Microbial resource management: introducing new tools and ecological theories

ir. Karen DE ROY

UNIVERSITEIT GENT

FACULTEIT BIO-INGENIEURSWETENSCHAPPEN

Promotor

Prof. dr. ir. Nico Boon
Department of Biochemical and Microbial Technology, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Co-promotor

Prof. dr. ir. Olivier Thas
Department of Mathematical modelling, Statistics and Bio-informatics, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Members of the examination committee

Prof. dr. Daniele Daffonchio
Department of Food, Environmental and Nutritional Sciences (DeFENS), University of Milan, Milan, Italy

Dr. ir. Tom Defoirdt
Department of Animal Production, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Prof. dr. David Gillan
Department of Proteomics and Microbiology (PROTMIC), Faculty of Science, Mons University, Mons, Belgium

Prof. dr. Hauke Harms
Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

Prof. dr. ir. Monica Höfte   *(Secretary of the committee)*
Department of Crop Protection, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Prof. dr. Colin Janssen
Department of Applied Ecology and Environmental Biology, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Prof. dr. Els Van Damme   *(President of the committee)*
Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Dean Faculty of Bioscience Engineering

Prof. dr. ir. Guido Van Huylenbroeck

Rector Ghent University

Prof. dr. Anne De Paepe

Microbial resource management: introducing new tools and ecological theories

Ir. Karen De Roy

Thesis submitted in fulfillment of the requirements for the degree of Doctor (PhD) in Applied Biological Sciences

Titel van het doctoraat in het Nederlands: Microbial resource management: de introductie van nieuwe methoden en ecologische theorieën.

Cover illustration by Tim Lacoere

AIC	Akaike's Information Criterium
CC	Cell Count
DGGE	Denaturating Gradient Gel Electrophoresis
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediamine Tetracetic Acid
FC	Flow Cytometry
FDA	Fisher Discriminant Analysis
FSC	Forward Scatter
GFP	Green Fluorescent Protein
HPC	Heterotrophic Plate Count
MRM	Microbial Resource Management
OD	Optical Density
PB	Probabilistic Binning
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PI	Propidium Iodide
RNA	Ribonucleic Acid
SG	SYBR Green
SSC	Sideward Scatter
TGGE	Temperature Gradient Gel Electrophoresis
TOC	Total Organic Carbon
T-RFLP	Terminal Restriction Fragment Length Polymorphism
TSB	Trypticase Soy Broth
YFP	Yellow Fluorescent Protein

TABLE OF CONTENTS

CHAPTER

1

LITERATURE REVIEW ON MICROBIAL ECOLOGY

LITERATURE REVIEW ON MICROBIAL ECOLOGY

Chapter partially redrafted after:

- **De Roy K**, Marzorati M, Van den Abbeele P, Van de Wiele T & Boon N (2013) Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environmental Microbiology (in press)*

- Wang Y, Hammes F, **De Roy K**, Verstraete W & Boon N (2010) Past, present and future applications of flow cytometry in aquatic microbiology. *Trends in Biotechnology* **28**: 416-424

1.   Introduction

Microorganisms are ubiquitous on earth, with an estimated amount of 10^6 bacterial species (Lopez-Garcia & Moreira, 2008) and 4×10^{30} microbial cells globally (Horner-Devine *et al.*, 2004). Their genetic and physiological diversity result in an enormous metabolic potential. They contribute to nearly all biogeochemical cycles as they are the drivers of global and local nitrogen, oxygen, carbon, sulphur and phosphorus cycles (Schmidt, 2006), what makes them essential for maintaining the earth's biosphere and for the survival of plants and animals.

Most of these processes are accomplished by joint effort of microorganisms with different functional roles. These microorganisms do not act as individuals, but rather as a dynamically changing microbial community, where all cells interact and communicate with one another (Little *et al.*, 2008; Klitgord & Segre, 2010). They influence each other's behaviour and possibly alter the biochemical phenotypes of the participating strains (Wintermute & Silver, 2010).

Understanding the factors that shape and influence these microbial ecosystems is essential from a microbiological, ecological and biotechnological point of view. According to Prosser *et al*. (2007), this knowledge can be achieved by using a theory driven approach: theories are

generated based on existing observational data, after which they are verified using quantitative research. A deliberate choice of the experimental setup, methodology and microbial model systems is indispensable for optimal hypothesis testing. Pure cultures and complex microbial communities are conventionally used, however synthetic ecosystems with intermediate complexity and high controllability are becoming increasingly popular.

The knowledge on the behaviour of microbial communities will allow us to predict and possibly counteract the negative effects caused by environmental changes like global climate change. Furthermore, this knowledge can be used to better understand ecological processes and steer microbial communities in biotechnological applications, this concept is better know as "Microbial Resource Management" or MRM (Verstraete *et al.*, 2007; Read *et al.*, 2011).

2.   Microbial resource management

Microorganisms that are working together in a microbial community require management as much as humans do that are working together in an organisation or company. Similar to human resource management, we could also introduce microbial resource management (MRM) for microorganisms. This MRM was suggested the first time by Verstraete *et al.* (2007). The key purpose is to control and steer microbial communities and microbial processes in different environments to improve the environment, human health and biotechnological applications in the most sustainable way (Read *et al.*, 2011). In order to manage a microbial community, different questions need to be answered: Who is present? How many of them are there? Are they equally abundant? Who is interacting with whom? Who is doing what? etc. Different microbiological techniques and molecular tools exist to analyse microbial communities (see Chapter 1, section 4) and different methodologies and experimental procedures have been developed to get an understanding of microbial communities (see Chapter 1, section 3). In order to deal with all this information and to allow interpretation of the obtained data, a set of MRM tools was developed by Marzorati *et al.* (2008). These MRM parameters are range-weighted richness (Rr), dynamics (Dy) and community organisation (Co). They were initially developed for data obtained from molecular fingerprinting techniques like DGGE, but the tools have been extended to other techniques like pyrosequencing (Read *et al.*, 2011; Marzorati *et al.*, 2013).

2.1.   Range-weighted richness

The range-weighted richness (Rr) gives an indication of the genetic diversity and the species richness of an ecosystem. For DGGE, which uses a denaturating gradient to separate DNA on a gel according to its GC-content, the number of bands (N) corresponds to the number of species or species richness. The denaturating gradient (D_g) that is required to separate the DNA is used to describe the genetic diversity. The Rr is calculated as follows:

$$Rr = Species\ richness^2 \times Genetic\ diversity$$
$$= N^2 \times D_g$$

High Rr-values indicate an environment with a high carrying capacity, this means that the environment is able to host several genetically diverse species (Marzorati *et al.*, 2008).

Even though Rr gives information on both species richness and genetic diversity, it is often only species richness that has been studied. The relation between richness and the invasion of non-native species has been studied in disparate terrestrial and aquatic environments, however there is no consensus on the this relation, both positive and negative correlations have been described (Stachowicz *et al.*, 1999; Jiang & Morin, 2004; Simberloff *et al.*, 2012). Next to invasion, also the relation between community functionality and species richness has been studied in different environments. Bell *et al.* (2005) used synthetic microbial ecosystems to show a positive relation between species richness and community functionality. A richness of 18 species result in high functionality, the addition of additional species only had a minor effect on community functionality.

2.2.   Dynamics

The second MRM parameter, community dynamics, is used to determine how much a community changes over fixed time intervals. It is visualised by a moving window analysis (Figure 1-1). This shows the percentage of change within a microbial community between consecutive time points (Marzorati *et al.*, 2008). This concept can be interpreted as the number species that on average come to dominance (Read *et al.*, 2011). Communities with low dynamics are regarded as closed, as they do not allow other species to become dominant. While in a highly dynamical community a lot of species can become dominant or leave the community.

A. Dynamics

B. Community organisation
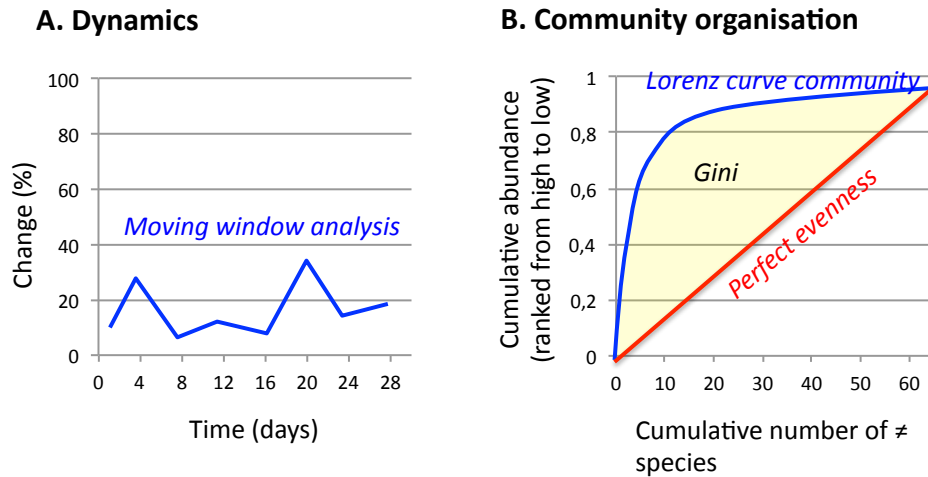
Figure 1-1. Visualisation of the dynamics (A) and community organisation (B). A. The dynamics is visualised by a moving window analysis. This shows the percentage of change of a microbial community between two consecutive time points. B. The community organisation is visualised by a Lorenz curve and the corresponding Gini coefficient. The Lorenz curve is constructed by plotting the cumulative number of species on the x-axis and the cumulative relative abundance on the y-axis. The surface between the perfect evenness line (red line) and the Lorenz curve (blue line) corresponds to the Gini coefficient and gives an indication of the community evenness.

2.3.   Community organisation

The third MRM parameter is community organisation (Co) or community evenness and is a measure of the organisation and structure within a microbial community. It describes the difference between the relative abundance of different species and can be visualised by Lorenz curves (Lorenz, 1905; Mertens *et al.*, 2005). These Lorenz curves are constructed by plotting the cumulative relative abundance of species, ranked from high to low abundance, on the y-axis and the cumulative number of species on the x-axis (Figure 1-1). For DGGE, the relative abundance of species is obtained from the relative band intensities on the DGGE-profile (Marzorati *et al.*, 2008). Lorenz curves deviating more from the diagonal, which is the complete evenness line, indicate more uneven structured microbial communities. In this case, a small number of species are dominant, while many others are resilient. This evenness is quantified by the Gini coefficient and is the surface between the Lorenz curve and the perfect evenness line. As such, high Gini coefficients correspond to uneven communities and low Gini coefficient to even communities.

Community evenness has been shown to be a key factor in preserving the functional stability of an ecosystem. By the use of more than 1000 synthetic ecosystems, it was shown that uneven communities are less resistant to stress, resulting in a decreased community functionality, compared to even communities (Wittebolle *et al.*, 2009).

3. Understanding microbial communities

3.1. Pure cultures

Culture-dependent methods allow the isolation of single microbial community members for in-depth analysis of their genetic and physiological characteristics. The body of literature on research with single microorganisms is tremendous (Jessup *et al.*, 2005). While such single microorganisms are the simplest microbial experimental systems in ecology, they are still not fully understood. Since the –omics era, a lot of knowledge on these simple model systems is gained. Over 4000 complete microbial genomes have been sequenced, while more than 12000 are in progress (www.genomesonline.org). Transcriptomics, proteomics and metabolomics gave further insight into their functionality, resistance to stress and adaptation. This increased understanding on how microorganisms function, led to the urge to steer and manipulate them. Synthetic biology, which is the application of engineering methodology to biology, was proven to be very useful (Endy, 2005; Leonard *et al.*, 2008). Microorganisms have been engineered to improve their resistance to stress, to have a higher productivity, to degrade toxic and recalcitrant compounds, to synthesize new chemical compounds or to have other particular – unnatural - characteristics (Benner & Sismour, 2005). The numerous capacities of both genetically engineered and wild-type microorganisms make them interesting for different applications. They are used as probiotics in the medical and food industry (Steidler *et al.*, 2000; Huibregtse *et al.*, 2012), as cell factories for valuable products in the food, pharmaceutical, chemical and agriculture industry, with products ranging from anticancer drugs to biofuels (Du *et al.*, 2011; Waegeman & Soetaert, 2011).

The fact that (i) only a small fraction of the microorganisms present in a microbial community can be cultured and (ii) the behaviour of microorganisms as pure cultures is different from their behaviour in a microbial community has caused a shift from single-organism studies to whole community studies.

3.2. Top-down approaches

Molecular fingerprinting and high-throughput sequencing techniques are used to characterise these microbial communities. These techniques use a top-down approach and target microbial communities as a whole. Metagenomics, metatranscriptomics and metaproteomics give information on the taxonomic and functional diversity, the population structure, the presence of genes as well as their expression levels and levels of translation into proteins

(VerBerkmoes *et al.*, 2009; Temperton & Giovannoni, 2012). A drawback is the complex post-processing of the big amount of data obtained by these high-throughput techniques (Raes & Bork, 2008). Even with the most advanced bioinformatics tools and sequencing technology, it is almost impossible to assign the (expressed) genes and proteins, and thus the functionality, to specific species (Temperton & Giovannoni, 2012; Zengler & Palsson, 2012). Furthermore, it is not possible to fully map and understand the microbial interactions, which are often the driving force of a community. All this makes that microbial community research mainly encompasses observational studies, while for more fundamental studies, like studying metabolic interactions, less complex systems are required.

3.3.   Synthetic ecosystems

Compared to the amount of literature available on single organisms and complex microbial communities, only a small fraction of microbial ecology research makes use of synthetic microbial communities. Synthetic microbial ecology is a collective noun for all assembled ecosystems that are created by a bottom-up approach where two or more defined microbial populations are assembled in a well-characterised and controlled environment (Figure 1-2). These synthetic ecosystems have a lower complexity, higher controllability, higher reproducibility and are a simplified representation or simulation of natural ecosystems. Synthetic ecosystems are used (i) to gain insight in fundamental principles such as metabolic processes, interactions, networking, diversity-functionality relation and nutrient cycling and (ii) to create interactions and communities with desired characteristics and functionality. Alternative terms for similar experimental setups are microcosms or artificial ecosystem, while other terms have been mistakenly used for synthetic ecology: (i) synthetic biology, which is the engineering of cells and (ii) systems biology, which considers the use of a top-down approach to understand a system by characterising the different parts.

3.3.1.   Ecological relevance

While a microbial community as such is already complex, numerous environmental factors further increase the level of complexity (Figure 1-3). Microorganisms live in close contact with each other as they continuously **interact** and **communicate (A)** with one another (Little *et al.*, 2008; Klitgord & Segre, 2010). These interactions may be unidirectional or bidirectional (West *et al.*, 2006). Molecules are produced that can be beneficial or detrimental for both the actor and recipient. Different kinds of interactions and cooperation are present in

nature: mutualism, syntrophy or cross-feeding (beneficial to the actor/beneficial to the recipient; +/+), selfishness or parasitism (beneficial to the actor/costly to the recipient; +/-), spite (-/-) and altruism (-/+) (West *et al.*, 2007a; Faust & Raes, 2012). Microorganisms can **communicate** with one another through mechanisms like quorum sensing, which allow them to express certain genes under favorable conditions in a coordinated manner (Manefield & Turner, 2002). Next to the abundant microorganisms that actively contribute to the **functionality** of the ecosystems, numerous species are present in lower abundance. They are regularly categorized as redundant and are responsible for the **resilience (E)** of the community (Bissett *et al.*, 2013). **Abiotic factors (C)** like temperature, salinity and pH can alter the environment in such a way that most of the abundant microorganisms cannot perform their role in the community anymore (Wu & Conrad, 2001; Sharma *et al.*, 2006). Under these circumstances, redundant species can take over and guarantee the ecosystem functionality. The resilience of a community is thus also strongly dependent on the **community diversity (B)** (Loreau *et al.*, 2001). Both the number of microorganisms (richness) and their relative abundance (evenness) influence the resistance to stress, invasion and predation (Wittebolle *et al.*, 2009; Saleem *et al.*, 2012; De Roy *et al.*, 2013). Next to the microbial diversity, also the **spatial organisation (F)** as it exists in a biofilm, can be of importance (Tolker-Nielsen & Molin, 2000). It allows only those species that are located in close proximity to interact and communicate with each other; furthermore, it provides microenvironments and niches for specific microbes.

CHAPTER 1
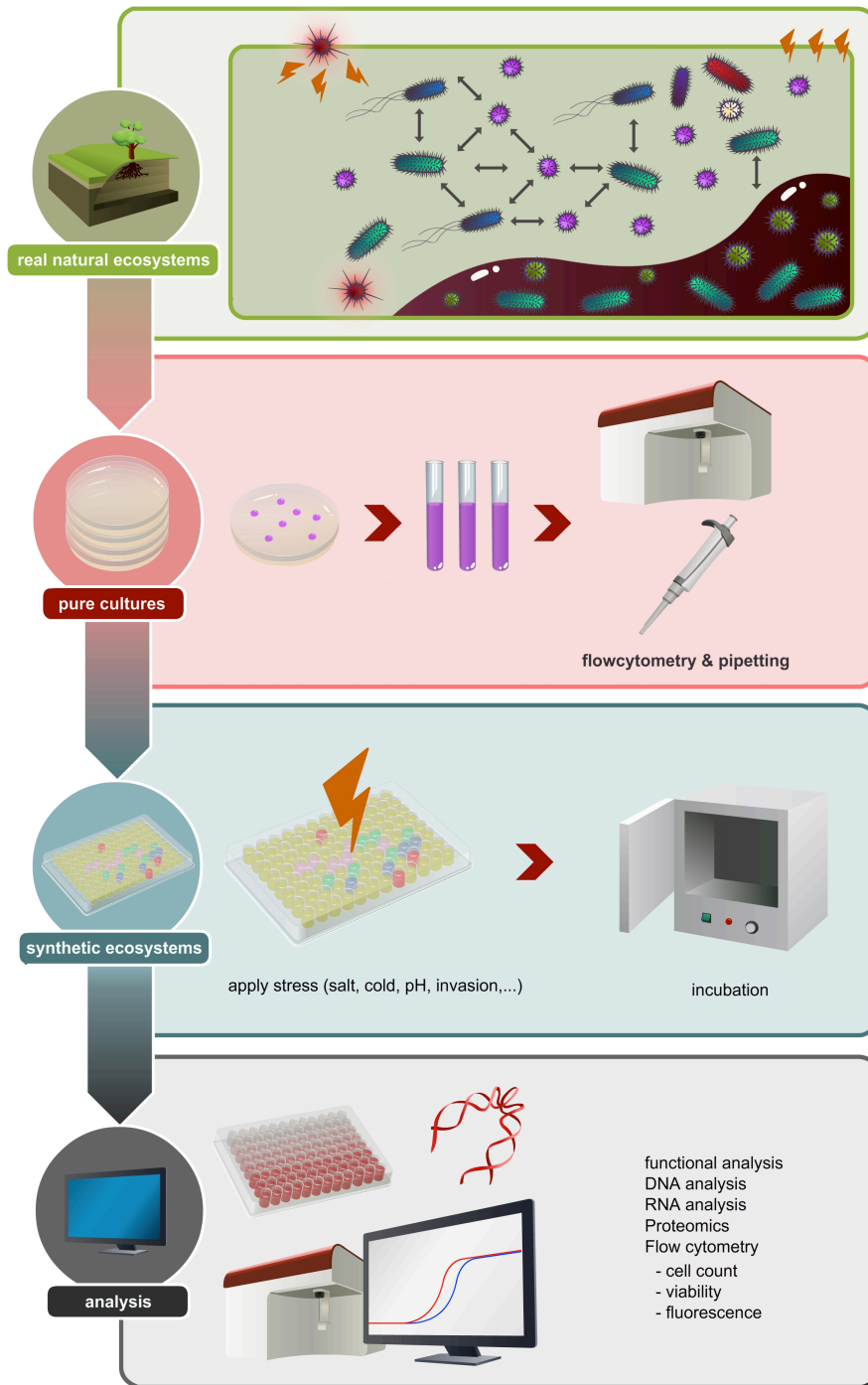
Figure 1-2. Flow chart of how to create synthetic ecosystems. *Synthetic communities are created by a bottom-up approach, this means one bacterial species at a time. For this, microorganisms first get isolated from their natural environment and grown in liquid medium. Subsequently they get measured by flow cytometry and diluted to the desired cell numbers. Synthetic communities are created by mixing the microbial species in correct proportions, the desired conditions are applied and the synthetic ecosystems get incubated. Eventually, all parameters of interest, like functionality and cell count, get analysed.*

CHAPTER 1
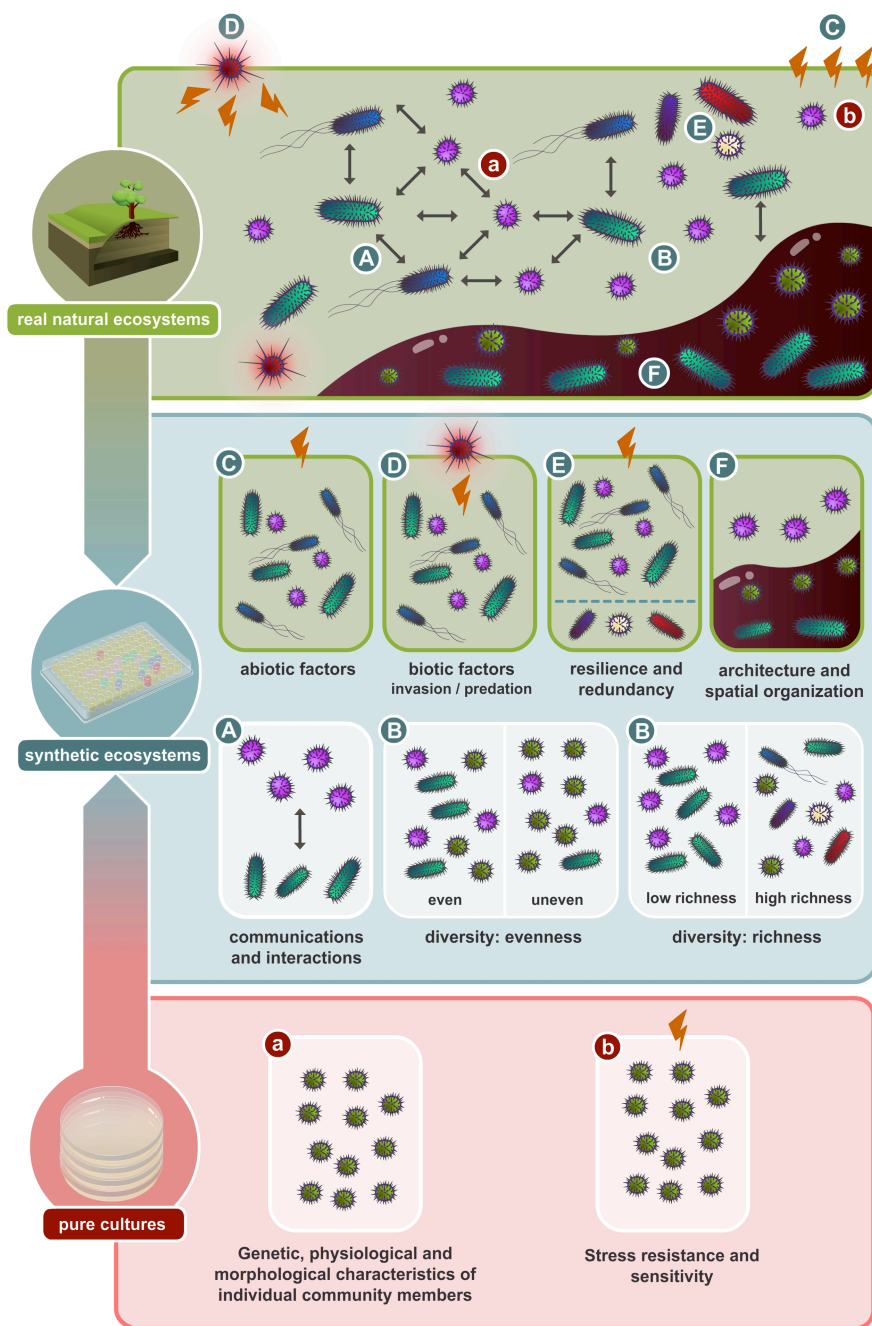
Figure 1-3. Synthetic ecosystems for research purposes. *Natural ecosystems are complex: different factor that influence and shape microbial communities are present and cannot be controlled. These factors are: A) Metabolic interactions, signalling and communication B) Diversity, C) Abiotic or environmental factors, D) Biotic factors like invasion and predation, E) Resilience and redundancy and F) Architecture and spatial organisation. Research with pure cultures provides a lot of information on individual population's genetic, physiological and morphological characteristics (a) and resistance and sensitivity to stress (b). However, they do not allow researchers to investigate the factors shaping and influencing microbial communities. For this, synthetic ecosystems are a powerful tool, they have a reduced complexity and higher controllability compared with natural ecosystems. They allow focusing on one of the parameters of interest while excluding other influencing factors.*

All these factors shape, characterise and influence an ecosystem and its functioning. By interfering with one of these parameters, a complete ecosystem might collapse. However, also the opposite might happen as an ecosystem may perform better or new functions can be introduced. By doing research and gaining knowledge on these fundamental principles, it will become possible to steer, manage and create ecosystems to optimise their performance.

In situ or *in vivo* models are complex systems in which nearly all of the above-mentioned influencing factors are present, thus giving a good representation of the real situation. The complexity of the microbiota in these systems is useful for the validation of different products or treatments, but may also be a confounding factor for research purposes, as most of the influencing factors cannot be controlled. Intrinsic system effects and reciprocal interactions may even lead to opposite conclusions on the role of a specific parameter in closely related ecosystems (Wilsey & Polley, 2002; Emery & Gross, 2007). For this reason, synthetic ecosystems are a powerful tool to investigate fundamental principles in natural and engineered systems. They limit the influencing factors to a minimum, allowing their management and tracking of the effects of the above-mentioned parameters. Furthermore, fully characterised microorganisms with a well-defined genetic background can be used in synthetic ecosystems. In the following paragraphs, we provide several examples of how synthetic microbial ecosystems have been used to study the role of specific influencing factors.

3.3.2. Ecological theory testing

The first synthetic ecosystems were used to study microbial **interactions and signalling**, as reviewed by Yu *et al.* (2012). For this type of research, communities mainly consist of only two or three microbial species, which are often also being genetically engineered to create the interaction of interest or to simplify tracking of the parameters of interest. In this way, hypotheses can be tested that would otherwise not be accessible (Wintermute & Silver, 2011). Next to creating an interacting community by genetically engineering the organisms, Klitgord and Segre (2010) showed it is also possible to create interactions by changing the environment: for every two species-consortium, a cooperation-inducing environment could be identified. **Environmental factors**, like the availability of nutrients, temperature, presence of toxic compounds and oxygen-level not only influence microbial interactions, but also influence the **resilience** of a community, which on its turn is influenced by the microbial

diversity. To get insight in the biodiversity-productivity relationship along different kinds of stress, researchers also opted for synthetic microbial ecosystem experiments. This allows controlling the evenness and richness, the applied stress and the follow up of the functionality, which is not possible in natural environments. Doing so, Wittebolle and coworkers investigated the effect of community evenness on the functionality of a denitrifying bacterial community in the presence and absence of salinity stress. They created over 1000 synthetic ecosystems in 96-well plates with the same 18 denitrifying strains, but with different levels of initial evenness. It was concluded that highly uneven communities (low biodiversity) are less resistant to environmental stress than even communities (high biodiversity). The latter could better retain their functionality under stress conditions (Wittebolle *et al.*, 2009). In another study regarding the effect of richness on resistance to cadmium pollution, 330 synthetic ecosystems differing in the number of algal species were created. It was shown that the conservation of biodiversity (richness) may reduce the future impacts of increasing environmental stresses (Li *et al.*, 2010). A positive relationship between richness and functionality was also shown by Bell *et al.* (2005) by using synthetic microcosms with up to 72 bacterial species. Finally, Gravel *et al.* (2011) showed that the loss of specialists - strains that exploit only few resources - has a stronger effect on ecosystem functioning, compared to loss of generalists, which are able to use a spectrum of substrates.

