecosystem. *Nat Rev Micro* **5**: 801–812.

Teeling, H., Fuchs, B.M., Bennke, C., Krüger, K., Chafee, M., Kappelmann, L., Reintjes, G., Waldmann, J., Quast, C., Glöckner, F.O., Wichels, A., Gerdts, G., Wiltshire, K.H. and Amann, R.I. Recurring patterns in bacterioplankton dynamics during coastal spring blooms. Manuscript in submitted for publication to e-life

Wallner, G., Amann, R., and Beisker, W. (1993) Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cyt* **14**: 136–143.

Wichels, A., Biel, S.S., Gelderblom, H.R., Brinkhoff, T., Muyzer, G., and Schütt, C. (1998) Bacteriophage diversity in the North Sea. *Appl Environ Microbiol* **64**: 4128–4133.

Wichels, A., Gerdts, G., and Schütt, C. (2002) *Pseudoalteromonas* spp. phages, a significant group of marine bacteriophages in the North Sea. *Aquat Microb Ecol* **27**: 233–239.

CHAPTER II:
DEVELOPMENT OF DIRECT-GENEFISH

II.1 - DIRECT-GENEFISH: A QUANTITATIVE PROTOCOL FOR SIMULTANEOUS FLUORESCENCE *IN SITU* HYBRIDIZATION (FISH) OF GENES AND RRNA IN ENVIRONMENTAL MICROORGANISMS AT THE SINGLE CELL LEVEL

JIMENA BARRERO-CANOSA, CRISTINA MORARU, LAURA ZEUGNER,
BERNHARD M. FUCHS, RUDOLF AMANN

Direct-geneFISH: A quantitative protocol for simultaneous fluorescence *in situ* hybridization (FISH) of genes and rRNA in environmental microorganisms at single cell level

Jimena Barrero-Canosa[a], Cristina Moraru[b], Laura Zeugner[a], Bernhard M. Fuchs[a], Rudolf Amann[a]

[a] Department of Molecular Ecology Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D- 28359 Bremen, Germany

[b] Institute for Chemistry and Biology of the Marine environment (ICBM), Carl-von-Ossietzky-Straße 9-11 D-26111 Oldenburg, Germany

Keywords: FISH, single gene detection, SR-SIM

Running title: direct geneFISH - microbial gene visualization at single cell level

Abstract

Linking microbial identity and potential function is an important task of microbial ecology. GeneFISH has been used successfully to link identity and gene presence within environmental samples, at the single cell level. Yet, the CARD signal amplification steps required for the detection of single genes are making the published protocol long and labor intensive and do not allow for the quantification of the gene copy number per cell. In this study we developed direct-geneFISH, which simultaneously detects rRNA and genes using fluorochrome labeled probes at the single cell level. Direct geneFISH improves the original protocol by i) being faster and more simple, ii) enabling quantification of the gene copy number per cell, and iii) allowing the sub-cellular localization of both rRNA and individual genes. Direct-geneFISH was first optimized on an *E.coli* clone with the target gene on a copy-control vector (*E. coli* EPI300 *unk*), which was induced to carry different copy numbers of the target gene per cell. Our new protocol was able to show the different copy numbers in the differently induced *E.coli* clones. Further on, the newly developed protocol was successfully applied for the detection of the sulfate thiolhydrolase (*soxB*) gene in SUP05 cells present in water samples from Rogoznica Lake, Croatia during a rapid anoxic inversion in the autumn of 2011. These developments open the door for routine detection of genes and identity at single cell level, in environmental samples.

1. Introduction

Microorganisms actively shape the ecosystems by catalyzing the cycling of all the biologically important elements (Fuhrman *et al.*, 2015). Therefore, understanding their roles is a fundamental task of microbial ecology. The best way to understand microbial physiology is through culture based techniques. Nevertheless, this is not a trivial task because a big fraction of the ecologically relevant microbes are not readily cultivable (Amann *et al*.,1995; Kaeberlein *et al.*, 2002). The advent of sequencing techniques opened the door to mining the genetic potential of microorganisms, with metagenomics generating a wealth of sequences at an ever increasing pace (Rodriguez-Valera, 2004). However, metagenomic predictions alone are not always sufficient for linking particular genes with their corresponding taxonomic clades. For example, one of the major challenges of metagenomics is to tease apart the genomes of different bacterial species and although different methods have been developed for this purpose (e.g., Albertsen *et al.*, 2013), these assignments can be difficult in complex communities. Furthermore, sequencing techniques rely on nucleic acids extracted out of their *in situ* and cellular context. On the other hand, fluorescence *in situ* hybridization (FISH) targets microbial nucleic acids while preserving both the morphology and the *in situ* spatial organization of microbial cells. Most often FISH targets ribosomal RNA (rRNA) (Amann *et al.*, 1990b), to identify, quantify and infer relationships between microorganisms. When rRNA and gene detection are combined, as e.g. in geneFISH (Moraru *et al.*, 2010), FISH can be used to visually link a particular gene to a specific bacterial group, complementing sequence based approaches.

GeneFISH has been successfully used to link microbial identity with gene presence in diverse samples, including marine sediment enrichments (Lenk *et al.*, 2012), marine bacterial–eukaryotic symbiosis (Petersen *et al.*, 2012; Bernhard *et al.*, 2012), planktonic microbial communities (Moraru *et al.*, 2010) and groundwater samples (Matturro and Rossetti, 2015). The use of geneFISH has been recently extended to the study of phage/host interactions (Allers *et al.*, 2013; Dang *et al.*, 2015), by increasing the gene detection efficiency to nearly 100% (Allers *et al.*, 2013). In geneFISH, the taxonomic identification of the microbial cells is achieved by catalyzed amplification reporter deposition (CARD)-FISH (Pernthaler *et al.*, 2002), using horseradish peroxidases (HRP)-labeled oligonucleotide probes targeting the ribosomal RNA. Subsequently, the gene detection is achieved by using double stranded (ds)DNA

polynucleotide probes multiple labeled with digoxigenin and a combined antibody-CARD signal amplification system (see Moraru *et al*., 2010 for a detailed description). While the CARD steps allow both for intracellular signal fixation, enabling thus combined detection of RNA and genes, and for the necessary sensitivity to detect single copy genes, it nevertheless has disadvantages. First, it does not allow per cell signal quantification, because the tyramide deposition is not linear with the number of targets (Tanke *et al.*, 1994), and second, it makes the protocol long and labor intensive.

There are CARD-free FISH techniques, which use dye-labeled probes to intracellularly localize and quantify low copy number nucleic acids (e.g. genes and mRNAs). While mainly applied on eukaryotic cells (reviewed in Crosetto *et al.*, 2015), these techniques have also been used on bacteria, e.g. to study plasmid distribution during cell division (Niki and Hiraga, 1997), stochastic gene expression (Maamar *et al.*, 2007) and to quantify mRNA (Coleman *et al.*, 2007; Taniguchi *et al.*, 2010; Skinner *et al.*, 2013). At the same time, the development of fluorescence super-resolution microscopy (SRM) platforms, such as structured illumination (SIM), photoactivated localization (PALM), stimulated emission depletion (STED*)* microscopy, is enabling the visualization of sub-cellular structures in microbial cells (reviewed in Coltharp and Xiao, 2012; Gahlmann and Moerner, 2013). SRM has been used in bacteria mainly in combination with autofluorescent proteins, for example to understand sub-cellular protein localization (Biteen *et al.*, 2008; Fu *et al.*, 2010; Eswaramoorthy *et al.*, 2011; Holden *et al.*, 2014), chromosome partitioning (Ptacin *et al.*, 2010; Helgesen *et al.*, 2015), nucleoid organization and transcription-translation coupling (Montero Llopis *et al.*, 2010; Cagliero *et al.*, 2014). The combination of SRM with rRNA-FISH (Moraru and Amann, 2012) allowed sub-cellular localization of ribosomes in *E. coli*, while eliminating the need for genetically engineered proteins. As consequence of this environment, rich in scientific and technological developments, we felt that the time is ready to take geneFISH to the next level. Therefore, we developed direct-geneFISH, which simultaneously detects rRNA and genes using flurochrome labeled probes at single cell level, and improves on the original protocol by allowing i) per cell quantification of the gene copy number, ii) sub-cellular localization of both rRNA and genes and iii) a much shorter and simplified protocol.

Direct-geneFISH was first optimized in a pure culture setting, using the *E.coli* EPI300 *unk* clone, which was induced to have various numbers of the target gene per

cell. Subsequently, the newly developed protocol was compared to the CARD-based geneFISH, for the detection of the sulfate thiolhydrolase (*soxB*) gene in SUP05 cells present in water samples from Rogoznica Lake, Croatia during a rapid anoxic inversion in the autumn of 2011.

2. Methods

2.1 Strains and clones

For the method optimization, *Escherichia coli* clones (Transformax™ EPI300™, Epicentre, Madison, Wisconsin) harboring a gene of unknown function (*unk*) of phage PSA-HP1 was used and ligated into a copy control plasmid, pCC1 (Epincentre, Madison, Wisconsin) (Allers *et al.*, 2013). The clones were induced to have different number of plasmid copies (Wild *et al.*, 2002) as described below.

All bacterial cultures used were grown in media containing 1% tryptone, 0.5% yeast extract, 1% NaCl amended with 12.5 μg ml^{-1} Chloramphenicol, and incubated at 37 °C shaking on a rotary shaker at a speed of 250 rpm. Overnight-cultures of *E.coli* clones were transferred into fresh media containing either 0.2% glucose (for low plasmid copy numbers) or plasmid multicopy induction solution (Epicentre, Madison, Wisconsin). The latter was used at different concentrations (0.1, 0.2, 0.6, 1, 2x stock solution) to gain cultures with different numbers of plasmid copies per cell. After another transfer and overnight incubation with 0.2% glucose or the corresponding concentration of the induction solution, clones were grown until late exponential phase (~300 min; OD ~0.2; Supplementary Figure S1). From the cultures subsamples were taken for total cell counts (TCC), geneFISH and plasmid DNA extraction. For total cell counts, 3 ml culture was fixed in paraformaldehyde (4% final concentration) for 14-18 h at 4°C and filtered onto polycarbonate filters (pore size, 0.2 μm, type GTTP; Millipore, Germany). TCC were determined by epifluorescence microscopy after 4', 6-diamidino-2-phenylindole (DAPI) staining. Filters were stained with 1 μg ml^{-1} DAPI for 5 min at room temperature, washed with ultrapure water (MilliQ, Merck-Millipore, Germany) and embedded with a mix of 3:1 Citifluor:Vectashield *(*Citifluor™, London, UK. Vectorlabs, Burlingame, California, USA). For geneFISH, 5 ml culture was fixed by adding paraformaldehyde to a final concentration of 3% for 14-18 h at 4 °C, concentrated through a Swinnex filter holder (25 mm diameter, Millipore, Germany) equipped with a 0.2 μm pore-size polycarbonate filter, and washed with 50 ml of 1X PBS. After recovery of the cell suspension, 96% ethanol was added in 1:1 ratio and the cells were stored at -20 °C. For plasmid DNA extraction, 5 ml culture was flash-frozen in liquid nitrogen and stored at -80°C. The plasmid DNA was extracted with Quick Plasmid Miniprep Kit (Invitrogen- Life technologies, USA).

The relationship between the number of genome equivalents (calculated from the growth curve, according to Bremer and Dennis, 1996) and the number of plasmids per cell according to Wild *et al*., 2002 was used to estimate a number of 2-6 *unk* genes per cell for the culture with the lowest copy number (no inducer added) (Table 1). For the induced cultures, the number of plasmids per cell was calculated from the ratio between the plasmid DNA (pDNA) concentration (corrected to cell numbers) in the 2-6 copies per cell culture and each of the induced cultures (See Supplementary Data). To confirm the increase of plasmid copies on the induced cultures, the extracted pDNA was fluorometrically quantified with the Qubit® dsDNA HS Assay Kit (Invitrogen- Life Sciences, USA). Additionally, a 2% agarose gel electrophoresis was run to confirm that the quantified DNA was solely plasmidic and contained no genomic DNA (Supplementary Figure S2). To estimate the increase of the number of plasmids of the different cultures, the DNA concentrations obtained with Qubit® were corrected by the cell density (cell ml^{-1}). Details are found in supplementary information (SI).

2.2 Probe design and labeling

To target the *unk* gene, 12 dsDNA polynucleotide probes (300 bp each) were designed and prepared as described by Allers *et al*., 2013. The dsDNA polynucleotides were synthesized by PCR, and purified with QiaQuick PCR purification kit (Qiagen, Germany), and eluted in TE buffer (5 mM Tris, 1 mM EDTA, pH 8.0). Further, the PCR products were chemically labeled with different dyes (Alexa488, Alexa546, Alexa594 and Alexa647), using the ULYSIS® Nucleic Acid Labeling Kit according to the manufacturer's instructions with the following modifications: 5, 10, 15 µl dye (Alexa546, Alexa594, Alexa647) or 1, 2, 3 µl dye per 1 µg DNA (Alexa488), respectively. The labeling reaction time was increased to 30 min. The unbound dyes were removed with Micro Bio-Spin Columns (Bio-Rad, California, USA). Base to dye ratios were calculated according the instructions given with the ULYSIS® Nucleic Acid Labeling Kit.

2.3 Direct- geneFISH protocol

Direct-geneFISH hybridizations were done on fixed cultures of *E.coli* EPI 300 clones carrying the *unk* gene on the pCC1 plasmid (see above). The cultures varied in the number of plasmids per cell ranging from low (2-6) to high plasmid copies (10-30).

An *E.coli* EPI300 clone that was transformed with the pCC1 plasmid but lacked the *unk* gene was used as a negative control.

Cell immobilization and permeabilization.

Volumes between 5 to 40 µl of the fixed cell suspensions were spotted on poly-l-lysine coated slides (Sigma-Aldrich, San Louis, Misuri, USA) and dried at 37 °C. Before spotting the cell suspensions, adhesive silicone isolators (Grace Biolabs, 9 mm diameter 0.5 mm depth) were adhered onto the glass slides in order to define the spotting areas and to avoid the evaporation of the hybridization mix during denaturation. The samples were dehydrated in an ethanol series of 50, 80 and 100% for 2 x 10 and 20 seconds, respectively, and permeabilized with lysozyme (0.5 mg ml^{-1} lysozyme in 1x PBS, 0.05 M EDTA, 0.1 M Tris-HCl pH 7.4), for one hour on ice, followed by 1 min wash in water, 1 min wash in 96% ethanol and air-drying.

Hybridization.

The gene and ribosomal RNA (rRNA) probes were hybridized simultaneously using a hybridization mix consisting of hybridization buffer (35% formamide, 5x SSC, 20% dextran sulfate, 0.1% SDS, 20 mM EDTA, 0.25 mg ml^{-1} sheared salmon sperm DNA, 0.25 mg ml^{-1} yeast RNA and 1% blocking reagent for nucleic acids) and rRNA probe and gene probe. The rRNA was targeted by 50 ng μl^{-1} EUB338 probe (Amann, *et al.*, 1990b) four times labeled with ATTO488 or ATTO637, whereas the gene detection required a mix of 3 to 10 polynucleotide probes targeting the *unk* gene. The gene probes tested were labeled with, Alexa488, Alexa546, Alexa594 and Alexa647. Different concentrations of individual polynucleotide probes were tested (5, 31 and 62 pg μl^{-1}).
To denature both, the target DNA and dsDNA probes, the cells were covered with hybridization mix and incubated for 40 min at 85 °C. After denaturation, the samples were transferred to 46 °C, for 2 h of hybridization. The excess probe was washed as previously described (Pernthaler *et al.*, 2002), in a washing buffer containing 80 mM NaCl, 100 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 10% SDS, for 15 min at 48 °, followed by 20 min wash in 1x PBS, 1 min water, 1 min 96% ethanol and air-drying.

Counterstaining and mounting.

The cells were counterstained with 4 μg ml^{-1} DAPI, washed quickly with MilliQ water and 96% ethanol and then let air-dry. After drying, the samples were mounted with a mix of 3:1 Citifluor Vectashield antifading reagents.

The simultaneous hybridization of rRNA probes with gene probes was tested under different stringencies. Low plasmid copy cells (2-6 plasmid copies) and an *E. coli* clone with pCC1 plasmid without insert as negative control were used. Gene hybridization buffers (composition mention above) were prepared containing increasing formamide concentrations from 0 - 50% in 5%-steps in duplicates. An EUB probe 4 times labeled with ATTO488 served as rRNA probe and for gene detection a mix of 10 *unk*-specific polynucleotide probes (unk-pr1 to unk-pr7 and unk-pr10 to unk-pr12) was used.

2.4 Microscopy

Slides were observed with an ELYRA PS.1 LSM 780 microscope (Carl Zeiss, Jena Germany). The gene detection efficiency was done on photomicrographs taken with an iXON897 EM-CCD camera with a 63x plan-apochromatic oil immersion objective coupled to the ELYRA PS.1 Microscope. The SR-SIM images were generated using ELYRA PS.I microscope (Carl Zeiss Germany) from Z-stacks of 0.6 μm thin containing 5 phase shifts and 3 rotations. The pictures were taken with a PCO edge sCMOS camera. For the excitation and emission of the dyes the following settings were used: 488 nm laser (100 mW) and beam splitters BP495-575 for ATTO488 and Alexa488; 561 nm laser (100 mW) and beam splitters BP570-650 for Alexa594 and Alexa546; 405 nm laser (50 mW) and beam splitter BP420-480 for DAPI; 642 nm laser (150 mW) beam splitters LP655 for Alexa647 and Cy5. The images were reconstructed using ZEN software (Black edition, 2011, Carl Zeiss, Germany) based on the structured illumination algorithm. Further analysis was performed on reconstructed super-resolution images in ZEN software blue edition (2012, Carl Zeiss, Germany).

The efficiency of gene detection by geneFISH was determined on the photomicrographs pictures taken. Per condition three replicates were taken and a total of 6 - 10 pictures per replicate were counted (a total of ≈1000 cells per tested culture). The detection efficiency (cells with gene signal / total cell x 100) obtained from each field of

view was used to calculate the average and variation. The statistical analysis and graphs were done with SigmaPlot 12.0.

The plasmids per individual cells were visualized and counted from reconstructed SR-SIM images. On average 200 cells per tested culture. Only cells with a strong rRNA FISH signal. When it was possible to visualize the gene copies as discrete dots, these were manually counted. The average and variation of number of gene copies per cell on each tested culture condition was calculated with SigmaPlot 12.0.

Intensity measurements of gene signal from cultures with different plasmid copy number

To measure the gene signal intensity, all the pictures were taken with iXON897 EM-CCD camera coupled to the ELYRA PS1 microscope. For all the fixed cultures, the laser intensity, gain and exposure time were maintained constant. A total of 12 pictures were taken from two duplicates (per experiment). The pictures were exported as png images. To measure gene signal intensities, we used two software tools, Ilastik v.1.1.6 (http://ilastik.org/, Sommer *et al.*, 2011) and CellProfiler v.2.1.1. (http://www.cellprofiler.org/, Kamentsky *et al.*, 2011). Ilastik was used to define cell shape based on the 16S rRNA signal and the results were imported in CellProfiler, for further processing. The previously defined cell shapes were overlapped with the gene signal images, in order to measure gene signal intensity per cell. Initially, two background corrections were performed. The first correction removed the fluorescence introduced by the microscope itself: for every picture a cell free region was defined, its mean intensity/pixel was calculated and then subtracted from each pixel in the image. The second correction was needed to remove the autofluorescence introduced in the fixation step. For this, the mean intensity/pixel of the cell occupied area was calculated, and this valued averaged in between several fields of view. Then, this value was used to correct the gene intensity of each cell in the *unk* cultures, by the following formula: total corrected intensity = total intensity per cell – (average intensity /pixel of the negative control * cell area in pixels). After applying the background corrections, the signal intensity for one plasmid was measured by marking individual plasmid signals and averaging their total intensities. To calculate the plasmid number per cell, for each cell the following formula was applied: total intensity per cell / intensity of one plasmid.

2.5 Application to environmental samples

Water samples were collected from 13 m depth of the Rogoznica Lake, Croatia (43°32′N, 15°58′E), in October 18, 2011. The samples were fixed with formaldehyde (2% final concentration) for approximately 12 h at 4 °C. Fixation was stopped by filtration of 100 ml of the sample onto polycarbonate filters (pore size, 0.2 μm, type GTTP; Millipore, Germany) and stored at -20 °C until analysis.

2.5.1 Probe design and synthesis for targeting soxB gene belonging to SUP05 clade

To target the sulfur oxidation gene marker *soxB* of the SUP05 clade, a single 721 bp gene probe was designed (Sequences are given in *Experimental procedures* in SI). The gene probe was designed based on a *soxB* clone library and alignment from Pjevac *et al*., 2015, using the PolyPro software (threshold 1 = 5% mismatches, th 2= 60% mismatches) (Moraru *et al*., 2011). A selected clone (*E.coli* TOP10-soxbC1) from the soxB library described in (Pjevac *et al.*, 2014) was used as template for the probe synthesis The clone was harvested from a culture grown overnight and the pDNA was extracted with the Quick Plasmid Miniprep Kit (Invitrogen- Life technologies, California, USA).

The probes were synthesized by PCR reaction and labeled with Alexa594 (direct geneFISH) or digoxigenin (CARD-geneFISH) (see Moraru *et al*., 2010 for details). The PCR mix contained 2 μM of each nucleotide, 1 μM per primer (SoxB_SUP05RL_F and SoxB_SUP05RL_R, see supplementary data), 1 x Taq buffer (that contributed 1.5 mM Mg^{2+}) and 0.1 U μl^{-1} Master Taq (5Prime). The thermo cycling conditions (see supplementary data) were determined *in silico* and crosschecked by gradient PCR. The PCR products were purified with QIAQuick PCR purification kit (Qiagen, Germany). The Alexa594 – labeled probes were chemically labeled as described for the *unk* probes (see above). For the labeling with digoxigenin (Dig), the probes were synthesized by the incorporation of Dig-labeled-dUTP (Digoxigenin-11-dUTP, Jena Bioscience) into the double stranded DNA during PCR. In this case, the 200 μM nucleotides were replaced by an unlabeled nucleotide mix (200 μM dATP, dCTP, dGTP and 130 μM DTTP, Invitrogen, California, USA), and 70 μM Dig-dUTP in the PCR reaction. For negative control we also synthesized a non-poly350, as described in Moraru *et al*., 2010. Dig incorporation was determined visually by gel electrophoresis (3% agarose) comparing labeled and non-labeled probes. The probe DNA concentration was determined

spectrophotometrically in a NanoDrop (Fisher Thermo Scientific, California, USA). Probes were stored frozen at -20 °C until use.

2.5.2 geneFISH (CARD protocol) vs Direct (protocol)

2.5.2.1 Sample preparation

Direct GeneFISH and CARD geneFISH were compared on environmental samples. The conditions for fixation, immobilization and permeabilization were the same for both protocols. Cells were fixed with 2% PFA, immobilized on polycarbonate membrane filters and subsequently permeabilized with 10 µg ml^{-1} lysozyme for 1 hour on ice. For CARD geneFISH endogenous peroxidases were inactivated with 0.5 M HCl for 10 min.

2.5.2.2 Direct geneFISH protocol

Direct geneFISH and rRNA hybridization were done simultaneously. The oligonucleotide probe GSO-LR183 (Pjevac *et al.*, 2014), four times labeled with ATTO488, was used to target the ribosomal RNA and a 721 bp long polynucleotide probe labeled with Alexa594 to target the *soxB* gene. For the hybridization a hybridization mix was used that consisted of hybridization buffer (30% formamide, 5x SSC, 20% dextran sulfate, 0.1% SDS, 20 mM EDTA, 0.25 mg ml^{-1} sheared salmon sperm DNA, 0.25 mg ml^{-1} yeast RNA and 1% blocking reagent for nucleic acids), 50 ng μl^{-1} of GSO-LR183 16S rRNA probe and 62 pg μl^{-1} of the Alexa594-soxB polynucleotide probe. The cells were covered with the hybridization mix and initially incubated for 40 min at 85 °C for the denaturation of the double stranded DNA. After denaturation, the samples were transferred to 46 °C for 2 h for probe hybridization. The excess of probe was washed for 15 min at 48 °C in a washing buffer containing 112 mM NaCl, 100 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 10% SDS. After washing, the samples were incubated in 1x PBS for 20 min, briefly washed in water and 96% ethanol and then air-dried. Finally the cells were counterstained with 2 µg ml^{-1} DAPI and mounted with a mix of 3:1 Citifluor Vectashield antifading reagent.

2.5.2.3 CARD geneFISH protocol

In the CARD geneFISH protocol gene and rRNA detection was done successively. First gene hybridization with DIG-labeled polynucleotide probes was

carried out followed by the ribosomal RNA detection with a horse-radish labeled oligonucleotide probe. For the gene detection, the samples were initially pre-hybridized using the gene hybridization buffer without any probe for 40 min at 46º C. After pre-hybridization, the samples were passed to the hybridization mix that contained the gene probe and incubated at 85ºC for 40 min to denature the probes and target-DNA. After denaturation, the samples were transferred to 46º C for 2 - 6 h. Excess of probe was washed by sequential washing steps: 2 times for 10 min in washing buffer I at 42º C (WBI) (2x SSC, 0.1% SDS); 2 times for 1 min in washing buffer II (WBII) (0.1% SSC, 0.1% SDS) and finally 1 h in WBII at 48º C. The washing buffer was removed with a quick wash in PBS. The detection of the DIG-labeled polynucleotide probes by HRP-labeled antibodies and CARD amplification with 2 μg ml^{-1} Alexa594 labeled tyramides was done as described in (Moraru *et al*., 2010).

To target the *soxB* gene a digoxigenin labeled probe of 721 bp length with a concentration of 5 pg μl^{-1} was used. A non-sense polynucleotide probe nonpoly350 (Moraru *et al*., 2010) and an *E coli* clone that lack the *soxB* gene served as negative controls. The *E.coli* TOP10-soxbC1 clone was used as positive control (see above) (Pjevac *et al*., 2015).

For the ribosomal RNA identification, the samples were hybridized with the GSO-LR183 probe that targets the SUP05-related clade identified in Rogoznica Lake. The hybridization was done as described in (Pernthaler *et al*., 2002).

3. Results

The direct geneFISH protocol was developed based on geneFISH and phageFISH methods (Moraru *et al*., 2010; Allers *et al*., 2013). The aim was to simplify the procedure by eliminating all the signal amplification steps and by simultaneously hybridizing the rRNA and the target genes.

3.1 direct-geneFISH development - Optimization of gene probe labeling

The optimization of the labeling step was crucial for the success of direct-geneFISH because the microscopic detection of the gene signal is based solely on the fluorochrome labels directly coupled to the polynucleotide probes. Therefore, our efforts were focused on increasing the number of fluorochromes per polynucleotide probe. To target the *unk* gene, 12 polynucleotide dsDNA probes, each 300 bp long, were synthesized (Supplementary Figure S3) (Allers *et al*., 2013). The gene probe synthesis consisted of two steps: first, the dsDNA polynucleotides were synthesized by PCR, and then, the amplicons were chemically labeled with fluorochromes using the Ulysis Nucleic Acid Labeling Kit. We tested different fluorochromes (Alexa488, Alexa546, Alexa594 and Alexa647), as well as different dye volumes in the labeling reactions. The best labeling conditions and efficiencies differed per fluorochrome tested. For Alexa488 the best conditions of labeling were using 6 µl per ng DNA with an average ratio of 2.41 dyes per 100 nucleotides (5-times increased from kit suggestion). For Alexa546 the best condition was 5 µl per ng DNA with approximately 6.97 dyes per 100 nucleotides. For Alexa594, the best labeling were achieved using 15 µl per ng DNA with an average of 3.41 dyes per 100 nucleotides. For Alexa647 the best labeling conditions were using 10 µl per ng DNA with an average 5.34 dyes per 100 nucleotides. For detailed information check SI Table S1.

3.2 direct-geneFISH development - Simultaneous rRNA and gene detection with fluorochrome labeled probes

For the protocol development, the model system used was an *E. coli* EPI300 clone, harboring a phage gene of unknown function (*unk*) in a copy control plasmid (pCC1) (Allers *et al*., 2013). The *unk* gene was targeted by multiple polynucleotide probes, directly labeled with many dyes per probe, as calculated from the labeling ratios

(Figure 1). The rRNA was targeted by a single oligonucleotide EUB338 probe, labeled with 4 Atto488 (in combination with gene probes labeled with Alexa546, Alexa594 or Alexa647) or 4 Cy5 dyes (in combination with gene probes labeled with Alexa488). The main steps of the protocol were sample preparation (cell fixation, immobilization on solid support and permeabilization), simultaneous gene and rRNA hybridization (denaturation, hybridization and washing), embedding and counterstaining and microscopy (Figure 1).

To find optimum hybridization conditions, we tested probes with 4 different fluorochromes (Alexa488, Alexa546, Alexa594 and Alexa647), varying the gene probe concentration and number of gene probes (and thus, the number of dye molecules per hybridized target) per hybridization. The *E. coli* cells with 2-6 plasmids per cell were used as positive control and *E. coli* EPI300 cells with a different insert from the *unk* gene were used as negative control. For the Alexa488 probes, we tested two hybridization mixes containing 9 and 12 unk-probes (with 103 and 132 dyes, respectively, Figure 2), both at a concentration of 31 pg μl^{-1}. The detection efficiencies were of 49 ± 9.4% and 79 ± 12.2% respectively, with 0.8 ± 0.8% and 1.4 ± 1.7% false positives (Figure 2). The gene signals were photo-stable. For the Alexa546 probes, we tested one hybridization mix with 9 unk-probes (352 dyes, Figure 2), at a concentration of 62 pg μl^{-1}, which resulted in a detection efficiency of 74 ± 3.2% and 0.30 ± 0.40% false positives. However, this dye bleached very fast. For the Alexa594 probes, we tested 5 hybridization mixes. The first two contained 2 and 7 unk-probes (with 46 and 124 dyes, respectively, Figure 2), both at a concentration of 31 pg μl^{-1}, and gave detection efficiencies of 34 ± 11.1% and 82 ± 8.5, respectively, with 1.38 ± 0.77 and 0.21 ± 0.53 false positives (Figure 2). The last three contained 10 unk-probes (168 dyes, Figure 2), at concentrations of 5 pg μl^{-1}, 31 pg μl^{-1} and 62 pg μl^{-1}, and gave detection efficiencies of 0%, 91 ± 5.5% and 96 ± 4.0%, respectively, with false positives for the last two hybridizations of 0.86 ± 1.04% and 0.99 ± 0.93% (Figure 2). Moreover, the gene signal was the strongest and the most photo-stable from all 4 dyes. For the Alexa647 probes, we tested two hybridization mixes containing 8 and 11 unk-probes (with 84 and 295 dyes, respectively, Figure 2), both at a concentration of 62 pg μl^{-1}. The detection efficiencies were 64± 13.2% and 75 ± 17.3%, respectively, with 1.69 ± 0.5% and 4.36 ± 4.58% false positives (Figure 2). This dye bleached fast and signals were weak (for detailed data of the hybridizations see Supplementary Table S2).

3.3 Gene melting curve to evaluate the flexibility of simultaneous hybridization of ribosomal oligonucleotide probes and gene polynucleotide probes

Further, we evaluated the possibility of combining the gene probes with rRNA probes requiring different formamide concentrations. To this purpose, we hybridized the *E. coli* unk clone 2-6 copies per cell, with the unk-probes (Hyb7 probe mix in Figure 2) and the EUB338 probe (Atto488) at different formamide concentrations in the range of 0% to 50% (as suitable for EUB338). The gene detection efficiency increased with the formamide concentration, starting from almost no detection in the range of 0% to 10% formamide, increasing to ~30% at 15% formamide and reaching a plateau (> 90% detection) at 30% formamide (Figure 3).

3.4 Quantification of the per cell gene copy number with direct geneFISH

Next, we evaluated the ability of the direct geneFISH to quantify the per cell gene copy number, by hybridizing cultures with different plasmid numbers per cell. To obtain these cultures, we controlled the copy number of the pCC1 plasmid by incubating with different concentrations of a chemical inducer (see Materials and Methods, and SI text). We obtained 4 cultures, which we labeled low copy, intermediate A, intermediate B and high copy. Using plasmid DNA extraction, we estimated the plasmid numbers per cell to be 2-6 (low copy), 3-9 (intermediate A), 4-12 (intermediate B) and 10-31 (high copy) (see materials and methods and SI text). For hybridization we used the *unk* probe mix (Hyb7, see Figure 2) and the EUB338 probe, Atto488 labeled. To resolve single gene signals, we used the SIM super-resolution microscopy technique. The gene signal appearance varied from a few, well defined dots per cell, for the low copy culture, to many, overlapping dots, almost cell wide, for the high copy culture (Figure 4). Therefore, two different gene quantification methods were used – i) by counting the single gene dots per cell, as resolved by SIM and ii) by using the fluorescence signal intensity, as measured in widefield mode.

Using SIM, most of the cells in the low copy culture had between 1 to 7 plasmids per cell, with an average of 3 (Figure 5). In contrast, 95% of the cells from the high copy culture had a cell wide gene signal, and only 5% had plasmid numbers countable by SIM (Figure 5). The intermediate cultures had 40% and 66% cells (intermediate A and intermediate B, respectively) with cell wide signals. The percentage

of cells in which the gene signal was countable was 60% and 34% for 3-9 and 4-12 respectively. The average of the countable fraction was 5.8 and 6.1 for intermediate A and intermediate B, respectively (Figure 5).

To quantify the plasmid numbers using fluorescence signal intensity, we first quantified the total fluorescence intensity of individual plasmids, by using cells with defined gene signals from the low copy culture. Then, we calculated the number of plasmids per cell by dividing the total intensity of each cell to the intensity of one plasmid (see materials and methods and supplementary information for details). The number of plasmids per cell were from 0 to 32, with an average of 4.5, for the low copy culture, from 0 to 48, with an average of 7.5, for the intermediate A culture, from 0 to 55, with an average of 10, for the intermediate B culture, and from 0 to 52, with an average of 15.6, for the high copy culture (Figure 5).

3.5 Application to environmental samples: detecting the soxB gene in SUP05 clade cells from the Rogoznica lake

To evaluate the potential for environmental application, we applied the direct-geneFISH and the CARD geneFISH protocols to water samples from the Rogoznica Lake. A single dsDNA probe of 721 bps was designed to specifically target the *soxB* allele cluster which was hypothesized in an earlier study that it belongs to the SUP05 clade (Pjevac *et al.*, 2014). To ensure specificity, the probe had a maximum of 5% mismatches with the targeted allele cluster, and at least 40% mismatches with the other *soxB* alleles retrieved in Rogoznika lake from other bacterial species (Supplementary Figure S4). The rRNA probe used for cell identification was GSO-LR183, which is specific for the SUP05 clade found in Rogoznica Lake. The gene probe was labeled either with Dig or with Alexa594 (46 dyes per ds probe).

The soxB signals co-localized with the GSO-LR183 signals. The percentage of SUP05 cells showing *soxB* probe signals was 24.2 ± 9.6% for CARD-geneFISH and 22.0 ± 10.7% for direct-geneFISH, from which 2.9 ± 3.6% and 1.2 ± 1.4%, respectively, were false positives, as detected by the negative control probe (see materials and methods, and Supplementary Figure S5).

The gene signals produced by CARD-geneFISH occupied almost all the cytoplasm (See Figure 6). For direct geneFISH visualized with SIM, the cells had one

gene dot-like signal that co-localized with the DNA signal, and an rRNA signal surrounding the nucleoid (Figure 7).

4. Discussion

4.1 Removing the CARD steps and overcoming the loss in sensitivity

In geneFISH, the high sensitivity is largely achieved due to the CARD steps, which enable the deposition of many fluorochromes per single 350 bp probe (Moraru *et al*., 2010). Alternative signal amplification approaches have been developed, e.g. hybridization based approaches (e.g. HCR-FISH, branchedFISH, etc.) (Collins *et al.*, 1997; Yamaguchi, *et al.*, 2015), or PCR / MDA based target amplification (e.g., *in situ* PCR, RCA-FISH) (Hoshino *et al.*, 2001; Maruyama *et al.*, 2005; 2006). While increasing sensitivity, signal amplification procedures usually are laborious, increase the chance for false positives, do not linearly amplify the number of targets, and thus prevent per cell signal quantification and most often results in a spread signal hampering sub-cellular localization.

Other approaches use no signal amplification, but rather directly labeled probes (with a variable number of fluorochromes per target, ranging from 2 to 50 in mRNA-FISH, e.g., Montero Llopis *et al.*, 2010; Coleman *et al.*, 2007; Skinner *et al.*, 2013) in combination with very sensitive microscope setups, especially very sensitive cameras. For instance, Coleman *et al*., 2007 suggests that high sensitive CCD camera and long exposure times facilitated the detection of mRNA in 49% of the cells with a mix of oligonucleotide probe (50 Alexa647-molecules) in *E.coli* cells with low to moderate target copies. Wang and collaborators (2006), had 70% efficiency in detecting a 3.1 kps target in maize, suggesting that the only necessary element was a very sensitive camera. Our approach for avoiding signal amplification, while still detecting single copy genes, involved optimization of several parameters: probe labeling, target size, probe concentration, fluorochrome stability and microscopy setup.

Gene probe labeling and target size.

As long as the fluorochrome density is not high enough to result in quenching or in inefficient hybridization, the signal intensity of a fluorescently labeled DNA increases linearly with the number of dyes introduced (Cox *et al.*, 2004; Cox and Singer, 2004). Alexa dyes have an optimal labeling density of 6-8 dyes per 100 bases, while higher densities, e. g., of ~10 dyes per 100 bases, were reported to decrease the intensity of hybridization signals (Cox *et al*., 2004, Cox and Singer, 2004). With a

maximum density of 7.7 dyes to 100 bases (see Supplementary Table S2), none of our probes should be subjected to quenching or inefficient hybridization. Our optimizations of the Ulysis chemical labeling resulted in significant increase in labeling densities generally (see Supplementary Table S1) and values as high as 7.7 and 6.3 were obtained for Alexa546 and Alexa647, respectively. However, for Alexa488 and Alexa594 the maximum labeling density was 2.3 and 4.2, respectively, still far from the reported optimum of 6-8 obtained by nick translation with amino reactive base analogs, followed by chemical coupling of Alexa-succinimidyl esters (Cox and Singer, 2004).

Another way of increasing the number of dyes per target is to increase the size of the probe-target region. For all fluorochromes tested (Alexa488, Alexa594 and Alexa647), increasing the number of probes (and thus, the number of dye molecules per hybridized target) per hybridization mixture, resulted in significantly higher detection efficiencies (Figure 2 and Supplementary Table 3). The increase in detection is due probably both to the increase of the number of dyes and to the number of probes itself. The higher the number of probes is, the higher the chance that the targeted region will be accessible to probe binding. On the other hand, the size of the target region is most often constrained by the size of the gene of interest, e.g. ~ 1 kbps is the average for the bacterial genes (Xu *et al.*, 2006), or, most important for environmental samples, by the size of the fragment which can be retrieved with PCR primers. From this point of view, it is preferable to have shorter target regions. The highest detection efficiency (~96%) was obtained with the Alexa594 dye, for a total of 168 dyes distributed over 10 probes and a target region of 3 kbps (Figure 2 and Supplementary Table S2). Assuming no interference from target accessibility, if all the probes would have a 4.2 labeling density (the highest density obtained for the Alexa594 dye with the Ulysis kit), only 7 probes, i.e. 2.1 kbps region, would be sufficient to give ~100% detection efficiency. Further increase in the labeling density e.g. by incorporation of amino reactive base analogs into PCR products, followed by chemical coupling of Alexa594-succinimidyl esters, would reduce even more the required target size. Previous results with CARD-geneFISH (Allers *et al*., 2013) showed detections efficiencies >90% detection efficiency starting with 4 probes, indicating that the number of as low as 4 probes is sufficient to avoid accessibility issues.

Fluorochrome stability.

Even if the probes with the highest number of fluorochromes were the ones labeled with Alexa546 and Alexa 647 dyes, their gene detection efficiency was not the highest (Figure 2). Both dyes bleached very fast under the microscopic conditions used, a behavior which has been previously reported (Panchuk-Voloshina *et al.*, 1999). The most stable and strong dye was Alexa594, which also gave the highest detection efficiency.

Probe concentration.

A crucial parameter is the probe concentration, with 5 pg μl^{-1} (the reference value for geneFISH and phageFISH) being insufficient for direct-geneFISH. Higher probe concentrations resulted in good detection, with 62 pg μl^{-1} being slightly, but significantly better than 31 pg μl^{-1} ($p = 0.014$), and allowing almost 100% detection efficiency (Figure 2).

Microscopy setup.

Using a highly sensitive EMCCD camera (see materials and method) we visualized as little as 50 Alexa594 dyes per target, although with lower detection efficiency (34% on *E.coli* clones and 22% for GSO-SUP05 cells).

Finally, the sensitivity lost by removing the CARD steps was overcome by using a mix of dsDNA gene polynucleotide probes each one carrying multiple Alexa594 dyes (170 dyes on a 3 kbps target region) and by using more sensitive microscopes.

4.2 Gene melting curve to evaluate the flexibility of simultaneous hybridization of ribosomal oligonucleotide probes and gene polynucleotide probes

During direct-geneFISH the gene and the rRNA are simultaneously hybridized. Since both probes are added in the same hybridization reaction, it is important that their requirements for stringency (i.e. formamide and salt concentrations in hybridization and washing buffers, hybridization temperature) are compatible. Generally, polynucleotide probes have higher melting points (Tm), a broad hybridization optimum at 20-30°C

below Tm, which translates in a broader optimum formamide range, and hybridize to similar, but not necessarily identical targets (max ~ 10-15% mismatches) (Wetmur, 1991; Moraru *et al.*, 2011). On the other hand, oligonucleotide probes have a lower Tm, and because most of the times they are designed to discriminate against one mismatch targets, their optimum formamide range is narrower. In the specific case of simultaneous FISH with Eub338 and *unk* probes, the stringency conditions are compatible (Eub338 – 0-50% Amann *et al*., 1990a, *unk* – 35%, - Allers *et al*., 2013). Therefore, we hybridized at 35% formamide. On the other hand, the washing conditions (20 mM Na^+) for the gene probe were too stringent for the rRNA probe (80 mM Na^+), and to get both the gene and the rRNA signal, we washed as corresponding for the rRNA probe.

To test the potential usability of the *unk* gene probe with other rRNA probes, we performed direct-geneFISH with increasing formamide concentrations, in the range 0% to 50%, always washing with the corresponding salt concentrations, resulting in a Na^+ range from 900 M to 28 mM (see Supplementary Table S5). The *unk* probe displayed maximum detection efficiency over a broad formamide range (30% - 50%), with the detection progressively decreasing with the decrease in formamide concentration (therefore distance from the T_m), in agreement with the theory (Figure 3). Furthermore, increasing the salt concentration in the washing buffer did not result in higher percentage of false positives, as could be expected for such a long probe. The need of multiple polynucleotides to achieve a visible signal is certainly favoring the low percentage of false positives, because it is unlikely that more unspecific probes will bind in close vicinity, and the signal from just one probe binding unspecific is too low to be detected.

Therefore, it is essential to keep the stringency conditions established for rRNA probes, combining the rRNA probe with gene probes which have optimum detection efficiency at the formamide concentration required by the rRNA. While the broad range of optimum formamide concentrations of polynucleotide probes allows a great flexibility, there will be cases where the rRNA and the gene probes do not match. In this case it is recommended to use either the CARD-geneFISH, or a subsequent hybridization of the rRNA and of the gene, in which the signal from the first step is fixed in the cell by CARD.

4.3 Per cell gene copy number quantification with direct geneFISH

Compared with the bulk plasmid DNA quantification, the direct-geneFISH offers a single cell level view, showing that, while the average plasmid content agrees with the bulk results, individual cells have a wide range of plasmid numbers. For example, in the low copy culture most of the cells have between 2 and 5 plasmids, however, a few cells have as high as 32 plasmids. In the same time, in the high copy culture most of the cells have between 9 and 18 plasmids, but a few cells have no plasmids (Figure 5).

The two approaches used for plasmid quantification gave comparable, and, in the same time, complementary results. While the increase in resolution gained from SIM allowed counting of individual gene signals in the cells with low plasmid numbers, the intensity method allowed quantification in cells with medium to high copy numbers. As expected, with increasing copy numbers, the individual plasmids could not be resolved by SIM. Therefore, different applications of the direct-geneFISH will use one or the other method, depending of the target gene copy number. For example, when targeting genes on chromosomes or low copy plasmids, the SIM method can be used. On the other hand, when targeting genes on lytic phages or high copy plasmids, the intensity method should be used.

The use of automatic cell recognition is greatly facilitating the quantification of cells presenting the genes signal and per cell gene signal intensities. There are several software which can be used for this purpose (Sommer *et al.*, 2011). Here we choose a workflow composed from Ilastik and CellProfiler (Sommer *et al*., 2011;Kamentsky *et al*., 2011). Ilastik enables recognition and segmentation of cells by interactive learning, being especially useful when cells are in close proximity to each other. CellProfiler can receive the input from Ilastik and further process the images in a user friendly and flexible environment. To help the direct-geneFISH user with the signal quantification work, we are making publically available the pipeline we created with CellProfiler for the purpose of per cell plasmid quantification in this study.

The double stranded nature of the gene probes can enable, in theory, binding to the corresponding mRNAs. In the case of the *unk* clones, we targeted a gene, which we know is not expressed in the growth conditions. Additionally, the denaturation step during the protocol is enhancing mRNA degradation. Occasionally, in the high copy culture we have noticed gene signals on the outer side of the rRNA signal. To confirm

that these signals were outside the cells, we have performed a triple labeling, with a membrane stain (Nile Red), rRNA probe and gene probe (Supplementary Figure S5). Indeed, some of the plasmids are leaking outside the cell, probably due to their small size and the permeabilization of the cell during the procedure. For quantification, we did not use the cells displaying plasmid leakage. We do not expect this phenomenon to happen with genes located on chromosomes or on other larger molecules. Alternatively, a milder permeabilization might prevent this.