The effect of trophic interactions - such as **predation** - on ecosystem functioning was investigated by altering the predator and prey richness. Predators were simulated by three bacterivorous protists, while five bacterial strains were used as model organisms of the prey. It was shown that the presence of multiple predators resulted in increased bacterial diversity, which had a positive effect on bacterial yields (Saleem *et al.*, 2012; Saleem *et al.*, 2013).

Finally, the **spatial organisation and architecture** of microbial community is also crucial to maintain a stable and functional community. By combining FISH with a digital image analysis software that quantifies the spatial localisation patterns of microorganisms in complex samples, it was shown that functionally linked species cluster together in a microbial community (Daims *et al.*, 2006). Kim *et al.* (2008) controlled the spatial organisation of a community by using a microfluidic device that controls the distance between three wild type soil bacterial populations with syntrophic interactions. In this community each species is required for the survival of the community. It was shown that spatial organisation is necessary to balance competition and beneficial interactions to create a stable community (Kim *et al.*, 2008). Brenner *et al.* (2011) used two genetically engineered *E. coli* populations to study the

benefits of the formation of physical structures like biofilms. Species associated in a biofilm were shown to be more productive than non-associated community members.

In conclusion, the use of synthetic ecosystems increased our knowledge regarding factors that shape and influence microbial communities. Such advances would have been almost impossible to obtain in natural ecosystems due to the presence of confounding factors that cannot be controlled or measured. As a result, the research regarding synthetic ecosystems initiates many opportunities to manage ecosystems. By changing one of the parameters, the community can be steered and a desired effect can be created.

4.   Analysing microbial communities

In order to be able to perform MRM and to allow the management of microbial ecosystems, more information about the current status of the microbial community, like its composition, structure and physiological characteristics, is required. Numerous techniques are available, all with their advantages and disadvantages. Conventionally, culture based and molecular methods are used, which are time-consuming and labor intensive. Therefore, the potential of flow cytometry as a fast and accurate methodology to study microbial communities is being investigated.

4.1.   Conventional techniques

Techniques to analyse microbial communities can be subdivided into two categories: culture-dependent and culture-independent techniques. Cultivation on general or selective solid media is a technique commonly used to assess the composition of microbial communities in environmental samples (Prakash *et al.*, 2013). It owes its widespread use to its user friendliness: it is easy to use and cheap. Nevertheless, it has some important disadvantages, the long incubation times (two to three days) make it a time-consuming technique. Furthermore it is highly selective, only about 1% of the microorganisms can be cultivated (Allen *et al.*, 2004). Although a lot of effort has been done to improve the cultivation efficiency by mimicking the original sample matrix and optimising the incubation conditions (Andreote *et al.*, 2009), the discrepancy between the culturable and total microbial fraction remains, plating still greatly underestimates microbial diversity. Therefore, culture-independent techniques, which do not require cultivation or isolation of microorganisms, became increasingly important. They can be categorized into two groups, direct and indirect methods (Dahllöf, 2002).
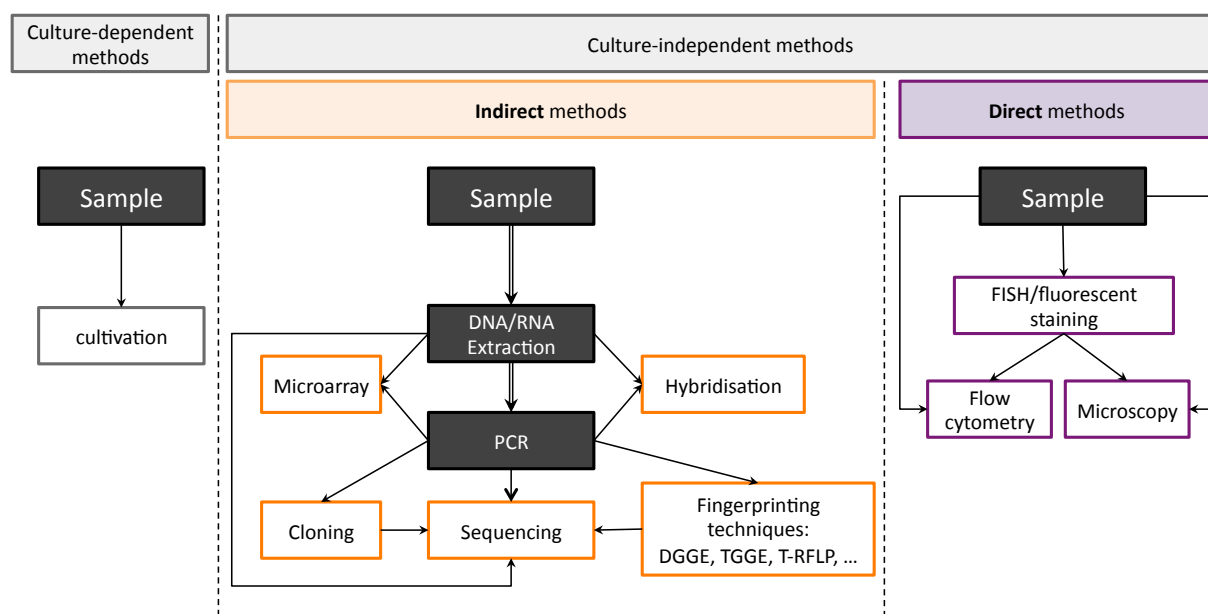
Figure 1-4. Overview of different techniques to study microbial communities *(Adapted from Dahllöf (2002)). Black boxes indicate preparatory steps and the product that is analysed by the detection techniques. Boxes with a coloured frame indicate detection techniques to analyse the sample, extracted nucleic acids or amplicons.*

With direct methods, the sample itself can be analysed directly. Prior to analysis, cells can be (selectively) labelled with fluorescent stainings or probes (FISH, fluorescent in situ hybridisation). This, however, does not require isolation of DNA and samples are analysed as such by cytometric techniques like microscopy (Amann, 1995).

Indirect methods require preparatory steps, like the extraction of nucleic acids and polymerase chain reaction (PCR) of target genes. The 16S rRNA gene is currently the gene mostly applied for this, however also functional genes can be used (Dahllöf, 2002). It is not the sample as such that is being analysed, but the extraction- or PCR-product (Figure 1-4). Molecular techniques like Denaturating Gradient Gel Electrophoresis (DGGE), Temperature Gradient Gel Electrophoresis (TGGE) and Terminal Restriction Fragment Length Polymorphism (T-RFLP) are DNA-based fingerprinting techniques. Following PCR, the amplicons are separated based on different intrinsic properties, which depend on the technique used. For T-RFLP this is the length of the terminal fragments after restriction. DGGE and TGGE apply respectively a denaturing and temperature gradient to separate the amplicons based on the GC-content of the sequences (Muyzer *et al.*, 1993). The obtained band- or peak pattern serves as a qualitative and semi-quantitative fingerprint of the microbial community. Each band corresponds to a different phylotype. The number of bands is thus a

measure for the community richness (i.e. the number of species present in a microbial community) (Marzorati *et al.*, 2008; Read *et al.*, 2011). Molecular fingerprinting techniques are especially useful for making an estimate of the diversity of a microbial community and to detect changes within or differences between microbial communities. Identification of the microbial species is however not directly possible. For this cloning and sequencing are required (Dahllöf, 2002).

The decreased cost and technological advances of high-throughput sequencing techniques such as 454 pyrosequencing and Illumina, greatly increased the popularity of sequencing approaches the last few years. They can be used to analyse the community structure and diversity and to identify individual species (Temperton & Giovannoni, 2012; Parkhill, 2013).

An important disadvantage of these DNA and RNA based fingerprinting techniques is the need for isolation of the nucleic acids, followed by PCR. Both steps introduce bias in the analysis of microbial communities (Becker *et al.*, 2000; Guo & Zhang, 2013). Furthermore, all these techniques are time-consuming. In the light of MRM, there is need for fast fingerprinting techniques that allow immediate intervention in case of problems or irregularities. The potential of flow cytometry was investigated in this context.

4.2. Flow cytometry

4.2.1. Flow cytometric measurement

Flow cytometry is an established technique in microbial ecology, it allows the analysis of microbial communities at community and single-cell level. To obtain this single-cell analysis, liquid samples are guided to the laser by hydrodynamic focusing in such a way that the particles pass the centre of the laser beam one by one. Each particle that passes the laser will cause scattering of the light in all directions. A part of this light is collected in the same direction as the laser beam at a low angle and is known as forward scatter (FSC). The intensity of the FSC is a measure of cell size. Another detector is positioned perpendicular to the laser beam and detects the light at an angle of 90°, which is known as sideward scatter (SSC). The intensity of the SSC is in relation with the internal complexity and granular structure of the particles (Diaz *et al.*, 2010). Fluorescent molecules present in the cell will absorb light energy of a specific wavelength and emit light of a longer wavelength. The excitation and emission wavelength vary between fluorochromes. This fluorescent light is detected by sensitive photomultiplier tubes (PMT) after it has been sorted by different

dichroic mirrors and optical filters. These mirrors and filters separate light with different wavelength and guide it to the correct PMT. This process of single-cell analysis is performed at a rate of thousands events per second.

For each particle, all parameters - being FSC, SSC and the different fluorescent signals - are recorded and analysed. These datasets are graphically visualised by the use of monoparametric histograms and biparametric dot plots.
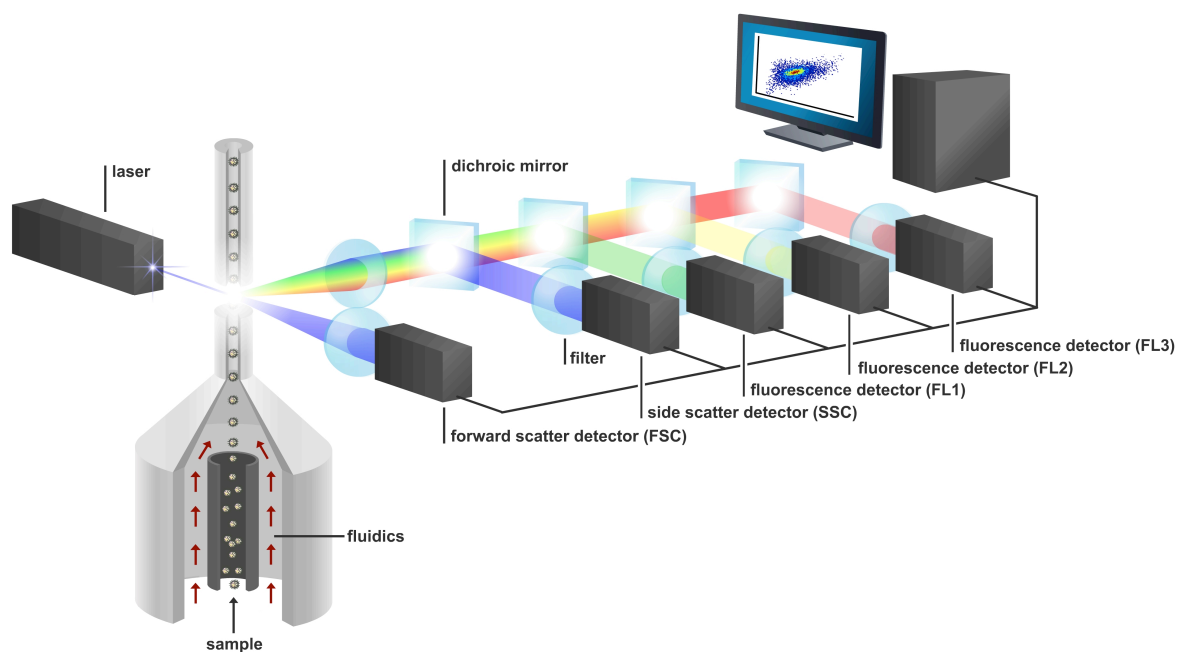
Figure 1-5. Scheme of the working mechanism of a standard flow cytometer. *A liquid sample is guided in a single beam of particles to the centre of a laser beam by hydrodynamic focusing. Each particles causes scattering of the light. Forward scatter light (FSC) is detected in the same direction as the laser, light in a 90° angle is detected by the sideward scatter detector (SSC). Fluorescent light is transported to different detectors by dichroic mirrors and filters. All signals are sent to a computer and processed.*

4.2.2. Fluorescent labelling

Abiotic particles, such as crystals and dust, also induce light scattering, just like prokaryotic and eukaryotic cells. Because of the small size of bacteria, it can be hard to discriminate them from these background signals. Therefore, bacteria are often labelled with fluorescent molecules. Fluorochromes allow performing cell differentiation based on morphological, structural or functional cell properties, such as membrane integrity, membrane potential and enzyme activity (Figure 1-6).

Membrane permeant nucleic acid stains like SYBR Green and SYTO 9 are conventionally used to stain all microbial cells, regardless of their physiological state (Wang *et al.*, 2010). They passively diffuse through the membrane of most cells and become fluorescent (SYTO) when bound to nucleic acids or their fluorescence is enhanced (SYBR Green) (Gregori *et al.*, 2001; Diaz *et al.*, 2010). The fluorescent intensity of a single cell is correlated to the amount of nucleic acids present in the cell. As such high nucleic acid (HNA) and low nucleic acid (LNA) bacteria can be distinguished (Wang *et al.*, 2009).

Membrane impermeant nucleic acid stains such as propidium iodide (PI) are not able to penetrate cells with an intact membrane, only cells with damaged membranes get stained. PI is the most commonly used dye to evaluate the membrane integrity and, as such, the viability of bacterial cells. Frequently, this type of stain is used in combination with the general stains SYBR Green or SYTO 9. These combinations are often erroneously called live/dead stainings, while they only differentiate between cells with damaged and intact cytoplasmic membranes (Berney *et al.*, 2007).

Membrane potential also provides information about the viability. It is a measure of the health of bacterial cells, as it is directly linked to ATP formation. Cells that are not able to maintain the potential get depolarised, have a decreased cell activity and are regarded as dead (Shapiro, 1994). DiBAC$_4$ is an anionic dye that accumulates in depolarised cells and stains cells that are not able to maintain the cell potential (Diaz *et al.*, 2010).

Enzyme activity can be used as a measure for metabolic activity. Non-fluorescent substrates passively diffuse through the membrane and are converted into membrane impermeable fluorescent products by intracellular enzymes. Carboxyfluorescein diacetate (CFDA) is used to evaluate esterase activity and gets converted into the fluorescent carboxyfluorescein (Hoefel *et al.*, 2003).

Next to this wide variety of fluorescent dyes to differentiate cells based on functional cell properties, fluorescent labelled probes and antibodies can be used to stain specific groups of microorganisms. Fluorescent in situ hybridisation (FISH) is mostly targeted towards 16S rRNA, it can however also be targeted to functional genes. Like this, specific taxonomic groups or cells with certain functionality can be detected (Amann, 2000; Porter & Pickup, 2000). Labelled antibodies can be used to detect specific microbial antigens (Hammes & Egli, 2010).
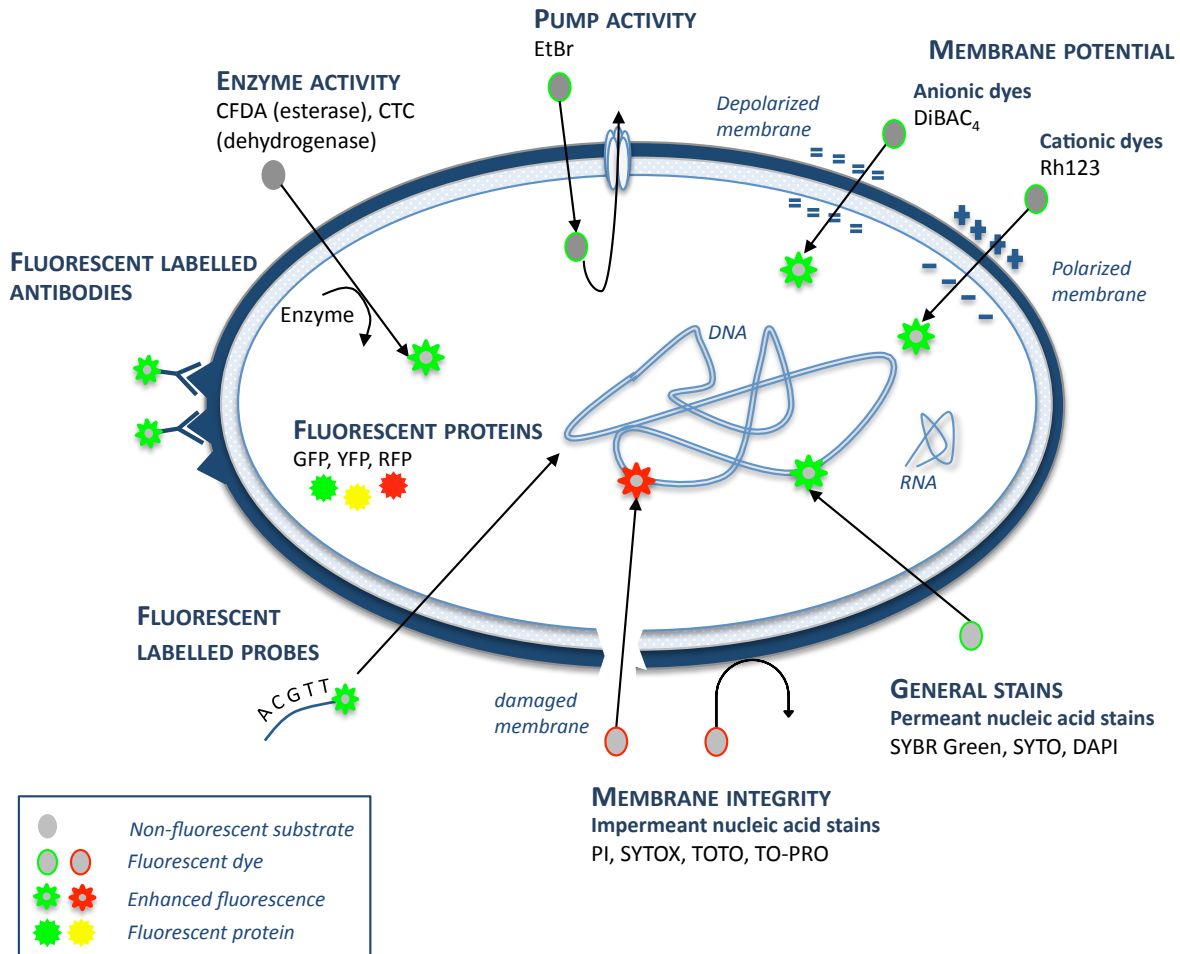
Figure 1-6. Fluorescent labelling of bacterial cells. *This scheme gives an overview of different types of fluorochromes, their working mechanisms and target sites (Adapted from Diaz et al. (2010)).*

4.2.3. Flow cytometry to date

The attractions of flow cytometry for the use in microbial ecology include the rapid analysis, increased accuracy and the availability of numerous types of fluorescent labelling methods. Furthermore, microbial cells can be detected irrespective of their culturability and no major pre-treatment steps like DNA extraction or PCR are required. Therefore, there is an increasing interest in using flow cytometry in different industries, like drinking water production and distribution (Hammes *et al.*, 2010), wastewater treatment (Guenther *et al.*, 2012), quality control of food and drinks (Ruszczynska *et al.*, 2007) and fermentation processes (Andorra *et al.*, 2011).

To date, flow cytometry is mainly used to count and differentiate cells based on their physiological and morphological characteristics. This can only be obtained by post-processing of the flow cytometry data. Conventionally, this is done by a visual interpretation of the histograms and dot plots. Approximated polygons are drawn around cells with specific

characteristics, in data-analysis terms: particles with higher or lower light intensities in the different scattered or fluorescent light channels. This process is not only very subjective (Maecker *et al.*, 2005), it also requires experienced personnel and is time consuming in case of a high number of samples. Furthermore, the combination of fluorescent dyes requires multi-parametric analysis. Three parameters can be visualised on a 3D dot plot, but more dimensions cannot be displayed on a single graph and multiple dot plots need to be created and analysed. Efforts have been done to create automated algorithms to analyse flow cytometry data, this involves automatic gating and feature extraction (Bashashati & Brinkman, 2009). But the comparison of different samples and detection of changes within a microbial community is still done manually.

4.2.4. The road to go: flow cytometric fingerprinting

Microbial communities are subjected to changes in various parameters in their environment, such as temperature, oxygen and pH. The short-term effects of these changes cannot be detected by standard molecular methods because the microbial response is not at the genetic level. Under such circumstances, flow cytometry in combination with fluorescence dyes and statistical analyses can provide fast and accurate assessment of physiological and functional changes in a community. During flow cytometric measurement, multiple parameters, such as fluorescence intensity and scattered light signals, are collected for individual cells. For example, for analysis of four separate parameters in one bacterial sample comprising 10000 cells, statistical analysis of the resultant 40,000 data points will generate a fingerprint for the bacterial community. Hence, each bacterial community under a certain condition could have a unique fingerprint signature and subtle changes might be detectable according to changes in these fingerprints. This method could serve as a sensor in treatment processes for drinking water, reused water and wastewater, or as a reliable tool for microbial resource management in the future (Marzorati *et al.*, 2008).

5.   Objectives and outline of this research

Knowledge is power! Knowledge of microbial ecosystems will allow us to predict and possibly counteract the negative effects caused by environmental changes like global warming. Knowledge of microbial ecosystems will allow us to perform microbial resource management and to steer microbial communities in numerous biotechnological applications. To make this possible, more research on the factors that are known to influence microbial

communities is necessary. In a first part of this thesis, synthetic microbial communities are used to investigate the effect of biological invasion and salt stress on the functionality of the community (Chapter 2). The establishment and progression of microbial interactions within such a community is investigated in Chapter 3. The results obtained in Chapter 2 and 3 allow us to better understand microbial communities. They bring us one step closer to the management of microbial communities. To allow this management approach, microbial community analysis techniques are required that are able to provide more information about the current status of the microbial community, like its composition, structure and physiological characteristics. In the second part of this thesis, the potential of flow cytometry as a fast and accurate methodology to study microbial communities is being investigated. A flow cytometric fingerprinting technique that is able to detect changes between and shifts within microbial communities is developed (Chapter 4). In chapter 5, the applicability of this technique is tested in a water distribution network.

Overview of the different chapters

In **Chapter 2**, biological invasion of non-native species in a microbial community was investigated under fully controlled starting conditions. We evaluated the degree of invasion and the effect on the community functionality in relation to the initial community evenness in presence of specific environmental stressors. We did this by using synthetic microbial communities consisting of 17 denitrifying bacterial species and a *gfp*-tagged invader. Denitrification was used as a measure of functionality, the number of invaders was measured by flow cytometry.

In **Chapter 3,** the effect of interspecies relatedness on the establishment of a mutualistic interaction was investigated. Synthetic microbial communities were created that were only able to survive upon the establishment of mutualism. For this, an ampicillin resistant strain auxotrophic to tyrosine was grown in combination with closely related and distantly related ampicillin sensitive species. The total fitness of the auxotroph was used as a measure of cooperation.

In **Chapter 4,** a flow cytometry based approach was developed for a fast and objective comparison of microbial communities based on the physiological status of single cells within these communities. Firstly fingerprint data was generated by the flow cytometric analysis of

different brands of bottled water and water that was exposed to different kinds of stress. Secondly a novel statistical pipeline for the analysis of flow cytometric data was developed.

In **Chapter 5,** the flow cytometric fingerprinting technique developed in chapter 4 was applied in the water distribution network. The water used during the rinsing procedure of water pipes was sampled and analysed every hour. Subsequently, the fingerprinting technique was used to determine the minimal rinsing time required to obtain water of the desired quality.

Finally, in **Chapter 6,** the obtained results are discussed in the framework of the research objective and conclusions are drawn.

ENVIRONMENTAL CONDITIONS AND COMMUNITY EVENNESS DETERMINE THE OUTCOME OF BIOLOGICAL INVASION

ENVIRONMENTAL CONDITIONS AND COMMUNITY
EVENNESS DETERMINE THE OUTCOME OF BIOLOGICAL
INVASION

Abstract

Biological invasion is widely studied, however conclusions on the outcome of this process mainly originate from observations in systems that leave a large number of experimental variables uncontrolled. Using a system, consisting of assembled bacterial communities, that controls the environmental factors and the initial community composition, we evaluate the degree of invasion and the effect on the community functionality in relation to the initial community evenness under specific environmental stressors. We show that evenness influences the level of invasion and that the introduced species can promote functionality under stress. The evenness-invasibility relationship is negative in the absence and neutral in the presence of stress. Under these stress conditions, the introduced species is able to maintain the functionality of uneven communities. These results indicate that communities, initially having the same genetic background, in the presence of the same invader, react in a different way with respect to invasibility and functionality depending on specific environmental conditions and community evenness.

1. Introduction

Biodiversity has been shown to significantly influence invasion in disparate aquatic and terrestrial environments (Stachowicz *et al.*, 1999; Simberloff *et al.*, 2012). Biodiversity correlates positively (Jiang & Morin, 2004), negatively (Wilsey & Polley, 2002) or neutrally (Emery & Gross, 2007) with the capacity of species to invade resident communities. Many factors, such as nutrient availability (Jiang & Morin, 2004), scale (Dunstan & Johnson, 2006), functional niches (Tilman *et al.*, 1997), environmental stressors (Kneitel & Perrault, 2006), facilitation or competition (Bruno *et al.*, 2003) and biodiversity (Davis, 2009) can co-occur, driving the invasion process and influencing the final outcome. The available studies, mainly based on observations, do not take into account intrinsic system effects and a large number of non-controlled variables are present (climatic conditions, soil type, soil microbial community, presence of symbiotic or antagonistic partners, etc.) (Rout & Callaway, 2009; Andonian *et al.*, 2011). These confounding factors and the reciprocal interactions may lead to opposite conclusions on the role of a specific parameter, even in closely related ecosystems (Wilsey & Polley, 2002; Emery & Gross, 2007). Experiments conducted under controlled conditions give the opportunity to target some of these confounding factors, eventually explaining the inter-system variability, despite the fact that this approach is a simplification of the reality. In this study, we applied the approach used by Wittebolle *et al.*, which is well suited for validating ecological theories (Naeem, 2009) and can be run under controlled conditions. The effect of an introduced species (termed the 'invader') on ecosystem functionality was investigated in relation to the initial evenness of the native community. We decided to focus on evenness and to maintain a high level of richness because natural and anthropogenic activities primarily influence the relative abundance of species long before a species is threatened by extinction (Chapin *et al.*, 2000; Bell *et al.*, 2005).

2. Material and methods

2.1. Experimental design

In this study, we used 210 mixtures with different evenness values, corresponding to unique Gini and dominant species combinations. These mixtures were created following a stochastic exchange search algorithm for D-optimal designs (Atkinson *et al.*, 2007) according to the following procedure. The first step was the random construction of a design. The relative

abundances of the 17 species were randomly assigned. Five 96-well plates were randomly filled with the mixtures, with adjacent duplicates of each mixture. Negative controls were positioned in the centre and the corners of each plate to assess potential row, column and plate effects. We computed the D-optimality criterion for this random design and a linear model, with factor effects for the row, column, invader and salt stress, a random effect for the plate, and a linear and quadratic effect for the Gini. In the next step, the mixtures were randomly exchanged between positions and plates, and the D-optimality criterion was computed for each new design; this procedure was repeated 1,000 times. The best design with the largest D-optimality criterion was selected as the quasi D-optimal design. The complete procedure, starting from the construction of the random design, was also repeated 10,000 times. From the 10,000,000 evaluated designs, the design with the largest D-optimality criterion was selected for this study. During the optimisation process, the convergence of the D-optimality criterion was monitored to ensure that the final selected design was sufficiently well converged.