4.4 Application to environmental samples: detecting the *soxB* gene in SUP05 clade cells from the Rogoznica lake

Direct geneFISH enabled the detection of the *soxB* gene in SUP05 cells from the Rogoznica lake, demonstrating the applicability of the method in environmental samples, even though a suboptimal *soxB* probe was used. With only 721 bps and 46 dyes, (the best *unk* probe had 3 kbps and 168 dyes), it detected the *soxB* gene in ~22% of the GSO/SUP05 cells. A similar *unk* probe (Hyb.4 in Figure 2) gave a detection efficiency of ~ 35% in the *E. coli* low copy culture. The difference in detection efficiency could be explained by the larger size of the soxB probe (one 721 bps polynucleotide, compared with 2 polynucleotides, 300 bps each), which would have a slower diffusion inside the cell, or by the lower soxB gene copy number per cell (1 gene copy per cell compared with 2-6 plasmids per cell) combined with the use of a suboptimal probe. Taking this in consideration, we suggest that all the GSO/SUP05 cells in the respective environmental sample had the *soxB* gene, and, therefore, the potential to oxidize thiosulfate. Furthermore, all *soxB* positive cells had one gene signal, as resolved by SIM, indicating that the gene is present in one copy per cell.

While the detection efficiencies of direct and CARD geneFISH were similar, (22% vs 24%), the direct-geneFISH outperformed the CARD protocol in terms of handling time (less than 1 day vs 3 days), cell preservation and resolution of both rRNA and gene signal (compare Figure 6 and Figure 7).

5. Conclusions and outlook

We developed a simplified and improved protocol, direct-geneFISH, to link cell identity and gene presence in less than a day. This, together with the development of functional gene databases, e.g. FunGene from RDP (http://fungene.cme.msu.edu/, Fish *et al.*, 2013), opens the door for its use on a routine basis in environmental samples. Furthermore, the ability to quantify per cell gene copy numbers, the improved sub-cellular localization and the potential for multiplexing (e.g. by simultaneous hybridization of rRNA and multiple genes), makes the protocol an excellent tool to study, for example, phage infection cycles and host-phage dynamics, bacterial and archaeal polyploidy, gene duplication and amplification, variations in plasmid numbers and plasmid partitioning

The choice between the CARD-geneFISH and the direct-geneFISH should be made as follows: i) highly sensitive microscope not available – CARD-geneFISH; ii) not matching formamide concentrations of the rRNA and gene probes – CARD geneFISH, or CARD for rRNA and direct labeling for gene, or viceversa; iii) per cell gene signal quantification – direct geneFISH; iv) sub-cellular localization of the gene signal – direct geneFISH.

6. Figures and Tables

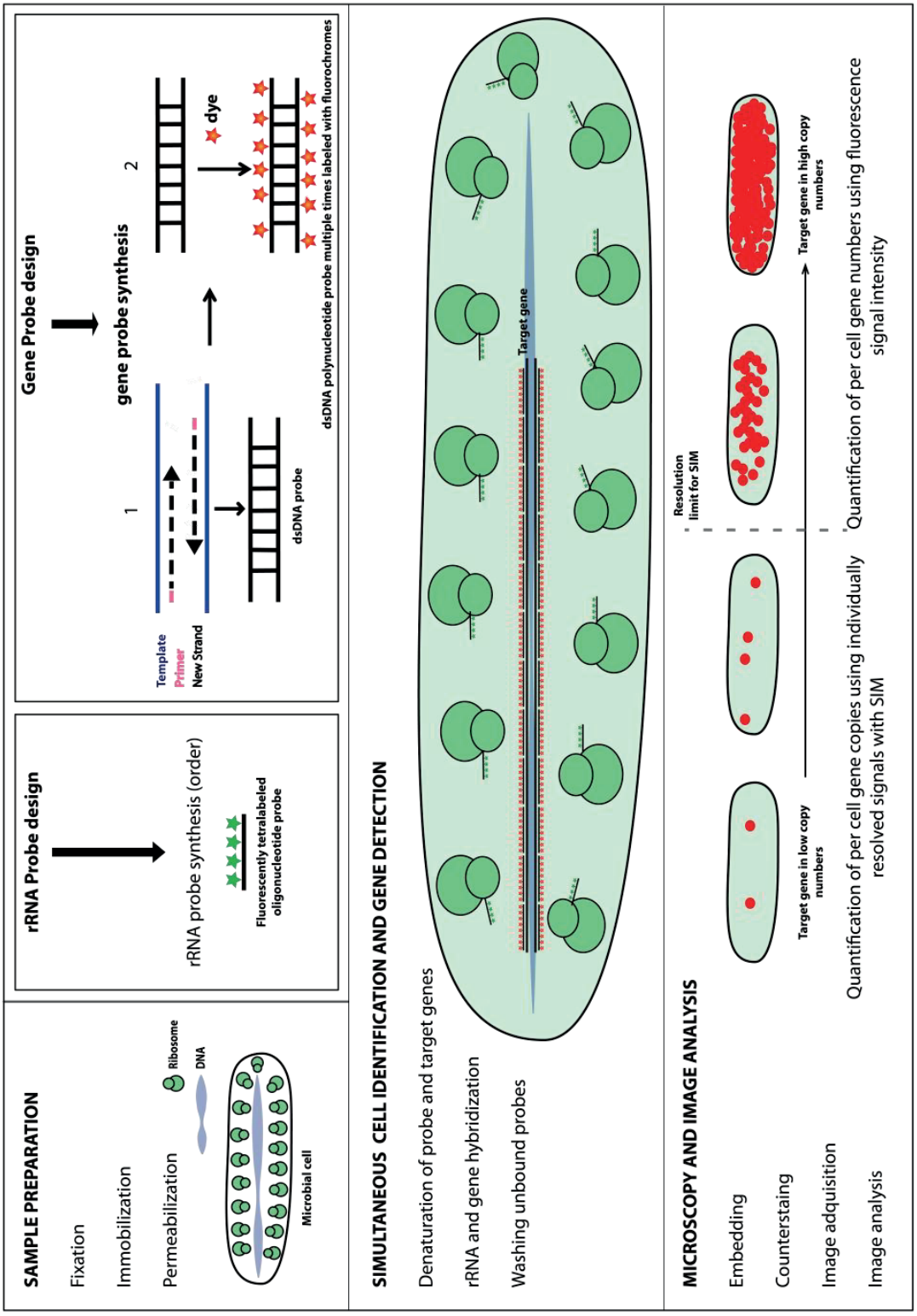

Figure 1. Schematics of the direct-geneFISH protocol, including sample and probe preparation, hybridization and microscopy

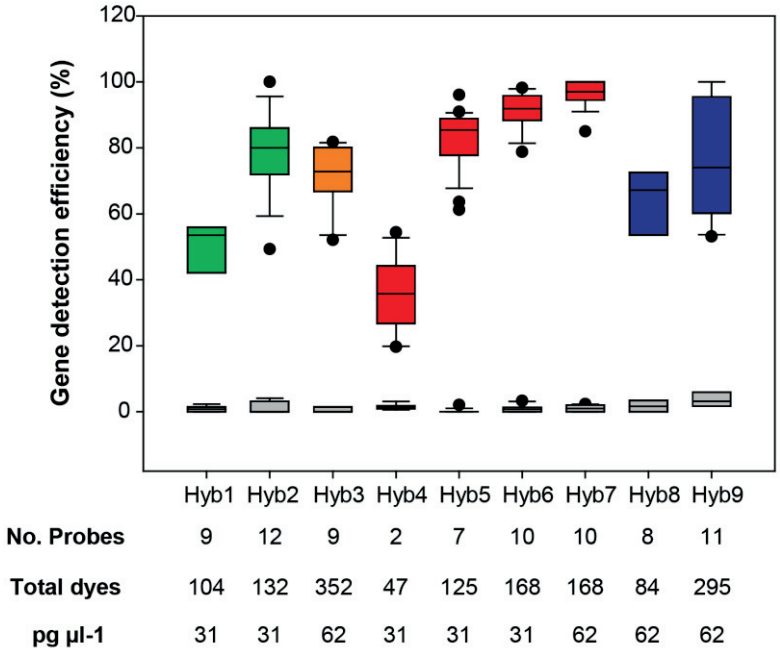

Figure 2. Gene detection efficiency of hybridizations targeting *unk* gene on *E. coli unk* clone, 2-6 target copies per cell (hyb1 to hyb9). Negative control – an *E. coli* EPI300 clone, without *unk* gene. The box plots show the median and the 25 -75% percentiles of the detection. The color of the boxes represents the label of the gene probes used in the hybridization. Boxes: green-Alexa488, orange-Alexa546 red-Alexa594, blue-Alexa647. All negative controls are represented with grey boxes. On x-axis different hybridizations including the tested parameters: number of gene probes in hybridization mix (No. Probes), the number of dye molecules per hybridized gene (total dyes) and the concentration of each probe (pg μl^{1})

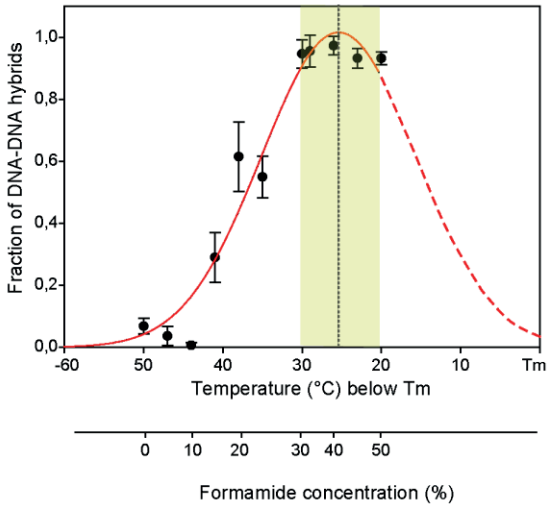

Figure 3. Gene melting curve modified after Moraru et al. 2010. On X- axis the temperature of hybridization, represented in ºC below the melting temperature. On a secondary x-axis the correspondent formamide concentration for *unk* probes at 46º C (0.6ºC%FA). The dots represent the experimental data. In yellow the broad optimum temperature for polynucleotide hybridization. Dash lines show theoretical data that was not measured.

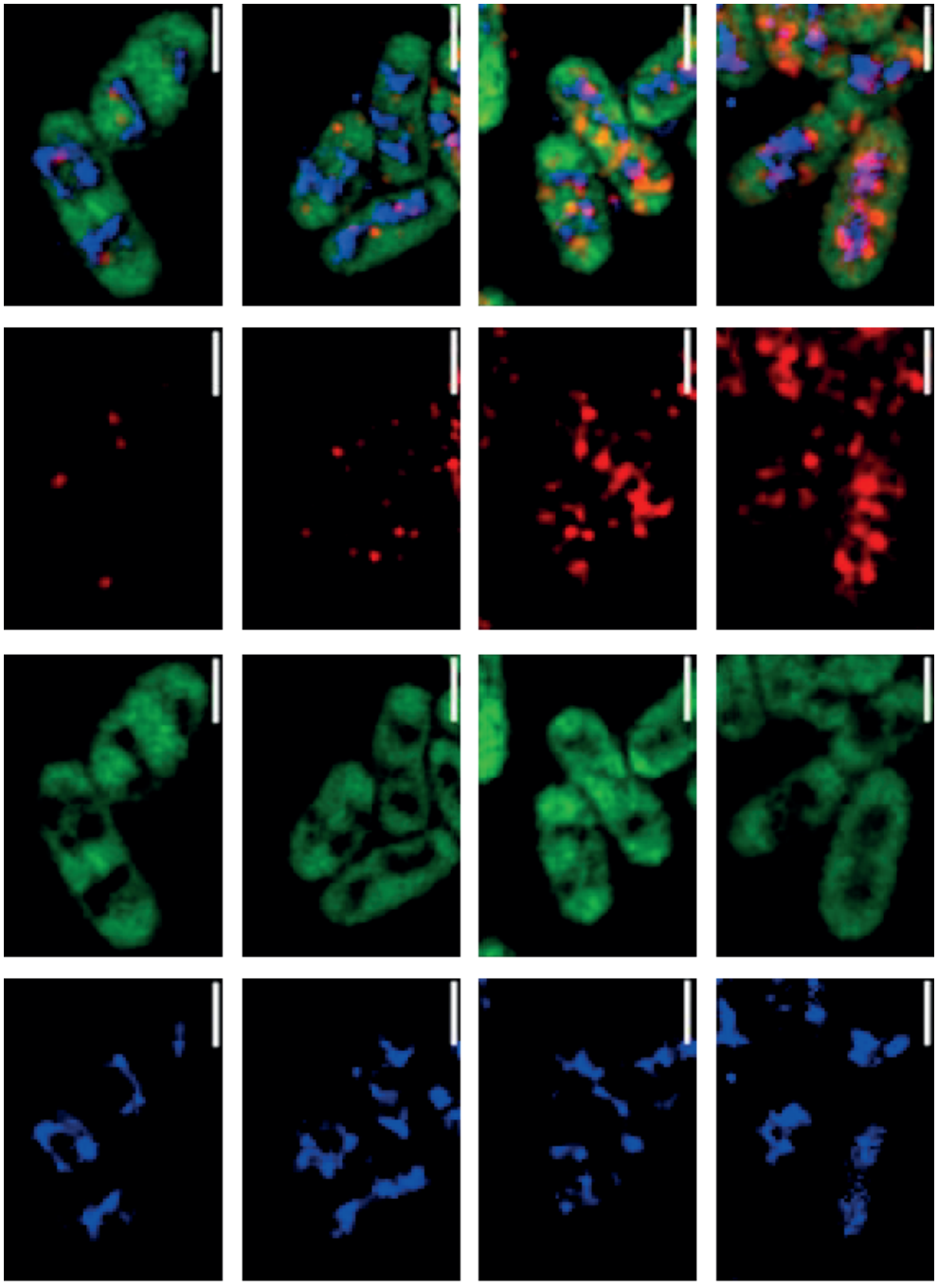

Figure 4. Structured Illumination Microscopy (SIM) images of direct geneFISH on cells induced to have different number of plasmids (gene target) per cell. Columns 1- 3: DAPI (blue), rRNA (green) and gene (red) signals. Fourth column shows merged images. Scale bar: 1µm.

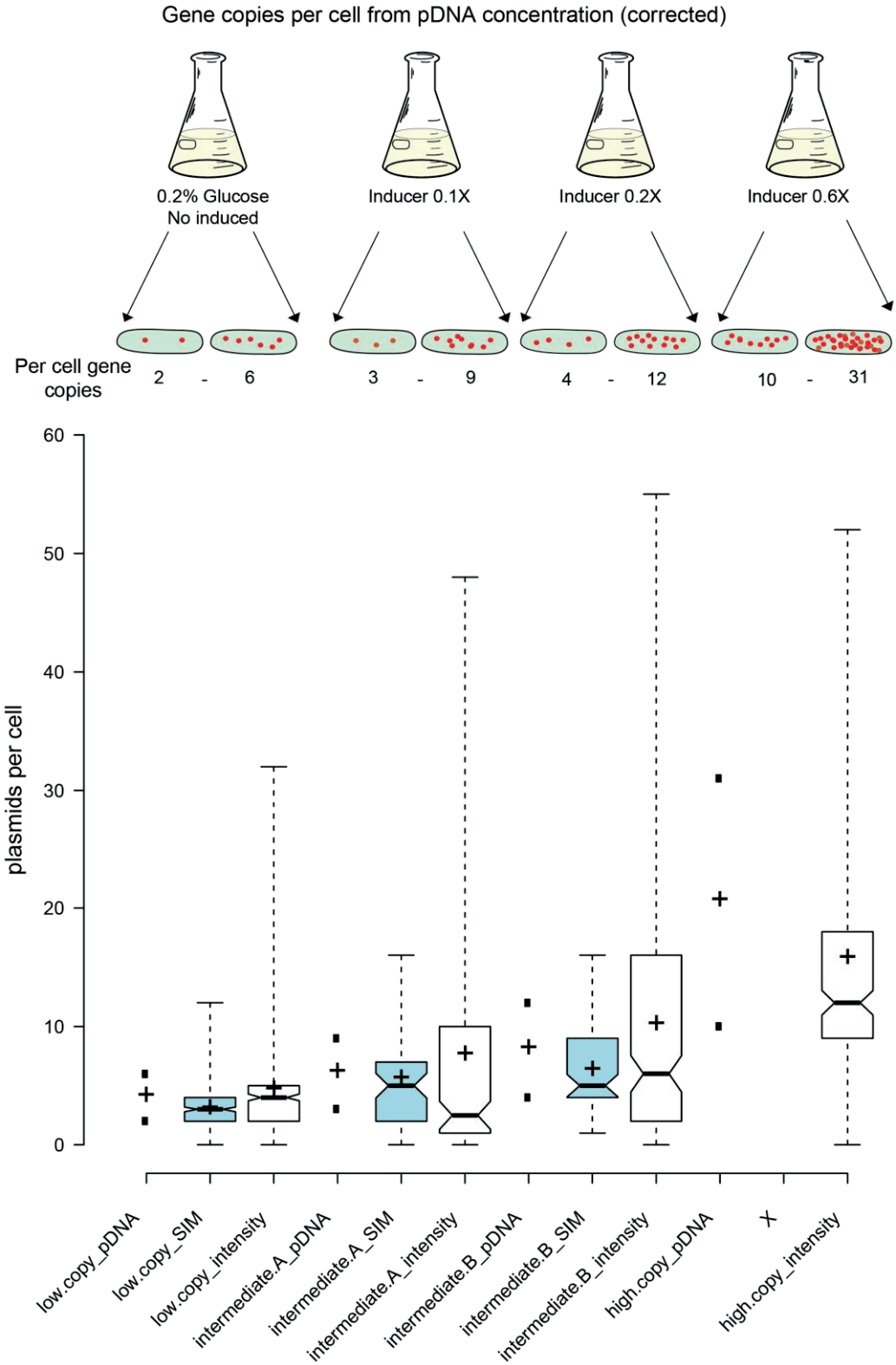

Figure 5. Quantification of the gene signal per cell with different approaches. On the top, the visual representation of the plasmid quantification based on the plasmidic DNA extraction corrected to the number of cells in the culture (See material and methods). For visual comparison these values were also plotted (pDNA) as a cross and lines in the blox plot graph. Blue box plot: gene copies per cell manually counted from SIM reconstructed images. White blox plot: measurements based on mean fluorescense intensity. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend to minimum and maximum values; crosses represent sample means. n = 2, 261, 271, 2, 56, 144, 2, 69, 208, 2, 0, 191 sample points

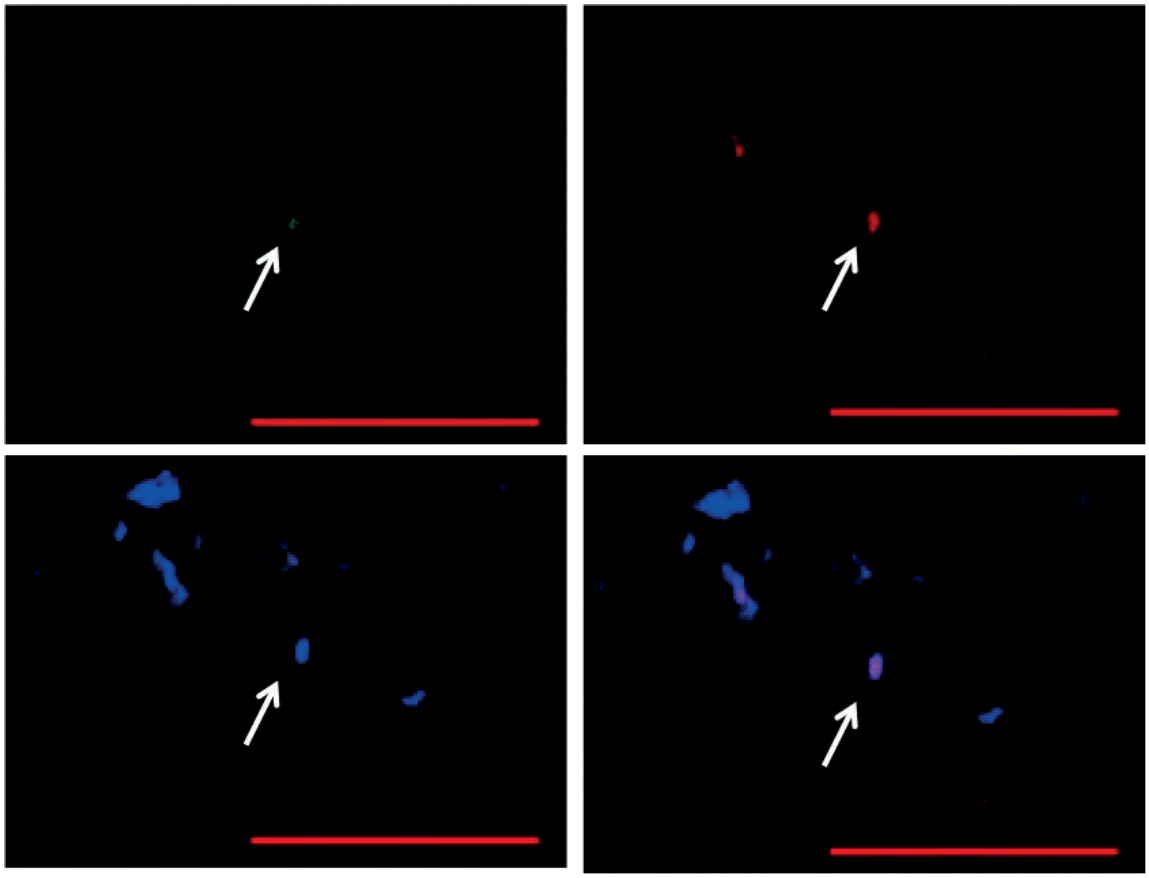

Figure 6. CARD-geneFISH on sample from Dragon Eye Lake. The arrows showing one cell. Green the rRNA-FISH: GSO/SUP05 clade. Red: soxB gene. Blue: DAPI staining. The fourth picture shows the merged image. Scale bar 10 µm.