-- Information box statistics --

Optimising the experimental design using the D-optimality criterion

The aim of the experiments is to model the effects of the Gini coefficient, salt stress and invader concentration on the response variables invasion, functionality and cell count community. Since only a limited number of experimental tests can be performed, it is important that the tests are designed in the most optimal way to allow (1) efficient estimation (i.e. small variance of estimators) of the parameters for the stress, invader and Gini effects and (2) large power of the related statistical tests. With a not optimal design, more runs would be required to obtain parameter estimates with the same precision and tests with the same power.

But next to the parameters of interest, also other parameters are present that can have an effect on the response variables, like row, column and plate effects. If, for example, all mixes with a high Gini would be positioned in one plate, while mixes with a low Gini are positioned in another plate, it would not be possible to determine if the observed effect is attributed to a plate or Gini effect (i.e. confounding between plate and Gini effects). Therefore, the design should guarantee that all parameters can be estimated (estimable).

So, to get an optimal design, both the composition of the mixes and the positioning within a plate and the distribution over the different plates need to be optimised so as to guarantee estimability and maximise efficiency and power.

The optimality of an experimental design depends on the statistical model, which is a linear model with effects for plate, row, column, invader, salt stress and Gini. The optimality is measured by a statistical criterion, in this case the D-optimality criterion. If only one parameter would be present, this criterion aims at minimising the variance of the estimator, which is equivalent to maximising the information. In the case of multiple parameters, as it is in this experimental setup, the D-optimality criterion is the determinant of the information matrix, which is the inverse of the variance-covariance

matrix of the vector of the parameter estimators. As such, the D-optimality criterion needs to be maximised, resulting in a design with overall the smallest variances and covariances of the parameter estimators. Because of the relation between variance and power, this design will also result in the largest overall power.

In a first stage of the experimental design, mixes with different relative abundance (and thus different Gini) are randomly created. These mixes are then randomly distributed over multiwell plates and the D-optimality criterion is calculated for the linear model. In the next step, the positions of the same mixes were randomly changed and again the D-optimality criterion calculated. The step of random position change was repeated 1,000 times and every time the criterion was calculated. Since also the composition of the mixes is of importance, the complete procedure of randomly creating the mixtures and shuffling the positions was repeated 10,000 times. Eventually 10,000,000 designs were evaluated and the design with the highest D-optimality criterion was retained and used for the experiment. This algorithm is a variation of a Federov algorithm for searching for optimal experimental designs.

2.2. Laboratory methods

A chromosomally *gfp*-tagged fermentative *Pseudomonas sp.* and seventeen denitrifying strains (Table 2-1) were stored in ready-to-use aliquots at -80°C. For each experiment, the strains were cultured for 48 h and subcultured for 40 h in Trypticase Soy Broth (TSB) at 28°C under aerobic conditions to obtain actively growing microorganisms. The denitrifying strains were diluted to 10^7 cells mL^{-1} and the invader to 1.5×10^7 cells mL^{-1}, as measured by flow cytometry (CyAn ADP LX, Dako, Heverlee, Belgium). Communities with different degrees of initial evenness were created by mixing the diluted denitrifying strains in different proportions. Each mixture was divided into three aliquots. No cells of the invader were added to the first aliquot (0% invader); to the second and third aliquot, the invader was added so that the number of invader cells in the mixtures was 0.1% and 1% of the total number of cells, respectively. These mixtures were used to assemble microcosms of 200 μL in duplicate in multiwell plates, according to the design, using a BioRobot 3000 (Qiagen, Venlo, The Netherlands). The mixtures were 1:1 diluted in TSB supplemented with 12 mM nitrite and optionally supplemented with 4% (w/v) NaCl (for the salinity stress) to obtain a final concentration of 6 mM nitrite and 2% NaCl. The final volume in each well of the plates, which have a maximum volume of 300 μL, was 200 μL. The plates were incubated anoxically in jars for 20 h at 28°C. Anoxic conditions were created by the Oxoid™ AnaeroGen™ Compact system (Oxoid, Basingstoke, UK).

After incubation, the optical density (620 nm) was measured and the residual nitrite concentration determined with the Montgomery reaction (Montgomery & Dymock, 1961) (Figure 2-1). The relative abundance of the invader cells and the total cell count of the

invaded communities were determined by flow cytometry. Since oxygen is required for the maturation of the Green Fluorescent Protein (GFP), the plates were incubated aerobically at 4°C prior to flow cytometric analysis. It was shown that a period of 2 hours resulted in an increase in fluorescence intensity of the invader cells to a level similar to aerobically grown invader cells (results not shown). As such the relative abundance of the invader cells was measured by flow cytometry after the plates were incubated at 4°C for 2 hours.

In order to identify a potential niche overlap for the electron donor, an anaerobic metabolic fingerprint (using 95 different carbon sources of the Biolog AN microplate (Hayward, USA) and 6 mM nitrite as electron acceptor) was made for 17 uneven communities (each with one different dominant strain), the perfectly even community and the invader. For each mix, we determined the unique available carbon sources for the invader by subtracting the carbon sources used by a single mix from those used by the invader itself.

Table 2-1. List of strains used to create the microcosms. ⋆ *This strain was used as invader and was obtained from Sternberg et al. (1999). All other strains were obtained from Heylen et al. (2006).*

Phylum	Collection code	ID	Functionality	Special characteristics	Reference
Firmicutes	R-32851	*Bacillus sp.*	denitrifier		Heylen *et al.* (2006)
α Proteobacteria	R-27049	*Paracoccus sp.*	denitrifier		Heylen *et al.* (2006)
	R-24665	*Paracoccus sp.*	denitrifier		Heylen *et al.* (2006)
	R-26895	*Brucella sp.*	denitrifier		Heylen *et al.* (2006)
	R-25018	*Ochrobactrum sp.*	denitrifier		Heylen *et al.* (2006)
	R-25203	*Ochrobactrum sp*	denitrifier		Heylen *et al.* (2006)
β Proteobacteria	R-24607	*Acidovorax sp.*	denitrifier		Heylen *et al.* (2006)
	R-25212	*Acidovorax sp.*	denitrifier		Heylen *et al.* (2006)
	R-24610	*Diaphorobacter sp.*	denitrifier		Heylen *et al.* (2006)
	R-26815	*Diaphorobacter sp.*	denitrifier		Heylen *et al.* (2006)
	R-26829	*Comamonas sp.*	denitrifier		Heylen *et al.* (2006)
	R-28220	*Comamonas sp*	denitrifier		Heylen *et al.* (2006)
	R-25060	*Comamonas sp*	denitrifier		Heylen *et al.* (2006)
	R-28413	*Comamonas sp*	denitrifier		Heylen *et al.* (2006)
ϒ Proteobacteria	R-25061	*Pseudomonas sp.*	denitrifier		Heylen *et al.* (2006)
	R-26828	*Pseudomonas sp*	denitrifier		Heylen *et al.* (2006)
	R-25343 /	*Pseudomonas sp* *Pseudomonas putida* SM1699⋆	denitrifier Non- denitrifier/ Fermentor	Gfp labeled: *gfp*mut3b-gene randomly inserted in the chromosome	Heylen *et al.* (2006) Sternberg *et al.* (1999)
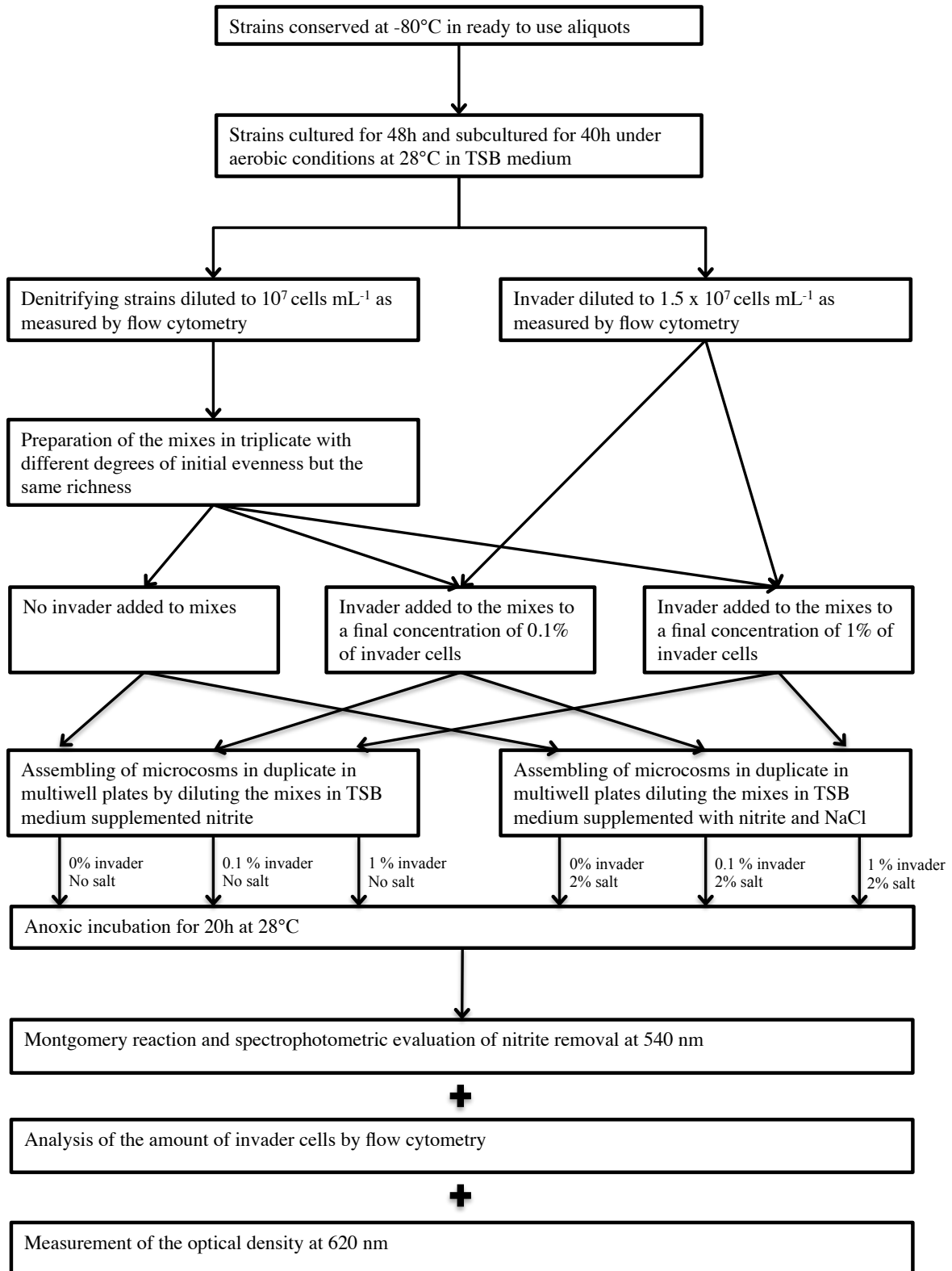
Figure 2-1. Overview of the experimental setup

2.3. Statistical analysis

Each of the response variables (functionality, invasion and $CC_{community}$) has been analysed with additive quantile regression models (Koenker, 2005). A separate analysis was performed for each invader/salt combination. The Gini effect was modelled non-parametrically using a smoother spline, either unconstrained or with a monotonicity constraint (increasing or decreasing). The smoothing parameter was optimised by minimising the Akaike's Information Criterion (AIC). All hypothesis tests were based on generalised likelihood ratio tests performed at the 5% level of significance. For forward model selection the testing sequence was: (1) no Gini effect versus a monotonic Gini effect (increasing or decreasing, depending on the AIC), (2) monotonic Gini effect versus an unconstrained Gini effect. Additive quantile regression models have been used to analyse the response variables at the 50%, 10% and 90% quantiles. Quantile regression extends ordinary regression models in the sense that a particular quantile of the conditional response distribution is modelled instead of the mean (Elsner *et al.*, 2008). The choice for this method is motivated by the variable shapes of the conditional response distributions as a function of the Gini coefficient. Results of the Gini effects are only reported for the analyses at the 50% quantile (median), unless Gini effects were only established at the 10% or 90% quantiles. Analyses were conducted with the quantile set at either 50% (invasion and $CC_{community}$) or 10% (functionality). The partial residuals, shown in the graphs of the model fits, were always constructed relative to the 50% quantile regression model that corrects for row, column and plate effects. The axis of the residuals of the functionality and invasion response variables were rescaled to obtain values between 0 and 1, with 0 the lowest and 1 the highest measured value.

-- Information box statistics --

Additive quantile regression analysis

Regression analysis is performed for estimating the relationship between the independent variable, which is here the Gini coefficient (the regressor), and the response variables, being invasion, functionality and cell count community. The regression models help us to understand how the response variables change when the Gini changes. Such a regression analysis is performed for each invader/salt combination for each of the response variables.

In conventional linear regression analysis, the mean of the response variable distribution is modelled as a linear function of the regressors. **Quantile regression** makes use of a different characteristic of the response variable distribution, for example the 50% quantile, which corresponds to the median. Quantile regression was shown to be very useful for ecology research, because the shape of the distribution of ecological data often changes with changing environmental conditions. As a

---

consequence, there may only be a weak or no predictive relationship between the mean of the response variable distribution and the regressor, while other quantiles might give more information (***Box Figure 2-1***)(Cade & Noon, 2003).
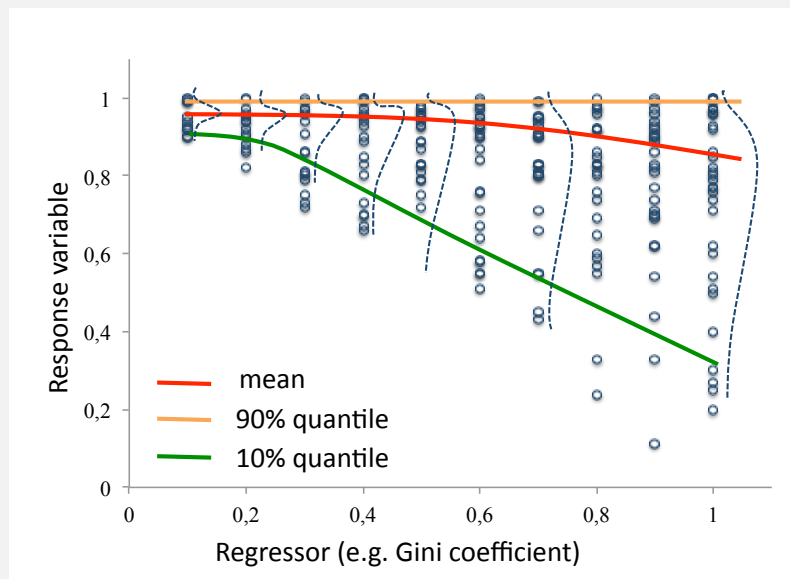
Box Figure 2-1. *Conventional regression analysis makes use of the mean of the response variable (red), while quantile regression uses a different characteristic of the response variable distribution (shown with dotted lines), like the 10% quantile (green) or the 90% quantile (orange). While there is only a weak predictive relationship between the mean of the response variable and the regressor, the 10% quantile is more sensitive to changes in the Gini.*

Non-parametric models do not make any assumptions regarding the relationship between the regressor and the response variable, unlike parametric models, such as linear regression models, which assume a linear relation between the regressor and the mean or quantile of the response variable.

The effect of the Gini coefficient was modelled non-parametrically. By design, the regression model also includes terms for row (R), column (C) and plate (P) effects, which are all additive to the Gini-effect.

$$Linear\ model: E\{Y|X\} = \beta_0 + \beta_1 X_1 + P + C + R + \cdots$$

$$Additive\ model: E\{Y|X\} = \beta_0 + f_1(X_1) + P + C + R + \cdots$$

where $f_1(X_1)$ is a smooth function of the X_1 regressor, but no analytical expression for f_1 need to be specified by the data-analyst. Non-parametric statistical methods (e.g. smoothing splines) can be used for the estimation of the regression function f_1. Smoothing splines require one tuning parameter that has to be specified by the data-analyst. This tuning parameter specifies the smoothness of the estimated regression function (***Box Figure 2-2***B).

Three types of relationship can occur between the Gini coefficient and the response variable: (i) constant, (ii) monotonic (either increasing or decreasing) and (iii) unconstraint (***Box Figure 2-2***A). For each of these non-parametric models, the tuning parameter of the smoother spline needs to be optimised. This tuning parameter is optimised by the Akaike's Information Criterion (AIC), which is a measure of the quality of the model predictivity. Optimising the AIC results in a compromise between the quality of the fit and the complexity of the model fit (***Box Figure 2-2***B).

Once the tuning parameter of each of the models is optimised, the models are compared and the

optimal model selected by forward model selection. Therefore, a constant Gini effect is initially compared to a monotonic Gini effect. If p>0.05 (based on likelihood ratio tests performed at the 5% level of significance), the constant model is the best. If p<0.05, the monotonic model was the best and needs to be compared to the unconstraint model. If in this comparison p>0.05, the monotonic model is the best, if p<0.05, the unconstraint model is the best.
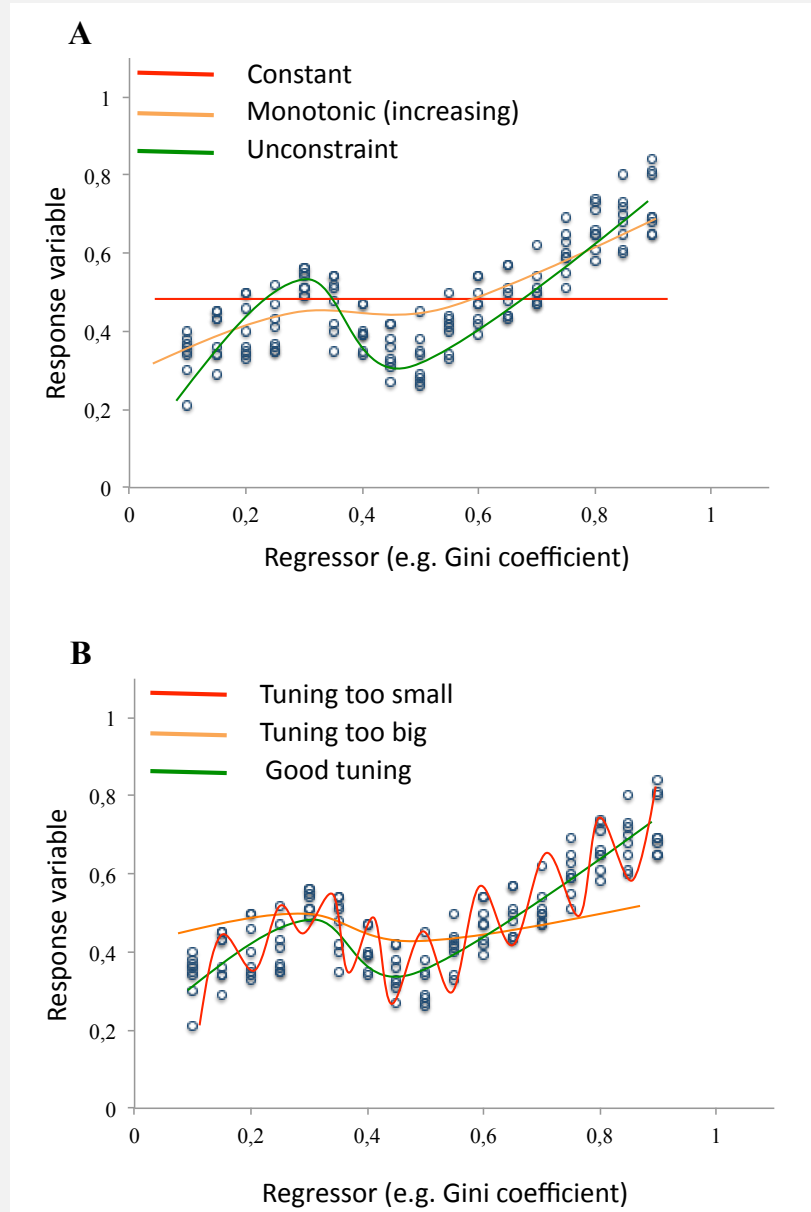
Box Figure 2-2. A. *Smoother splines of three different models: constant (red), monotonic increasing (orange) and unconstraint (green). In this example, the unconstraint model would be selected by forward model selection.* ***B.*** *Smoother splines of the unconstraint model using different tuning parameters. The optimal tuning parameter is selected using the Akaike's Information Criterion (AIC), which looks for a compromise between complexity and goodness of fit. In this case, the tuning parameter of the green spline is optimal.*

3. Results

3.1. Rationale of the experimental setup

The impact of initial community evenness on invasibility and the effect of the invader on the functionality of the resident community, i.e. denitrification, were assessed by the use of assembled denitrifying bacterial communities. To exclude other confounding factors that could influence invasion, our tests were conducted with (i) a complex medium to avoid nutrient limitation, (ii) an assembled community composed of bacteria occupying the same functional niche (the capability for nitrite respiration), isolated from the same sample of homogenised activated sludge and without prior history of adaptation to prevailing environmental conditions, and (iii) an invader incapable of denitrification but able to grow under anaerobic conditions by fermentation, it uses organic compounds as electron acceptor (Figure 2-2 and Figure 2-3). As such resident species from one niche are used against an invader from another niche to test the effect of the evenness of the resident community on invasion and the effect of invasion on functionality. The use of an invader within the same functional niche of the resident community would only increase the richness of the community, without being able to pinpoint the effect of evenness. Besides, it would not have been possible to investigate how invasion influences the functionality of the resident community, since the invader would then have contributed to this functionality.
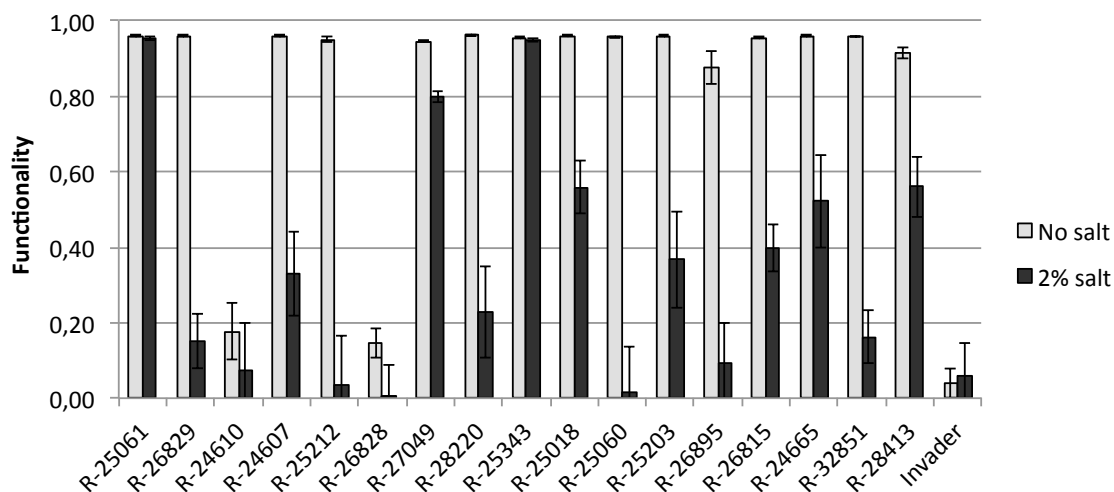
Figure 2-2. Functionality of the individual strains in the absence and presence of salt stress. *The bars show the relative amount of nitrite removed after 20 hours anaerobic incubation in the presence of 6 mM nitrite. 1.0 corresponds to complete nitrite removal and 0,0 to no removal. (mean ± SD, n=2)*
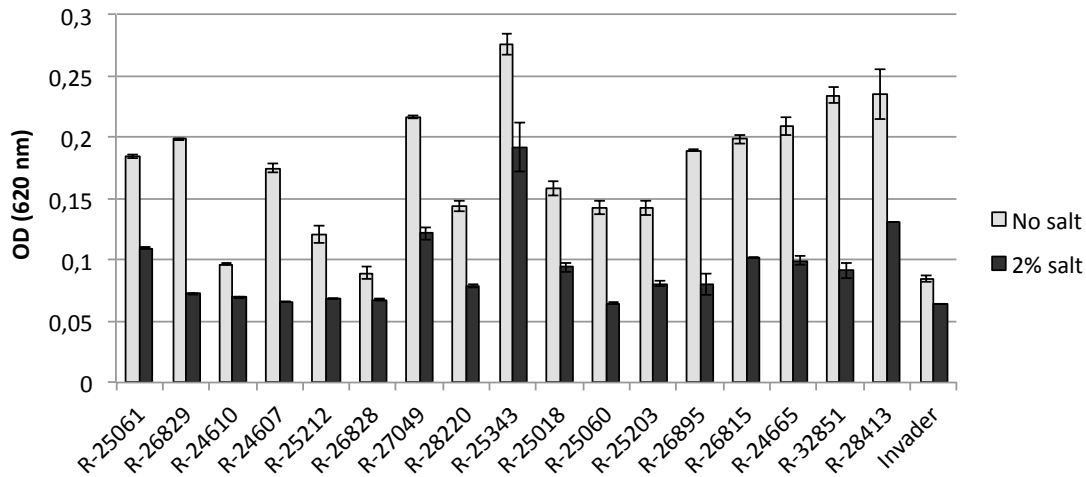
Figure 2-3. Growth of the individual strains in the absence and presence of salt stress. *The bars show the optical density (620 nm) of the denitrifiers and the invader after 20 hours of anaerobic growth in the presence of 6 mM nitrite. (mean ± SD, n=2)*

A total of 17 denitrifying strains from four phyla (Table 2-1) were mixed in different proportions to create 3192 microcosms with different levels of initial evenness but with the same richness. This number of strains represents a high richness, which ensures a good functionality (Bell *et al.*, 2005); the complete range of evenness, expressed by the Gini coefficient (i.e. 0 being a complete even community and 1 the most uneven community), was covered (Figure 2-4).
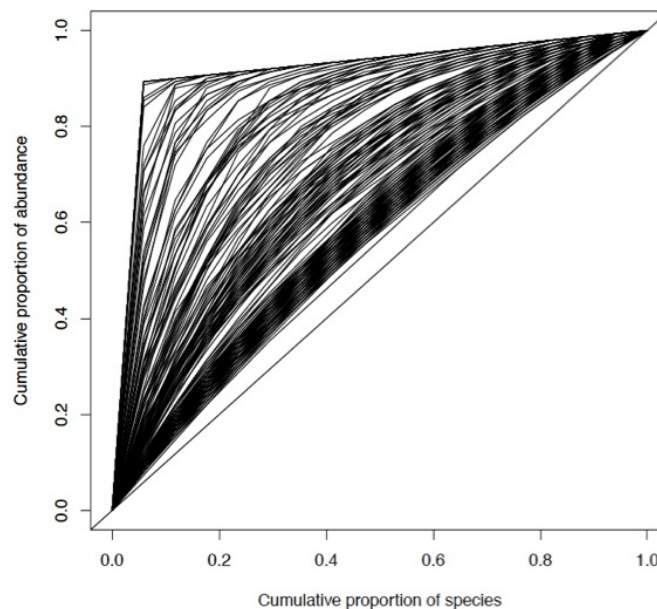
Figure 2-4. Lorenz curves of the mixes used in the experiment. *The curves span the entire region between high evenness and high dominance. Each curve can be numerically described by a Gini coefficient, being the area between the Lorenz curve and the line of the perfect evenness.*

The microcosms were arranged in 96-well plates and incubated under two distinct conditions: no stress and salinity stress. Salinity stress was chosen because it cannot be readily altered by the microorganisms. Furthermore, it was shown to have a significant impact on the functionality (Figure 2-2) and growth (Figure 2-3) of the individual strains. The impact of salt stress on community functionality was also shown before by Wittebolle *et al.* (2009). An open ecosystem (i.e. a system that has an input of matter, e.g. microorganisms (McArthur, 2006)) was simulated by challenging the assembled community with an introduced *gfp*-tagged, salt-resistant, non-denitrifying, fermentative species - the invader - at a concentration of 0.1% and 1% of the initial total cell number (Figure 2-1). In contrast, the control experiment with no introduced species simulated a closed ecosystem. After 20 hours of anaerobic incubation, the percentage of nitrite removal was used as a measure of functionality of the community. The total number of cells and the number of *gfp*-tagged invader cells were analysed by flow cytometry to determine the invasion coefficient and the resident community cell number. The invasion coefficient, ranging from 0 (not invaded) to 1, corresponds to the proportion of invader cells to the total cell count, while the community cell count ($CC_{community}$) is equal to the total number of cells subtracted by the number of invader cells. Each of the response variables - functionality, invasion and $CC_{community}$ - has been analysed with additive quantile regression models (Koenker, 2005). By design, all of the models included terms for the row, column and plate effects. The effect of the Gini coefficient was modelled non-parametrically with a smoother spline, either unconstrained or with a monotonicity constraint (increasing or decreasing) (Table 2-2).