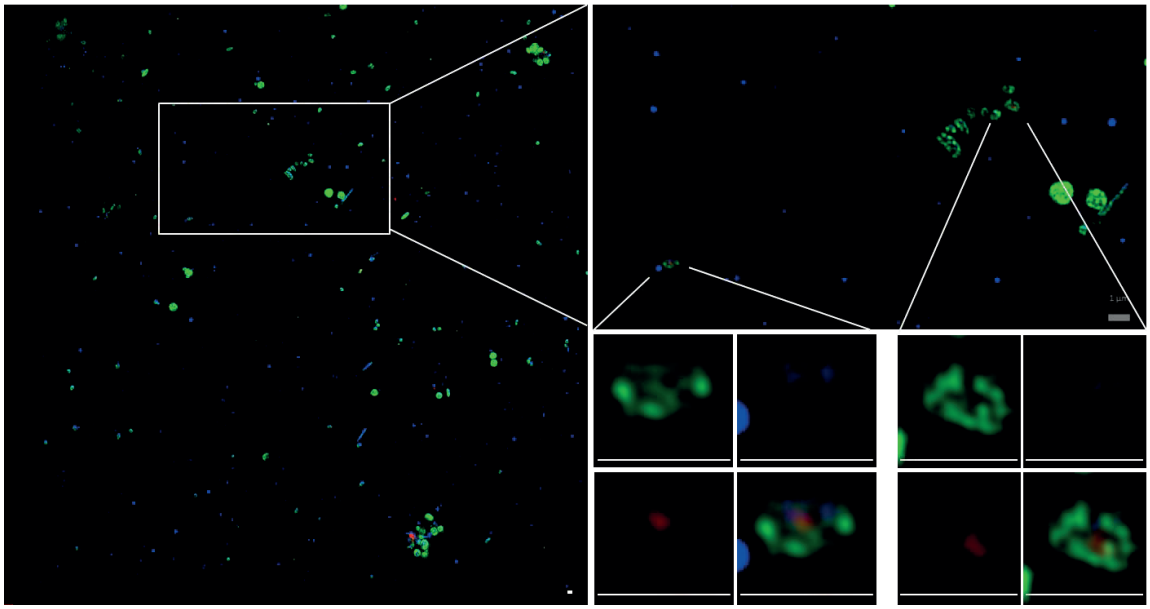

Figure 7. SIM visualization of Direct-geneFISH on samples from Rogoznica Lake, in Croatia. Green: rRNA SUP05/GSO clade. Red: soxB gene Blue: DAPI. Scale bar ~1 µm.

References

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnol* **31**: 533–538.

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R and M.D Sullivan. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15**: 2306–2318.

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

Amann, R., Krumholz, L., and Stahl, D.A. (1990b) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol* **172**: 762–770.

Bernhard, J.M., Edgcomb, V.P., Casciotti, K.L., McIlvin, M.R., and Beaudoin, D.J. (2012) Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. *ISME J* **6**: 951–960.

Biteen, J.S., Thompson, M.A., Tselentis, N.K., Bowman, G.R., Shapiro, L., and Moerner, W.E. (2008) Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable EYFP. *Nat Meth* **5**: 947–949.

Bremer, H. and Dennis, P.P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In Neidhart FC Editor, *Escherichia coli* and *Salmonella*. ASM press; Washington: 1553-1569

Cagliero, C., Zhou, Y.N., and Jin, D.J. (2014) Spatial organization of transcription machinery and its segregation from the replisome in fast-growing bacterial cells. *Nucl. Acids Res.* **42**: 13696–13705.

Coleman, J.R., Culley, D.E., Chrisler, W.B., and Brockman, F.J. (2007) mRNA-targeted fluorescent in situ hybridization (FISH) of Gram-negative bacteria without template amplification or tyramide signal amplification. *J. Microbiol Methods* **71**: 246–255.

Collins, M.L., Irvine, B., Tyner, D., Fine, E., Zayati, C., Chang, C., Horn, T., Ahle, D., Detmer, J., Shen, L., Kolberg , J., Bushnell, S., Urdea, M.S., and Hol, D.D. (1997) A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucl. Acids Res.* **25**: 2979–2984.

Coltharp, C. and Xiao, J. (2012) Superresolution microscopy for microbiology. *Cell Microbiol* **14**: 1808–1818.

Cox, W.G. and Singer, V.L. (2004) Fluorescent DNA hybridization probe preparation using amine modification and reactive dye coupling. *Biotechniques* **1**:114-122

Crosetto, N., Bienko, M., and van Oudenaarden, A. (2015) Spatially resolved transcriptomics and beyond. *Nature Reviews Genetics* **16**: 57–66.

Dang, V.T. and Sullivan, M.B. (2014) Emerging methods to study bacteriophage infection at the single-cell level. *Front Microbiol* **5**: 724.

Dang, V.T., Howard Varona, C., Schwenck, S., and Sullivan, M.B. (2015) Variably lytic infection dynamics of large Bacteroidetes podovirus phi38:1 against two *Cellulophaga baltica* host strains. *Environ Microbiol*. Early view. Online version DOI: 10.1111/1462-2920.13009

Eswaramoorthy, P., Erb, M.L., Gregory, J.A., Silverman, J., Pogliano, K., Pogliano, J., and Ramamurthi, K.S. (2011) Cellular architecture mediates DivIVA ultrastructure and regulates min activity in *Bacillus subtilis*. *mBio* **6**: e00257–11

Fish, J.A., Chai, B., Wang, Q., Sun, Y., Brown, C.T., Tiedje, J.M., and Cole, J.R. (2013) FunGene: the functional gene pipeline and repository. *Front Microbiol* **4**: 291.

Fu, G., Huang, T., Buss, J., Coltharp, C., Hensel, Z., and Xiao, J. (2010) *In vivo* Structure of the *E. coli* FtsZ-ring revealed by photoactivated localization microscopy (PALM). *PLoS ONE* **5**: e12680.

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat Rev Micro* **13**: 133–146.

Gahlmann, A. and Moerner, W.E. (2013) Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nat Rev Micro* **12**: 9–22.

Gregory Cox, W., Beaudet, M.P., Agnew, J.Y., and Ruth, J.L. (2004) Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. *Anal Biochem* **331**: 243–254.

Helgesen, E., Fossum-Raunehaug, S., Sætre, F., Schink, K.O., and Skarstad, K. (2015) Dynamic *Escherichia coli* SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. *Nucl Acids Res* **43**: 2730–2743.

Holden, S.J., Pengo, T., Meibom, K.L., Fernandez, C.F., Collier, J., and Manley, S. (2014) High throughput 3D super-resolution microscopy reveals *Caulobacter crescentus* in vivo Z-ring organization. *PNAS* **111**: 4566–4571.

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 Microbiol* **67**: 5261–5266.

Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.

Kamentsky, L., Jones, T.R., Fraser, A., Bray, M.-A., Logan, D.J., Madden, K.L., Ljosa,V., Rueden, C., Eliceiri, K.W., and Carpenter, A.E. (2011) Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**: 1179–1180.

Lenk, S., Moraru, C., Hahnke, S., Arnds, J., Richter, M., Kube, M., Reinhardt, R., Brinkhoff, T., Harder, J., Amann, R., and Mußmann, M. (2012) Roseobacter clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *ISME J* **6**: 2178–2187.

Maamar, H., Raj, A., and Dubnau, D. (2007) Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317**: 526–529.

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 Microbiol* **72**: 6248–6256.

Matturro, B. and Rossetti, S. (2015) GeneCARD-FISH: detection of *tceA* and *vcrA* reductive dehalogenase genes in *Dehalococcoides mccartyi* by fluorescence *in situ* hybridization. *J of Microbiol Methods* **110**: 27–32.

Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2010) Spatial organization of the flow of genetic

information in bacteria. *Nature* **466**: 77–81.
Moraru, C. and Amann, R. (2012) Crystal ball: fluorescence *in situ* hybridization in the age of super-resolution microscopy. *Syst Appl Microbiol* **35**: 549–552.
Moraru, C., Lam, P., Fuchs, B.M., Kuypers, M.M.M., and Amann, R. (2010) GeneFISH - an in situ technique for linking gene presence and cell identity in environmental microorganisms. *Environ Microbiol* **12**: 3057–3073.
Moraru, C., Moraru, G., Fuchs, B.M., and Amann, R. (2011) Concepts and software for a rational design of polynucleotide probes. *Environ Microbiol Rep* **3**: 69–78.
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.
Panchuk-Voloshina, N., Haugland, R.P., Bishop-Stewart, J., Bhalgat, M.K., Millard, P.J., Mao, F., Leung,W.Y., and Haugland, R.P. (1999) Alexa Dyes, a Series of New Fluorescent Dyes that Yield Exceptionally Bright, Photostable Conjugates. *J Histochem Cytochem* **47**: 1179–1188.
Pernthaler, A., Pernthaler, J., and Amann, R. (2002) Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68**: 3094–3101.
Petersen, J.M., Zielinski, F.U., Pape, T., Seifert, R., Moraru, C., Amann, R., Hourdez, S., Girguis, P.R., Wankel, S.D., Barbe, V., Pelletier, E., Fink, D., Borowski, C., Bach,W., and Dubilier, N (2012) Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176–180.
Pjevac, P., Korlevic, M., Berg, J.S., Bura-Nakic, E., Ciglenecki, I., Amann, R., and Orlic, S. (2014) Community shift from phototrophic to chemotrophic sulfide oxidation following anoxic holomixis in a stratified seawater lake. *Appl Environ Microbiol* **81**: 298–308.
Ptacin, J.L., Lee, S.F., Garner, E.C., Toro, E., Eckart, M., Comolli, L.R., Moerner, W.E., and Shapiro, L. (2010) A spindle-like apparatus guides bacterial chromosome segregation. *Nat Cell Biol* **12**: 791–798.
Rodriguez-Valera, F. (2004) Environmental genomics, the big picture? *FEMS Microbiol Lett* **231**: 153–158.
Skinner, S.O., Sepúlveda, L.A., Xu, H., and Golding, I. (2013) Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat Protoc* **8**: 1100–1113.
Sommer, C., Straehle, C., Köthe, U., and Hamprecht, F.A. (2011) Ilastik: Interactive learning and segmentation toolkit," in *Biomedical Imaging: From Nano to Macro* :230–233.
Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., and Hearn, J. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**: 533–538.
Tanke, H., Vrolijk, J., and Raap, A.K. (1994) Image analysis of FISH stained specimens: A new diagnostic tool in genetics and oncology. *Proc - Annu Meet, MSA* 78-79
Wang, C.J.R., Harper, L., and Cande, W.Z. (2006) High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *Plant Cell* **18**: 529–544.
Wetmur, J.G. (1991) DNA probes: applications of the principles of nucleic acid hybridization. *Crit. Rev. Biochem. Mol. Biol.* **26**: 227–259.
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.

Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z., and Luo, Z.W. (2006) Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Mol Biol Evol* **23**: 1107–1108.

Yamaguchi, T., Kawakami, S., Hatamoto, M., Imachi, H., Takahashi, M., Araki, N., Yamaguchi, T., and Kubota, K. (2015) *In situ* DNA-hybridization chain reaction (HCR): a facilitated in situ HCR system for the detection of environmental microorganisms. *Environ Microbiol* **17**: 2532–2541.

II.1.1 SUPPLEMENTARY DATA

Supplementary Materials and Methods.

Determining the number of plasmids in different induced cultures

For the induced cultures, the plasmid number per cell was calculated as follows: from the extracted pDNA the concentration per ml of culture was determined: [pDNA] (ng ml^{-1}) x Volume used for elution (0.075 ml) / initial volume for extraction (5 ml). Next, the number of plasmid per ml was calculated using the formula [pDNA] (ng ml^{-1}) x avogradro's number ($6.022*10^{23}$ molecules mol^{-1}) / plasmid molecular weight ($8.15*10^{15}$ ng mol^{-1}). The molecular weight of the plasmid was calculated by multiplying the plasmid length 12,539 bp (plasmid size 8139 bp + 4400 bp gene fragment) with the average molecular weight of a nucleic acid base pair (650 g mol^{-1}). The molecular weight of our plasmid was $8.15*10^{6}$ g mol^{-1} or $8.15*10^{15}$ ng mol^{-1}. The number of plasmids per ml was divided by the cell abundance (cell ml^{-1}). The cells were counted after DAPI staining with an epifluorescence microscope and the abundance was calculated with the formula: Average cells per field of view (cells) x filtered area (mm^2) / grid area (mm^2) x volume filtered (ml) x dilution factor. We expected cell/plasmid loss during the pDNA extraction; therefore, the plasmid per cell value was corrected as follows: plasmid $cell^{-1}$ of induced culture / plasmid $cell^{-1}$ of low copy culture (according pDNA concentration and cell abundance) x plasmid $cell^{-1}$ (according genome equivalents; 2-6).

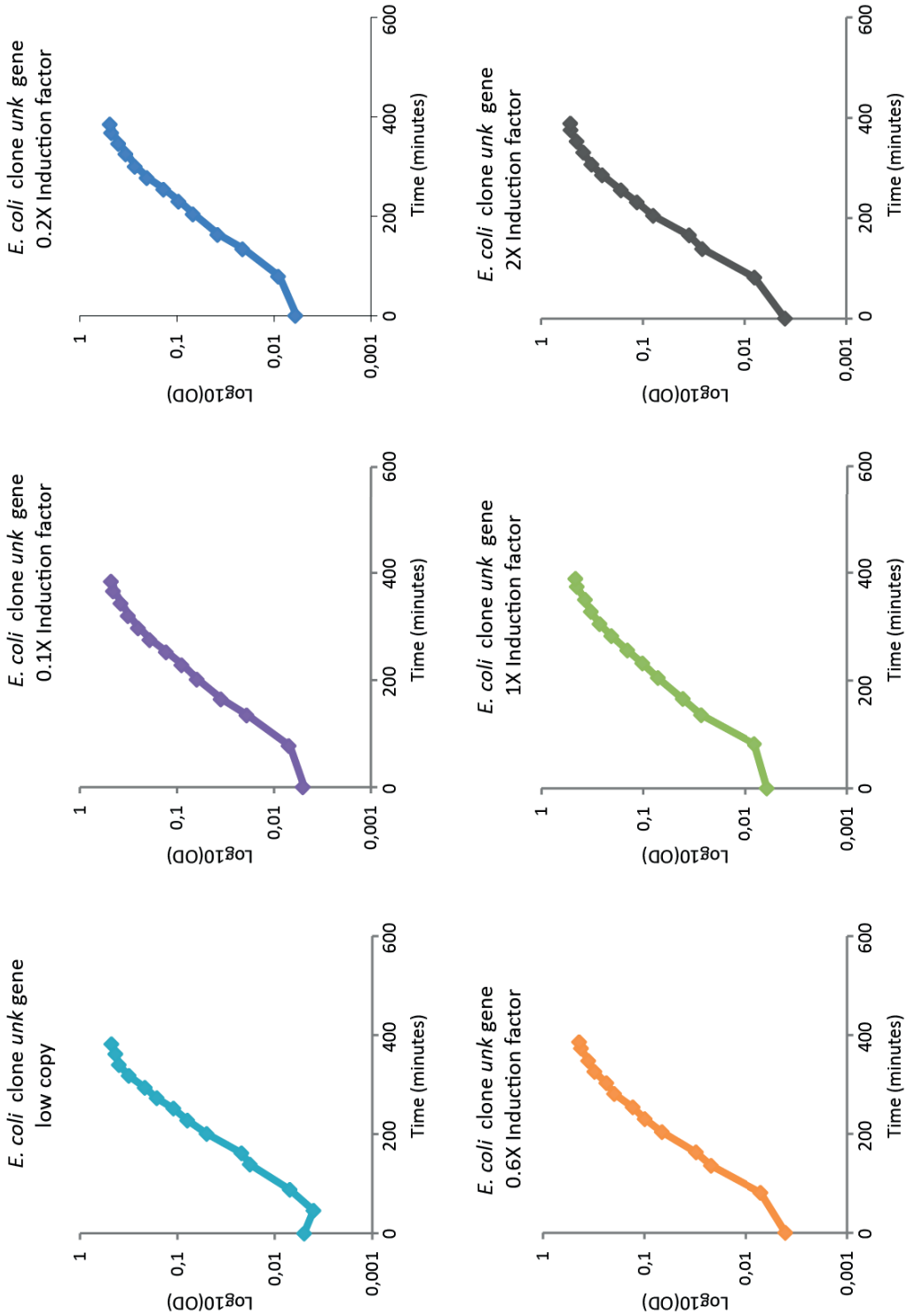

Figure S1. *E. coli* growth curves in time of different cultures used to induce the cell with different induction factor (Epicentre). Initial concentration 1000X according to Epicentre. The samples where fixed after approximately 300 minutes of growth.

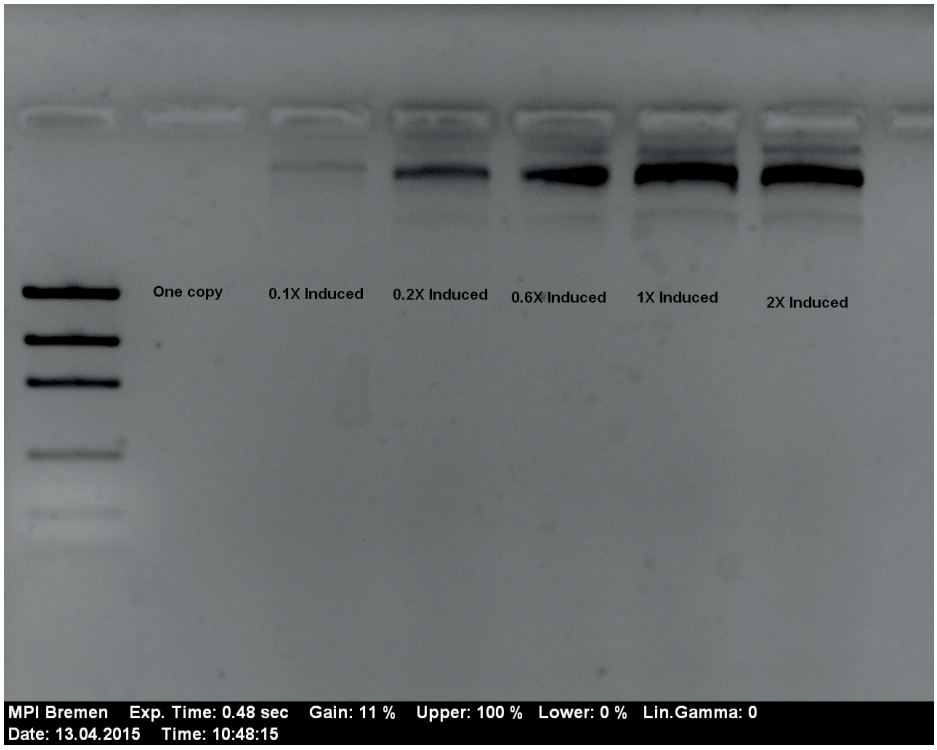

Figure S2. Gel electrophoresis of the plasmidic DNA extractions from the different cultures showing, the three different plasmidic configuration migrations and the increase in concentration in the cultures

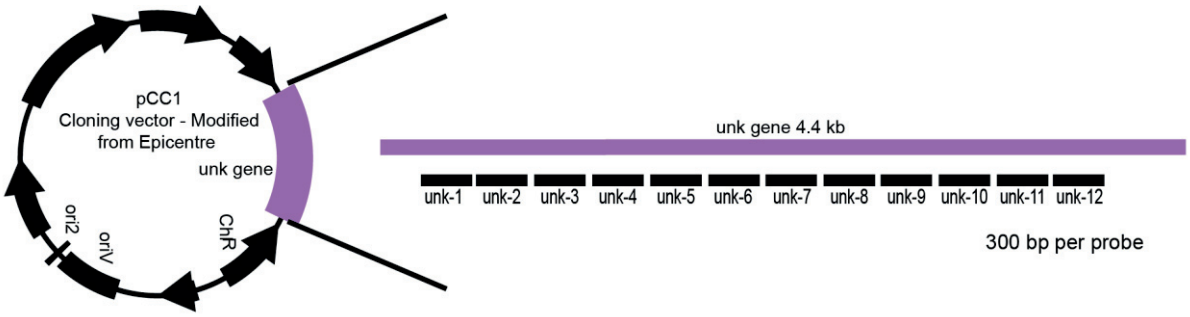

Figure S3. Diagram showing the target gene in a copy control plasmid and the organization of the probes use for the different hybridizations.

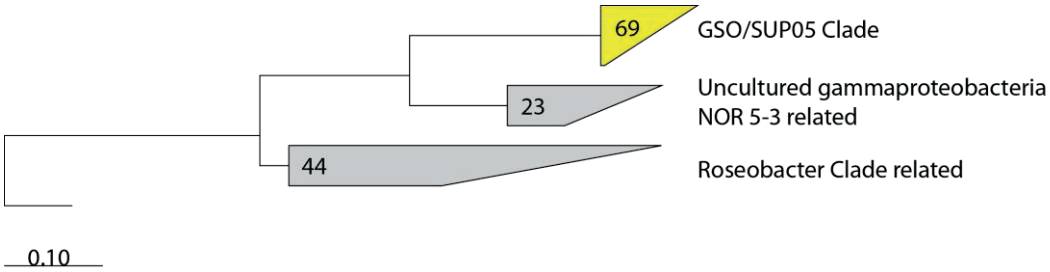

Figure S4. Phylogenetic tree based on the *soxB* gene showing the gene diversity from samples taken inRogoznica lake (October 2011). In yellow is highlighted the target group of the gene probe SoxB_SUP05RL.

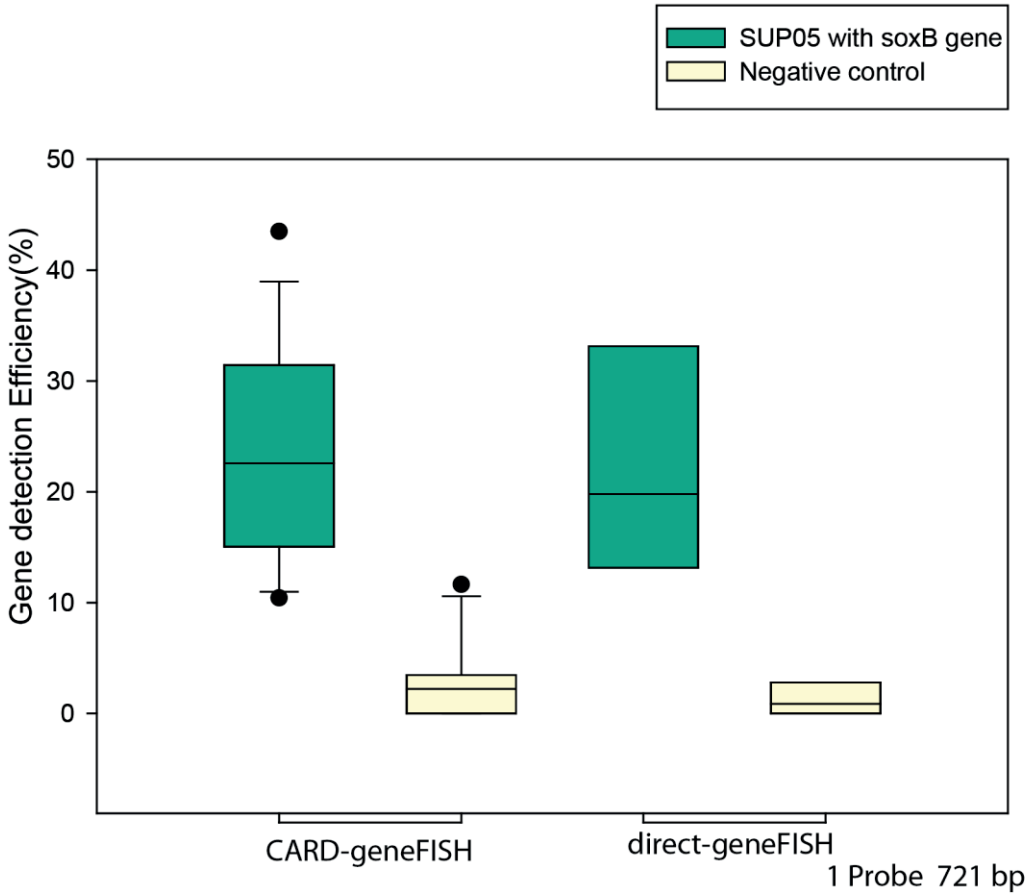

Figure S5. Comparison of Card- vs direct- geneFISH on Environmental samples. The 2 protocols were applied on samples collected in Rogoznica lake during an anoxic inversion (October, 2011). The target gene (soxB) was hybridized with a 721 bp probe. For negative control the samples were hybridized with NonPoly350 (Moraru *et al.,* 2010). Efficiency based on percentage of cells identified with rRNA probe that also presented gene signal.

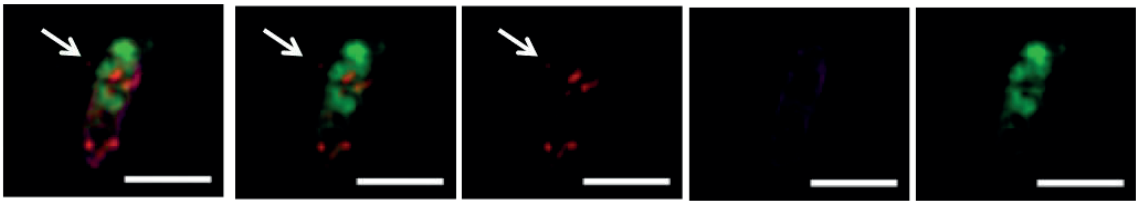

Figure S6. E.coli high copy culture visualized with SIM, the samle was prepared with direct-geneFISH and the membranes were stained with Nile Red. The first image shows the merged image of Nile Red (purple), ribosomes (green) and plasmids (in red). Next picture shows the merged images of the rRNA + genes. The last three pictures each individual channel. The arrow show the plasmid out of the cell. Pictures acquired with Elyra PS.1 microscope (Carl Zeiss, Jena, Germany). Scale bar: ~2 µm.

Table S1. Individual ratios of dye per 100 nucleotide of probes used for direct geneFISH optimization and labeling conditions.

Dye per 100 bases per individual probe					
Dye	Labeling conditions µl dye per ng	Probe	dye per 100 bases	Average	SD
Alexa 488	1	Unk PR1	0	0	0
		Unk PR1	0		
		Unk PR1	0		
		Unk PR1	0		
		Unk PR1	0		
		Unk PR1	0		
		Unk PR1	0		
		Unk PR1	0		
	2	UNK PR3	1,84	1,49	0,31
		UNK PR9	1,35		
		UNK PR10	1,62		
		UNK PR11	1,14		
	3	UNK PR1	1,97	1,86	0,34
		UNK PR3	1,20		
		UNK PR4	1,94		
		UNK PR5	2,14		
		UNK PR6	1,71		
		UNK PR7	1,60		
		UNK PR8	2,11		
		UNK PR12	2,23		
	6	UNK PR2	2,41	n.a	n.a
Alexa546	5	UNK PR7	6,25	6,97	1,02
		UNK PR12	7,69		
	10	UNK PR1	7,69	6,31	1,21
		UNK PR2	6,25		
		UNK PR5	6,25		
		UNK PR6	5,00		
		UNK PR8	5,00		
		UNK PR11	7,69		
	15	UNK PR10	6,67	n.a	n.a
Alexa 594	5	UNK PR1	1,07	1,03	0,15
		UNKPR1	0,81		
		UNK PR1	1,06		
		UNK PR12	1,17		
	10	UNK PR1	1,45	2,31	0,54
		UNK PR4	2,64		
		UNK PR5	2,15		
		UNK PR10	2,49		
		UNK PR11	2,83		
	15	UNK PR1	2,26	3,41	0,59
		UNK PR2	3,57		
		UNK PR2	3,57		
		UNK PR3	3,15		
		UNK PR6	3,56		
		UNK PR6	3,62		
		UNK PR7	4,17		
Alexa647	5	UNK PR1	2,61	1,91	0,76
		UNK PR2	2,53		
		UNK PR3	2,04		
		UNK PR4	2,76		
		UNK PR6	1,35		
		UNK PR7	1,04		
		UNK PR8	1,04		
	10	UNK PR5	4,55	5,34	0,74
		UNK PR6	5,00		
		UNK PR7	5,56		
		UNK PR8	6,25		
	15	UNK PR10	6,25	5,92	0,57
		UNK PR11	6,25		
		UNK PR12	5,26		

Table S2. Detailed data of hybridizations carried out on E.coli clones (low copy culture) as part of the method development. Unk1 to Unk12 refers to the polynucleotide probes labeled with 4 dyes (Alexa 488, Alexa546, Alexa594 and Alexa647). Each probe targets a 300 bp region of the unk gene. Two values representing the labeling degree are given for each probe: the total number of dyes per probe (in black) and the number of dyes per 100 bases (in red). The last two columns are the mean detection efficiency and the values obtained with the negative control. On average 800 cell were counted per hybridization per control.

Hyb.	Dye	Number of dye molecule per dsDNA probe												Total dyes / average dye # to 100 bases	Total no. probes in mix	Probe concentration pg µl-1	Detection efficiency (%)	
		UNK1	UNK2	UNK3	UNK4	UNK5	UNK6	UNK7	UNK8	UNK9	UNK10	UNK11	UNK 12				E.coli (2-6 copies)	Neg control
1	A488	12 2	14 2.3	7 1.2	12 2	13 2.2	10 1.7	10 1.7	13 2.2				13 2.2	104 1.9	9	31	49±9	0.85±0.80
2		12 2	14 2.3	11 1.8	12 2	13 2.2	10 1.7	10 1.7	13 2.2	8 1.3	10 1.7	6 1	13 2.2	132 1.8	12	31	79±12	1.41±1.73
3	A546	46 7.7	38 6.3	38 6.3			30 5	38 6.3	30 5		40 6.7	46 7.7	46 7.7	352 6.5	9	62	71±9	0.51±0.94
4	A594						22 3.7	25 4.2						47 3.9	2	31	35 ±11	1.38±0.77
5		14 2.3	21 3.5	19 3.2	16 2.7		22 3.7	25 4.2					8 1.3	125 3.0	7	31	83±8	0.21±0.53
6		14 2.3	22 3.7	18 3	16 2.7	13 2.2	22 3.7	25 4.2			14 2.3	16 2.7	8 1.3	168 2.8	10	31	92±5	0.86±1.04
7		14 2.3	22 3.7	18 3	16 2.7	13 2.2	22 3.7	25 4.2			14 2.3	16 2.7	8 1.3	168 2.8	10	62	96±4	0.99±0.93
8	A647	16 2.7	16 2.7	12 2	16 2.7	4 0.7	8 1.3	6 1	6 1					84 1.7	8	62	64±13	1.69±1.95
9		16 2.7	16 2.7	12 2	16 2.7	26 4.3	30 5	35 5.8	38 6.3		32 5.3	38 6.3	36 6	295 4.5	11	62	75±17	4.36±4.58

Table S3. T-test probability comparisons of the detection efficiencies values from different hybridizations.

Comparison	T-test (P,two tailed)
Hyb1 vs Hyb2	1E-05
Hyb4 vs Hyb5	8E-10
Hyb5 vs Hyb6	0.001
Hyb6 vs Hyb7	1E-02
Hyb8 vs Hyb9	2E-01
Hyb2 vs Hyb6	2E-03
Hyb2 vs Hyb5	3E-01
Hyb7 vs Hyb9	6E-04
Hyb3 vs Hyb7	5E-07

Table S4. Primer for amplification of soxB gene polynucleotide probe of SUP05 clade from Dragon Eye Lake.

Primer Name	Sequence Primer	Characteristics		
		Size	Tm	%GC
SoxB_SUP05RL_F	TCTTCCCAGATTGGTCAT	18 nt	59,1	54,1
SoxB_SUP05RL_R	GTTGTGAACCCTTCATCT	18 nt	58,9	53,9

Table S5. NaCl concentration in the washing buffer for washing at 48°C after hybridization at 46°C. Modified from Silva FISH database.

% formamide in hybridization buffer	[NaCl] in M (final concentration)	µl 5 M NaCl in 50 ml
0	0.900	9000
5	0.636	6300
10	0.450	4500
15	0.318	3180
20	0.225	2150
25	0.159	1490
30	0.112	1020
35	0.080	700
40	0.056	460
45	0.040	300
50	0.028	180
55	0.020	100
60	0.014	40
65	-	-
70	-	-

Table S6. *E.coli* copy control data of fixed cultures for geneFISH experiments indicating the culture growth rate (μ) and the doubling time (G) and genome equivalents of the cultures grown to harbour different number of plasmids per cell. Ind Sln: Plasmid induction solution (Epicentre)

Culture (E.coli EPPI300 pCC1- unk)	pDNA (ng ml^{-1})	plasmids ml^{-1}	Cell density cell ml^{-1}	plasmid cell $^{-1}$	Corrected on non induced	μ (h^{-1})	Genome equivalents	Plasmids per cell according Wild et al.2002	
Not induced - Low Copy	0.03	1.93E+06	2.73E+06	0.71	1	1	1.9	2	6
0.1X Ind Sln – Intermediate A	0.06	4.68E+06	4.47E+06	1.05	1.48	1	1.9	3	9
0.2X Ind Sln – Intermediate B	0.17	1.30E+07	9.28E+06	1.4	1.97	1	1.8	4	12
0.6X Ind Sln – High Copy	0.3	2.28E+07	6.15E+06	3.7	5.23	1	1.9	10	31

Gene soxB- SUP05RL PROBE SEQUENCE 721bp:

TCTTCCCAGATTGGTCATTTGGCCTTCGTGAGGATGAAATGAGCGAATTGGT
TGCAGACATTCGCGAGAATGAAAAACCTGATGCGGTGATTGTGGTTTCTCAT
AATGGTATGGATGTTGATATTAAGATGGCTGGCAATGTTGTTGGTATTGATG
CGATCTTTGGTGGTCATACGCACGATGGTATGCCAAAGCCAGTTGAAGTTCA
AAATGCTGGTGGCATGACAGTAGTTACTAATGCAGGTTGTTCTGGTAAGTAC
ATTGGCGTGATGGACTTAGAAATTAAAGACCATAAAATGGTAGGCTATAAC
TACAAGATGCTGCCAATTATCACGAATTTTATTAAGCCAGATCCGGCTGTGG
TTTCGTTTATTGATAAGATGCGTAATACTAAATACGACAAAAACGTTATTGA
GGCACGTAACTCAGCAATGAGTAACAACCCATCTCGTTTAGGTAAAACTTA
CGATGAAATTCTAACTGAAAAGCTTTGCACAACGGAGCAAACACTTTATCG
TCGTGGTAACTTTATGGGCACCTGGGATCAGGTATTGGTAAATTCATTACGT
GAAGAACACGACGCTGATTTTGCGATGTCTGCAGGGGTTCGCTGGGGTACT
TCGGTGCTGGCTGGTCATGATGTAACGATGGAAGATTTGATGACTAACACTT
CAATGACGTATGGCGAGACTTATGTTTCAGAGATGAAGGGTTCACAACT

CHAPTER III:
DIRECT-GENEFISH ON *FORMOSA* STRAIN B

III.1 - *IN SITU* DETECTION OF A GLYCOSYL-HYDROLASE 92 IN *FORMOSA* STRAIN B - A CASE STUDY

JIMENA BARRERO-CANOSA, KAREN KRÜEGER, TANJA FISCHER,
JENS HARDER, HANNO TEELING, BERNHARD M. FUCHS, RUDOLF AMANN

In situ detection of a glycosyl-hydrolase 92 in Formosa strain B - a case study

Preliminary results

Jimena Barrero-Canosa[a], Karen Krüeger[a], Tanja Fischer[b], Jens Harder[b], Hanno Teeling[a], Bernhard M. Fuchs[a**], Rudolf Amann[a]

[a] Department of Molecular Ecology. Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

[b] Department of Microbiology. Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

**To whom correspondence should be addressed:

Max Planck Institute for Marine Microbiology

Department of Molecular Ecology

Celsiusstr. 1, 28359 Bremen, Germany

Office: +49 421 2028935, email: bfuchs@mpi-bremen.de

Running title: Revealing ecological niche with direct geneFISH

Abstract

Phytoplankton spring blooms are seasonal events in which diatoms massively grow over a short period of time. In the German bight, successional blooms of diverse clades of bacteria follow the end of these phytoplankton blooms. Metagenomic analysis of the carbohydrate active enzymes (CAZymes) demonstrated that the substrate spectra of the dominant bacterioplankton clades after the bloom are notably distinctive, indicating that the successional appearance of certain taxonomic clades depends on the substrate availability. This study focuses on *Formosa* strain B, a flavobacterium isolated at Helgoland after the spring bloom in 2010. Genomic analysis suggests that *Formosa* strain B is a specialist for mannan degradation. Furthermore, the 16S rRNA phylogenetic affiliation and comparative analyses of the strain genome with metagenomes indicates that the strain is clonal in the spring bacterioplankton at Helgoland. We used direct-geneFISH to show the direct link between *Formosa* strain B and mannan degradation. Direct-gene-FISH is a technique based on fluorescence *in situ* hybridization that allows the detection of functional genes at the single cell level. We successfully hybridized a glycosyl hydrolase gene, GH92, in cultured *Formosa* strain B cells. GH92 is specific for mannan degradation. It turned out, however, that Formosa cells were quite delicate and cells tended to lyse upon hybridization. Some preliminary optimizations were carried out on pure cultures to simultaneously detect the 16S rRNA and the GH92 gene but resulted in broken Formosa strain B cells. Further optimizations needs to be done to adapt direct-geneFISH for GH92 in *Formosa* strain B cells and in environmental samples.

Introduction

Phytoplankton spring blooms are seasonal events characteristic for the coastal zones of higher latitudes, in which autotrophic microorganisms massively grow for a short period of time (Townsend *et al.*, 1992). The blooms end as a result of nutrient depletion, viral lysis and predator grazing (Banse, 1992), leading to a massive release of algae organic matter of different complexity (Teeling *et al.*, 2012 Hahnke *et al.*, 2014). Successional blooms of heterotrophic bacteria, able to degrade organic matter of diverse complexity, follow the ending of the phytoplankton bloom. The ability to degrade diverse organic matter resides on the presence of polysaccharide utilization loci (PUL) in the bacterial genomes (Larsbrink *et al.*, 2014). PULs are bacterial operons composed of susCD-like polysaccharide binding proteins and carbohydrate active enzymes (CAZymes) (Koropatkin *et al*., 2012)

In Helgoland (German bight), diatoms usually dominate the phytoplankton bloom. The successional bacterioplankton blooms are dominated by flavobacteria and gammaproteobacteria clades (Teeling *et al.*, 2012). Furthermore, metagenomic analysis of the carbohydrate active enzymes-CAZymes demonstrated that the substrate spectra of the dominant clades are notably distinctive, indicating that the successional appearance of certain taxonomic clades depends on the substrate availability (Teeling *et al*., submitted). Additionally, it has been observed that some of the heterotrophic bacteria that appear after the spring bloom are recurrently appearing over the years. This observation was based on metagenomic analysis and comparisons with pure cultures isolated during the spring bloom of 2010 (Hahnke *et al.*, 2014).

The Formosa sp. Hel1_33_131 was isolated after the spring bloom of 2010 in Helgoland. The isolate is a flavobacteria that appears after the phytoplankton bloom and, in this study, it is named *Formosa* strain B. The metagenomic analysis suggested that these bacteria specialized in the degradation of mannan. Additional analysis, based on comparison of the strain 16S rRNA gene with sequences retrieved from the environment, the comparison of the draft genome with metagenomes of the spring bloom 2009 (Hahnke *et al.*, 2014), and CARD-FISH counts, suggest that the isolate is highly abundant and clonal in the environment. Yet, the direct link showing Formosa strain B clonality and its potential capacity to degrade mannose in the environment is still missing.

Our purpose was to use direct-geneFISH to show the direct link between *Formosa* strain B and mannose degradation, and that clonal populations of this strain recurrently appears after the spring phytoplankton blooms in Helgoland. Direct-gene-FISH is a method based on fluorescence *in situ* hybridization that allows linking a functional gene to a particular taxonomic clade at single cell level. The method simultaneously hybridizes rRNA and gene probes that are labeled with multiple fluorescent dyes. Additionally, the method is simple, straightforward and results can be obtained within a day (Chapter II). To reach our goal, we designed a direct-geneFISH experiment to link the presence of a glycosyl hydrolase gene, GH92, to *Formosa* strain B cells.

Materials and Methods

Bacterial strains and environmental samples

The *Formosa* strain B pure culture was grown as described by Hanhke and collaborators 2014. The environmental samples that will be used were collected during the spring bloom from 2009 to 2012 in Helgoland (German Bight) at the station 'Kabeltonne' (54° 11.3' N, 7° 54.0' E).

16S rRNA *Formosa* Strain B Phylogenetic Affiliation

16S rRNA gene sequences from *Formosa* were retrieved from the Silva database (release Ref 119) (Quast *et al.*, 2013). The analysis included 16S rRNA gene sequences originating from environmental samples (Helgoland) and gene sequences of isolated cultures (described in Hanke *et al.*, 2014). The sequences were aligned with the SINA aligner (Pruesse *et al.*, 2012), and the alignment was manually curated. Phylogenetic analyses were done in ARB (Ludwig *et al.*, 2004) with maximum likelihood and neighbor joining using a filter of 50% similarity, which resulted in 1300 nucleotide positions. Bootstrap support was calculated with 100 replicates for both methods.

GH92 gene target selection for probe design

The closed genome of *Formosa* strain B was automatically annotated with Artemis (Rutherford *et al.*, 2000). Four genes coding for GH92 (GH92-1 to -4) were retrieved and compared with BLAST against GH92 sequences from 10 metagenomes of free-living bacterioplankton (0.2 - 3 µm) of the spring phytoplankton bloom of 2010 - 2012 in Helgoland.

GH92-1 gene probe synthesis

The genomic DNA from *Formosa* strain B was extracted with DNeasy Blood & Tissue Kit (Qiagen). The genomic DNA concentration was determined spectrophotometrically in a NanoDrop (Fisher Thermo Scientific) and used for the GH92-1 gene PCR amplification and cloning. For the amplification of the GH92-1 gene, the PCR mix contained 2 µM of each nucleotide, 1 µM per primer (GH92-2345bpF and GH92-2345bpR, see Table 1), 1 x HF buffer, and 0.1 U μl^{-1} Phusion® DNA Polymerase (New England BioLabs). The thermo cycling conditions were

determined *in silico* and crosschecked by gradient PCR. The PCR product was purified with QIAQuick PCR purification kit (Qiagen), eluted in TE buffer (5 mM Tris, 1 mM EDTA, pH 8.0) and chemically labeled with Alexa594 using the ULYSIS® Nucleic Acid Labeling Kit (Molecular Probes). The labeling was done according to the modifications suggested in Chapter II. The unbound dyes were removed with Micro Bio-Spin Columns (Bio-Rad, cat. no. 2320). The base to dye ratio was calculated according the instructions given with the ULYSIS® Nucleic Acid Labeling Kit. The probe was kept a -20 ºC until used.

Furthermore, the GH92-1 amplicon was also cloned into a pCC1 vector and transformed in *E.coli* Epi300 cells (Epicentre) for further synthesis of probes GH92-1 PR1 –PR5.

Direct geneFISH Protocol

Fixation

The *Formosa* strain B was fixed on middle logarithmic phase by adding paraformaldehyde to a final concentration of 3% for 14-18 h at 4 °C, concentrated through a Swinnex filter holder (25 mm diameter; Millipore) equipped with a 0.2 µm pore-size polycarbonate filter, and washed with 50 ml of 1X PBS. After recovery of the cell suspension, 96% ethanol was added in 1:1 ratio and the cells were stored at -20 °C.

Cell immobilization and permeabilization

Formosa strain B fixed cells were immobilized on polycarbonate filters (diameter, 25 mm; pore size, 0.2 µm; type GTTP; Millipore). For the sample preparation, different conditions were tested including non and lysozyme permeabilization. Lysozyme permeabilization was carried out with 2 enzyme concentrations (0.5 and 10 mg ml^{-1}) for 10 or 60 min and at 4 or 20 ºC (See details in Table 2).

Hybridization

The gene and ribosomal RNA (rRNA) probes were hybridized simultaneously using a hybridization mix consisting of hybridization buffer (35% formamide, 5x SSC, 20% dextran sulfate, 0.1% SDS, 20 mM EDTA, 0.25 mg ml^{-1} sheared salmon sperm DNA, 0.25 mg ml^{-1} yeast RNA and 1% blocking reagent for nucleic acids) and rRNA probe and gene probe. In pure cultures the rRNA was targeted by 50 ng μl^{-1} EUB338

(Amann *et al*., 1990) or FORMB181 (Telling *et al*.,2012) probes. In environmental samples the rRNA was targeted with FORMB181 probe. The rRNA probes were four times labeled with ATTO488. The gene was targeted with 62 pg μl^{-1} Alexa594 labeled GH92-1_2345bp probe (**Table 1**).

To denature both, the target DNA and dsDNA probes, the cells were covered with hybridization mix and incubated for 40 min at 85 °C. After denaturation, the samples were transferred to 46 °C, for 2, 8 or 16 h of hybridization (Table 2). The excess probe was washed as previously described (Pernthaler *et al*., 2002), in a washing buffer containing 80 mM NaCl, 100 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 10% SDS, for 15 min at 48 °, followed by 20 min wash in 1x PBS, 1 min water, 1 min 96% ethanol and air-drying.

Counterstaining and mounting

The cells were counterstained with 1 µg ml^{-1} 4', 6-diamidino-2-phenylindole (DAPI), washed quickly with ultrapure water (MiliQ; Milipore,) and 96% ethanol and then let air-dry. After drying, the samples were mounted with a mix of 3:1 Citifluor - Vectashield antifading reagents (Citifluor™. Vectorlabs).

Results and discussion

The purpose of this study was to show *in situ* that *Formosa* strain B is a specialist for mannan degradation and that clonal populations of this strain recurrently appear after spring phytoplankton blooms in Helgoland. We, therefore, designed a direct-geneFISH experiment to link the presence of the GH92 gene, involved in mannan degradation, to *Formosa* strain B cells. Here we discuss the experimental design and preliminary experiments for the adaptation of a direct-geneFISH protocol on pure cultures of the mentioned strain.

Formosa strain B: an ideal target group

During the spring bloom of 2010 in Helgoland, three Formosa strains, Formosa sp. H3_A48, Formosa sp. Hel1_33_131, Formosa sp. Hel1_31_27, were isolated (Hahnke *et al*., 2014). In this study, the isolates are called *Formosa* strain A, *Formosa* strain B and *Formosa* strain C, respectively. To assess their 16S rRNA phylogenetic affiliation, a phylogenetic tree was reconstructed with maximum likelihood and neighbor joining algorithms. The two reconstructed trees agreed in their branching patterns, therefore the maximum likelihood tree is shown in Figure 1. *Formosa* strain B clustered with similarity above 98.9% to sequences retrieved from Helgoland during the spring bloom of 2009 (Figure 1). Furthermore, the strain's genome was nearly completely recovered from metagenome sequences from the same bloom (92.9%) (Hahnke *et al*., 2014). This suggests that *Formosa* strain B is a clonal member of Helgoland bacterioplankton. Furthermore, CARD-FISH counts, with an rRNA probe that targets specifically *Formosa* strain B cluster (FORM181B; Figure 1), indicate that this clade appears in high abundance (up to 4% of the total cell counts) after the spring phytoplankton bloom (Teeling *et al*., 2012).