3.2. Invasion in the absence of stress

Under different environmental conditions, the intrinsic characteristics of a microbial community influence its susceptibility to invasion and its functional stability. In the absence of salt, invasion increased with an increasing Gini coefficient (Figure 2-5a) and was observed both at a low and high initial concentration of the invader: 0.1% ($p<0.001$) and 1% ($p<0.001$) of the total cell count, respectively (Table 2-3). A higher initial concentration of the introduced strain produced a higher level of invasion ($p<0.001$). Since the total number of invader cells was always higher than the number initially added to the communities, the differential invasion was not the result of differential die-off but due to differential growth. The presence of the invader affected the performance of the community by lowering the overall denitrifying functionality ($p<0.001$), independent of the degree of evenness (Figure

2-5b) without influencing the growth of the community (Figure 2-5c). Therefore, under conditions of no stress, the Gini coefficient was positively correlated with the degree of invasion and neutrally correlated with the $CC_{community}$.

Table 2-2. Model selection for the three response variables. *The effect of Gini on the $\tau=0.5$ or the $\tau=0.1$ quantile of each response variable was modelled non-parametrically. Forward model selection was based on likelihood ratio tests performed at the 5% level of significance (p). A constant Gini effect (C) was tested versus a monotonic Gini effect (M), either decreasing (MD) or increasing (MI). If $p>0.05$, the constant model was the best, and the model selection was completed. If $p<0.05$, the unconstrained model (UC) was compared with the monotonic model (M vs UC). In the latter case: if $p>0.05$, the monotonic model was selected; the unconstrained model was selected if $p<0.05$. "n/a": not applicable because no invader was added in the control, "-": statistical test not conducted because the previous model was selected. Upper line percentages are initial invader concentrations*

Response variable	Model	0% No Salt	0.1% No Salt	1% No Salt	0% Salt	0.1% Salt	1% Salt
Invasion ($\tau=0.5$)	p (C vs M)	n/a	0	0	n/a	0.86	0.99
	p (M vs UC)	n/a	0.07	1	n/a	-	-
	Selected model	n/a	MI	MI	n/a	C	C
Functionality ($\tau=0.1$)	p (C vs M)	0.29	0	0.15	0	0	0.16
	p (M vs UC)	-	0.03	-	0	0.97	-
	Selected model	C	UC	C	UC	MD	C
Log $CC_{community}$ ($\tau=0.5$)	p (C vs M)	n/a	0.30	0.74	n/a	0	0
	p (M vs UC)	n/a	-	-	n/a	0.14	0
	Selected model	n/a	C	C	n/a	MD	UC

3.3. Invasion in the presence of stress

In the presence of salt (Figure 2-5d), the degree of evenness did not influence the invasibility of the community (p=0.86 and p=0.99 for 0.1% and 1% invader, respectively). However, the functionality of the community was strongly influenced by the introduction of the invader (p<0.001) (Figure 2-5e). Under salinity stress and in the absence of invasion, nitrite was only partially reduced by the denitrifying communities with a high Gini coefficient. This data confirmed what was previously shown, that communities with a high initial evenness have a higher potential to counteract the effect of a sudden selective stress than communities with a low initial evenness (Wittebolle *et al.*, 2009). If the same communities were exposed to an invader, no negative correlation between functionality and the Gini coefficient was observed.

The functionality was maintained at a high level over the complete range of evenness if the communities were challenged with the highest invader concentration. Similar to the effect observed under the conditions without salt, the presence of the invader under the stress condition had no effect on community growth (p=0.365) (Figure 2-5f). The functionality of invaded communities was always lower under non-stressed conditions compared to stress conditions (p<0.001), while invasion was higher under non-stressed conditions and community growth was similar, all independently of the degree of evenness.

Table 2-3. The tables show the p-values of the statistical tests for comparing the semiparametric models of the effect of Gini for the different invader and stress conditions. The part above the diagonal of each table shows the comparison of the two conditions for the response variable, while the part below the diagonal shows its p-value. "n/a": Not applicable because the response variable was not determined for at least one of the conditions; "/": Not useful to compare the two conditions for this response variable; "I": Response variable 'Invasion'; "F": Response variable 'Functionality'; "C": Response variable 'CCcommunity'; "S": Salt; "NS": No salt; ★ For Gini values higher than 0.4.

Invasion

	No Salt 0% inv	No Salt 0.1% inv	No Salt 1% inv	Salt 0% inv	Salt 0.1% inv	Salt 1% inv
No Salt, 0% inv		n/a	n/a	n/a	n/a	n/a
No Salt, 0.1% inv	n/a		$I_{(NS,0.1\%)}<I_{(NS,1\%)}$	n/a	$I_{(NS,0.1\%)}>I_{(S,0.1\%)}$	n/a
No Salt, 1% inv	n/a	p <0.001		n/a	/	$I_{(NS,1\%)}>I_{(S,1\%)}$
Salt, 0% inv	n/a	n/a	n/a		n/a	n/a
Salt, 0.1% inv	n/a	n/a	/	n/a		$I_{(S,0.1\%)}<I_{(S,1\%)}$
Salt, 1% inv	n/a	p <0.001	p <0.001	n/a	p <0.001	

Functionality

	No Salt 0% inv	No Salt 0.1% inv	No Salt 1% inv	Salt 0% inv	Salt 0.1% inv	Salt 1% inv
No Salt, 0% inv		$F_{(NS,0\%)}>F_{(NS,0.1\%)}$	$F_{(NS,0\%)}>F_{(NS,1\%)}$	$F_{(NS,0\%)}>F_{(S,0\%)}$ ★	/	/
No Salt, 0.1% inv	p <0.001		$F_{(NS,0.1\%)}>F_{(NS,1\%)}$	/	$F_{(NS,0.1\%)}<F_{(S,0.1\%)}$	/
No Salt, 1% inv	p <0.001	p <0.001		/	/	$F_{(NS,1\%)}<F_{(S,1\%)}$
Salt, 0% inv	/	/	/		$F_{(S,0\%)}<F_{(S,0.1\%)}$	$F_{(S,0\%)}<F_{(S,1\%)}$
Salt, 0.1% inv	/	/	/	p =0.048		$F_{(S,0.1\%)}<F_{(S,1\%)}$
Salt, 1% inv	/	/	/	p <0.001	p <0.001	

CC$_{community}$

	No Salt 0% inv	No Salt 0.1% inv	No Salt 1% inv	Salt 0% inv	Salt 0.1% inv	Salt 1% inv
No Salt, 0% inv		n/a	n/a	n/a	n/a	n/a
No Salt, 0.1% inv	n/a		$C_{(NS,0.1\%)}=C_{(NS,1\%)}$	n/a	$C_{(NS,0.1\%)}>C_{(S,0.1\%)}$	n/a
No Salt, 1% inv	n/a	p =0.365		n/a	/	$C_{(NS,1\%)}>C_{(S,1\%)}$
Salt, 0% inv	n/a	n/a	n/a		n/a	n/a
Salt, 0.1% inv	n/a	n/a	/	n/a		$C_{(S,0.1\%)}=C_{(S,1\%)}$
Salt, 1% inv	n/a	p <0.001	p <0.001	n/a	p =0.609	

Figure 2-5. Invasion, community functionality and community growth in the absence and presence of salinity stress. The individual plots show invasion (a and d), its effect on functionality (b and e) and the growth of the community (c and f) in relation to the Gini coefficient in the absence (a, b and c) and presence (d, e and f) of salinity stress. Black crosses, blue triangles and red squares indicate the residuals at an initial invader concentration of 0%, 0.1% and 1%, respectively, after correction for the row, column and plate effects. The axis of the residuals of the functionality and invasion were rescaled to obtain values between 0 and 1. Black (0%), blue (0.1%) and red (1%) lines show the fits after model selection (Table 2-2).

4. Discussion

Several previous studies attempted to correlate invasion with the composition of the invaded community and some of these studies experimentally addressed the effect of invasion on ecosystem functioning (Sousa *et al.*, 2011). In our work, the degree of evenness was negatively correlated with the susceptibility of the community to invasion in the absence of an external stress, as previously shown in grassland communities (Wilsey & Polley, 2002; Tracy & Sanderson, 2004; Fink & Wilson, 2011). However, under stress conditions, the evenness-invasibility relationship became neutral, and the invasion potential was only associated with the initial amount of the invader. The effect of invasion on the functionality depended strongly on the environmental conditions: the functionality at a high Gini coefficient was enhanced under stress and lowered under non-stress conditions.

The diversity-invasibility hypothesis states that a high species richness confers a high degree of invasion resistance (Kennedy *et al.*, 2002). However, it has also been reported that a positive relationship between diversity and invasibility can occur (Jiang & Morin, 2004; Dunstan & Johnson, 2006; Davis, 2009). In the present study, we found that, even at a fixed high richness, the relative abundance of the species (i.e., evenness) is an essential factor that determines the invasibility of the community (Figure 2-5a). This effect can be explained by the potential niche overlap for carbon sources (electron donor) between the community and the invader. As shown in Figure 2-6, the niche overlap (consumption of carbon sources) was much higher with an even community as compared to most of the uneven mixes. This means that with an uneven community in a nutrient rich environment with a lot of different carbon sources, like TSB, more carbon sources are present that can solely be used by the invader. Therefore, invasion is facilitated in uneven communities. In addition to biodiversity, environmental stress influenced the invasion potential. In fact, the level of invasion in two identical communities (same richness, evenness and species composition) in the presence of the same resources depended strictly on the presence of a stress.

CHAPTER 2
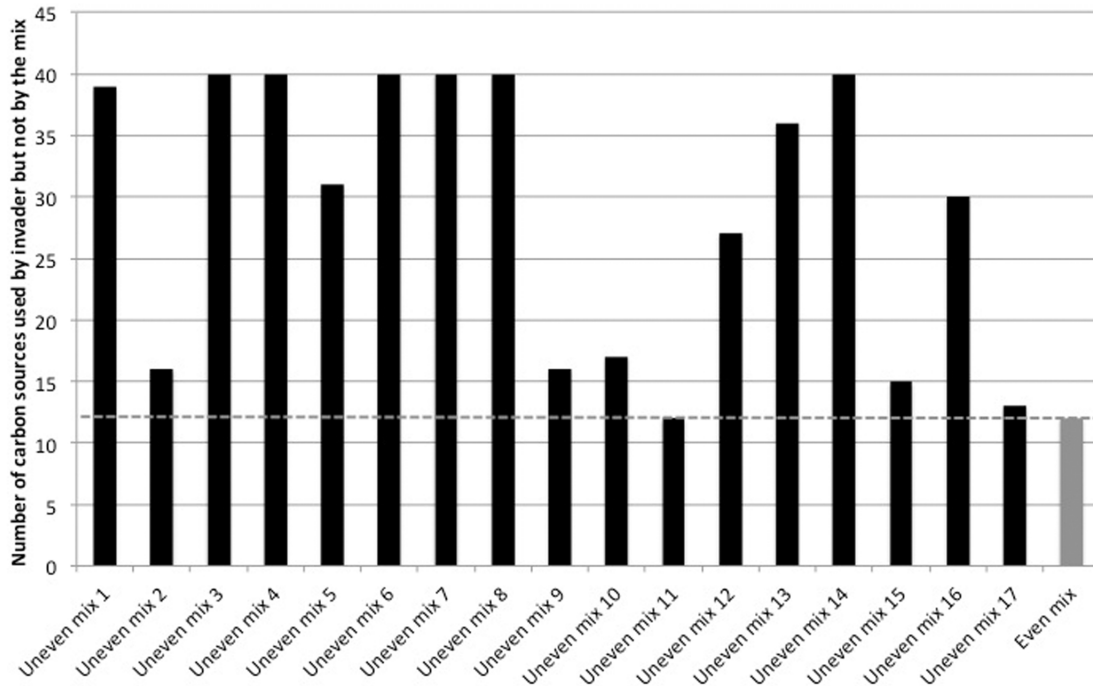
Figure 2-6. Niche overlap of the available carbon sources. *Number of carbon sources available for the invader that are not used by the even (grey bar)/uneven (black bars) mixes. The grey dotted line represents the level of carbon consumption of the even mix. The amount of available carbon sources for the invader – which was able to use 40 out of 95 carbon sources - was much lower with an even community as compared to most of the uneven mixes.*

Current debate also focuses on the effect of the invasion of an ecosystem on the functioning of the system (Davis *et al.*, 2011; Lambertini *et al.*, 2011; Simberloff *et al.*, 2013). In this study, we confirmed that the degree of evenness is a key element for preserving the overall functionality in a closed community (Wittebolle *et al.*, 2009). In fact, the functionality of an even community could be maintained under both stressed and non-stressed conditions. Conversely, the functional stability of a highly uneven community is endangered by salinity stress (Figure 2-5e, black line). The same effect on functionality was also observed in an open community. In this situation, however, the overall functionality of an uneven community was preserved at high levels of the invader, although the introduced species, a non-denitrifier, made no direct contribution to the existing functionality (Figure 2-5e, red line). This result suggests that a closed system is sensitive to functionality loss when exposed to fluctuating environmental conditions (Curtis & Sloan, 2004; Naeem, 2009) (Figure 2-7, bottom right). An open system in which non-native species have the ability to invade the resident community may be more resistant to stress than a closed system and can maintain or even improve its functionality when an invader is present.
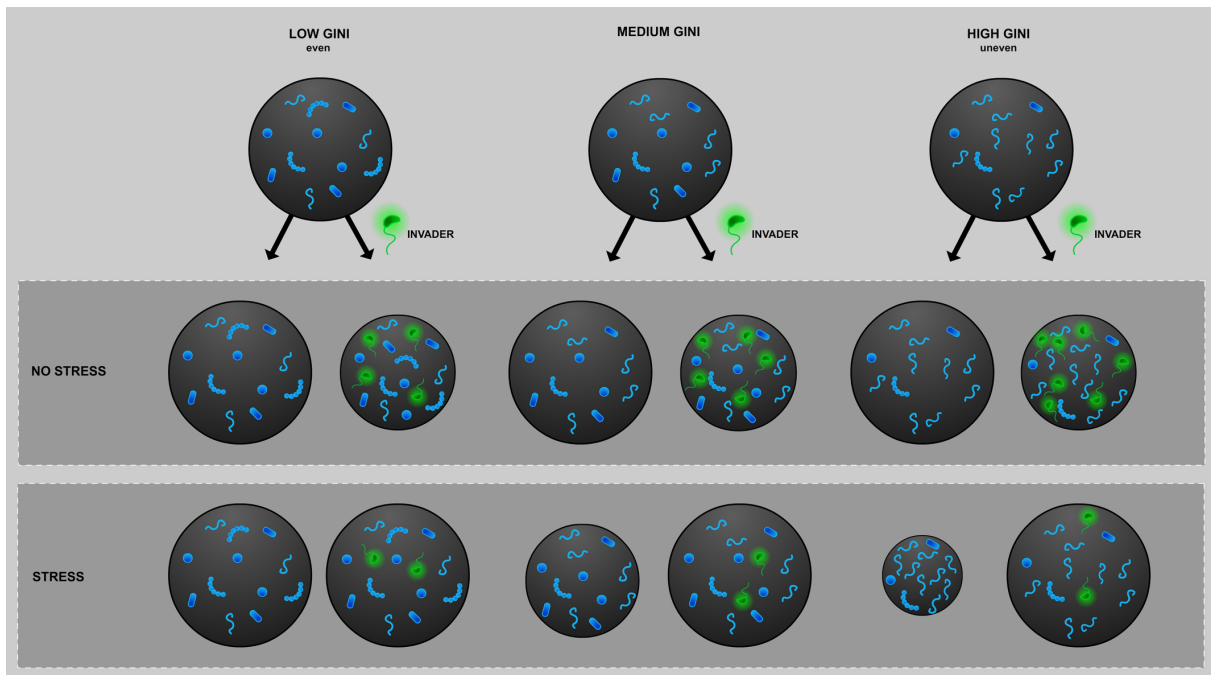
Figure 2-7. Graphical summary of the main observations. *Microcosms, composed of seventeen denitrifying strains (blue cells, different shapes), have been assembled with different levels of initial evenness (different proportion of cells), from a low to high Gini coefficient. These microcosms were challenged with a non-denitrifying invader (green cells) under stressed and non-stressed conditions. The ecosystem functionality (size of circles) and invasibility (number of green cells) were measured. Invasion was generally lower in the presence of salt stress than in the absence of salt. Under no-stress conditions, the presence of the invader negatively affected the community and its functionality, whereas the same relationship was positive under stress conditions.*

5.   Conclusions

Our aim was to evaluate what could be the impact of human activities - that tend to modify the composition of natural communities - on invasibility and functionality under different environmental conditions. We showed that the same community, with the same genetic background, in the presence of the same invader behaved differently depending on specific environmental conditions and its evenness. Despite our conclusions rising from these observations are firstly pertinent to our ecological system, we suggest to consider it as a possible interpretation model that could have analogies in other systems including the macro-ecology systems. Under stress conditions an invasive species can preserve the indigenous functionality, while under non-stress conditions the functionality can be threatened. In the latter case evenness plays a crucial role in determining the community resistance to invasibility and in preserving ecosystem functionality. Thus, on the one hand, invasion can support ecosystem resilience and services (Davis *et al.*, 2011), while on the other hand, it can be considered negative towards conservation biology (Lambertini *et al.*, 2011; Simberloff *et*

al., 2013). Hence, in case of putative imposition of ecological management, both potential impacts of invasion warrant careful consideration.

6.   Acknowledgements

CHAPTER

3

SPECIES RELATEDNESS DETERMINES THE SUCCESS OF MICROBIAL MUTUALISTIC INTERACTIONS

SPECIES RELATEDNESS DETERMINES THE SUCCESS OF MICROBIAL MUTUALISTIC INTERACTIONS

Abstract

Relatedness has been shown to be a key reason for individuals to perform an altruistic behaviour, both for microorganisms and higher organisms. Its importance in mutualistic interactions - which is a two-way beneficial interaction - has however never been studied in microbial communities. In this study, synthetic microbial communities were used to test the importance of interspecies relatedness during the establishment of mutualism between previously non-interacting microorganisms. Obligatory mutualism was created by making pairwise combinations of an ampicillin resistant, tyrosine auxotrophic *E. coli* and ampicillin sensitive strains. Initially, no cooperation could be detected in any of the mixes. Closely related species were able to adapt their phenotype after two transfers and successfully established mutualisms, while distantly related species were not able to support growth of the other strain and were consequently threatened with extinction.

1.   Introduction

Microorganisms are the main drivers of numerous biogeochemical cycles on earth (Schmidt, 2006). These processes are established by the joint effort of microorganisms with different functional roles. They cooperate and are often dependent on each other for survival or growth. By interacting, microorganisms are able to occupy niches that would not be accessible otherwise (Stolyar *et al.*, 2007). They can provide each other with (essential) metabolites, e.g. amino acids and siderophores (West & Buckling, 2003), or with nutrients (Bull & Harcombe, 2009; Estrela *et al.*, 2012) or metabolise antibiotics like ampicillin (Dugatkin *et al.*, 2005). When such an act is costly to the actor but beneficial to the recipient this is called altruism, while mutually beneficial acts are beneficial to both the actor and recipient (West *et al.*, 2007a). Mutualism is especially important when multiple species are involved in the degradation of organic compounds (Stolyar *et al.*, 2007). A more stringent form of mutualism is obligatory mutualism, meaning that species rely on each other for growth. This mainly occurs among anaerobic bacteria (Schink, 2002). However, more mutualistic interactions may evolve due to environmental fluctuations (Klitgord & Segre, 2010). Environmental changes, like antibiotics in the human gut or toxins in the soil, may require microorganisms to start cooperating for survival.

The development of mutualism has mainly been studied between mutants and genetically engineered populations of the same strain. It has been shown both with auxotrophs of *Escherichia coli* (Hosoda *et al.*, 2011) and auxotrophs of *Saccharomyces cerevisiae* (Shou *et al.*, 2007) that organisms have the potential to adapt their phenotype after the first encounter with another organism to establish mutualism before extinction. Furthermore, some auxotrophic pairs were shown to be metabolically complementary, while others were not able to survive (Wintermute & Silver, 2010). The effect of relatedness between non-interacting microorganisms on the development and establishment of a mutualistic cooperation has however never been studied, while relatedness has been shown multiple times to be the key reason for an individual to perform a costly altruistic behaviour that is beneficial to another individual (Griffin *et al.*, 2004; West *et al.*, 2007b): by helping a related individual reproduce, it is indirectly passing on its own genes to the next generation (Hamilton, 1963; Hamilton, 1964; West *et al.*, 2006).

To explore the effect of phylogenetic relatedness on the development of a mutualistic interaction, obligatory mutualism was created between a tyrosine auxotrophic, ampicillin resistant *E. coli* and a set of closely and distantly related ampicillin sensitive strains.

The ampicillin resistant tyrosine auxotrophic *E. coli* carries the plasmid pAK5 that contains the *bla*$_{TEM-1}$ gene that confers the ampicillin resistance (Lee & Keasling, 2005). It encodes ß-lactamase TEM-1, this enzyme cleaves the ß-lactam ring of ampicillin, and this way inactivates ampicillin. ß-Lactamases produced by Gram-negative bacteria are generally retained in the periplasm, such that only the cells producing it or cells in close vicinity are protected from ampicillin. However, membrane permeability can cause ß-lactamase to leak out of the cells, causing a declined ampicillin concentration in the environment and as such, protect all cells (Petrosino *et al.*, 1998; Dugatkin *et al.*, 2005).

Tyrosine is an aromatic amino acid that is required for bacterial growth. Bacteria have the metabolic potential to synthesize this aromatic compound from simple carbon sources. As the biosynthesis of tyrosine and other aromatic amino acids is metabolically costly, bacterial cells do not over-produce these compounds naturally. They have a regulatory mechanism for producing only the amount necessary for bacterial growth. The enzyme chorismate mutase/prephenate dehydrogenase, encoded by the *tyrA* gene, has two enzymatic functions in the biosynthesis pathway and is essential for the synthesis of tyrosine (Lutke-Eversloh & Stephanopoulos, 2007; Chavez-Bejar *et al.*, 2012). Next to the ability to produce tyrosine, bacteria can also take up tyrosine using two systems: the general aromatic transport system (AroP), encoded by the aroP gene, and the specific transport system (TyrP), encoded by the *tyrP* gene. However excretion of tyrosine is limited, it can happen by simple diffusion or by the aromatic amino acid transporter YddG (Burkovski & Kramer, 2002; Doroshenko *et al.*, 2007). The auxotrophic knockout mutant used within this study is lacking the *tyrA* gene and is not able the produce tyrosine. For growth, it is dependent on the supply of tyrosine from its environment.

Pairwise combinations of the tyrosine auxotroph and ampicillin sensitive strains were made in minimal medium with ampicillin but lacking tyrosine. In this environment, the ampicillin sensitive strain provides the auxotroph with tyrosine, while the auxotroph degrades ampicillin and thus protects the sensitive strains. Only when cooperating, both strains are able to survive.

2. Material and methods

2.1. Strain info

E. coli Y3 (K12 *ΔtyrA::kanR intC::yfp* pAK5) is a tyrosine auxotrophic knockout mutant lacking the *tyrA* gene and chromosomally tagged with *yfp,* this strain originates from *E. coli*

K12. The pAK5 plasmid carries the bla_{TEM-1} gene that confers ampicillin resistance (Kerner *et al.*, 2012). This plasmid originates from plasmid pPRO24 (Kerner *et al.*, 2012), which was derived from the plasmid pBAD24 (Lee & Keasling, 2005). pBAD24 is a commercial expression vector and does not contain an origin of transfer or transfer genes. No plasmid transfer has been reported within earlier works.

The partner strains, all soil isolates indicated with P and a number, are ampicillin sensitive. For all strains, partial 16S rRNA gene was amplified using 27F and 1492R primers (Hendrickx *et al.*, 2006a). PCR was performed as previously described and PCR products were sequenced using Sanger sequencing (Hendrickx *et al.*, 2006a) (Table 3-1). The strains were identified using BLAST and the phylogenetic distance to the auxotroph calculated using maximum likelihood. All strains were stored in ready-to-use aliquots at -80°C.

2.2.   Growth experiments

The strains were cultured for 48 hours and sub-cultured for 40 hours in Luria Bertani medium (10 g L^{-1} trypton, 5 g L^{-1} yeast extract and 5 g L^{-1} NaCl) under aerobic conditions at 28°C to obtain actively growing microorganisms. After cultivation, the strains were washed three times with physiological solution (5000 × g, 7 minutes) and the pellet was suspended in minimal M9 medium (48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8 mM NaCl, 19 mM NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 5 mM glucose). The number of intact cells was measured by flow cytometry (Accuri C6, BD) after a viability staining with propidium iodide and SYBR® Green I. Dilutions were made to obtain 10^7 intact cells mL^{-1}. Subsequently, pairs were made by mixing each partner strain with the auxotroph in equal abundances. These pairs and the individual strains were used to assemble microcosms in microwell plates, they were 1:1 diluted in M9 medium, optionally supplemented with double concentrations of ampicillin and/or tyrosine. The initial cell count in each mixture was thus $5×10^6$ cells mL^{-1}. The final volume in each well, with a maximum volume of 300 μL, was 200 μL. The plates were incubated at 28°C, every hour the optical density (620 nm) was measured. After incubation, the proportion of auxotrophic (yellow fluorescent cells) and partner cells was determined by flow cytometry with an Accuri C6 flow cytometer (BD Biosciences). This flow cytometer was equipped with a 488 nm solid-state laser and signals were detected in fluorescent channels FL1 and FL2, respectively equipped with a 510/15 nm and 540/20 nm bandpass filter. Prior to flow cytometric analysis, the mixtures were diluted 100 times in physiological solution (8.5 g L^{-1} NaCl) to obtain cell numbers within the detection range of the flow cytometer.