A mannosidase-coding gene (GH92) as the functional FISH target

The closed genome of *Formosa* strain B was automatically annotated to find indications that the strain is a specialist in polysaccharide degradation. We identified one polysaccharide utilization locus (PUL) at the 1198 – 1227 kb genome position (Figure 2). This PUL was predicted to encode a tonB sugar receptor, an outer membrane sugar binding protein (susC-like), five sulfatases, one phospho-mannose isomerase and

six glycosyl-hydrolases (GH) from two different mannosyl hydrolase families (two GH99 and four GH92). The characteristics of the PUL strongly suggest that *Formosa* strain B is a specialist for the degradation of mannose-rich sulfatated polysaccharides (e.g. mannane sulfate, xylomannan), that are common components of marine red algae cell walls (Mandal *et al.*, 2008; Ghosh *et al.*, 2009; Pérez Recalde *et al.*, 2012).

Furthermore, the metagenomic analysis of three Helgoland spring blooms (2010-2012) linked the GH92 family almost exclusively to *Formosa* spp. (Telling *et al*., submitted). Based on this knowledge, we considered that the GH92 family were good candidates to select as the gene targets for direct-geneFISH. GH92s are exo-acting glycosyl hydrolases with four different cleaving modes described(α 1-2, α 1-3, α 1-4 and α 1-6) (Zhu *et al*., 2010). *Formosa* strain B has four GH92 genes (GH92-1 to 4; see Figure 2) that could perhaps code for each of the cleaving modes described therefore maximizing the utilization of mannose-rich polysaccharides. While this is very speculative and based only on our genomic analysis, it could be tested by cloning the genes into expression vectors, purifying the protein products and measuring their individual affinity to different substrates (Zhu *et al*., 2010).

For geneFISH, however, we had to choose between the four genes all with similar length and %GC content (~2250 bp; ~36% GC). We first compared the four genes with each other and with sequences retrieved from the mentioned metagenomes. Our comparison indicated that: (i) the four GH92 genes of *Formosa* strain B were very different among each other and (ii) among the GH92s sequences retrieved from metagenomes, 24%, 26%, 23% and 27% were clustered with similarity ≥90% to GH92-1, GH92-2, GH92-3, and GH92-4 respectively. This suggests that Formosa cells in the environment also harbor the four GH92 genes. Furthermore, we tried to predict if the gene products could be anchored to the membrane by looking for signal peptides in the amino-acid sequences with LipoP 1.0 (Junker *et al*, 2003) and Sig4P (Petersen *et al*., 2011) software. The predictions indicated that GH92-2, GH92-3 and GH92-4 most likely are anchored to *Formosa* strain B membranes, as the three presented a predicted Type II lipoprotein signal sequence. To prove that these proteins are indeed anchored to the membranes, fluorescently-labeled antibodies for each specific protein could be generated allowing their respective protein localization on the cell membranes (Rogowski *et al*., 2014). Interestingly, the GH92-1 (See Figure 2), did not present any type II lipoprotein signal indicating that this protein could either be secreted or located

in the periplasm. Based on these results, at nucleic acid level, any of the four genes could be a good candidate for the gene probe design. We selected GH92-1 since it is the first appearing on the locus and, therefore, having a higher chance to be expressed. In the future, probes targeting the remaining three GH92 could be designed to simultaneously co-localize the four genes.

Gene probe synthesis

A set of primers were designed to amplify the GH92-1 gene (2256 bp) including short flanking regions (Figure 2; Table 1). Genomic DNA extracted from a pure culture of *Formosa* strain B was used as a template for the PCR amplification. The PCR product was divided in two fractions: one fraction was chemically labeled with Alexa594 (~3 dyes per 100 base) and used directly as a gene probe, while the second fraction was cloned into a vector for future probe synthesis.

Furthermore, five polynucleotide probes of ~445 bp each were designed to target consecutive regions of the GH92-1 gene (Figure 2; Table 1). The vast amount of A and T repetitions in the gene hindered the optimal primer design (See supplementary data). This did not represent a problem for the probe amplification using the cloned DNA as template. Nevertheless, the usage of these primers to amplify DNA extracted from environmental samples will require further optimization as they can generate un-specific amplicons. The usage of the designed five probes in direct-geneFISH will be subject of future research.

Preliminary optimizations of direct-geneFISH protocol on pure cultures

Optimizations were carried out on a pure culture of *Formosa* strain B fixed with 3% PFA. For the initial tests, the rRNA was targeted with the EUB probe tetra-labeled with ATTO488, while the GH92-1 gene was targeted with the GH92-2345bp probe (Table 1). The purpose of using the full fragment of the gene as a long polynucleotide probe, instead of the multiple ~450bp probes, was to investigate the possibility of using DNA amplified directly from environmental samples as gene probes.

As part of the protocol optimization, we tested different permeabilization procedures to find the conditions that allow the penetration of the long probe without disturbing the cell morphology. Furthermore, different incubation times were tested to find the most suitable one for the hybridization of the long probe. Indeed, due to the

variation of the probe length, the time to reach and access the gene could vary as well. (Figure 2; Details in Materials and Methods section).

Direct-geneFISH was originally developed on *Escherichia coli* cells and tested on sulfur oxidizing gammaproteobacterial cells (see Chapter II). *Formosa* strain B cells turned out to be more fragile than gammaproteobacterial ones. After carrying out the direct-geneFISH experiments, many of the cells broke or bundled together. That hindered the possibility to calculate the gene detection efficiency. Hence, we created an artificial system to evaluate the quality of the rRNA and gene signals (Figure 2). For the experiments with long hybridization times, the gene signal was in most of the cases bright and sharp. However, all the cells were broken and lacked a rRNA signal. In contrast, the experiments with short hybridization times the rRNA signal was detectable but the gene signal was often unclear (Figure 2). The following conditions were identified as the ones offering the best compromise between gene and rRNA signals: cells permeabilization with 10 mg ml^{-1} lysozyme for 60 min at 4ºC and hybridization time of 2 h (Figure 3). Nevertheless, the results were not completely satisfactory as the quality of the cell signal was still suboptimal. We also noticed that probes clustered on the polycarbonate filters (Figure 3), most likely due to the tendency of long probes to stick together.

Based on the obtained results, we were able to draw the following conclusions: (i) short hybridizations (2 hours) are required in order to preserve *Formosa* strain B cell morphology; (ii) for an optimal hybridization of the long probes (2345bp), a strong permeabilization and longer incubation times are particularly necessary; (iii) the quality of the preparations is decreased by the tendency of long probes to clump together outside the cells. Therefore, it is recommended to use multiple short probes that will help to avoid penetration and breakage issues, as suggested by (Allers *et al.*, 2013).

Outlook

Direct geneFISH optimization

The next steps will focus on the preservation of *Formosa* strain B cell morphology. Based on the conclusions drawn above, we plan to do further optimizations using the mentioned ~450 bp dsDNA polynucleotide probes (see Table 1 for the details). The usage of shorter probes will permit the utilization of a gentler permeabilization protocol and reduction of the incubation times (denaturation and hybridization).

Furthermore, the current results indicate using full length genes amplified directly from environmental samples as gene probes is not recommendable. A possible solution, which could lead to the usage of amplified genes directly from the environment, could be the fragmentation of long amplicons with restriction enzymes or with ultrasonication. The restriction enzymes should be preferred since they cut at specific sites (See Supplementary Figure S1). Ultrasonication, instead, generates random fragments that could lead to the fragmented probes acting as competitors and, additionally, makes the control of the hybridization stringent conditions more difficult.

The optimized protocol will be used on environmental samples taken during the spring blooms of 2009 to 2012 in Helgoland, to show that the environmental *Formosa* strain B identified with an rRNA specific probe (FORM181B) harbor the GH92-1 gene with at least 90% similarity to the *Formosa* strain B pure culture.

Formosa strain B as a model system for understanding the polysaccharide degradation

In the natural environment, *Formosa* strain B has three appealing characteristics: it is abundant, recurrent and clonal. While the first characteristic was demonstrated with CARD-FISH, the second two will be confirmed with direct-geneFISH. The advantages of having an isolate with these characteristics, is that it allows the *in vitro* understanding of an ecologically relevant microorganism's physiology (Kaeberlein *et al.*, 2002) which can be extrapolated to their natural environments, allowing a better understanding of their ecological role. Additionally, if an uncultured bacteria harbors similar genes, their functionality could also be inferred. For instance, *Formosa* strain B can serve as a

model for understanding polysaccharide degradation processes by doing an in-depth molecular characterization of all the genes of the PUL including reverse genetics, *in vitro* protein chemistry, enzymology and structural biology (Larsbrink *et al*., 2014; Rogowsky *et al*., 2014).

Figures and Tables

Figure 1. Maximum likelihood phylogenetic tree based on the 16S rRNA gene. Phylogenetic analyses were done in ARB g using a filter of 50% similarity, which resulted in 1300 nucleotide positions. Bootstrap support was calculated with 100 replicates. In red are highlighted the 16S rRNA gene sequences corresponding to Helgoland isolates from Hahnke *et al*., 2014. The green stars show the targets of FORM181B probe.

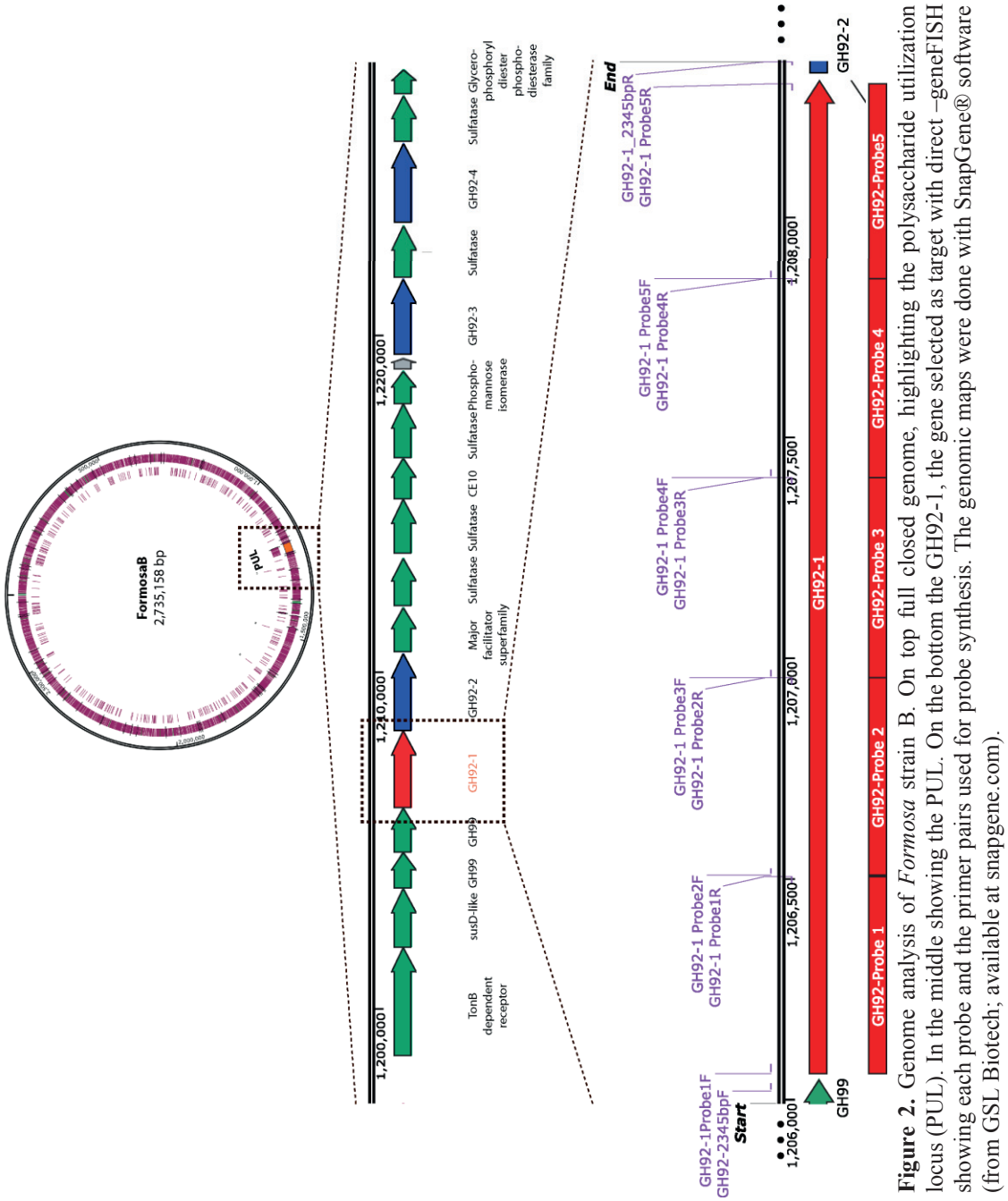

Figure 2. Genome analysis of *Formosa* strain B. On top full closed genome, highlighting the polysaccharide utilization locus (PUL). In the middle showing the PUL. On the bottom the GH92-1, the gene selected as target with direct –geneFISH showing each probe and the primer pairs used for probe synthesis. The genomic maps were done with SnapGene® software (from GSL Biotech; available at snapgene.com).

Table 1. Polynucleotide probes targeting a GH92-1 gene of *Formosa* Strain B (genome position 1206060-1208315 bp). *Fragment amplified and cloned for future probe synthesis. The PCR probe was also labeled and used for direct geneFISH. ** PCR product ready for labeling.

GH92-1 Probes and Primers										
PROBE	Region covered in the genome	Length (bp)	% GC	PCR Forward Primers				PCR Reverse Primers			
				Primer Name	Sequence	Length	Tm	Primer Name	Reverse	Length	Tm
GH92-1 2345bp FULL *	1206021-1208365	2345	35,9	GH92-2345bpF	GCAGCCCGTTTATCCTT	17	53	GH92-2345bpR	CTGTAAGAAGTAGAAAGTAG	20	45
GH92-1 PR1	1206061 - 1206506	446	34,9	GH92-1 Probe1F	TGAATTTAAAGACCTTTATA	20	42	GH92-1 Probe 1R	TGATTGCGCTGCGGGTGT	18	61
GH92-1 PR2**	1206509 - 1206955	447	33,8	GH92-1 Probe2F	TTTCTTTAAATTTACCT	17	37	GH92-1 Probe2R	ATCATTCGTTCATAGGAA	18	46
GH92-1 PR3	1206956-1207411	456	39,5	GH92-1 Probe3F	AAATATTTCTGAAGACGG	18	45	GH92-1 Probe3R	TGAACTCTAGGCTCTGTTT	19	51
GH92-1 PR4**	1207413 - 1207862	450	35,1	GH92-1 Probe4F	GATTTTGAAAAAAGAC	16	38	GH92-1 Probe4R	ACTGGGTTCATTTCCCATA	19	51
GH92-1 PR5**	1207863 - 1208304	442	37,5	GH92-1 Probe5F	TTTCATATCCCCTATT	16	40	GH92-1 Probe5R	AATCGTTTGGTATCAG	19	43

Table 2. Direct geneFISH optimizations on pure cultures of *Formosa* Strain B. The experiments were carried out with GH92-1 2345bp gene and EUB probes. The table shows the steps and the conditions tested (permeabilization and hybridization). The majority of the cases, the cells broke and clumped. Consequently, we could not determine the gene detection efficiency (DE = cell presenting gene signal / total cell x100%). Therefore a qualitative system was used in which the FISH signals were evaluated. The system has four different possible gradings: +++ for bright and defined gene and cells signals; + blurred signal that still permit the identification for genes and cells; +/- blurred signal that does not clearly permit identification for genes and cells; - no signal

Permeabilization				Hybridization incubation time (h)	Signal Quality	
Enzyme	**Concentration (mg ml^{-1})**	**Time (min)**	**T (°C)**		**Gene Signal**	**rRNA signal**
NON				16	-	-
Lysozyme	0.5	10	20		n.a	-
	0.5	60	4		+/-	-
	10	10	20		+++	-
	10	60	4		+++	-
Lysozyme	10	60	4	8	+++	-
	0.5	10	20	2	+/-	+/-
	0.5	60	4		+/-	+/-
	10	10	20		+/-	+
	10	**60**	**4**		+	+
NON					+/-	+/-

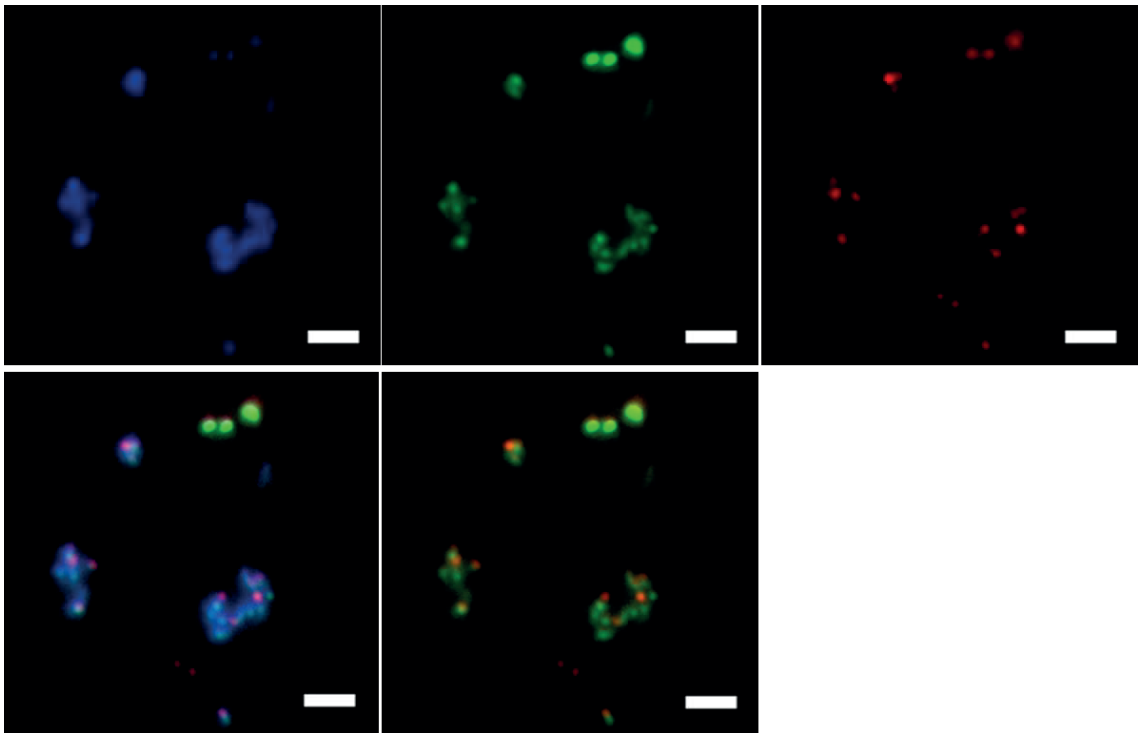

Figure 3. Direct-geneFISH on *Formosa* strain B pure culture. The pictures were taken with an iXON897 EM-CCD camera with a 63x plan-apochromatic oil immersion objective coupled to the ELYRA PS.1 Microscope. The images were analyzed with Zen software blue edition (Carl Zeiss). On the top, individual signals from DAPI (blue) EUB rRNA (green) and GH92-1 gene (red). On the bottom merged images: left - DAPI, rRNA and gene, right-rRNA and gene. Scale bar ~2 µm.

References

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R and M.D Sullivan. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15**: 2306–2318.

Amann, R., 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.

Banse, K. (1992). Grazing, temporal changes of phytoplankton concentrations, and the microbial loop in the open sea. In *Primary productivity and biogeochemical cycles in the sea*. Falkowski, P., Woodheard, A., and Vivirito, K (eds): Springer US:409-440

Ghosh, T., Pujol, C.A., Damonte, E.B., Sinha, S., and Ray, B. (2009) Sulfated xylomannans from the red seaweed *Sebdenia polydactyla*: structural features, chemical modification and antiviral activity. *Antivir Chem Chemother* **19**: 235–242.

Hahnke, R.L., Bennke, C.M., Fuchs, B.M., Mann, A.J., Rhiel, E., Teeling, H., Amann, R. and Harder, J. (2014) Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environ Microbiol* **10:** 3515–3526

Juncker, A.S., Willenbrock, H., von Heijne, G., Nielsen, H., Brunak, S., and Krogh, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **8**:1652-62

Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129

Koropatkin, N.M., Cameron, E.A., and Marten, E.C. (2012) How glycan metabolism shapes the human gut microbiota. *Nat Rev Micro* **10**: 323-335

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinter, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies. G.J., Martens, E.C., and Brumer, H. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**: 498–502.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A and Schleifer, K.H. (2004) ARB: a software environment for sequence data. *Nucl. Acids Res.* **32**: 1363–1371.

Mandal, P., Pujol, C.A., Carlucci, M.J., Chattopadhyay, K., Damonte, E.B., and Ray, B. (2008) Anti-herpetic activity of a sulfated xylomannan from *Scinaia hatei*. *Phytochemistry* **69**: 2193–2199.

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

Pérez Recalde, M., Carlucci, M.J., Noseda, M.D., and Matulewicz, M.C. (2012) Chemical modifications of algal mannans and xylomannans: Effects on antiviral activity. *Phytochemistry* **73**: 57–64.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high-throughput

multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785-786

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* **41**: D590–6.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona,T., Terrapon, N., Lowe, E.C., Baslé, A., Morland, C., Day., A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J., and Bolam, D.N. (2014) Glycan complexity ditactes microbial resource allocation in the large intestine. *Nat commun* **6**:7481

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., Kassabgy, M., Huang, S., Mann, A.J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F.D., Callies, U., Gerdts, G., Wichels, A., Wiltshire, K.H., Glöckner, F.O., Schweder, T. and Amann, R. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.

Teeling, H., Fuchs, B.M., Bennke, C., Krüger, K., Chafee, M., Kappelmann, L., Reintjes, G., Waldmann, J., Quast, C., Glöckner, F.O., Wichels, A., Gerdts, G., Wiltshire, K.H. and Amann, R.I. Recurring patterns in bacterioplankton dynamics during coastal spring blooms. Manuscript in submitted for publication to e-life

Townsend, D.W., Keller, M.D., and Sieracki, M.E. (1992) Spring phytoplankton blooms in the absence of vertical water column stratification. *Nature* **360**: 59-62.