Table 3-1. List of partner strains and the auxotroph. *The species were identified after sequencing of the 16S rRNA genes and BLAST analysis. The best match in the BLAST analysis and the corresponding identity scores (similarity) are shown. The phylogenetic distance to the auxotroph was calculated by maximum likelihood estimation.*

Strain	Best match in BLAST analysis	Identity scores (Similarity)	Phylogenetic distance to the auxotroph	Phylum
P1	*Escherichia/Shigella flexneri*; X96963	0.960	0.007	Ɣ Proteobacteria
P2	*Bacillus thuringiensis*; IAM 12077; D16281	0.986	0.243	Firmicutes
P3	*Escherichia/Shigella flexneri*; X96963	0.999	0.000	Ɣ Proteobacteria
P4	*Rhizobium daejeonense*; L61T; KCTC 12121; AY341343	0.994	0.201	α Proteobacteria
P5	*Paracoccus pantotrophus*; ATCC 35512T; Y16933	0.960	0.241	α Proteobacteria
P6	*Escherichia/Shigella dysenteriae*; X96966	0.956	0.007	Ɣ Proteobacteria
P7	*Escherichia/Shigella dysenteriae*; X96966	0.985	0.000	Ɣ Proteobacteria
P8	*Escherichia/Shigella dysenteriae*; X96966	0.963	0.007	Ɣ Proteobacteria
P9	*Rhizobium radiobacter*; IAM 12048; AB247615	0.927	0.213	α Proteobacteria
P10	*Rhizobium daejeonense*; L61T; KCTC 12121; AY341343	0.984	0.197	α Proteobacteria
Auxotroph	*Escherichia/Shigella flexneri*; X96963	0.990	0.000	Ɣ Proteobacteria

CHAPTER 3

Subsequently, pairs were made by mixing each partner strain with the auxotroph in equal abundances. These pairs and the individual strains were used to assemble microcosms in microwell plates, they were 1:1 diluted in M9 medium, optionally supplemented with double concentrations of ampicillin and/or tyrosine. The initial cell count in each mixture was thus 5×10^6 cells mL^{-1}. The final volume in each well, with a maximum volume of 300 µL, was 200 µL. The plates were incubated at 28°C, every hour the optical density (620 nm) was measured. After incubation, the proportion of auxotrophic (yellow fluorescent cells) and partner cells was determined by flow cytometry with an Accuri C6 flow cytometer (BD Biosciences). This flow cytometer was equipped with a 488 nm solid-state laser and signals were detected in fluorescent channels FL1 and FL2, respectively equipped with a 510/15 nm and 540/20 nm bandpass filter. Prior to flow cytometric analysis, the mixtures were diluted 100 times in physiological solution (8.5 g L^{-1} NaCl) to obtain cell numbers within the detection range of the flow cytometer.

For the partnership tests, microcosms with only the auxotrophic strain and microcosms with the partner-auxotrophic pairs, created as described above, were incubated during 47 hours at 28°C. After incubation, the number of auxotrophic and partner cells was analysed by flow cytometry and 10% of the liquid cultures was transferred to fresh medium supplemented with ampicillin and again incubated. This procedure was repeated five times.

Sensitivity of the partner strains to ampicillin

After culturing and sub-culturing, the partner strains were grown in M9 medium supplemented with 0, 1, 5, 10, 20 or 100 mg L^{-1} ampicillin in multiwell plates. The plates were incubated at 28°C, every hour the optical density (620 nm) was measured.

2.3. Supernatant tests

Supernatant of the partner strains

The partner strains and auxotroph were grown respectively in M9 and LB medium after being cultured and sub-cultured. After 40 hours incubation, the partner strains were centrifuged (5000 × g, 7 min) and the supernatant filtered (0.22 µm). The auxotroph was washed three times with physiological solution and the pellet was suspended in M9. The filtered supernatant of the partner strains was supplemented with glucose and ampicillin to a final concentration of 5 mM and 100 mg L^{-1}, respectively. The partner supernatant was (i) inoculated with the *yfp*-tagged auxotrophic strain, (ii) not inoculated (negative control) and

(iii) supplemented with different concentrations of tyrosine, ranging from 0.01 µg mL^{-1} to 100 µg mL^{-1} and inoculated with the auxotrophic strain (positive controls). Growth curves were made in duplicate for all these supernatant cultures and after incubation, cell numbers were determined by flow cytometry.

Supernatant of the auxotroph

After culturing and sub-culturing, the partner strains and auxotroph were respectively grown in LB medium and M9 medium. The M9 medium was supplemented with 100 mg L^{-1} ampicillin. After 15 hours, 20 hours and 40 hours of incubation, the auxotroph was centrifuged (5000 × g, 7 min) and the supernatant filtered (0.22 µm). These time-points correspond to the logarithmic, early stationary and late stationary phase of the auxotroph. After 40 hours incubation, the partner strains were washed three times with physiological solution and the pellets suspended in M9. The filtered supernatants of the three time-points were supplemented with glucose to a final concentration of 5 mM and (i) inoculated with the partner strains or (ii) not inoculated (negative control). Growth curves were made in duplicate for all these supernatant cultures.

3.   Results

3.1.   Characterisation of the auxotroph and partner strains

The 'auxotroph', which is a *yfp*-tagged, ampicillin resistant and tyrosine auxotrophic *E. coli,* was not able to grow in absence of tyrosine. A minimal concentration of 0.5 µg mL^{-1} tyrosine was necessary to obtain growth. A logarithmic relation between tyrosine concentration and maximal optical density was observed, with saturation at higher tyrosine levels. Concentrations higher that 20 µg mL^{-1} did not result in additional growth (Figure 3-1).

A

B
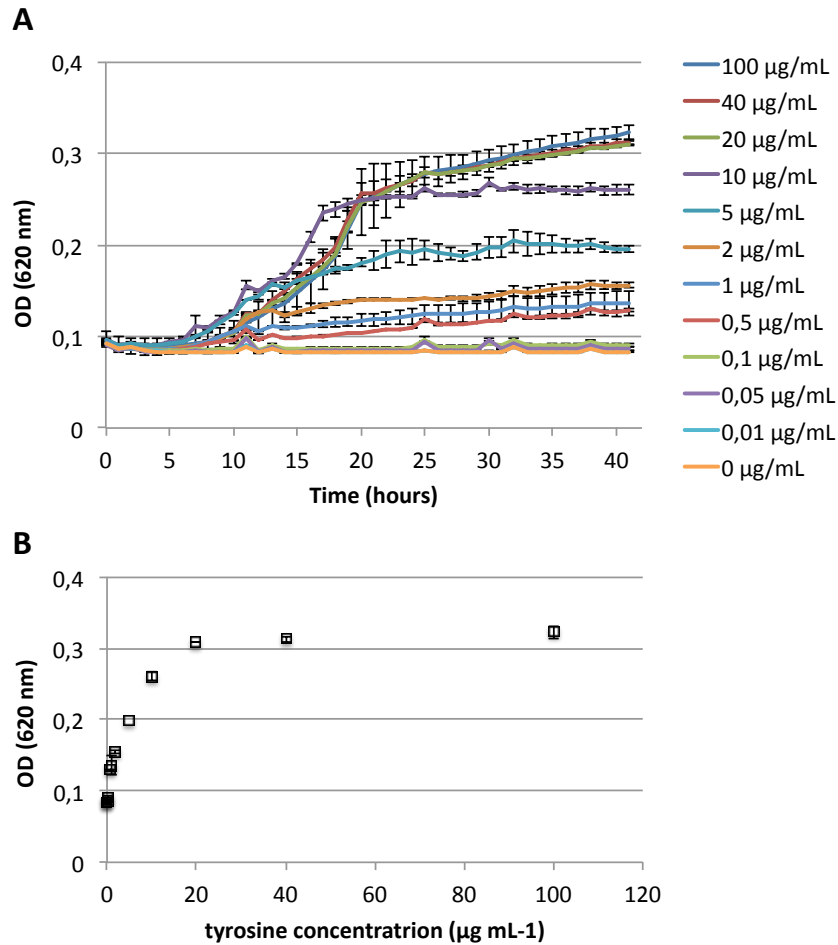
***Figure 3-1. Growth curves of the auxotrophic strain.** (A) Growth curves of the auxotrophic strain at different tyrosine concentration, ranging from 0 μg mL^{-1} to 100 μg mL^{-1}. The optical density (OD$_{620}$) was measured every hour during 40 hours. (B) The relation between the maximum OD at different tyrosine concentration in relation to the tyrosine concentration. All tests were performed in duplicate (n=2), standard deviations are indicated by error bars.*
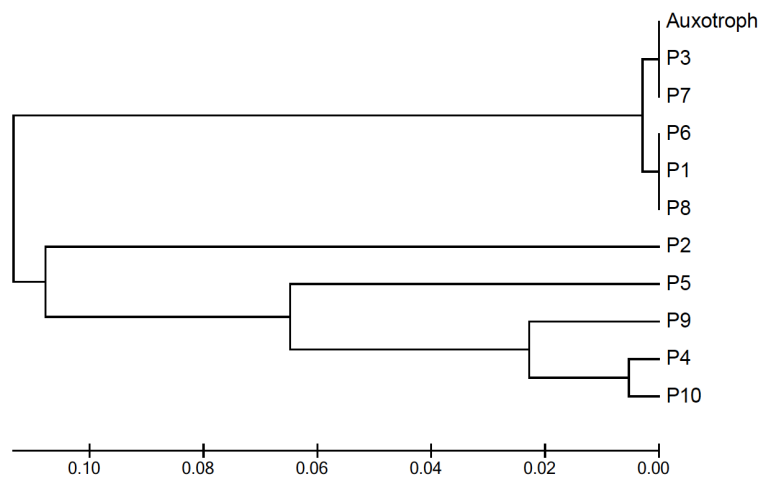
***Figure 3-2.** Phylogenetic tree based on the 16S rRNA gene of the partner strains and the auxotroph. Phylogenetic distances are calculated using maximum likelihood.*

Ten ampicillin sensitive strains (referred to as partner strains) were selected based on the genetic relatedness to the auxotroph, five are closely related to the auxotroph, while five are more distantly related (based on the similarity of the 16S rRNA gene) (Table 3-1) (Figure 3-2). None of these partner strains were able to grow in minimal medium with 100 mg L^{-1} ampicillin. Growth in presence of 5 mg L^{-1} ampicillin was already significantly lower for all strains (except for P6) compared to the situations without ampicillin. All strains, except P6, are approximately as sensitive to ampicillin. The growth of P6 is not affect by a concentration lower than 20 mg L^{-1} (Figure 3-3).
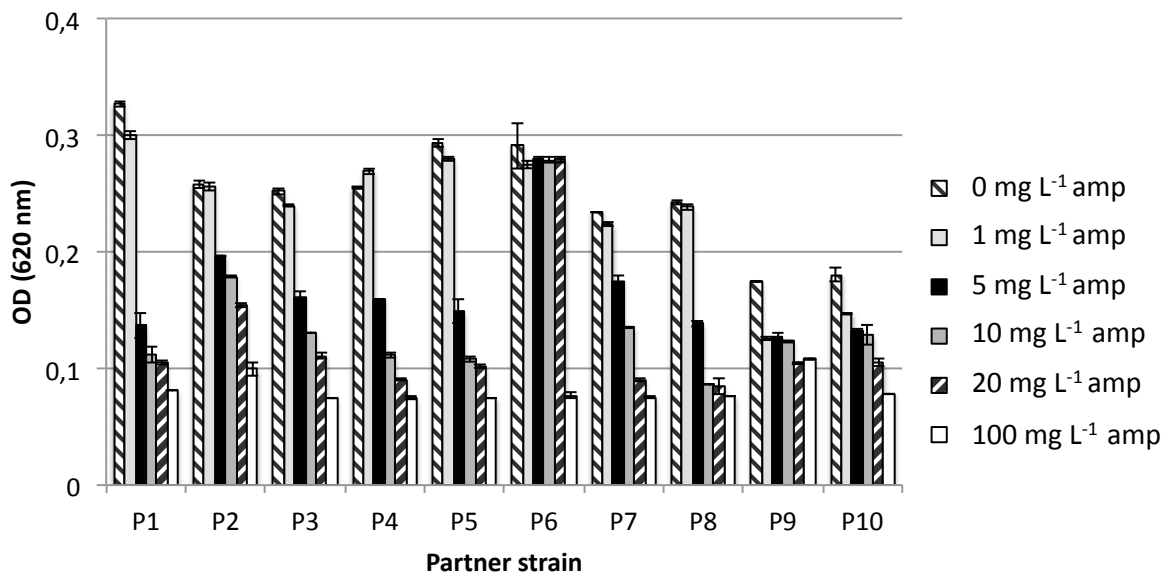
Figure 3-3. Partner growth M9 medium with different concentrations ampicillin. *The bars show the optical density (620 nm) of the different partner strains after 42 hours incubation in (i) minimal medium supplemented with 0 mg L^{-1} ampicillin (◨), 1 mg L^{-1} ampicillin (▢), 5 mg L^{-1} ampicillin (■), 10 mg L^{-1} ampicillin (▦), 20 mg L^{-1} ampicillin (◪) and 100 mg L^{-1} ampicillin (▢). (mean ± SD, n=2)*

To determine if the auxotroph degrades a sufficient amount of ampicillin to allow growth of the partner strains, the supernatant of the auxotroph was sampled at three time points and was inoculated with the partner strains. It was tested if the partner strains could grow in this supernatant. After 15 hours, which corresponds to the logarithmic growth phase of the auxotroph, already enough ampicillin was removed by the auxotroph to allow growth of all partner strains. This growth was higher when supernatant was sampled at later time points (Figure 3-4). Surprisingly, the growth of partner strain 7 and 8 was higher in the supernatant sampled at later time points compared to M9 without ampicillin. This indicates that the ampicillin concentration was below the inhibitory concentration and growth-promoting compounds were produced by the auxotroph.
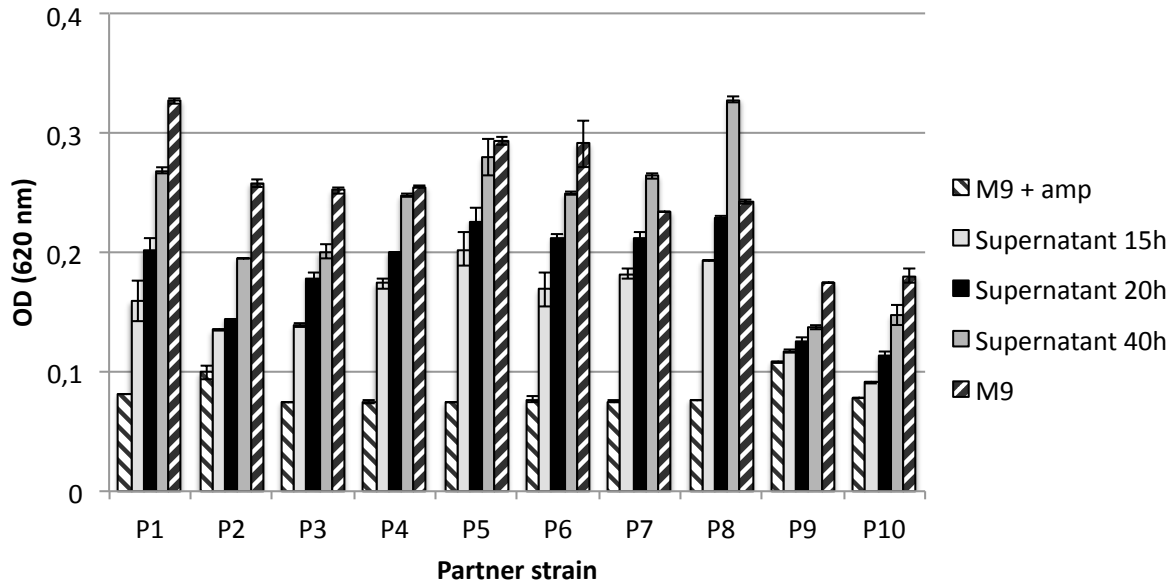
Figure 3-4. Partner growth in supernatant of the auxotroph and minimal medium. *The bars show the optical density (620 nm) of the different partner strains after 42 hours incubation in (i) minimal medium supplemented with 100 mg L^{-1} ampicillin (▨), (ii) supernatant of the auxotroph sampled after 15 hours (▢), 20 hours (■) and 40 hours incubation (▨) and (iii) minimal medium (▨). (mean ± SD, n=2)*
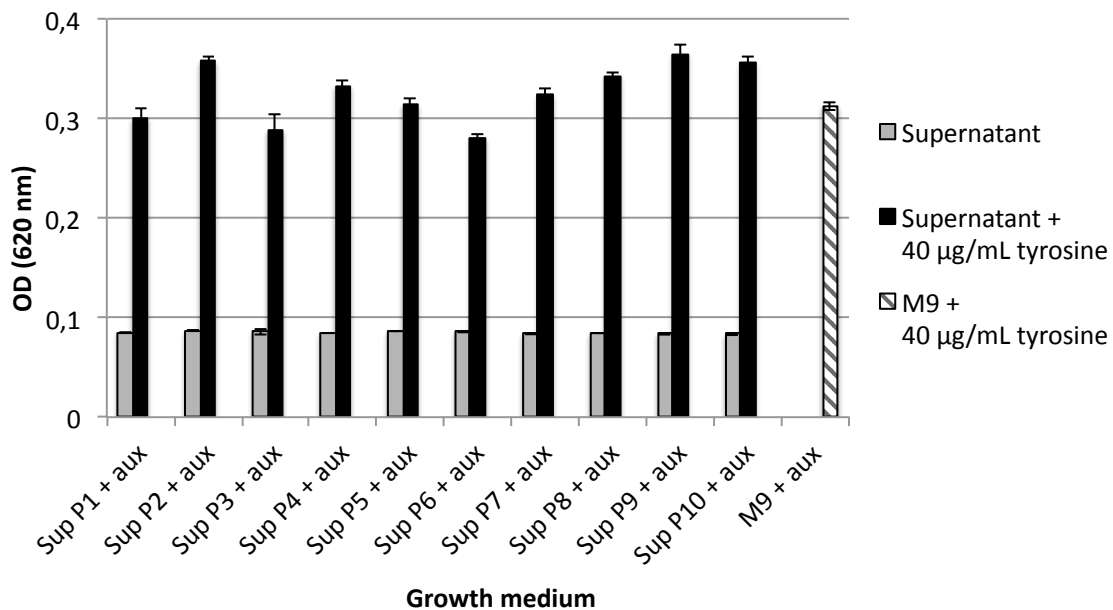
Figure 3-5. Auxotroph growth in supernatant of the partner strains. *The bars show the optical density (620 nm) of the auxotroph after 42 hours incubation in (i) supernatant of the different partner strains (▨), (ii) supernatant of the partner strains supplemented with 40 µg mL^{-1} tyrosine (■) and (iii) minimal medium supplemented with 40 µg mL^{-1} tyrosine (▨).(mean ± SD, n=2)*

The auxotroph was not able to grow in the supernatant of any of the partner strains, which was supplemented with additional glucose to prevent a lack of carbohydrates. Addition of

tyrosine to the supernatant resulted in growth similar to the growth in M9 supplemented with tyrosine (Figure 3-5). Thus, the supernatant of all partner strains neither contains enough tyrosine to allow growth of the auxotroph, nor does it contain compounds that inhibit growth.

3.2.   Partnership growth

In minimal medium, lacking tyrosine but supplemented with 100 mg L^{-1} ampicillin, the auxotroph was washed out of the system when after incubation 10% of the liquid cultures was transferred to fresh medium. There was the expected 10 times decrease of the number of auxotrophic cells during each transfer (Figure 3-6). After three transfers, the cell count was below the detection limit of the flow cytometer, which is 10^5 cells mL^{-1}.
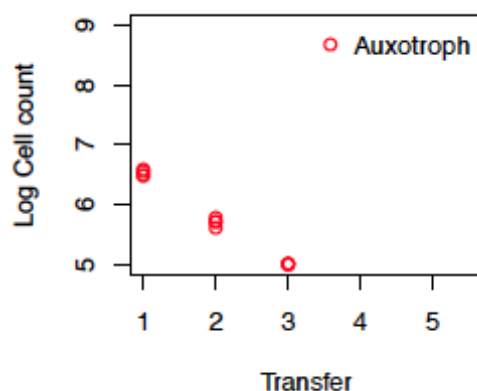
Figure 3-6. Log cell count (cells mL^{-1}) of the auxotroph after 47 hours incubation after the different transfers. Each transfer, 10% of the liquid culture was transferred to fresh medium and incubated. Cell counts were measured by flow cytometry. Tests were performed in quadruplicate (n=4).
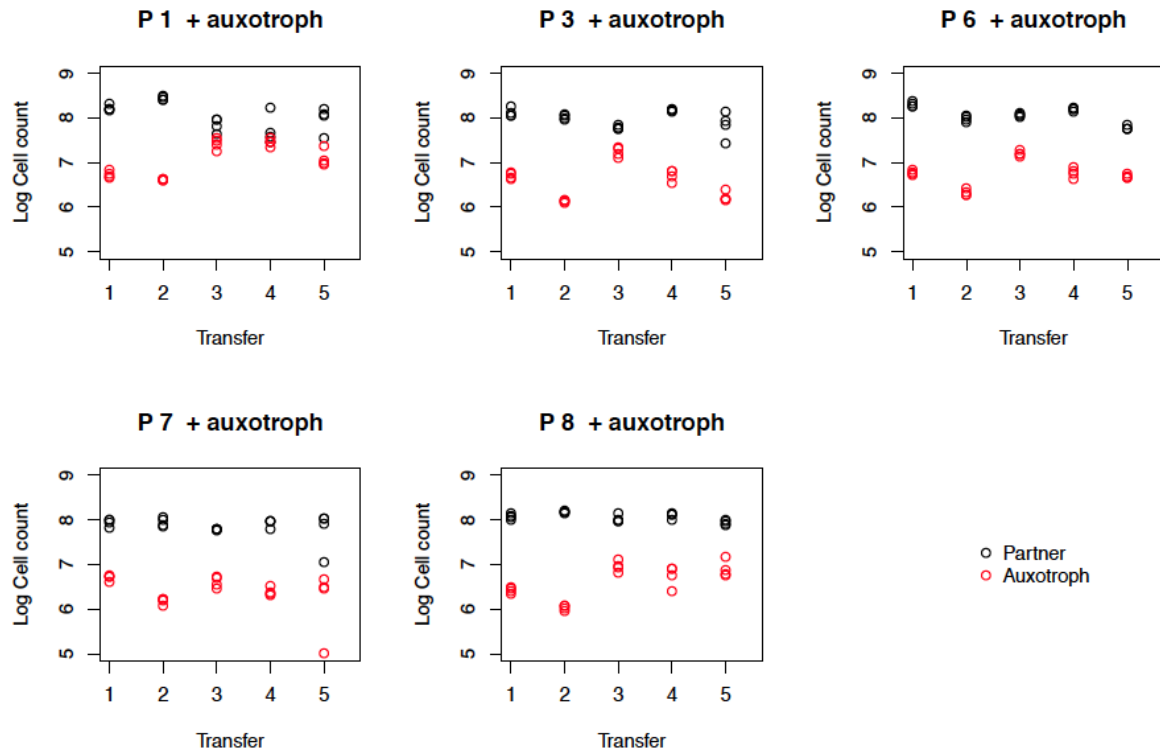
Figure 3-7. Log cell counts of the auxotroph and closely related partner strains after pairwise incubations. *Pairwise combination of the closely related partner strains and auxotroph were incubated in quadruplicate (n=4) after which 10% was transferred to fresh minimal medium and the number of partner (○) and auxotrophic cells (○) analysed by flow cytometry. The process of incubation and transfer was performed five times.*

Pairwise combinations of the auxotroph and the partner strains were made in the same environment (Figure 3-7 and Figure 3-8). During the first incubation period, before the first transfer, the auxotroph did hardly grow in any of the mixes. The cell number of the auxotroph was for all pairs below 10^7 cells mL^{-1}, while the initial cell count was 2.5×10^6 cells mL^{-1}. Most partner strains were however able to grow within this period. The cell counts of the auxotroph were lower after the second incubation period compared to the first incubation period for all mixes. Taking into account the ten-fold dilution, this corresponds to minor or no growth. During the third incubation period, an increase of auxotrophic cells could be detected for the mixes with partner 1, 3, 6, 7, 8 and 10. For partner 10, this was however still very low. For partner 2, 4, 5 and 9 growth was always less than ten-fold, resulting in a decreasing cell number over the different transfers, eventually followed by a washout of the auxotroph.
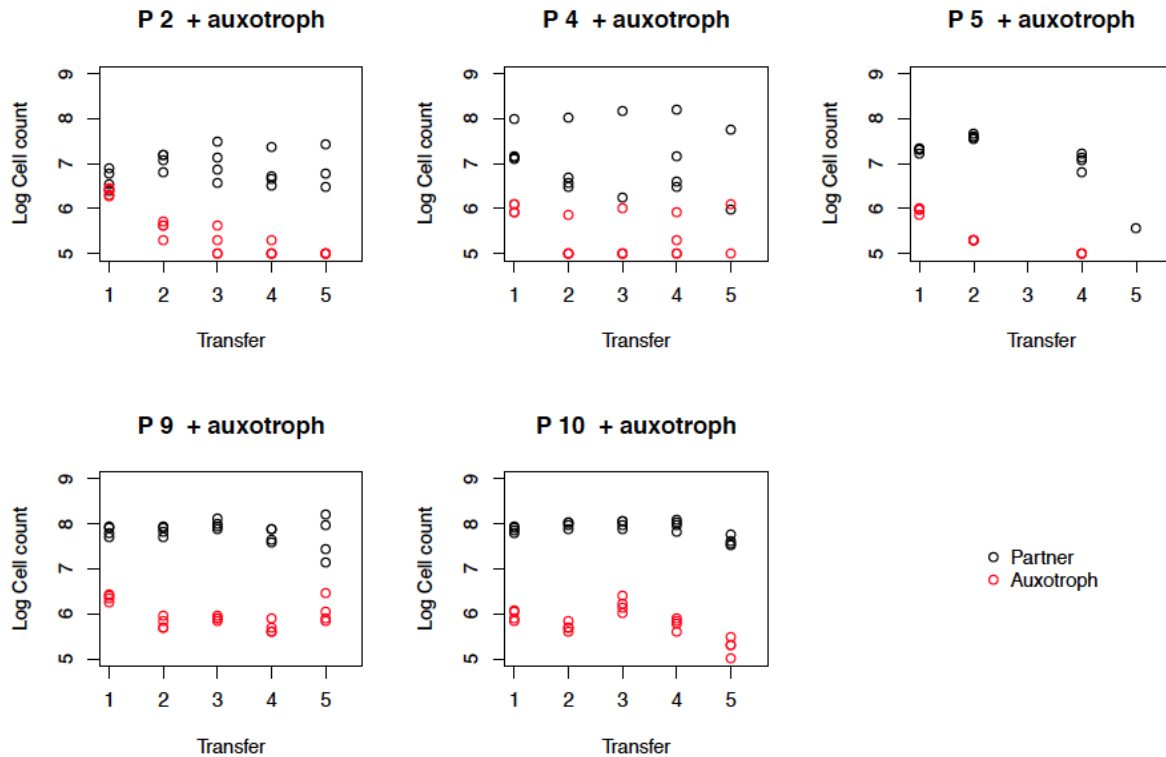
Figure 3-8. Log cell counts of the auxotroph and distantly related partner strains after pairwise incubations. *Pairwise combination of the distantly related partner strains and auxotroph were incubated in quadruplicate (n=4) after which 10% was transferred to fresh minimal medium and the number of partner (○) and auxotrophic cells (○) analysed by flow cytometry. The process of incubation and transfer was performed five times. (Data points for partner 5 after the third transfer are missing due to analytical problems)*

3.3. Cooperation-relatedness

The fitness of the auxotroph over the five transfers was used as a measure of the cooperation potential of the partner strains. The fitness equals the reproduction capacity; the total fitness is thus calculated by making the summation of the produced biomass during the different incubations and transfers:

$$Total\ fitness = (Cell\ count_1 - 2.5{\times}10^6) + \sum_{i=2}^{5}\left(Cell\ count_i - \frac{Cell\ count_{i-1}}{10}\right)$$

At the 5% level of significance, cooperation with strains closely related to the auxotroph resulted in a significantly higher fitness than with strains distantly related (p<0.0001 with the two-sided Wilcoxon rank sum test, Figure 3-9). Indicating that cooperation is more successful between closely related species.
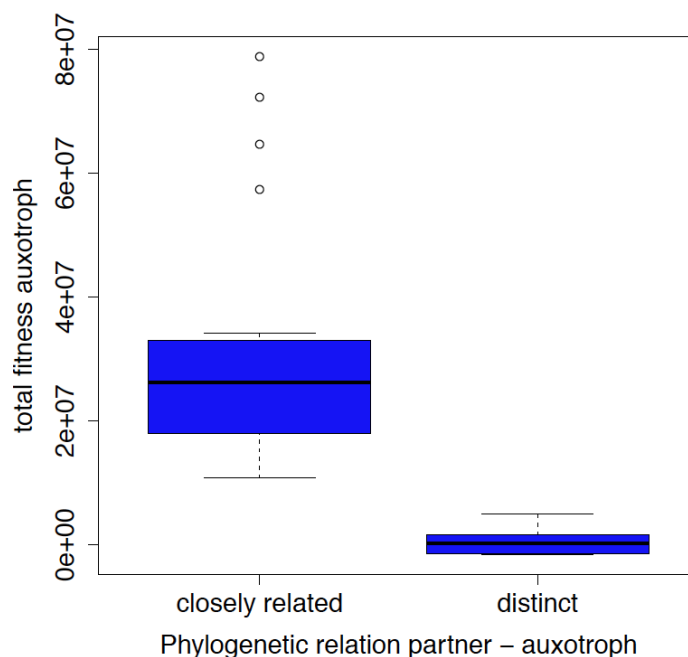
Figure 3-9. Boxplots of the total fitness of the auxotroph. *The boxplots are shown for the auxotroph incubated with closely or distantly related partner strains.*

4. Discussion

In this paper, the role of phylogenetic relatedness on the establishment of a mutualistic interaction was investigated. For this, synthetic obligatory mutualistic communities were created with on the one hand ampicillin sensitive strains and on the other hand a tyrosine auxotrophic ampicillin resistant *E. coli*. In an environment lacking tyrosine, the auxotroph is dependent on the other strain for the supply of the amino acid tyrosine, as this strain is not able to produce it itself. Thus, only when both strains cooperate they will be able to survive and grow in an environment with ampicillin and lacking tyrosine. Therefore growth was followed over several serial passages and the final cells counts measured by flow cytometry. The total fitness of the auxotroph was used as a measure of cooperation.

The auxotroph was shown to need a minimal concentration of 0.5 µg mL^{-1} tyrosine to allow growth. It was not able to grow in the supernatant of any of the partner strains and thus, none of the partner strains produced or secreted a sufficient amount of tyrosine when grown individually in the absence of ampicillin. A monoculture of the auxotroph on the other hand degraded sufficient ampicillin to make it possible for the partner strains to grow. Consequently, the partner strains are here the limiting factor in the evolution of a mutualistic interaction. Only the presence of the partner strains is not enough for the auxotroph to grow. The partner strains need to adapt their phenotype and produce/secrete more tyrosine to

support growth of the auxotroph (Hosoda *et al.*, 2011). If not, the auxotroph will be washed out of the system and the ampicillin concentrations will remain too high for the partner strains to grow, leading to extinction of both the auxotroph and partner strain. On possible way to alter their phenotype is by a stringent response to environmental stress (Miethke *et al.*, 2006; Kümmerli *et al.*, 2009; Persky *et al.*, 2009). The stringent response is a regulatory mechanism that enables bacteria to adapt to stress by the production of signal molecules, called alarmones. These alarmones increase the transcription of genes involved in the biosynthesis of amino acids (Manuel *et al.*, 2012). The overproduction of tyrosine as a stringent response to stress has already been reported in *E.coli* (Santos *et al.*, 2012). Other possibilities for bacteria to overproduce tyrosine is by eliminating the transcriptional regulation of genes involved in the tyrosine production. For example by deletion or mutation of the *tyrR* gene. This gene encodes the tyrR protein which regulates the transcription of the *tyrA* gene (Lutke-Eversloh & Stephanopoulos, 2007).

Five strains that are closely related to the auxotroph and five that are distantly related were examined for their ability to establish a mutualistic interaction with the auxotroph. Pairwise combinations were made and followed over serial passages. During the first encounter, none of the partner strains were able to provide the auxotroph with sufficient tyrosine to allow its growth. Over several passages, some strains were able to support growth of the auxotroph, while others were not. The fitness or the number of offspring of the auxotroph determines the level of cooperation of each partner strain. Cooperation with strains closely related to the auxotroph resulted in a significantly higher fitness than with strains distantly related. For some distantly related strains, no cooperation could be detected causing extinction of both the auxotroph and partner strains. Phylogenetic relatedness was shown here to be a key factor in the partner choice. Within-species relatedness was already shown to be an important factor for partner choice in the evolution of mutualism with higher organisms (Frank, 1994). Also for altruistic behaviours in microbial communities, relatedness is a key factor (Griffin *et al.*, 2004; West *et al.*, 2006; West *et al.*, 2007b; Chuang *et al.*, 2010). As such, the observed positive relation between relatedness and mutualism could be expected. However, as related species are more likely to use the same resources, also a higher degree of competition is expected between closely related species (Hibbing *et al.*, 2010). Both strains are therefore expected to converge to a stable situation in which cooperation and competition are balanced (Allen & Nowak, 2013; Sanchez & Gore, 2013). The cell counts of the auxotroph over the different transfers gives a first indication of this spiralling to a stable situation, however this needs to be confirmed by long-term experiments.

The obtained knowledge is important from an ecological point of view. It was shown before that new mutualistic interactions might arise more readily due to environmental fluctuations (Klitgord & Segre, 2010). Environmental changes and stress in different ecosystems, like the human gut, soil or wastewater treatment plant, may require the formation of mutualistic interactions for the microbial community to remain operational and functionally stable.

The results presented here also have important implications for the development of different applications. Synthetic microbial ecosystem with two or multiple populations have been proposed a number of times as a highly controllable and tuneable system for different applications, like biodegradation or fermentation (Dunham, 2007; Shou *et al.*, 2007; Kerner *et al.*, 2012). By splitting the complex multiple pathways over a number of species, there is a division of labour. Furthermore, pathways can be more easily manipulated and problems of feedback-loops are circumvented. The knowledge gained within this work can be used for development of such synthetic systems, which are more likely to succeed when closely related species are used.

5.   Conclusions

Using synthetic microbial communities with obligatory mutualism, interspecies relatedness between previously non-interacting species was shown to be an important factor for the establishment of a successful mutualistic interaction. Closely related species resulted in a mutualistic interaction with a high fitness of the auxotroph, while most distantly related species were threatened with extinction after a few generations.

6.   Acknowledgements

FLOW CYTOMETRY FOR FAST MICROBIAL COMMUNITY
FINGERPRINTING

FLOW CYTOMETRY FOR FAST MICROBIAL COMMUNITY FINGERPRINTING

Abstract

Characterising the microbial communities is important in different domains, ranging from food and beverage production to wastewater treatment. Conventional methods, such as heterotrophic plate count, selective plating and molecular techniques, are time consuming and labour intensive. A flow cytometry based approach was developed for a fast and objective comparison of microbial communities based on the distribution of cellular features from single cells within these communities. The method consists of two main parts, firstly the generation of fingerprint data by flow cytometry and secondly a novel statistical pipeline for the analysis of flow cytometric data. The combined method was shown to be useful for the discrimination and classification of different brands of drinking water. It was also successfully applied to detect changes in microbial community composition of drinking water caused by changing environmental factors. Generally, the method can be used as a fast fingerprinting method of microbial communities in aquatic samples and as a tool to detect shifts within these communities.

1. Introduction

The microbial community of water can serve as an indicator of the general microbiological quality (Bartram *et al.*, 2003; WHO, 2011). Characterisation of this microbial community is conventionally done by heterotrophic plate counts (HPC) and selective plating to test the presence of so called "indicator organisms" of pathogens (Dufour, 2003; Allen *et al.*, 2004). The main disadvantages of these quantitative techniques are the long incubation times (two to three days) and the discrepancy between the number of cultivable and non-cultivable cells, since less than 1% of the bacteria present in aquatic environments are culturable (Amann *et al.*, 1995). Therefore, cultivation independent techniques like the fingerprinting technique DGGE has become more popular. These techniques are generally more qualitative; they provide extra information about the species present and their relative abundance, furthermore changes in community composition can be detected by these techniques. But they are less quantitative than HPC, since no cell numbers are obtained. Besides, also the molecular techniques are time consuming and labour intensive (Amann *et al.*, 1995; Dewettinck *et al.*, 2001; Seurinck *et al.*, 2005). An easy method that is increasingly used in drinking water analysis, is the quantification of adenosine tri-phosphate (ATP), the energy-carrier in viable cells (Kucnerowicz *et al.*, 1982). It is a very fast and straightforward method, but an in depth interpretation is not possible (Hammes *et al.*, 2008).

Hence, there is need for a rapid, cultivation independent, objective and easy to use method to characterise the microbial community of drinking water. A promising tool is flow cytometry. This method has already been successfully used for the analysis of total bacterial cell counts in water treatment processes (Hammes *et al.*, 2008) and the assessment of the bacterial viability in drinking water (Berney *et al.*, 2008). The short analysis times are the main advantage of this method. For some applications they can even drop below 20 minutes post sampling. Besides, a wide range of fluorescent dyes is available for analysing different bacterial features (Vives-Rego *et al.*, 2000; Wang *et al.*, 2010). The main disadvantage is however the interpretation of the flow cytometry patterns. The more flow cytometric parameters that are taken into account, the harder it gets to analyse the results. Three parameters are possible to be visualised using a 3D dot plot, but higher dimensional problems can no longer be displayed on a single graph. Multiple lower dimensional projections have to be used to represent the data. Another limitation is the subjective handling of the results. In many contributions, approximated polygons are drawn to distinguish between bacterial groups (Li *et al.*, 2007; Wang *et al.*, 2009; Hammes & Egli, 2010).

In the last decade, a large body of flow cytometry literature is dedicated towards automated algorithms for FC data analysis. For a review we refer to Bashashati and Brinkman (2009). The data analysis procedures for FC data can be divided into two distinct parts: 1) Preprocessing and feature extraction from the multivariate distribution spanned by the FC variables, i.e. outlier removal and automatic gating. 2) Interpretation: comparison and classification of flow cytometric profiles from multiple samples. Within this context, clustering and classification methods are often used. Most contributions, however, focus on automatic procedures for fingerprinting and gating, they provide algorithms that automatically extract relevant features from a multivariate FC distribution, e.g. (Lo *et al.*, 2009; Luta, 2011). But the interpretation based on the extracted features between multiple samples (FC profiles) is mostly done manually (Bashashati & Brinkman, 2009). Rogers et al. (2008) developed an algorithm for cytometric fingerprinting that provides an efficient representation of the multivariate FC data and facilitates quantitative comparisons between samples. In this paper, we build upon the work of Rogers et al. (2008) and provide an automated method for comparing FC profiles across different treatments. We use their algorithm for establishing the FC fingerprint for each sample and extend their approach to deal with the multiple class problem and to provide statistical hypothesis testing for assessing differences between groups. Our tool also provides an efficient visualisation of the results and allows an interpretation of the observed differences between treatments in terms of the original components of the multivariate FC distribution.

The method is illustrated in two case studies: 1) a comparison of the flow cytometric fingerprints of six brands of bottled water, which is used to tune the parameters of the pipeline and 2) a flow cytometric assessment of the effect of specific treatments influencing the water microbiota. In the latter case study, the results of our flow cytometric method were also compared with the DGGE fingerprint method.

2.   Material and methods

2.1.   Sampling and treatment

Six commercially available brands of natural mineral water were used: Chaudfontaine, Evian, Spa, Mont Calm, Romy and Vittel. From each brand, 6 bottles were purchased and sampled.

In a second experiment, two types of treatments were conducted on Evian water to simulate changing physico-chemical conditions: temperature and nutrient treatment. For the heat

treatment, 1 L bottles were incubated for 3 and 24 hours at 37°C. For the nutrient treatment, 1 mL of water was replaced by 1 mL of a 1/10 dilution of autoclaved Luria Bertani broth (10 g tryptone, 5 g yeast extract and 10 g NaCl per L) to a final TOC of 0.65 mg L^{-1}. The bottles were incubated for 3 and 24 hours at room temperature. Incubations were started at different time points so all incubations finished at the same time and samples could be analysed together. Next to the treatments, the control without nutrient or heat treatment was kept at 4°C. All 4 treatments and the control were conducted in 6 replicates, each replicate being a different 1 L bottle Evian water. Finally, six different bottles (330 mL) of Evian water were sampled and analysed each day for 5 consecutive days, each day 6 new bottles from the same batch were used.

2.2.   Staining procedure and flow cytometric measurements

Two fluorescent dyes, SYBR$^{®}$ Green I and Propidium Iodide (PI), were used in combination as a viability indicator that differentiates between cells with intact and damaged cytoplasmic membranes (Berney *et al.*, 2007). The staining solution was prepared as follows: PI (20 mM in dimethyl sulfoxide (DMSO), from the LIVE/DEAD BacLight Kit, Invitrogen) was diluted 50 times and SYBR$^{®}$ Green I (10 000 times concentrate in DMSO, Invitrogen) was diluted 100 times in 0.22 μm filtered DMSO. Water samples were stained with 10 μL mL^{-1} staining solution and 10 μL mL^{-1} EDTA (pH 8, 500 mM) for outer membrane permeabilization. Before staining, all 1 mL samples were kept at room temperature for 30 minutes to minimize the effect of staining temperature. Prior to flow cytometric analysis, the stained samples were incubated for 15-20 minutes in the dark at room temperature.

Flow cytometry was performed using a CyAn™ ADP LX flow cytometer (Dakocytomation, Heverlee, Belgium) equipped with a 50 mW Sapphire solid-state diode laser (488 nm). The stability and performance of the flow cytometer was checked and controlled prior to the experiment by the use of Cyto-Cal Alignment Beads and Cyto-Cal multifluor Fluorescent Intensity Calibrator (Distrilab, Leusden, The Netherlands). Green and red fluorescence were collected with photomultiplier tubes using 530/40 and 613/20 bandpass filters, respectively. Forward (FSC) and side light scatter (SSC) were collected with a 488/10 bandpass filter. Milli-Q water was used as the sheath fluid. All samples were collected as logarithmic signals and were triggered on the green fluorescence channel. For each sample run, data for 20 000 events were collected.

2.3. Total bacterial community analysis by PCR-DGGE

Prior to PCR-DGGE, bacteria were first concentrated by filtering the water sample (1 L) over a 0.22 μm cellulose based filter (MilliPore) and stored at -20°C until further processing. DNA extraction was carried out using the UltraClean® Water DNA Isolation Kit following manufacturers instructions (Mo Bio Laboratories, Inc). Following the DNA extraction, the DNA was concentrated 10 times and purified using the Wizard® DNA Clean-Up System (Promega). The 16S rRNA genes for all bacteria were amplified by PCR using the Taq-polymerase kit (Fermentas) with general bacterial primers P338F and P518R and a GC-clamp of 40 bp on the forward primer (Muyzer *et al.*, 1993). DGGE was performed using the Bio-Rad D gene system (Bio-Rad, Hercules, CA) according to Boon *et al.* (2002). Clustering was based on the densitometric curves.

2.4. Statistical analysis of flow cytometric data

Our contribution presents a data analysis pipeline for an automated comparison of FC profiles originating from K different treatment groups. The data analysis pipeline consists of the following steps: 1) Creating a quantitative fingerprint from the multivariate distribution, 2) Extracting features from the fingerprint that are informative for the differences between the treatment groups, 3) Perform statistical hypothesis tests for finding significant differences across treatment groups within the reduced feature space and interpreting these results with respect to the original characteristics of the multivariate FC distributions (Figure 4-1).
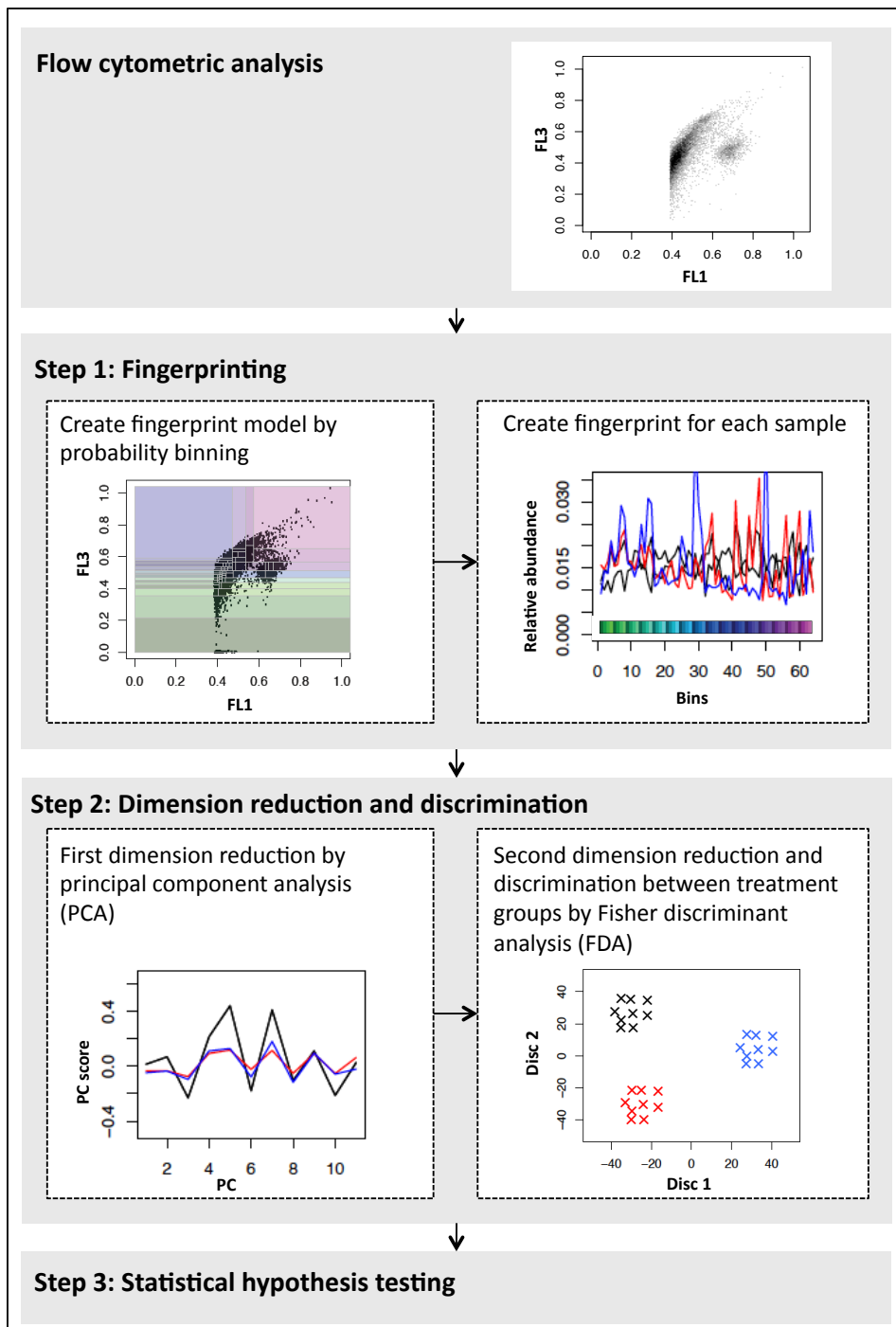
Figure 4-1. Flow chart of the flow cytometric fingerprinting technique. *After flow cytometric analysis, the data is extracted and used for flow cytometric fingerprinting. The first step is creating the fingerprint model by probability binning en determining the relative abundance of cells in each bin for all samples (64 dimensional fingerprint). The second step is a dimension reduction by principal component analysis followed by a second dimension reduction and discrimination between treatment groups by Fisher discriminant analysis (2 dimensional discriminating fingerprints). As a result, a data reduction was achieved from more than 10 000 data points per sample to a few discriminant values with minimal loss of information. The last step is statistical hypothesis testing to detect significant differences between treatment groups. All statistical methods are explained in more detail in the individual statistical information boxes at the end of material and methods section.*

In the first step, a p-dimensional quantitative fingerprint is derived from the multivariate FC distribution using the recursive probability binning (PB) algorithm for flow cytometry data that is implemented in the BioConductor package FlowFP (Rogers & Holyst, 2009). At the first level of the algorithm, the population is divided into two bins. Then, each of the two "parent" bins is divided into two "daughter" bins, and so forth. The final number of bins n_{bin} is determined by the number of recursive subdivisions l, such that $n_{bin} = 2^l$. Note, that the algorithm constructs the bins in such a way that they contain a nearly equal number of cells from the FC samples that were used to build the PB model. This provides an efficient representation of the structure in the multivariate data space by hyper-rectangular regions (bins) of varying size and shape. We pool the data of all samples together for constructing the PB model. The obtained PB model is then applied to each individual sample, which results in a feature vector of counts for each bin of the model. The latter is also referred to as the fingerprint. More details on the PB algorithm can be found in Rogers et al. (2008) and Rogers and Holyst (2009).

In step 2, a dimension reduction of the fingerprint is obtained by using Fisher discriminant analysis (FDA). FDA aims at understanding how K groups differ from one another in terms of a p-dimensional variate X. FDA is not based on any distributional assumptions and interprets the difference among the K-groups by projecting the input variables onto discriminants. In one dimension a good discrimination is obtained when the between class variance of the K experimental groups is large with respect to their within class variance. The discriminants are exactly those linear combinations of the input variables, which maximise the between class variance with respect to the within class variance after projecting the data onto the particular discriminant. It can be shown that K group means in the p-dimensional feature space of the fingerprints span at most a K-1 dimensional subspace of orthogonal discriminants. Hence, a huge dimension reduction is realised when K$<<$ p. Quantitative measures exist for the relative potential of the i-th discriminant function to discriminate the K-groups. They are often used for deciding how many discriminant functions are really necessary to get a good discrimination and can provide a further dimension reduction. Because discriminants are linear combinations of the input variables, they often provide a very useful interpretation within the original data space. We will use this property for displaying the leading differences in the multivariate distributions of the FC profiles.

Principal component analysis (PCA) is another method that is often used for the purpose of dimension reduction. In the result section we show that a dimension reduction with PCA prior to FDA is beneficial and reduces the influence of spurious features in the fingerprint. PCA

performs a rotation of the input variables without loss of information and essentially provides a decomposition of the variance-covariance matrix of the input variables. It allows for dimension reduction as q PCs contain more information than any q of the original variables or linear combinations thereof. We establish the dimension reduction by retaining the q principal components (PC) so that the portion of the variance that is explained by them exceeds a certain threshold δ. As both PCA and FDA are essentially linear combinations of input variables, the resulting discriminants after dimension reduction with PCA maintain their interpretation feature within the original multivariate FC data space.

The original K-1 dimensional discriminant space or any of its reduced d-dimensional subspaces (d < K-1) can also be used for classification purposes: a new sample is classified to the class, which has the centroid with the lowest distance to the new sample after projecting its fingerprint in the considered discriminant space. We exploit this link for calibrating the tuning parameters that are involved in step 1 and 2.

In step 3, statistical hypothesis tests are performed after dimension reduction. Within the discriminant space pairwise tests are used for assessing if the observed differences between the groups are significant. The Holm procedure is used to address the problem of multiple hypothesis testing. As the data was already used within the dimension reduction procedure, standard statistical hypothesis testing in the discriminant space does control the type I error of the statistical tests at their nominal significance level, α=0,05. Therefore we adopt permutation-based procedures for deriving the null distribution of the test statistics. For more details on permutation tests we refer to Good (2005).

The following procedure is used for deriving the permutation null distribution:

a) Permute class labels,

b) Adopt the PCA and FDA dimension reduction procedure from step 2 to the permuted data,

c) Construct the permutation based pairwise test statistics t^* within the discriminant space established in b,

d) Repeat steps a)-c) B times.

Permutation based p-values are then defined as the fraction of the permuted test statistics that are more extreme than observed test statistic: $p = \dfrac{\#[\,|t^*| > |t|\,]}{B}$.

2.5. Statistical analysis of DGGE fingerprints

The DGGE profiles were digitized according to Zhang and Fang (2000). Hence, the results of the DGGE analysis can be summarized in a matrix, say Y_{kl}, with $k=1, ..., m*g$ rows, with g the

number of groups and *m* the number of replicates within each group, and *l=1, ..., L* columns that correspond to *L* distinct bands that were identified from the *m*g* samples. The matrix *Y* can be used as input of a Fisher discriminant analysis (FDA). Score plots in the discriminant space may reveal differences between the *g* groups and again the discriminants can be interpreted in terms of the bands in the DGGE fingerprint.

Similar to the proposed analysis for flow cytometric data, permutation tests are used to test the null hypothesis that the fingerprints of the *g* groups are equal and for a pairwise comparison between the groups. The reported *p*-values are based on 10 000 random permutations.

CHAPTER 4

-- Information box statistics --

Creating quantitative fingerprints

The objective of creating a quantitative fingerprint is to transform the information that is captured within the multidimensional flow cytometry data (which is a matrix with the signals of the scattered and fluorescent light from each channel for every cell within a sample) to a form that enables quantitative comparison among samples. As such, each flow cytometric sample in transformed into a simplified fingerprint (Rogers et al., 2008).

Probability binning model

The first step in creating a quantitative fingerprint is creating a fingerprinting model. The model is the core of the algorithm that transforms the raw data to the fingerprints. As the model needs to be able to capture most of the variability between samples, all samples are combined into one big sample to create the fingerprinting model. The same fingerprinting model is then used for creating the fingerprints of all individual samples.

The fingerprinting model is created by probability binning (***Box Figure 4-1***). This algorithm divides the multivariate flow cytometry space in (hyper)rectangular regions or bins. In case when only two parameters are of interest, for example FL1 and FL3, these regions are rectangles. When multiple parameters are of interest, for example FL1, FL2, FSC and SSC, these bins are hyperrectangular regions.

In the initial step of this algorithm, all data points (i.e. cells) are within a single rectangle which spans the full space. During the first recursion (l=1) of the algorithm, this rectangle (or bin) is split in two daughter bins in such a way that each daughter bin contains an equal number of cells. During the second recursion (l=2), each of these two daughter bins is again divided into two, resulting in four bins with an equal number of cells but with varying size and shape. This is repeated six times (l=6) resulting in 64 bins. After the final recursion, the fingerprint model is obtained and used to create the fingerprints of each individual sample.