Zhu, Y., Suits, M.D.L., Thompson, A.J., Chavan, S., Dinev, z., Dumon, C., Smith, N., Moremen, K.W.., Xiang, Y., Siriwaderna, A., Williams, S. J., Gilbert, H.J. and Davies, G.J. (2010) Mechanistic insights into Ca2+ -dependent family of α-mannosidades in a human gut symobiont. *Nat chem biol* **6**:125-132

Supplementary data

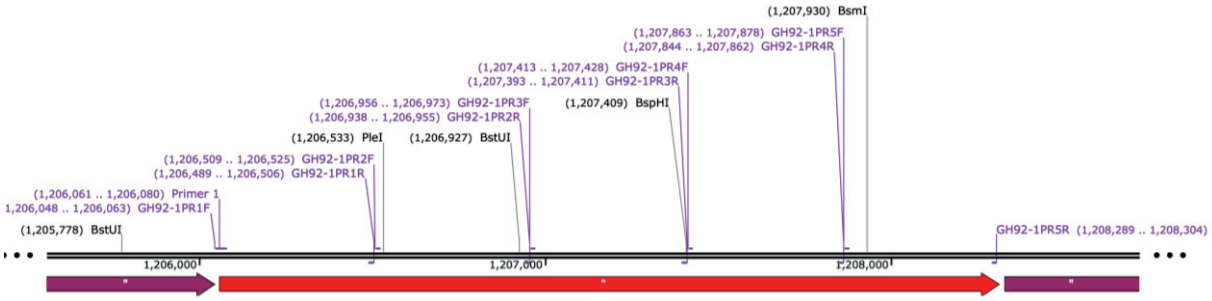

Figure S1. The GH92-1 gene (red) showing the primer positions (in light purple) designed in this study and the recognition sites of for restriction enzymes, BstUI, PleI, BspHI and BsmI. These restriction enzymes could potentially be used after the full gene amplification to generate gene probes of 400-500 bp lengths. The gene visualization was done with SnapGene® software (from GSL Biotech; available at snapgene.com).

GH92-1 Cloned Fragment 2345 bp:
Primers for cloning
Starting & ending GH92-1

GCAGCCCGTTTATCCTTACGTAAAATAATTCCACAGTTTATGAATTTAAAGA
CCTTTATACCAAAAATAGAATTTATTATTTATATAAAAATAGTAGTGATTTC
TCTGCTCATATTTTCATGTAACTCAACAAAAGGCACCTCAAAAAATAGTGAT
AATGTAAACTATGTTGATCCACAAATTGGAGGCGTTGCGCCTTTTTTACAAC
CGACAAGAACACGCATACATTTGCCAAATTCAATGGTACGAATGTATCCAG
AAAGAAAAGACTACAGAGACGATCAAATCACTTCTTTTCCATTAACAACAC
GAAAGCATCGATCTGATTTACTTTTTAATATTATGCCTGTTTCGGGGGCTAT
AAATGAAGCCGATGAACCGATTTCTGCTTGGGATCAAGAGTTAGAGATTGC
ACACCCTCACTACTACTCTACTTGGCTAGAAGATTATAATATTACTACAGAA
TTTACACCCGCAGCGCAATCAGGTTTCTTTAAATTTACCTATCAAAACAACA
CGACTCGAAAACTATATCTCAGCGATTTAACGGGTGATGATTGGCAATTAA
ATGAAGATGGTTCTCTTACGGGAACCGAATCTTTTGAAGGAATGTTGGGCTA
TGTGTATGCCGTAACAGACAAAAAAGGAGTGTTTACATCTGTAGAAACTCA
AGATAACACTCTCAACTCTTGGATGACTTGGAACGATACAGATATCAATCA
AGTTCAATTTAAATATGGAATCTCATTTATAAGTAGTGAGCAAGCAAAGAA
AAATTTAGAAAGAGAAATTCCAGATTTCAATTTTGAAAATGTAAAAGAAAA
AGCGCATAAAAAATGGTCTAATTTGATGAATCAAATTGAAGTTGAAGGCGG
CACAGATGCGTACAAAAGAACTTTTTATACCGCGCTTTATCGTTCCTATGAA
CGAATGATAAATATTTCTGAAGACGGAACCTATTACAGTAATTACGACCAA
AAAATTCACGAAGGGGAACAAGACTTTTATGCCGATGATTGGGTTTGGGAT
ACGTTTTTAGCATTGCATCCACTACGCTATATTTTACAACCAGACGTGGAAT
CAGATATGATGGCTTCTTACGTAAGAATGTACGAACAATCGGGTTGGTTGCC
TCAATTTCCTCAAATTCATGGAGATGCACCTCCAATGAATGGCTTCCACTCT
ACCATCATGTTATTGGACGCTTACAGAAAAGGAGTTACTAATTTTGATGCTG
AAATAGCATTTCAGGCAATGAAAAAAAATGCTTTAGAAGGCACCATGATTC
CTTGGAAAATGGGCGCACATGCCCCTCTTGACAAACTTGCTTTAGAACTTGG
TTATTTTCCAGCATTGTACCCAGAAGAAAAAGAAACAGAGCCTAGAGTTCA
TGATTTTGAAAAAAGACAATCGGTTGCGGTTACCTTAGCTCAAGGTTATGAT

GATTGGGCATTAGCAGAATTTGCTAAAGAATTAGGTAAAGAGGAAGATTAT
AAATACTTCTCCAAAACAGCACAGAATTACAAAAATCTATATTGGAAGGAA
AGAGGATTTTTTATGCCAAAAGATGCCAAGGGCAATTGGATTGATATCGAT
CCAAAATTTGATGGTGGAATGGGTGGTAGAGATTATTATGATGAAAATAAT
GGTTGGACCTACCTATGGAATGTACAACAAGATATTTTAGGATTGCAAGAA
TTAATGGGCGGAAGAAAAACGTTTGAAGAACGTTTAGACCAAATGTTTAGA
GAAGATTTAGGAAGATCTACCTATGCTTTAAATGCAAGATTTCCAGATTTCA
CAGGGATTGTTGGGCAATATTCTATGGGAAATGAACCCAGTTTTCATATCCC
CTATTTATACAATTTTACAGATTCGCCTTGGAAAGCGCAAAAGAAAATAAG
AATGCTTTTGAACACATGGTTTAAAGATAATATTTTTGGTATTCCAGGAGAT
GAAGATGGTGGAGGCATGTCTGCTTTTGTGGTGTTTTCTGCAATGGGCTTTT
ATCCCATAACTCCAGGAATCCCCGTCTATACGATTGGAAGTCCTTTGTTTTC
AAAAATCAGCATCAACCTTCCAAATGGTAAAAAATTCACCATAAACGCACC
CAACTGTAATGAAACTAACAAATACATACAGTCTGCAACCTTAAACGGAAA
AGAGCTAAAAGGCCCTTGGTTTACACACAATGATTTGATCAATGGCGCAAC
CATCACATTGCAAATGGGGAAATACCCAAATATAAATTGGGGTACAGATCC
ATCTCTGATACCAAACGATTTTAAAAACTAACAACCAATTACTTATGAAAA
ATATAAAAATCTACTTTCTACTTCTTACAG

GENERAL DISCUSSION AND OUTLOOK

The principal aim of this thesis was the optimization of the geneFISH method developed by Moraru and collaborators in 2010. The main efforts were focused both on increasing the sensitivity and on simplifying the method. The strategies adopted to address these issues were (i) the employment of multiple probe-target regions, (ii) the reduction on incubation times, (iii) the removal of CARD, and (iv) the implementation of highly sensitive detection systems for the signal acquisition (Figure 1). The synergistic effect of these strategies allowed the development of a simple and sensitive protocol, named direct-geneFISH. The method allows the detection and potential quantification of genes at the community level (i.e., which genus harbors a target gene), population level (i.e., how many cells of a genus harbor a target gene), and single cell level (i.e., how many gene copies does a cell harbor). In the following sections, the results of this thesis will be briefly summarized and will be used as basis for the definition of future directions of research.

In direct-geneFISH, the employment of multiple probe-target regions (targeting a region 3 kbp long) permitted to reach almost 100% gene detection efficiency. However, direct-geneFISH has limited detection efficiency for shorter gene fragments (e.g., 450 bp). This could be improved by increasing the number of labels per probe, for instance by implementing other labeling strategies such as click chemistry ligation (Kolb *et al.*, 2001). Also, the probes can be labeled with brighter molecules like quantum dots (QD) (Resch-Genger *et al.*, 2008). Nevertheless, the QD are large molecules (6 - 60 nm) compared to the organic dyes (~0.5 nm) (Resch-Genger *et al.*, 2008), therefore they can introduce penetration-related difficulties (Pawley, 2010). Alternatively, other approaches can be implemented such as the hybridization chain reaction-HCR (Yamaguchi, *et al.*, 2015a; Yamaguchi *et al.*, 2015b), which allows the simultaneous amplification of multiple linker-probes (e.g., gene and rRNA) (Choi *et al.*, 2014), or oligo–paint based approach, in which the oligo-paint probes carry a region for the hybridization of a second tetra-labeled probe (Beliveau *et al.*, 2015). Nevertheless, oligo-paints are limited to organisms to which the full genome is sequenced and could

only be applied on environments with low allele diversity since they use short probes (Beliveau *et al.*, 2012).

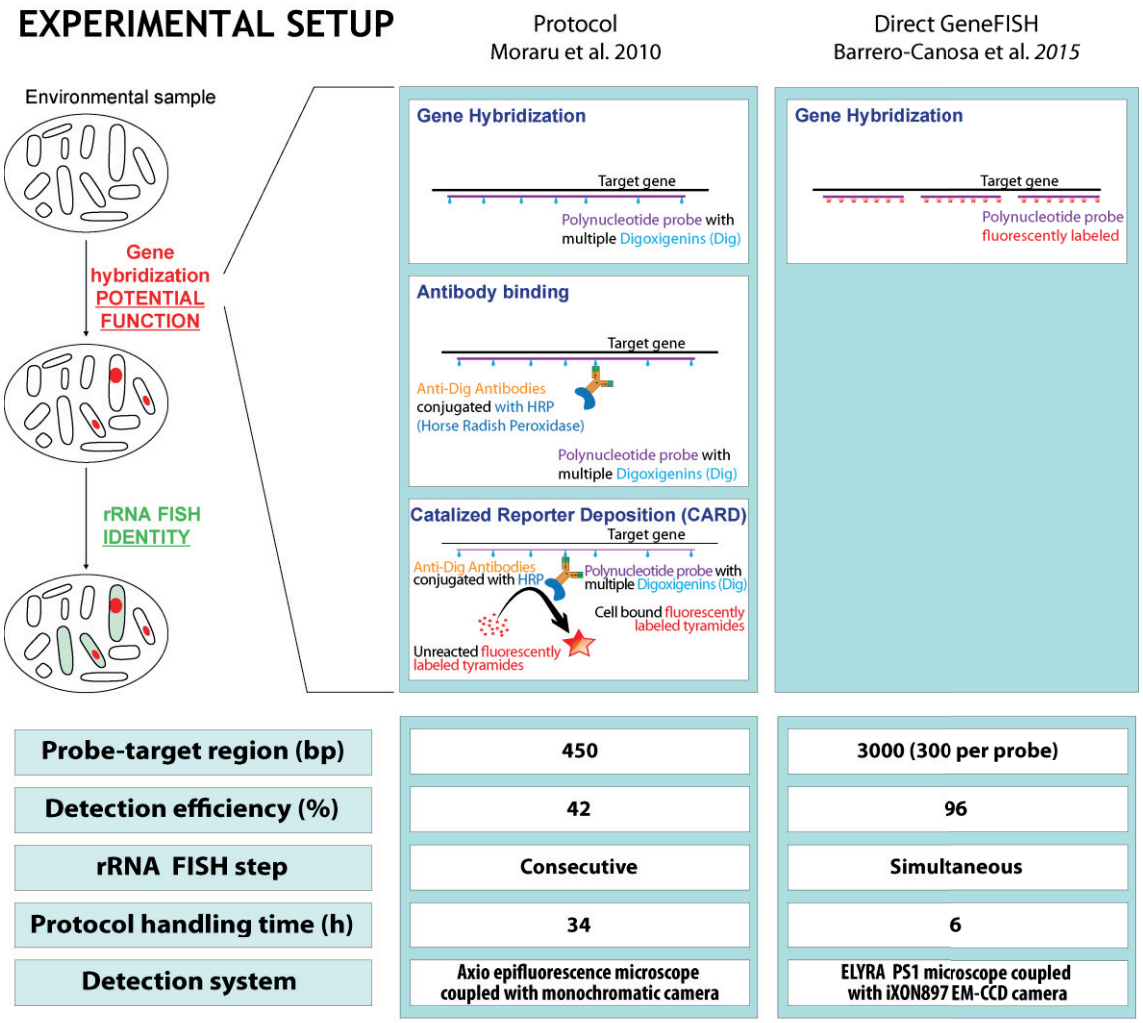

Figure 1. Evolution of GeneFISH protocol from the protocol of Moraru *et al.*, 2010 and the protocol developed in this thesis. The detection system refers to the microscope used for the detection of signals and for the estimation of the detection efficiency on pure cultures.

So far, direct-geneFISH was established and used on gammaproteobacterial cells (see Chapter II). Future efforts should focus on the adaptation of the method to other important bacterial groups in marine environments, like *Flavobacteria* and *Roseobacter* (Eilers *et al.*, 2000 Stepanauskas and Sieracki, 2007; Teeling *et al.*, 2012). Furthermore, while the simultaneous hybridization of gene and rRNA probes considerably shortens the protocol, it can limit its usage when the stringencies required for the hybridization of the gene and rRNA probes are very different. In such cases, the flexibility of the

protocol allows to combine CARD-FISH for rRNA detection followed by gene hybridization with dye-labeled probes.

DIRECT-GENEFISH AND SUPER-RESOLUTION MICROSCOPY

The developed direct-geneFISH protocol was combined with structured illumination microscopy (SIM). SIM was chosen among all super-resolution (SR) technologies since the sample preparation that it requires is the same as for wide-field fluorescence microscopy in contrast to other SR technologies, which have more specific requirements. For instance, PALM and STORM need specific dyes and buffers for photo-switching, and STED requires that the fluorochromes used are excited and depleted by spectrally separated lasers (Coltharp and Xiao, 2012). The combination with SIM allowed a more detailed localization of molecules than wide-field fluorescence microscopy; for instance, with SIM it was possible to visualize some plasmids presumably anchored to the membranes of *E.coli* cells (Figure 2). Nevertheless, SIM has a spatial resolution of 100 nm that limited the discernment of single gene signals (dot-like) when the targets were located within 100 nm apart, as observed in the cell of Figure 2 (See CHAPTER II for details). Future investigations could focus on increasing the resolution for single gene detection by using technologies such as STORM, which can resolve molecules that are ~20 nm apart (Huang *et al.*, 2009). However, this approach might have the tradeoff of loosing the simplicity of direct-geneFISH protocol for the following reasons: (i) the blinking behavior of common dyes still needs to be fully understood (Fernández-Suárez and Ting, 2008); (ii) due to the high sensitivity of the detection systems, any minor molecule drifting might be reflected in the image acquired; (iii) tuning the embedding medium for permitting the “blinking” of two different fluorescent-molecules in double hybridizations (genes and rRNA) might be challenging (Dempsey *et al.*, 2011).

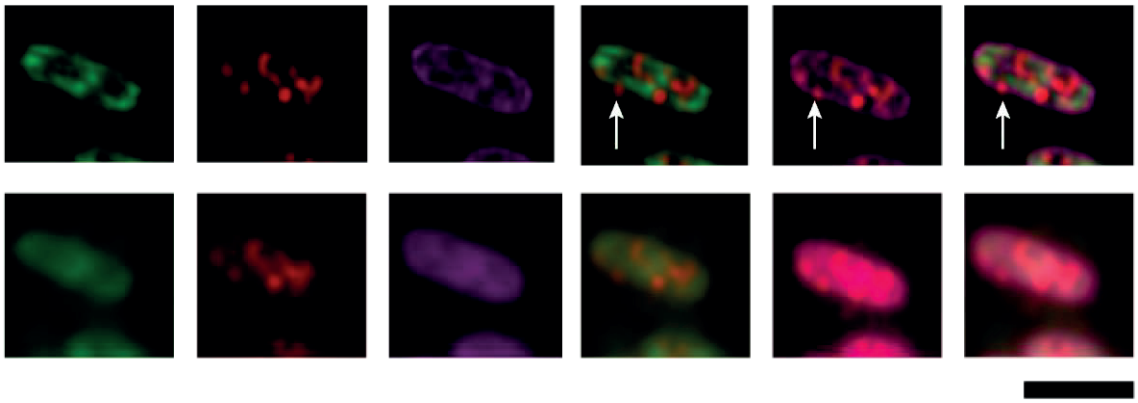

Figure 2. Top-line: SIM images of an *E.coli* cell harboring multiple plasmid (target) copies. Bottom-line: Corresponding wide-field image. Sample prepared with direct-geneFISH and membranes were stained with Nile Red. Columns from left to right: 1. rRNA-FISH/Cy5-EUB probe. 2 Genes (plasmid)/Alexa488-labeled probes. 3. Nile Red Staining. 4. Merged image of rRNA (green) + genes (red). 5. Merged image of Nile red (purple) + genes (red) . 6 Merged image of rRNA (green) + gene (red) + Nile red (purple). Pictures acquired with Elyra PS.1 microscope (Carl Zeiss, Jena, Germany). Black line-scale bar: ~2 µm. Note: During image acquisition of direct-geneFISH preparations on cells with multiple targets, it was noticed that some plasmids look like located outside the cell based on the rRNA signal (column 4), therefore for some preparations the membrane was stained with Nile Red (colum 5 and 6). The majority of the gene signals apparently "outside" the cell co-localized with the membrane stain.

DIRECT-GENEFISH: ENVIRONMENTAL MICROBES AND POTENTIAL APPLICATIONS

Direct-geneFISH is an attractive tool for microbial ecology because it is flexible, it is simple, it is faster and the samples are better preserved compared to CARD protocols. The potential of direct-geneFISH, as a tool for the study of environmental microbes, is very broad. Next, some examples of fields in which geneFISH shows promising potential, as well as further developmental steps will be mentioned.

DIRECT-GENEFISH AND SUPER RESOLUTION VISUALIZATION.

The visualization of cells shows the spatial organization of microbes *in situ* and, in combination with SIM, it improves the subcellular localization of genes and ribosomes. As an example, direct-geneFISH is a tool that might be used to show polyploidy and horizontal gene transfer in environmental microbes. Additionally, the visualization of target genes with 3D-SIM or CLSM can provide evidence for the direct genomic linkage of the metabolic dependence in syntrophic consortia, such as the anaerobic oxidation of methane coupled to sulfate reduction (Boetius *et al.*, 2000). The

protocol can also help to identify and differentially localize functional genes carried by bacteria living in symbioses with a eukaryotic host (Petersen *et al.*, 2012; Bernhard *et al.*, 2012). Furthermore, it can allow teasing apart symbiotic sub-populations that present heterogeneous genomic potential by targeting different functional genes (Ikuta *et al.*, 2015). For instance, recent metagenomic analysis of *B. puteoserpentis* endosymbiotic bacteria suggests that certain strains of the sulfur-oxidizer endosymbiont harbor a methanol dehydrogenase-coding gene. Thereof, these strains could potentially use the methanol released by the co-ocurring methane oxidizing symbiont as an energy source, reducing the competition for the available electron donors. In this respect, direct-geneFISH will permit the direct linkage between the methanol dehydrogenase-coding gene and sulfur-oxidizer strains differentially located in the host (Sayavedra *et al*., in prep).

DIRECT-GENEFISH TOWARDS MULTIPLE GENE DETECTION

The usage of directly dye-labeled probes will permit the simultaneous hybridization of multiple genes. The limitation in the number of dyes that can be simultaneously visualized can be overcome by using direct-geneFISH in combination with Combinatorial Labeling and Spectral Imaging - CLASI-FISH (Valm *et al.*, 2012) or Multi-Labeled approach – Mil-FISH (Schimak *et al.*, 2015). The *in situ* simultaneous detection of multiple genes can directly show if a bacterium harbors all the genes necessary to complete particular metabolic pathway, or if environmental microbes carry multiple genes with a similar function. As an example, it could detect if all *Formosa* strain B in the environment carry the four GH92 gene variations annotated in the genome of one isolated strain (CHAPTER III). Additionally, targeting multiple functional bacterial core genes enables "*in situ* strain genotyping" based on the average nucleotide identity (Konstantinidis *et al.*, 2006), allowing a better taxonomically resolved visualization of bacteria.

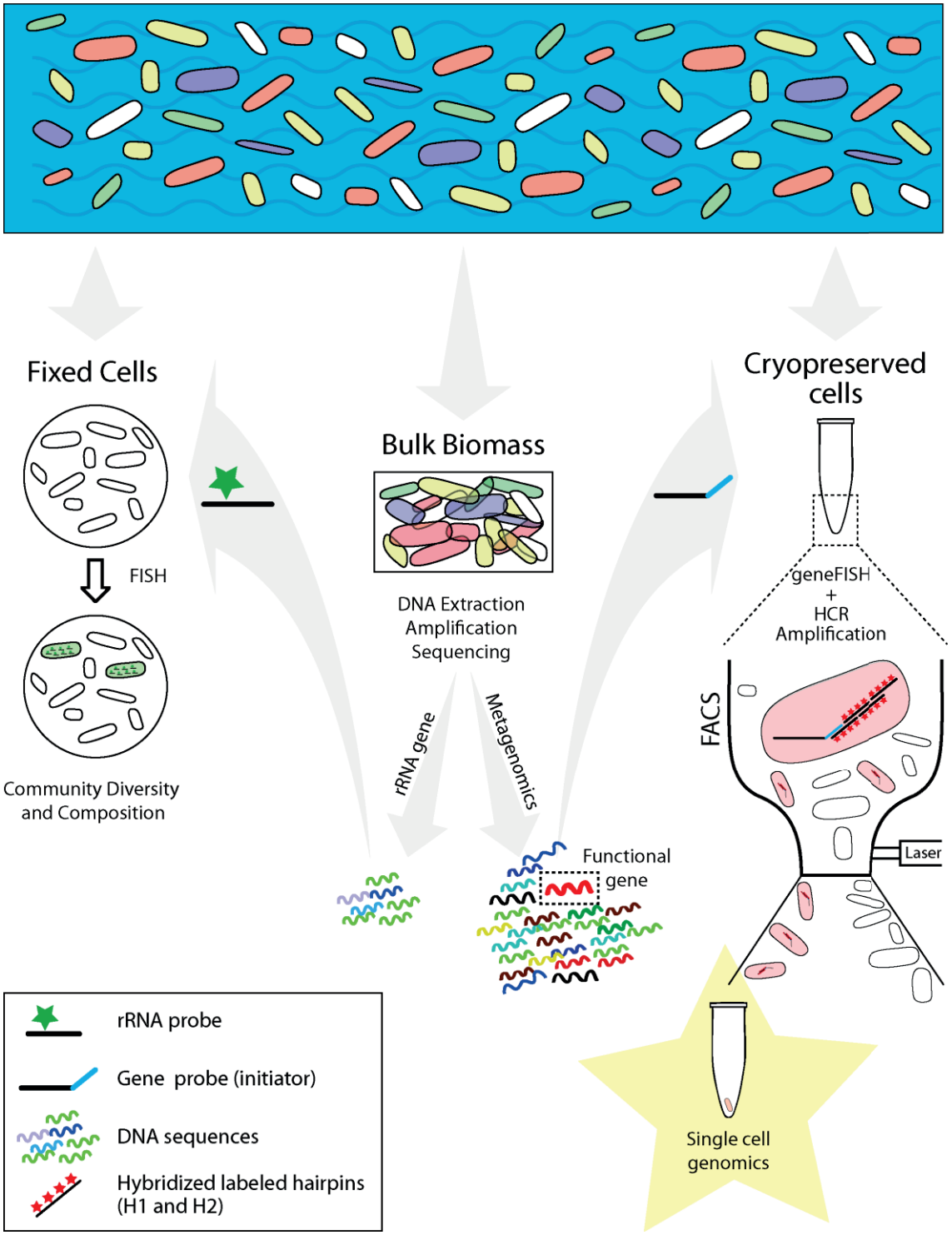

Figure 3. Full overview of modified protocol for using geneFISH as a tool for SCG. In which a environmental samples are collected for different culture independent techniques, including FISH and metagenomics. The implementation of SCG will required specific sampling, here by cryopreservation, as the common fixatives hinders SCG. For functional based single cell genomics, the target gene can be selected from the metagenomic data collected from the sampling place. The gene probe will need an initiator (in blue) that will allow a cascade of hybridization of two complementary hairpins (probes labeled with red star) (HCR is briefly explained in the introduction). The amplification of the gene signal will permit cell sorting with FACS for SCG

DIRECT-GENEFISH TOWARDS SINGLE CELL GENOMICS

In direct-geneFISH the DNA degradation decreases by removing CARD (Raap *et al.*, 1986). Therefore, direct-geneFISH combined with fluorescent activated cell sorting (FACS) can become a key approach for function-targeted single cell genomics (SCG) (Figure 3). The two immediate limitations are the low signal-to-noise ratio of single gene (Chapter II), and the fixation with aldehydes (Stepanauskas, 2012; Clingenpeel *et al.*, 2014). On one hand, the signal-to-noise ratio can be increased by HCR amplification: since the amplification sequences are repetitive and known (Yamaguchi, *et al.*, 2015a; Yamaguchi, *et al.*, 2015b), they can be easily tracked and filtered out of the assembly. On the other hand, for successful SCG amplification direct-geneFISH should be tested on unfixed cells (Yilmaz *et al.*, 2010) and the environmental samples can be cryo-preserved before FISH (Clingenpeel *et al.*, 2014). Alternatively, a modified metharcan fixative, that was proven to preserve nucleic acids (Cox *et al.*, 2006), could also be tested.