Fingerprints

Once the fingerprint model is created, the model can be applied to each individual sample. For a single sample, the model transforms the raw data into a vector of 64 cell counts, corresponding to the 64 bins. These cell counts get transformed to relative abundances to allow comparison between samples with an unequal number of cells. As such, a fingerprint is obtained for each sample with relative abundances for each bin (***Box Figure 4-2***).
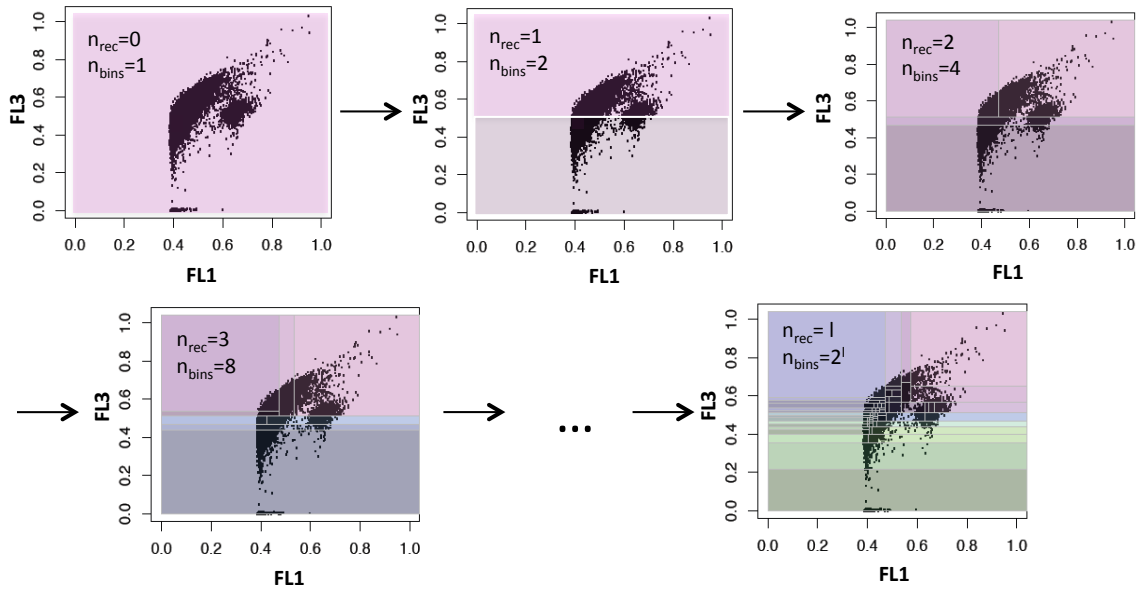
Box Figure 4-1. Recursive probability binning. *All data is initially contained within one bin. This parent bin in divided in two daughter bins in such a way that each bin contains an equal number of cells. Each of these two daughter bins is again split in two daughter bins during the second recursion resulting in 4 bins. This is repeated l times, resulting in 2^l bins with an equal number of cell but with different size and shape.*
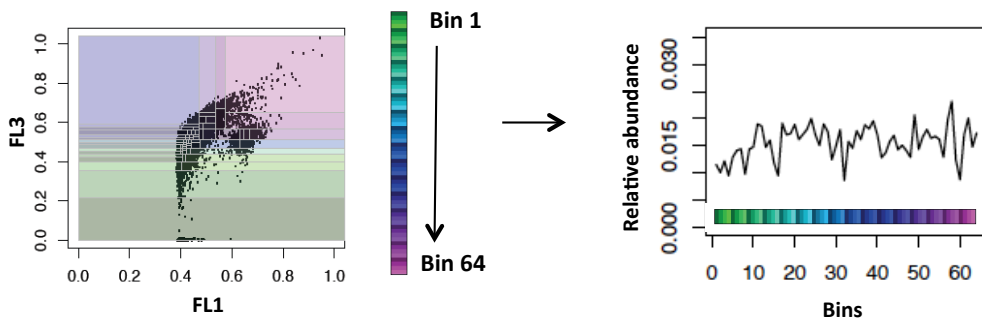
Box Figure 4-2. Creating fingerprints for each individual sample. *To obtain the fingerprint, the relative number of cells within each bin is determined.*

CHAPTER 4

-- Information box statistics --

Dimension reduction and discrimination

Principal Component Analysis (PCA)

PCA is a tool used for dimension reduction. It transforms *p* variables to *p* principal components (PC) without loss of information. Although this does not look like a dimension reduction, the PCs are constructed in such a way that *q* (< *p*) PCs always contain the largest variance that can be contained in *q* linear combinations of the original variables. As such a dimension reduction can be performed by only using *q* PCs.

PCA transforms possibly correlated variables into uncorrelated PCs. It searches for the directions in the data that have the largest variance and subsequently project the data onto it. The first PC is the direction along which projections have the largest variance. This PC is the eigenvector of the covariance matrix corresponding with the highest eigenvalue. The second PC is the direction that maximizes variance among all directions orthogonal to the first and is the eigenvector of the second largest eigenvalue, etc. (***Box Figure 4-3***)
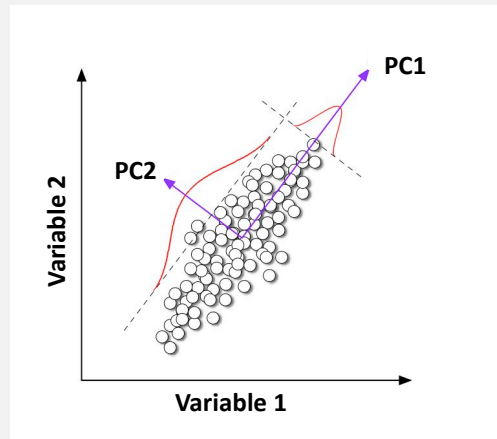
Box Figure 4-3. Graphical representation of PCA transformation in only two dimensions.

In case of our flow cytometric fingerprinting method, the fingerprint dataset contains a set of 64 variables, corresponding to the bins. A dimension reduction is established by retaining the *q* PCs so that the proportion of variance that is explained by them exceeds the threshold value of δ=0,95. ***Box Figure 4-4***A shows the variance that is explained by the each PC, and ***Box Figure 4-4***B shows the cumulative proportion of variance that is explained by the PCs. In this example, it can be seen that by retaining 11 PCs, more than 95% of the variability is captured. By performing PCA, the dataset can be reduced from 64 variables to 11 (***Box Figure 4-5***).
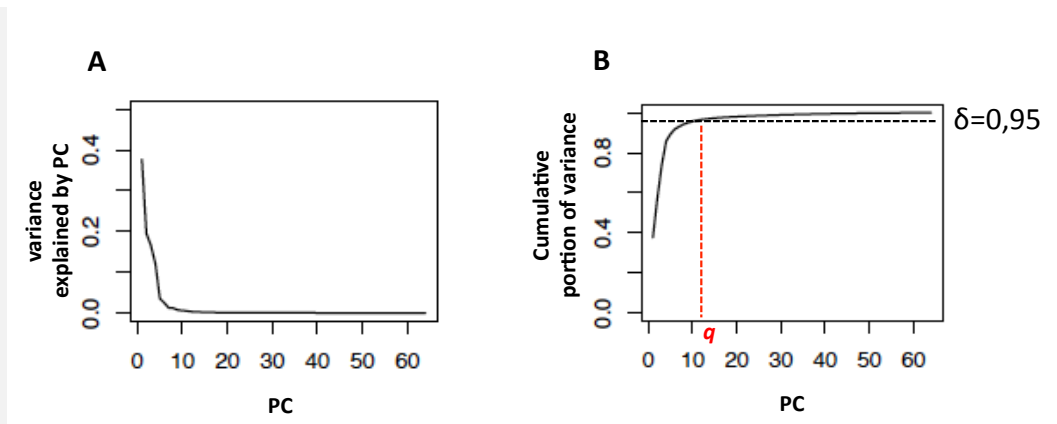
Box Figure 4-4. Variance (A) and cumulative variance (B) that is explained by the different PCs. *95% of the variance is explained by the first q PCs. A dimension reduction is established without loss of much information if these q PCs are retained.*
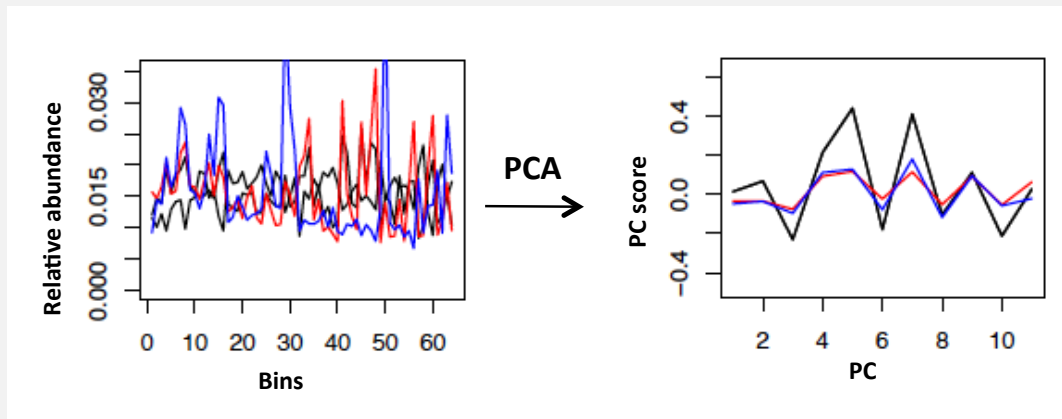
Box Figure 4-5. Example PCA applied to the fingerprinting data. *The fingerprinting dataset consisting of 64 bins is reduced by PCA from 64 to 11 dimensions.*

Fisher Discriminant Analysis (FDA)

FDA is another method used for dimension reduction. Just like PCA, the data is transformed to a new coordinate system, called the discriminants. But for FDA these discriminants, which are again linear combinations of variables, are defined in another way than principal components. While PCA is an unsupervised technique that searches for the directions in the data that has the largest variance without taking into account label information, FDA is a supervised technique that includes label information (for example different treatment groups) and aims at maximising the variance between the treatment groups while minimising the variance within the treatment groups (***Box Figure 4-6***). In case of K treatment groups, K-1 orthogonal discriminants can be found. When K<<p, a huge dimension reduction is realised. As such, FDA can be used for dimension reduction while maximally discriminating between different treatment groups.

CHAPTER 4
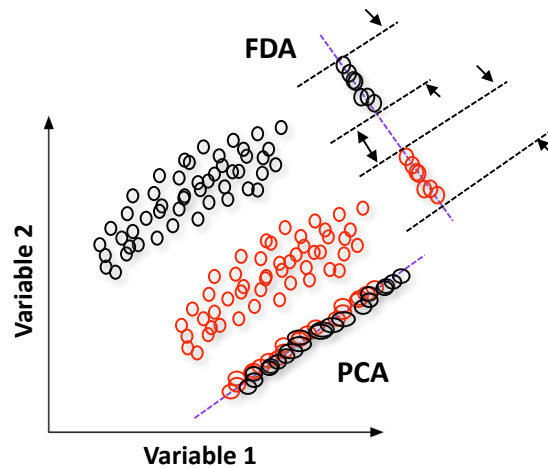
Box Figure 4-6. Comparison between FDA and PCA. *While PCA looks for the directions with the largest variance, FDA looks for the directions that maximises the between class variance and minimises the within class variance.*

In the fingerprinting method, FDA is performed after PCA. Initially PCA is performed to remove the "noisy" directions and remove spurious features from the fingerprinting data. The principal components are the new variables and are subsequently transformed into discriminants by FDA. In the example in ***Box Figure 4-7***, it can be seen that a dimension reduction from 11 PCs to two discriminants is performed, which allows discrimination between the three treatment groups.
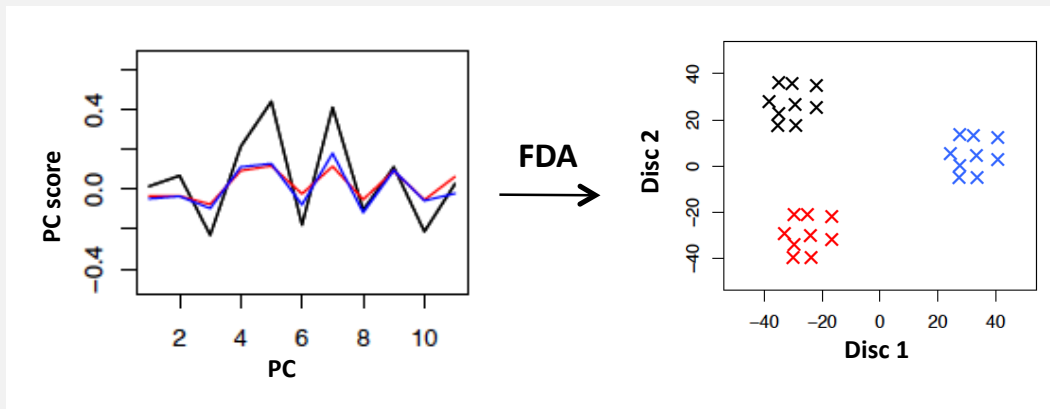
Box Figure 4-7. Example of FDA applied to PCs obtained after PCA of the original fingerprinting dataset. *The q dimensional dataset is reduced to a K-1 dimensional dataset that is able to discriminate between K treatment groups.*

Classification

The K-1 dimensional discriminant space can also be used for classification purposes. A new sample is projected onto the same discriminants and classified to the group which has the lowest distance to the new sample

-- *Information box statistics* --

Statistical hypothesis test

To test if treatment group means are significantly different, the test statistic *t* is calculated for each pairwise combination of groups.

$$t = \frac{\overline{X_1} - \overline{X_2}}{s_{\overline{X_1} - \overline{X_2}}}$$

In a classical statistical *t*-test, the *p*-value corresponding to each test statistic *t* is obtained from the theoretical null distribution of the test statistic (**Box Figure 4-8**), which is only correct if the data satisfies a set of assumptions (e.g. normality). The *p*-value is the probability of obtaining a test statistic *t* that is at least as extreme as the observed test statistic *t*, given that the null hypothesis is true. When *p* is smaller than the nominal significance level α=0.05, it is highly unlikely that the null hypothesis is true. It this case, the groups are regarded as significantly different.
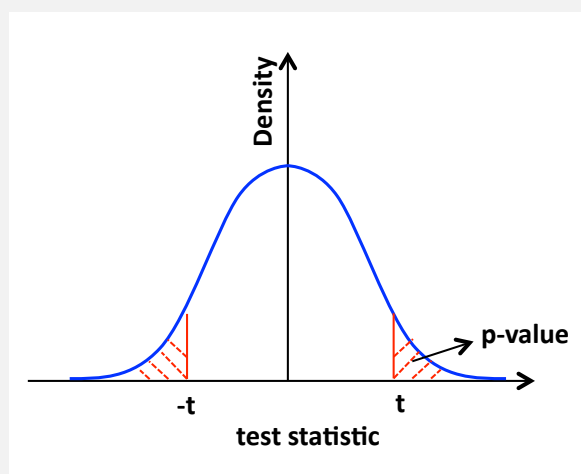
Box Figure 4-8. Theoretical null distribution of the test statistic with the test statistic t and the corresponding p-value (shaded area under the curve) for a two-sided t-test.

For many classical statistical tests, a permutation version exists. The permutation based statistical *t*-test uses the same test statistic *t* as for the classical test, but it does not make any stringent distributional assumptions about the data. Without such assumptions, not theoretical null distribution can be found (for finite sample sizes). The null distribution of the test statistic is now obtained by repeatedly calculating the test statistic *t** when the group labels of the data points are randomly rearranged. When the null hypothesis is true, which states that data does not depend on the group label, it is allowed to randomly rearrange (i.e. permute) the group labels. After permutation of the group labels, the test statistics *t** is calculated (**Box Figure 4-9**). This is repeated B times (B must be large), the set of these *t** forms a good approximation to the exact null distribution of the test statistics under the null hypothesis that the class labels do not matter (**Box Figure 4-10**). The *p*-value is calculated as the proportion of permutations for which the permuted test statistic t* is larger than the test statistic *t* calculated from the original dataset:

$$p = \frac{\#[\,|t^*| > |t|\,]}{B}$$

When *p* is smaller than the nominal significance level α=0.05, it is highly unlikely that the null hypothesis is true. It this case, the group means are regarded as significantly different.
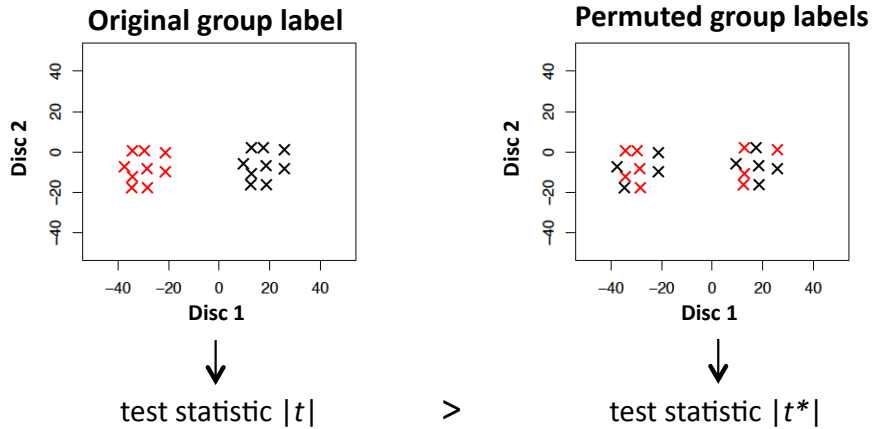
Box Figure 4-9. Permutation of the group labels. *After permutation, the test statistic t of the corresponding to the original group labels and the test statistic t* corresponding to the permuted group labels can be calculated and compared. A higher value for the test statistic indicates a better separation between the treatment groups.*
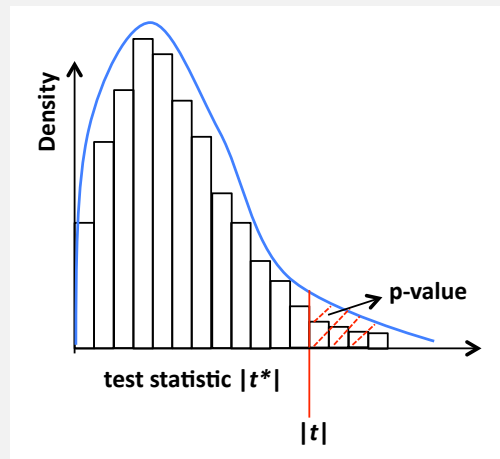
Box Figure 4-10. Permutation null distribution. *This distribution is obtained by making a density plot of the B test statistics t*, which were obtained after permutation of the group labels. The original test statistic t and the corresponding p-value are shown in the figure.*

3.   Results

The goal of this research was to develop a statistical tool that uses flow cytometric data to quickly compare fingerprints of microbial communities. The application of the statistical pipeline for flow cytometric evaluation should allow us to detect differences between microbial communities. The proposed pipeline consist of 3 stages: 1) creating a quantitative fingerprint from the multivariate distribution, 2) Extracting features from the fingerprint that are informative for the differences between the treatment groups, 3) Perform statistical hypothesis tests for finding significant differences across treatment groups within the reduced feature space and interpreting the results.

The probabilistic binning algorithm (PB) that is used in stage 1 and the dimension reduction with PCA and FDA of stage 2 involve two parameters that needs to be tuned: the number of recursions in the PB algorithm and the percentage of the variability that is explained by the first q PCs. This task is performed in section 3.1. In Section 3.2 the tuned data analysis pipeline is used for assessing changes in the microbial community structure related to environmental factors. Finally the robustness of the proposed method is assessed in section 3.3.

3.1.   Analysis of microbial communities in different brands of bottled water

The microbial communities of 6 different brands of water were analysed by flow cytometry: i.e. Chaudfontaine, Vittel, Evian, Spa, Mont Calm and Romy and raw FC data are presented in Figure 4-2. The communities were stained with a viability indicator consisting of propidium iodide and SYBR Green and data were recorded for 20 000 cells in the SSC, FL1 (530/40 nm) and FL3 (613/20 nm) channel. Six independent replicates (different bottles) were analysed for each water brand.
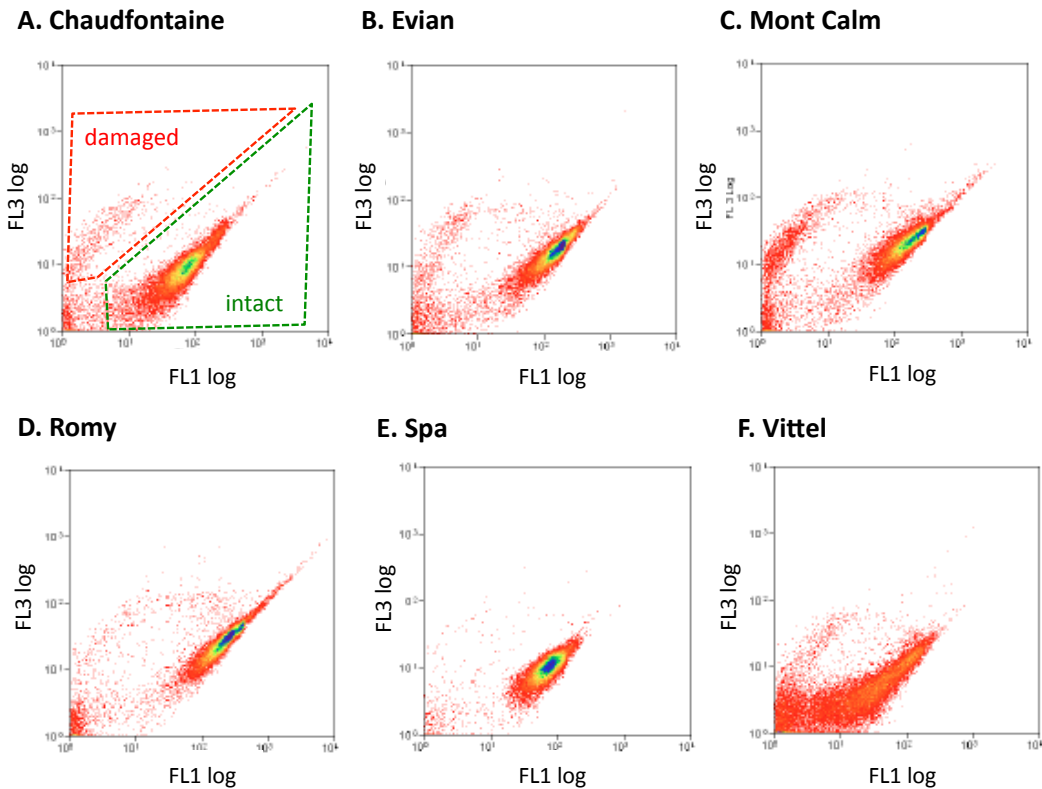
Figure 4-2. Flow cytometric dot plot of one replicate of 6 different water brands*: A. Chaudfontaine, B. Evian, C. Mont Calm, D. Romy, E. Spa and F. Vittel after straining with SYBR Green I and Propidium iodide. The different colours represent the density of cells at a given position, with each red dot representing one cell and an increasing number of cells from red to yellow, green and blue. Cells with high intensity in FL1 and low intensity in FL3 are intact cells (indicated by a green polygon), cells with high intensity in FL3 en low intensity in FL1 are cells with a damaged cytoplasmic membrane (indicated with a red polygon).*

The multivariate density of the flow cytometric measurements in the FL1, FL3 and SSC channel were used for calibrating the tuning parameters of the first two stages of the pipeline that impact the discrimination and dimension reduction with FDA. We will exploit the link between FDA and classification: i.e. optimal parameters values are selected that minimize the misclassification error estimated by leave-one-out cross validation. Differences in the microbial population across water brands have already been established with other microbial fingerprinting methods (Dewettinck *et al.*, 2001). Hence, it is a good setting to train our method for applications on microbial fingerprinting of microbial communities in bottled water. With several parameter combinations a perfect classification could be obtained. We have chosen the combination with the lowest number of bins to obtain a parsimonious model that also favors a more efficient visualisation. The optimal parameter values were respectively 64 bins and 96%. A dimension reduction is needed for efficient discrimination between the FC fingerprints for the different water brands. The results of the feature extraction with

discriminant analysis are given in Figure 4-3. The first discriminant accounts for 64.5% of the discrimination potential and allows discrimination between Spa, Vittel and the remaining brands. The second discriminant discriminates Romy from MontCalm, Evian and Chaudfontaine and the third discriminant allows for a further discrimination between the latter 3 brands. Hence, all brands could be discriminated from one another based on three discriminants.

The FDA can also be used for classifying unknown water samples to one of the 6 brands. The corresponding classification rule is evaluated by leave-one-out cross validation (CV). The estimated misclassification rate is 0%.
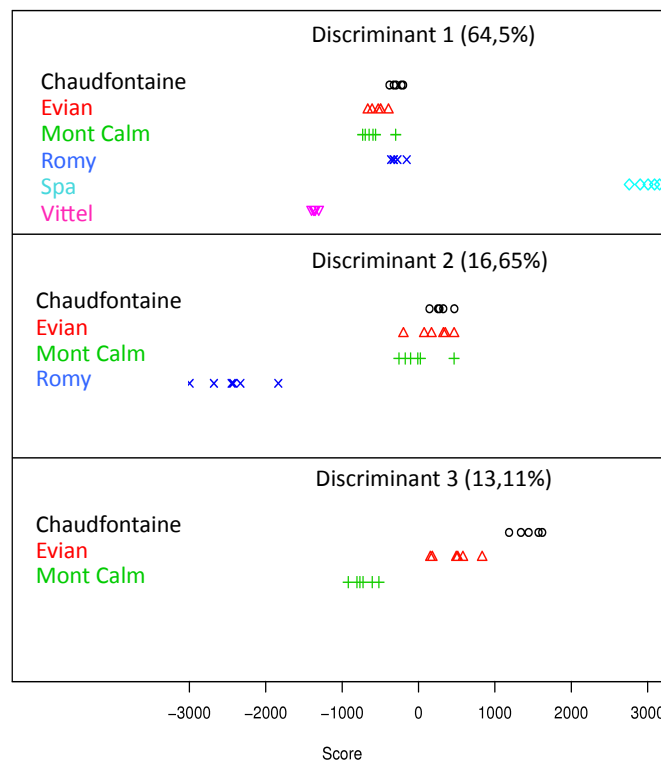
Figure 4-3. Functional discriminant analysis on flow cytometric measurements of different water brands*. The scores on the first three discriminants are displayed. (Chaudfontaine: black circles, Vittel: purple triangles, Evian: red triangles, Spa: light blue diamonds, Mont Calm: green + and Romy: blue x)*

3.2. Analysis of microbial communities in bottled water under different treatments

Our method is illustrated on a study that assesses changes within a community that were caused by different environmental factors. We evaluate if the method can quickly detect shifts in microbial community composition. Effects of temperature and nutrients on the bacterial community of bottled Evian water were evaluated: a control treatment (c); 3 hours (h3) and 24

hours (h24) of incubation at 37°C; 3 hours (n3) and 24 hours (n24) incubation after addition of nutrients (0.65 mg L^{-1} TOC). The impact of five different treatments were both assessed by flow cytometry and validated by Denaturing Gradient Gel Electrophoresis (DGGE), which is a commonly used fingerprinting technique. For flow cytometric analysis, six independent biological replicates were analysed for each treatment (Figure 4-4). PI-negative cells with intact cytoplasmic membranes are clearly separated from PI-positive cells with damaged cytoplasmic membranes.
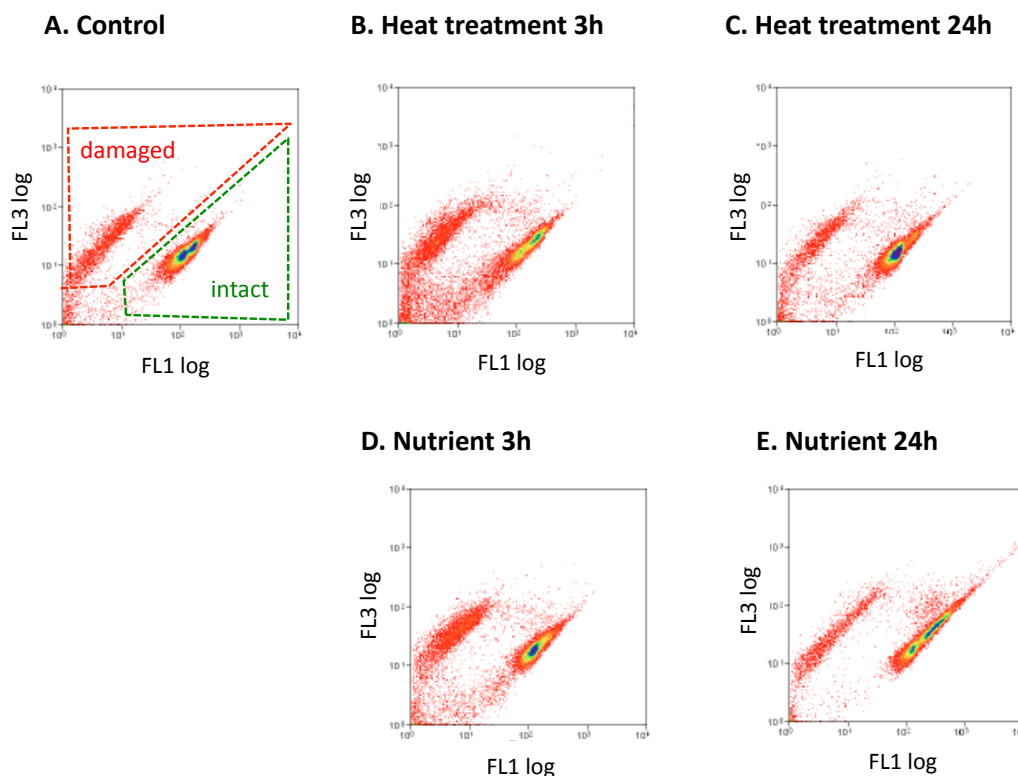
Figure 4-4. Flow cytometric dot plot of one replicate of Evian after different treatments: *A. control, B. heat treatment 3h, C. heat treatment 24h, D. nutrient treatment 3h, E. nutrient treatment 24h and straining with SYBR Green I and Propidium iodide. The different colours represent the density of cells at a given position, with each red dot representing one cell and an increasing number of cells from red to yellow, green and blue. Cells with high intensity in FL1 and low intensity in FL3 are intact cells (indicated by a green polygon), cells with high intensity in FL3 en low intensity in FL1 are cells with a damaged cytoplasmic membrane (indicated with a red polygon).*