DIRECT-GENEFISH TOWARDS ENVIRONMENTAL PHAGE DETECTION

PhageFISH (Chapter I) is a variant of geneFISH in which the gene allows to identify and track the infection of a bacteriophage. Further application of direct-geneFISH for the phage detection will allow the absolute quantification of the phages based on the fluorescence intensity (Chapter II). Furthermore, the increased resolution with SIM permits to collect more information regarding the phage-host interaction. For instance, Figure 4 shows *Pseudoalteromonas* cells on an early stage of infection with PSA-HP1 phage, visualized with direct-geneFISH and SIM on a preliminary test. The super-resolution image shows the phages (in red) located in close proximity to the bacterial ribosomes (green), indicating that most likely the phages are taking advantage of the bacterial cellular machinery to produce their own proteins during this infection stage.

The usage of (direct or CARD) PhageFISH as a tool for the identification and quantification of environmental phages will require the development of an approach similar to the “full rRNA-cycle”. A potential target could be the ribonucleotides reductases (RNR) genes, which recently have been suggested as markers of viral

diversity in aquatic environments, due to their wide distribution in many viral taxa, including naturally abundant phages in marine environments (Sakowski *et al.*, 2014).

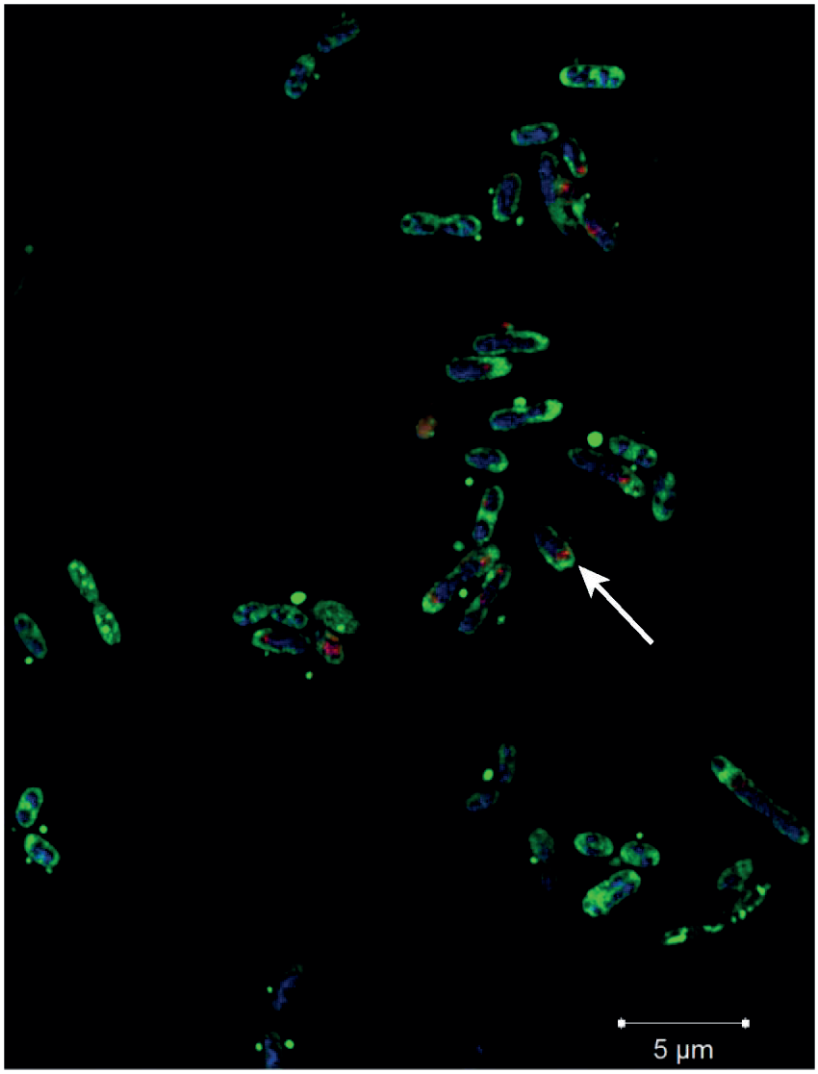

Figure 4. Direct-geneFISH on *Pseudoalteromonas* sp. cells, in green, on an early stage of infection with the PSA-HP1 phage, in red. In Blue- DAPI staining. Cells visualized with SIM of ELYRA PS1 Microscope (Carl Zeiss, Jena-Germany). The samples for direct-geneFISH were kindly provided by Dr. Elke Allers and Dr. Matthew Sullivan.

ADAPTATIONS TOWARDS mRNA DETECTION

The direct-geneFISH protocol could also be adapted for mRNA detection (dmRNA-FISH). Similar to direct-geneFISH, the advantages that this protocol could have in comparison to CARD-based mRNA-FISH (Pernthaler and Amann, 2004) is that

the target molecules are better preserved and, since rRNA and mRNA would be simultaneously hybridized, the leaking out of mRNA would decrease. The *in situ* mRNA detection would help to understand cellular regulation processes and partially infer cell activity on environmental microbes. dmRNA-FISH could for instance show *in situ* differential gene expression (Coleman *et al.*, 2007). For example, with respect to polysaccharide degraders bacteria, dmRNA-FISH could permit the localization mRNA from genes involved in polysaccharide degradation, for instance if they are preferentially expressed closed to the cell membranes, or to see differential gene expression of GHs during polysaccharide degradation processes.

For the adaptation of direct-geneFISH towards dmRNA-FISH, the following issues need to be considered: (i) the samples and all chemicals in use should be free of RNAses; (ii) the endogenous RNAses of bacteria need to be inactivated; (iii) the permeabilization step needs to be carefully considered as the mRNA can easily leak out (a second fixation step can be included to avoid the leaking out of hybridized molecules); (iv) the denaturation step of geneFISH has to be removed because mRNA is quickly degraded by heat; (v) only the antisense probe will hybridize: the reduction of the labels to half the number (since in geneFISH probes are double stranded) will be overcome by a higher number of target copies.

REFERENCES

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

Amann, R., 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.

Amann, R., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.

Beliveau, B.J., Boettiger, A.N., Avendaño, M.S., Jungmann, R., McCole, R.B., Joyce, E.F., C Kim-Kiselak, C., Bantignies, F., Fonseka, M., Erceg, C.Y., Hannan, M.A., Hoang, H., Colognori, D., Lee,J.T., .Shih, W. M., Yin,P., Zhuang, X., and Wu. C (2015) Single-molecule super-resolution imaging of chromosomes and *in situ* haplotype visualization using oligopaint FISH probes. *Nat Commun* **6**: 7147.

Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang ,Y., Lia, J.B., Senaratnea, T.N., Williams, B.R., Rouillard, J.M and Wua, J. (2012) Versatile design and synthesis platform for visualizing genomes with oligopaint FISH probes. *Proc Natl Acad Sci U.S.A.* **109**: 21301–21306.

Bernhard, J.M., Edgcomb, V.P., Casciotti, K.L., McIlvin, M.R., and Beaudoin, D.J. (2012) Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. *ISME Jl* **6**: 951–960.

Bidnenko, E., Mercier, C., Tremblay, J., Tailliez, P., and Kulakauskas, S. (1998) Estimation of the state of the bacterial cell wall by fluorescent *in situ* hybridization. *Appl Environ Microbiol* **64**: 3059–3062.

Bobrow, M.N., Harris, T.D., Shaughnessy, K.J., and Litt, G.J. (1989) Catalyzed reporter deposition, a novel method of signal amplification application to immunoassays. *J. Immunol Methods* **125**: 279–285.

Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jørgensen, B.B., Ursula Witte, U., and Pfannkuche, O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**: 623–626.

Choi, H.M.T., Beck, V.A., and Pierce, N.A. (2014) Next-Generation *in situ* hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano* **8**: 4284–4294.

Clingenpeel, S., Schwientek, P., Hugenholtz, P., and Woyke, T. (2014) Effects of sample treatments on genome recovery via single-cell genomics. *ISME J* **8**: 2546–2549.

Coleman, J.R., Culley, D.E., Chrisler, W.B., and Brockman, F.J. (2007) mRNAtargeted fluorescent in situ hybridization (FISH) of Gram-negative bacteria without template amplification or tyramide signal amplification. *J Microbiol Methods* **71**: 246–255.

Coltharp, C. and Xiao, J. (2012) Superresolution microscopy for microbiology. *Cell Microbiol* **14**: 1808–1818.

Cox, M.L., Schray, C.L., Luster, C.N., Stewart, Z.S., Korytko, P.J., M Khan, K.N., Paulauskis, J.D., and Dunstan, R.W. (2006) Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity. *Exp Mol Pathol* **80:** 183–191.

Davenport, R.J., Curtis, T.P., Goodfellow, M., Stainsby, F.M., and Bingley, M. (2000) Quantitative use of fluorescent *in situ* hybridization to examine relationships between mycolic acid-containing actinomycetes and foaming in activated sludge plants. *Appl Environ Microbiol* **66**: 1158–1166.

Dempsey, G.T., Vaughan, J.C., Chen, K.H., Bates, M., and Zhuang, X. (2011) Evaluation of fluorophores for optimal performance in localization-based super resolution imaging. *Nature Methods* **8**: 1027–1036.

Dirks, R.M. and Pierce, N.A. (2004) Triggered amplification by hybridization chain reaction. *Proc Natl Acad Sci U.S.A.* **101**: 15275–15278.

Eilers, H., Pernthaler, J., Glöckner, F.O., and Amann, R. (2000) Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl Environ Microbio* **66**: 3044–3051.

Fernández-Suárez, M. and Ting, A.Y. (2008) Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Bio* **9**: 929–943.

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat Rev Micro* **13**: 133–146.

Gahlmann, A. and Moerner, W.E. (2013) Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nat Rev Micro* **12**: 9–22.

Hell, S.W. and Wichmann, J. (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt Lett* **19**: 780–782.

Huang, B., Bates, M., and Zhuang, X. (2009) Super resolution fluorescence microscopy. *Annu Rev Biochem* **78**: 993–1016.

Ikuta, T., Takaki, Y., Nagai, Y., Shimamura, S., Tsuda, M., Kawagucci, S., Aoki, Y., Inoue,K., Teruya, M., Katou, K., Teruya, K., Shimoji, M., Tamotsu, H., Hirano, T.,Maruyama, T., and Yoshida, T (2015) Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *ISME J*. Advance online publication doi: 10.1038/ismej.2015.176

Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V. (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**: 587–590.

Jost, A. and Heintzmann, R. (2013) Superresolution Multidimensional imaging with structured illumination microscopy. *Annu Rev Mater Res.* **43**: 261–282.

Kawakami, S., Hasegawa, T., Imachi, H., Yamaguchi, T., Harada, H., Ohashi, A., and Kubota, K. (2012) Detection of single-copy functional genes in prokaryotic cells by two-pass TSA-FISH with polynucleotide probes. *J Microbiol Methods* **88**: 218–223.

Kolb, H.C., Finn, M.G., and Sharpless, K.B. (2001) Click Chemistry: diverse chemical function from a few good reactions. *Angew Chem Int Ed* **40:** 2004–2021

Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl Environ Microbiol* **72**: 7286–7293.

Kubo, K., Lloyd, K.G., F Biddle, J., Amann, R., Teske, A., and Knittel, K. (2012) Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J* **6**: 1949–1965.

Kwon, S. (2013) Single-molecule fluorescence *in situ* hybridization: Quantitative

imaging of single RNA molecules. *BMB Rep* **46**: 65–72.
Lenk, S., Moraru, C., Hahnke, S., Arnds, J., Richter, M., Kube, M., Reinhardt, R., Brinkhoff, T., Harder, J., Amann, R., and Mußmann, M. (2012) Roseobacter clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *ISME J* **6**: 2178–2187.
Lichtman, J.W. and Conchello, J.A. (2005) Fluorescence microscopy. *Nature Methods* **2**: 910–919.
Lloyd, K.G., May, M.K., Kevorkian, R.T., and Steen, A.D. (2013) Meta-analysis of quantification methods shows that archaea and bacteria have similar abundances in the subseafloor. *Appl Environ Microbiol* **79**: 7790–7799.
Loy, A., Maixner, F., Wagner, M., and Horn, M. (2007) probeBase--an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucl Acids Res.* **35**: D800–4.
Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüssmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., and Schleifer, K.H. (2004) ARB: a software environment for sequence data. *Nucl Acids Res* **32**: 1363–1371.
Maamar, H., Raj, A., and Dubnau, D. (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317**: 526–529.
Matturro, B. and Rossetti, S. (2015) GeneCARD-FISH: detection of *tceA* and *vcrA* reductive dehalogenase genes in *Dehalococcoides mccartyi* by fluorescence *in situ* hybridization. *J Microbiol Methods* **110**: 27–32.
Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature* **466**: 77–81.
Moraru, C. and Amann, R. (2012) Crystal ball: fluorescence in situ hybridization in the age of super-resolution microscopy. *Syst Appl Microbiol* **35**: 549–552.
Moraru, C., Lam, P., Fuchs, B.M., Kuypers, M.M.M., and Amann, R. (2010) GeneFISH - an in situ technique for linking gene presence and cell identity in environmental microorganisms. *Environ Microbiol* **12**: 3057–3073.
Moraru, C., Moraru, G., Fuchs, B.M., and Amann, R. (2011) Concepts and software for a rational design of polynucleotide probes. *Environ Microbiol Rep* **3**: 69–78.
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.
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.
Nikolakakis, K., Lehnert, E., McFall-Ngai, M.J., and Ruby, E.G. (2015) Using hybridization chain-reaction fluorescent *in situ* hybridization (HCR-FISH) to track gene expression by both partners during initiation of symbiosis. *Appl Environ Microbiol* **81**: AEM.00890–15–4735.
O'connor, C. (2008) Fluorescence in situ hybridization (FISH). *Nat Educ* **1:** 171.
Pathak, S., Choi, S.K., Arnheim, N.A., and Thompson, M.E (2001) Hydroxylated quantum dots as luminescent probes for *in situ* hybridization. *J Am Chem Soc* **17**: 4103–4104.
Pawley, J. (2010) Handbook of Biological Confocal Microscopy Pawley,J.B.(ed) Springer Science & Business Media, Boston, MA.

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

Pernthaler, A. and Amann, R. (2004) Simultaneous fluorescence *in situ* hybridization of mRNA and rRNA in environmental bacteria. *Appl Environ Microbiol 7*0: 5426–5433.

Petersen, J.M., Zielinski, F.U., Pape, T., Seifert, R., Moraru, C., Amann, R., Hourdez, S., Girguis, P., Wankel, S.D., Barbe, V., Pelletier, E., Fink, D., Borowski, C., Bach, W., and Dubilier, N. (2012) Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176–180.

Raap, A.K., Marijnen, J.G.J., Vrolijk, J., and Van der Ploeg, M. (1986) Denaturation, renaturation, and loss of DNA during *in situ* hybridization procedures. *Cyt* **7**: 235–242.

Resch-Genger, U., Grabolle, M., and Cavaliere-Jaricot, S. (2008) Quantum dots versus organic dyes as fluorescent labels. *Nature Methods* **5**:763 - 775

Sakowski, E.G., Munsell, E.V., Hyatt, M., Kress, W., Williamson, S.J., Nasko, D.J., Shawn W Polson, S.W., and Wommack, K.E. (2014) Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. *Proc Natl Acad Sci U.S.A*. **111:** 15786–15791.

Schermelleh, L., Heintzmann, R., and Leonhardt, H. (2010) A guide to super-resolution fluorescence microscopy. *J Cell Biol.* **190**: 165–175.

Schimak, M.P., Kleiner, M., Wetzel, S., Liebeke, M., Dubilier, N., and Fuchs, B.M. (2015) MiL-FISH: Multi-labelled oligonucleotides for fluorescence in situ hybridisation improve visualization of bacterial cells. *Appl Environ Microbiol* Early view. Doi: 10.1128/AEM.02776-15

Sekar, R., Pernthaler, A., Pernthaler, J., Warnecke, F., Posch, T., and Amann, R. (2003) An improved protocol for quantification of freshwater Actinobacteria by fluorescence *in situ* hybridization. *Appl Environ Microbiol* **69**: 2928–2935.

Skinner, S.O., Sepúlveda, L.A., Xu, H., and Golding, I. (2013) Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent *in situ* hybridization. *Nat Protoc* **8**: 1100–1113.

Stagars, M., Ruff, E., Amann, R., Knittel, K. High diversity of anaerobic alkane-degrading microbial communities in marine seep sediments based on their (1-methylalkyl)succinate synthase genes. Manuscript submitted for publication in *Front microbiol*

Stepanauskas, R. and Sieracki, M.E. (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *PNAS* **104**: 9052–9057.

Stepanauskas, R. (2012) Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* 15: 613–620.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., Kassabgy, M., Huang, S., Mann, A.J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F.D., Callies, U., Gerdts, G., Wichels, A., Wiltshire, K.H., Glöckner, F.O., Schweder, T. and Amann, R. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611

Teeling, H. and Glöckner, F.O. (2012) Current opportunities and challenges in microbial metagenome analysis—a bioinformatic perspective. *Brief Bioinform* **13**: bbs039–742.

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.

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

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

Wu, S.-M., Tian, Z.-Q., Zhang, Z.-L., Huang, B.-H., Jiang, P., Xie, Z.-X., and Pang, D.-W. (2010) Direct fluorescence *in situ* hybridization (FISH) in *Escherichia coli* with a target-specific quantum dot-based molecular beacon. *Biosens Bioeletron* **26**: 491–496.

Valm, A.M., Mark Welch, J.L., and Borisy, G.G. (2012) CLASI-FISH: Principles of combinatorial labeling and spectral imaging. *Syst Appl Microbio*l **35:** 496–502

Yamaguchi, T., Fuchs, B.M., Amann, R., Kawakami, S., Kubota, K., Hatamoto, M., and Yamaguchi, T. (2015b) Rapid and sensitive identification of marine bacteria by an improved in situ DNA hybridization chain reaction (quickHCR-FISH). *Syst Appl Microbiol* **38**: 400–405.

Yamaguchi, T., Kawakami, S., Hatamoto, M., Imachi, H., Takahashi, M., Araki, N., et al. (2015a) *In situ* DNA-hybridization chain reaction (HCR): a facilitated in situ HCR system for the detection of environmental microorganisms. *Environ Microbiol* **17**: 2532–2541.

Yilmaz, S., Haroon, M.F., Rabkin, B.A., Tyson, G.W., and Hugenholtz, P. (2010) Fixation-free fluorescence *in situ* hybridization for targeted enrichment of microbial populations. *ISME J* **4:** 1352–1356

ACKNOWLEDGEMENTS

I want to acknowledge all the people that in one way or another were involved in the development of this thesis. First I want to thank my supervisors, Rudi Amann, for the support and guidance during my PhD. To Bernhard Fuchs for all the patience, the good advices and for always being willing to help me with my doubts and to Cristina Moraru, for all the trust, the interesting discussions and the support.

I would like to acknowledge, Prof. Dr. Rudolf Amann, Prof. Dr. Andreas Schramm, Prof. Dr. Michael Friedrich, Dr. Cristina Moraru, David Probandt and Dolma Michellod for being part of my thesis defense committee.

I want to thank the Max Planck Institute for Marine Microbiology and my colleagues for the nice and friendly working environment. I also want to thank to all the nice people I met in there, many of which, I consider my good friends. I am very grateful with Jennifer Bachmann for helping me with the German translations of this thesis and to my friends Nadine Lehnen, Oliver Jäckle and Greta Reintjes for all the times that have helped me with German translations during these years. I also want to thank my friends, Sara Cregeen, Clara Martinez Perez, Cameron Callbeck and Lizbeth Sayavedra for their comments and feedback related to this thesis.

I also want to thank my Belgian family for all their support and warmness for so many years, thanks for always made me feel at home, I am very fortunate to have two families!

I want to thank to Giorgito for the unconditional support, the patience (specially while I was finishing this thesis), for all the help and for the many times you made me smile while I was stressed, you are simply amazing, I love you a nanometer.

A mis amigos y primos en Colombia que los extraño y los llevo siempre en mi corazón, gracias por siempre darme la energía y por estar siempre presentes a pesar de la distancia y las ocupaciones.

Finalmente, quiero agradecer a mi familia que son lo más importante para mi y quienes me han dado toda la fuerza para seguir adelante en los momentos difíciles. Gracias por el apoyo incondicional, la confianza tan grande, por todo el amor y todas las oportunidades que me han dado. Son mi mayor y más grande bendición, todos mis logros se los debo a ustedes.

Universität Bremen

Name: Jimena Barrero Canosa

Ort, Datum: Bremen, 10.11.2015

Anschrift: Georg-Gröning Strasse 28, 28209, Bremen

ERKLÄRUNG

Hiermit erkläre ich, dass ich die Doktorarbeit mit dem Titel:

.

Systematic optimization of the detection of single genes in microorganisms by Fluorescence *in situ* Hybridization (geneFISH)

selbstständig verfasst und geschrieben habe und außer den angegebenen Quellen keine weiteren Hilfsmittel verwendet habe.

Ebenfalls erkläre ich hiermit, dass es sich bei den von mir abgegebenen Arbeiten um drei identische Exemplare handelt.

_____

(Unterschrift)