A functional Fisher Discriminant Analysis (FDA) was performed after dimension reduction of the FC fingerprint with PCA. Again the SSC, FL1 and FL3 channel are considered. 64 bins were used for constructing the fingerprint and δ is set at 0.96. The result of the discriminant analysis is displayed in Figure 4-5. The first discriminant discriminates between three groups: i) c and h24, ii) h3 and n3, iii) n24 and it captures 54.6% of the discrimination potential. The second discriminant discriminates between n24 versus the remaining conditions and accounts

for 27.3% of the discrimination potential and the third discriminant separates h3. Significant differences between each treatment and the control are assessed by the use of contrasts (average distances) in the discriminant space. There is a significant difference between the control and heat treatment of 3 hours, $\Delta_{1,c\text{-}h3}$ and the control and both nutrient treatments, $\Delta_{1,c\text{-}n3}$ and , $\Delta_{1,c\text{-}n24}$ based on the first discriminant (p-values are $p = 0.001$; $p = 0.001$ and $p = 0.01$). On the second discriminant, only the difference between, $\Delta_{2,c\text{-}h3}$ is significant ($p = 0.002$). None of the contrasts involving the control and the heat treatment of 24 hours were significant.
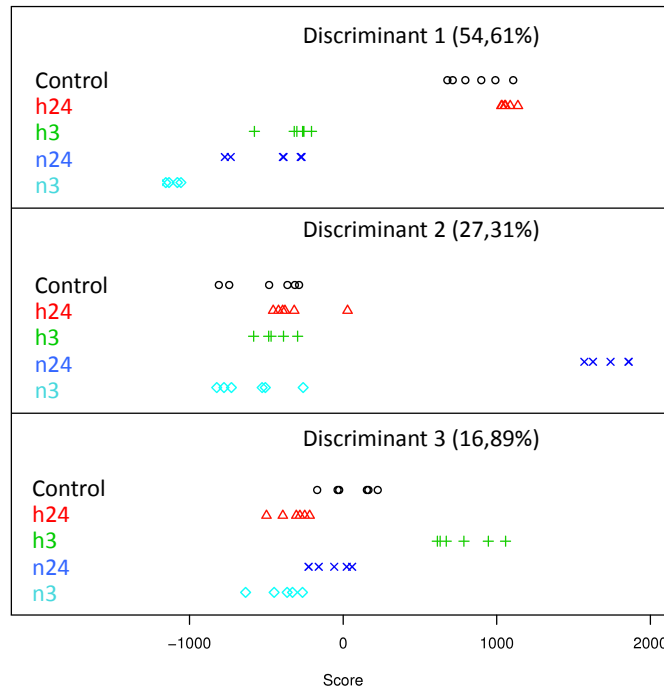
Figure 4-5. Functional discriminant analysis on flow cytometric measurements of Evian after different treatments. The scores on the first three discriminants are displayed. (c: control, black circles; h24: 24 hours temperature treatment, light blue triangles; h3: 3 hours temperature treatment, blue +; n24: 24 hours with nutrient treatment, red x; n3: 3 hours with nutrient treatment, brown diamonds)

The scores of the discriminants are linear combinations of the original features of the fingerprint. Therefore, the contrasts can also be used for the interpretation of the original FC distribution. The graph for the significant contrast c-n3 is shown in Figure 4-6. The graphs display the regions with cells that contribute to the discriminant by a) a blue colour if the region contains less cells in the bin than the control and b) a red colour for regions that contain more cells than the control. The intensity of the colouring of the bins is proportional to the magnitude of their contribution to the first discriminant (note: the colouring intensity is not proportional to the number of cells in the bins, this can be derived from the fingerprint plot). The top-panels and the left bottom panel correspond to 2D projections of the multivariate

distribution spanned by SSC, FL1 and FL3. The right-bottom-panel displays the contrast between the average fingerprint of the treatment and the control. Negative contributions correspond to bins with a lower cell density than the control and positive contributions correspond to bins with more observations than the control. We first concentrate on the contrast plots of the fingerprint. For all significant contrasts, bin numbers 16-26 seem to contribute to the first discriminant. In this region, the h3, n3 and n24 share the property that there are less cells in their average FC profile than for the control. In the FL1-FL3 plot it can be seen that these bins correspond to the regions around 0.4-0.6 in the FL1 channel and 0.15-0.3 in the FL3 channel. The density mass of the cells in these regions is shifted to towards the higher bins, which correspond to high FL1 and FL3 intensities. For n3 the density mass is also partly shifted towards bins 13-16, which correspond to the higher intensities in the FL3 channel and the lower intensities of the FL1 channel, i.e. to more damaged cells.
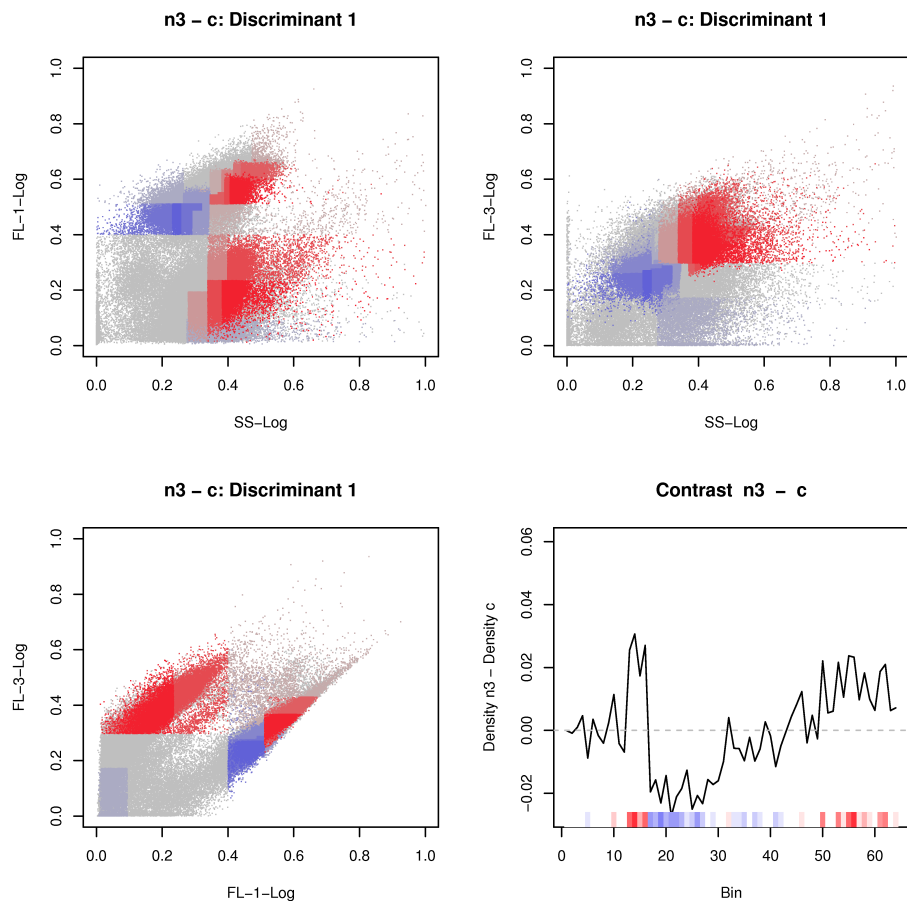
Figure 4-6. Interpretation of the first functional discriminant of the n3 treatment. *The blue and red regions display the regions with cells that contribute to the discriminant. Blue regions have a lower number of cells compared to the control, red regions a higher number. The intensity of the colours is proportional to the magnitude of their contribution to the first discriminant. The bottom-right-panel displays the contrasts between the average fingerprint of the control and the treatment.*

The results obtained by flow cytometry were confirmed by the use of DGGE (Figure 4-7). For each treatment three independent replicates were analysed. A total of 35 different bands were detected in the DGGE analysis.
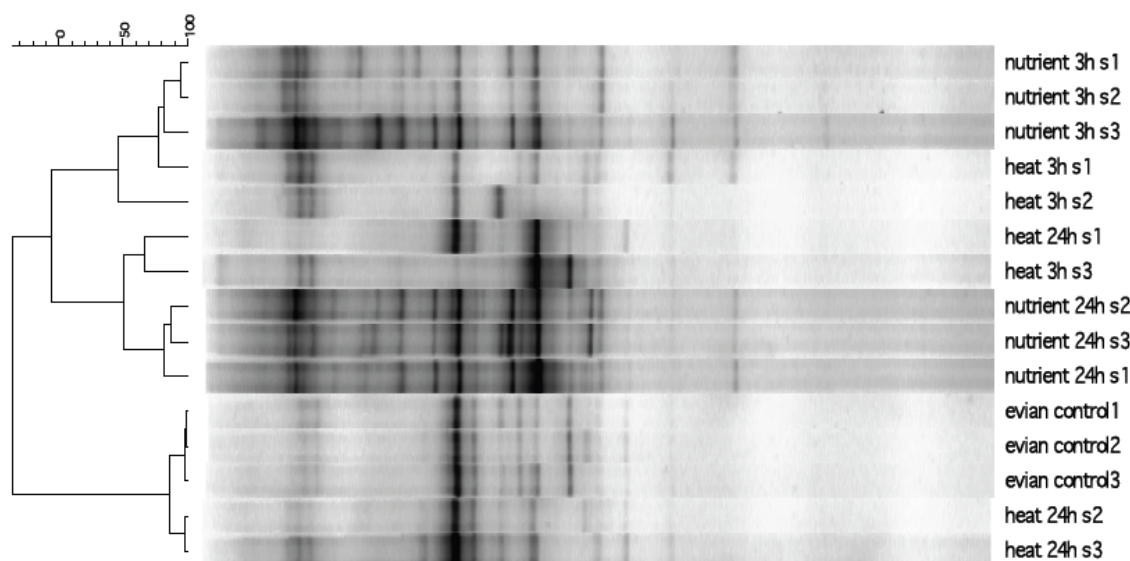
Figure 4-7. *Cluster analysis of the DGGE profile of Evian water after nutrient and temperature treatment.*

Based on these DGGE results, a discriminant analysis was performed for dimension reduction and the same method as for the FC data was used for hypothesis testing. The results of the discriminant analysis are given in Figure 4-8. The first discriminant differentiates between c-h24 versus h3, n3 and n24 and accounts for 76.8% of the discrimination potential. The second discriminant allows further discrimination between h3, n3, and n24 and it captures 14.4% of the discrimination potential. Similar to the analysis on flow cytometric data, tests on contrasts between the control and the treatments in the discriminant space were performed. The contrasts based on the first discriminant $\Delta_{1,c-h3}$, $\Delta_{1,c-n3}$ $\Delta_{1,c-n24}$ are again very significant (all p-values are smaller than $p < 0.01$). The contrasts based on the second discriminant $\Delta_{2,c-h3}$ and $\Delta_{2,c-n3}$ are also significant ($p < 0.01$ and $p < 0.04$ respectively). The contrasts between the control and the heat treatment of 24 hours, $\Delta_{1,c-h24}$ and $\Delta_{2,c-h24}$, are not significant (both values equal to p=1). The discriminant analysis shows a remarkable resemblance to the functional discriminant analysis on flow cytometric data. It confirms that the bacterial population of h3, n3 and n24 diverge from the control sample and it also cannot discriminate between the heat treatment h24 and the controls. The interpretation of the contrasts for the DGGE analysis is in terms of the influential bands. The difference based on the first discriminant is dominated by the impact of the band that is located on position 32.1%. The relative intensity of this band in

the control and the heat treatment of 24 hours fluctuates around 50%, which indicates that this organism dominates the bacterial population in these samples. For the other treatments, however, the relative intensities for this band lays within the range of 3-19% indicating that this organism is much less dominant in the samples of the h3, n3 and n24 treatments. The contrasts on the second discriminant are again highly influenced by band 32.1%. But, the bands on 41.8%, 39.0%, 14.0% and 11.9% also contribute considerably and allow for discriminating among h3, n3 and n24.
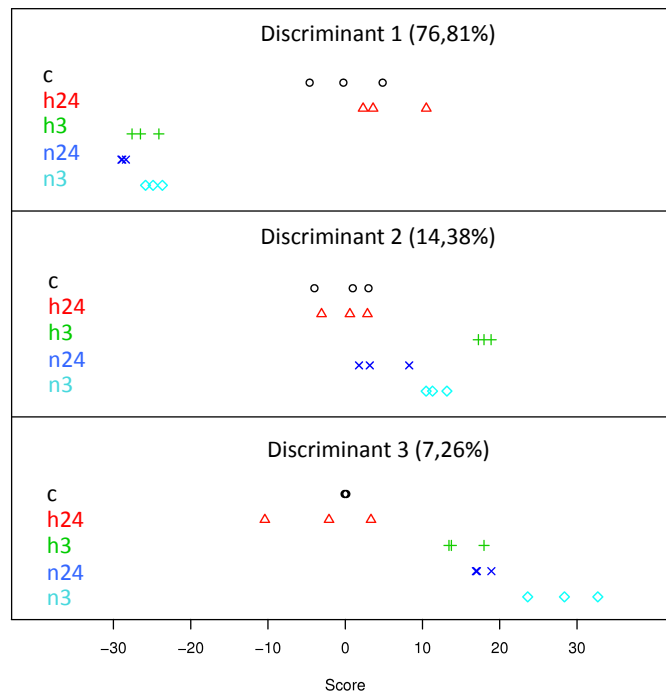
Figure 4-8. Discriminant analysis on DGGE of Evian water after nutrient and heat treatment. *The scores on the first two discriminants are displayed. c: control, black circles; h24: 24 hours heat treatment, red triangles; h3: 3 hours heat treatment, green +; n24: 24 hours with nutrient treatment, blue x; n3: 3 hours with nutrient treatment, light blue diamonds.*

3.3. Robustness evaluation: day-to-day variation

A daily variation experiment was also performed in which untreated water samples were analysed on 5 consecutive days. On each day 6 independent water samples were analysed by flow cytometry. The results of the treatment experiment and the daily variation experiment were jointly analysed by FDA. The samples in the daily variation experiment originated from a different water batch than the water used in the treatment experiment. The first three discriminants capture 72.5%, 16.1% and 6.25% of the discrimination potential, respectively. The water samples from the daily variation experiment (c1-c5) clearly cluster together and are different from the samples of the treatment experiment that originate from a different water

batch. Contrasts among the samples from the daily variation experiment were not significant (with p-values between $p \in [0.9; 1]$). The contrasts based on the first discriminant between the daily variation experiment (c1-c5) and the treatment experiment (c, n3, n24, h3, h24) were all significant (all p-values are $p \ll 0.01$). Based on the first discriminant a significant difference was observed between the control treatment (c) and the 24 hours nutrient treatment, $\Delta_{1,c-n24}$ (p =0.01). On the second discriminant differences between the control treatment (c) and 3 hour heat treatment, $\Delta_{2,c-h3}$, and the control and both nutrient treatments, $\Delta_{2,c-n3}$ and $\Delta_{2,c-n24}$ are significant (all p-values are $p \ll 0.01$). The contrasts between the control treatment and the 24 hour heat treatment, $\Delta_{1,c-h24}$ and $\Delta_{2,c-h24}$, were non-significant (p = 1 and p=0.54).

4. Discussion

In this paper, we developed an objective method to quickly and automatically create and compare fingerprints of microbial communities. These fingerprints are based on cellular features of the single cells in a microbial community using flow cytometry with a tailored statistical data analysis pipeline. Based on the similarities and differences between the microbial fingerprints, we were able to discriminate different brands of bottled water, classify unknown samples and detect shift of microbial communities caused by changing physico-chemical conditions.

The last few years, the potential of flow cytometry as a fast methodology for investigating the microbial community in aquatic samples has been explored (Hoefel *et al.*, 2005; Wang *et al.*, 2010). It is used for the routine analysis of aquatic samples or pure cultures (Pinder *et al.*, 1990). The total cell number is the most commonly used parameter for characterising the microbial community of water samples. But also cell size, DNA and RNA content, viability and enzyme activity of the bacterial cells are used (Hammes & Egli, 2010). These characteristics are also used for creating the FC fingerprint. Prior to flow cytometric analysis, the samples are stained with a suited dye or a combination of dyes to visualise cellular features of single cells. The post-processing of the FC data used to consist of several steps: It starts with the selection of the FC parameters of interest. Histograms with one parameter and/or dot plots with two or three parameters are made. These figures are then used for the interpretation of the data. Groups of cells are selected based on the density distributions, gates are made to focus on a certain group of cells and cell numbers are measured (Diaz *et al.*, 2010). Selection of the cells is done manually by the formation of regions which is very subjective (Bashashati & Brinkman, 2009). Subjective handling of the data is one of the main

drawbacks of flow cytometry, the interlaboratory variability can go up to 44% (Maecker *et al.*, 2005). Comparison of microbial communities in environmental samples is done visually by comparing dot plots or is based on absolute or relative cell numbers. Changes or shifts in community structure are hard to detect by conventional flow cytometry software. Recently, many efforts have been made for providing automated and more objective preprocessing and gating (Lo *et al.*, 2009; Luta, 2011)**.** However, the interpretation of FC profiles is mostly done manually. With the new method there was no need for any subjective post-processing of the data. The raw data were loaded into the statistical data analysis pipeline to create an automatic comparison and discrimination of different fingerprints of the microbial community across K treatment groups. The data collection and analysis could be done within 30 minutes and allowed an objective comparison of a sample with other samples or a database of previously analysed samples.

In this research, the number of parameters was set at three to create a multi-dimensional fingerprint. The first two channels, FL1 and FL3, were selected based on the emission wavelength of the used dyes, here propidium iodide and SYBR Green (Berney *et al.*, 2008). These dyes reflect the viability of the bacterial cells. The third parameter, the sideward scatter, reflects the morphological characteristics of the cells. These three parameters are most relevant for the used staining. Lowering of the number of parameters to two for the statistical analysis did not change the main conclusions of the analysis (results not shown). The viability staining can be substituted by other staining methods to make a fingerprint based on other cellular features (Wang *et al.*, 2010).

The concept of using flow cytometry in combination with the developed statistical tool for comparing FC fingerprints of the bacterial communities in different aquatic samples needs to be confirmed by a fingerprinting method that is accurate, reproducible and sensitive. PCR-DGGE is a suitable method for this, it has been successfully used for bacterial community profiling of groundwater, bottled mineral water (Dewettinck *et al.*, 2001) and wastewater treatment plants (Boon *et al.*, 2002; Ding *et al.*, 2011). Nutrient supplementation and elevated ambient temperatures were selected in this research to look at changes in the microbial community of drinking water due to environmental factors. The availability of nutrients is often the cause of regrowth (Morton *et al.*, 2005) and nitrification (Zhang & Edwards, 2010) in water distribution systems and elevated ambient temperatures (37°C) represent optimal growth conditions for a wide range of microorganisms. Both treatments were conducted for 3 and 24 hours to compare short and long-term effects. FC fingerprints were created with the aim of detecting changes within a microbial community and a second set of fingerprints were

made that differentiate between six brands of bottled mineral water for tuning the parameters of the data analysis pipeline. For both experiments visual discrimination of the flow cytometric dot plots of the samples (Figure 4-2 and Figure 4-4) was difficult and was prone to subjective interpretation. The statistical data analysis pipeline was used for differentiating between the FC fingerprints. In the treatment experiment, a significant separation between the control samples and three out of four treatments could be observed (Figure 4-5). The presence of nutrients in water and a temperature treatment caused only after 3 hours a change in the microbial community that was detectable by our method. Although there is a significant difference between the control and the heat treatment of 3 hours, the control treatment and the heat treatment of 24 hours were not significantly different. This could also be observed in the discriminant analysis of the DGGE (Figure 4-8) and the DGGE itself (Figure 4-7). Consequently, the lack of significant results between the control and the heat treatment of 24 hours is not because of the insensitivity of the system but probably due to the growth of some species. This might be explained by the so-called r/K selection theory. In unstable environments, r-strategist are known to become more dominant because of their fast growth (Reznick *et al.*, 2002). After a longer time of incubation, the environment gets more stable again, which benefits the K-strategists. This can cause the community to revert back to its initial situation. This phenomenon of intermediary situation and reversions to the original community was already seen before (Boon *et al.*, 2003).

The similarity between the discriminant analysis of the FC results and DGGE for all treatments confirms that our method can be used as a fast fingerprinting method for detecting changes in the microbial community of drinking water. DGGE was already successfully used to show that failure of the ammonia oxidation process in wastewater treatment plants can be linked to shifts in bacterial communities (Wittebolle *et al.*, 2005) and that stagnation of tap water in water pipes induces microbial growth and changes in community composition (Lautenschlager *et al.*, 2010b). This is an indication that our fingerprinting method can also be used for a fast detection of problems in wastewater treatment plants or water distribution systems. Experiments that confirm this are necessary. The experiment with the different water brands showed a clear and significant separation between the brands of bottled water (Figure 4-3). These differences in community composition were already seen by Dewettinck and coworkers by the use of DGGE (Dewettinck *et al.*, 2001). In case of a trained system with an internal database, the method can also be used for the classification of samples of unknown origin. This was shown by leave-one-out cross validation, as all samples were classified correctly.

Reproducibility and stability of the system are important characteristics that need to be guaranteed. With the aim of using the tool for the follow up of setups or quality control of products, the variability introduced needs to be negligible. One needs to be sure that changes detected by the method can be ascribed to changes in the microbial community and are not caused by i) changing laboratory procedures or ii) the variability of the equipment or the system. To prevent changes because of changing laboratory conditions or procedures, a standardised protocol needs to be followed during the preprocessing of the samples. The daily variation experiment, where water samples originating from the same batch were analysed on 5 consecutive days, shows the stability of the system. All water samples from the same batch cluster together, although they were analysed on different days, while the samples of the treatment experiment are clearly separated. A poor discrimination between the samples of the daily variation experiment and the control of the treatment experiment was expected, however this is not the case. This can possibly be explained by the difference in bottle volume used for both experiments. This "volume effect" was shown before to influence the bacterial community of bottled water (Hunter, 1993).

The use of our flow cytometry pipeline is not restricted to the analysis of drinking water. Automated flow cytometry systems, like the flow injection flow cytometer (Zhao *et al.*, 1999), make it possible to automatically sample, pre-process and analyse bioreactors. The developed data analysis method also allows the comparison of new samples with results of previous samples that are stored in an internal database. This opens new opportunities for building completely automated objective on-line warning systems for bioreactors, which are susceptible to contamination. Such systems are useful in different industries in which processes follow up and product quality control are important. Flow cytometry is already used within this context, e.g. for the evaluation of the quality of lactic acid bacteria starter cultures, raw materials and the production processes itself in the dairy industry (Riis *et al.*, 1995; Ruszczynska *et al.*, 2007), for the follow up of fermentation processes during the production of beer (Hutter, 2002), quality control in winemaking (Couto & Hogg, 1999) and in the different stages of drinking water production (Hammes *et al.*, 2010) or waste water treatment (Forster *et al.*, 2003). Our data analysis pipeline can contribute to optimising these applications.

The development of portable flow cytometers will make it possible to go from the laboratory to the field for the analysis of samples (Song *et al.*, 2011). This allows a real-time analysis at the place of sampling (Diaz *et al.*, 2010). Fast objective data analysis pipelines will boost portable flow cytometers to give an indication of the microbial quality within one hour so that

they can be used as an early warning system. This will enable an immediate intervention in case of problems and will provide clear economic benefits.

5.   Conclusions

Flow cytometry followed by our statistical data analysis pipeline can be used as a fast objective method for the construction, interpretation and comparison of microbial community fingerprints of aquatic samples. The results were confirmed with DGGE, which resulted in a similar discrimination between the water samples.

6.   Acknowledgement

CHAPTER 4

CHAPTER

5

FLOW CYTOMETRIC FINGERPRINTING FOR THE FOLLOW UP OF THE RINSING PROCESS OF WATER PIPES

FLOW CYTOMETRIC FINGERPRINTING FOR THE FOLLOW UP OF THE RINSING PROCESS OF WATER PIPES

Abstract

After reparation or maintenance works in a water distribution network, it is required to rinse these water pipes to ensure that only save and clean drinking water reaches the consumer. The conventional procedure consists of disinfection by the addition of chlorine, followed by rinsing with a quantity of water that equals three times the volume of the water pipe. This is a time consuming and costly process that water companies would like to reduce. Therefore, the flow cytometric fingerprinting approach was used to determine the time needed for the rinse water to be again of the same quality as the clean and safe reference water. It was shown that the rinsing time could be drastically reduced and that flow cytometry fingerprinting is a valuable technique for the follow up of this rinsing procedure.

1. Introduction

Before water pipes in a water distribution network are put back into use after reparation works, maintenance or installation, the pipes are rinsed to ensure that only save and clean water reaches the consumer. This rinsing process consists of several steps; first chlorine is added in high concentrations to the pipes for disinfection. Next, the pipes are filled with water and rinsed with a volume equal to three times the pipe-volume (De Gusseme & Van de Velde, 2013). Known that the average diameter of such a pipe is 1 m, 2.4 m^3 water is needed per running meter. The length of the pipe can go up to 2000 m, what brings the total rinsing volume to 4800 m^3 and a water cost of € 4800. The time of rinsing is generally between 15 and 20 hours. To make sure that water meets the quality standards after rinsing, a water sample is taken and different chemical and biological parameters are analysed (WHO, 2011). For example, no coliforms, *E. coli* and enterococci can be detected in 100 mL of water to meet the standards. Due to the long incubation times of heterotrophic plate counts, the process of quality control takes at least 18 hours (Allen *et al.*, 2004). After confirmation of the water quality, the pipes can be put back into use. This makes that the total rinsing procedure, including quality control, takes at least 30 hours. During this period, the water pipe cannot be used and an alternative water supply has to be organised for all consumers in the network. Water distribution companies raised the concern of this time demanding rinsing procedure and its high cost (De Gusseme & Van de Velde, 2013). Whereas it is not possible to get around or replace the standard procedure for quality control, it could be possible to reduce the rinsing time and volume by testing when the water reaches the desired quality. The flow cytometric fingerprinting method (developed in Chapter 4) - which was shown to be able to differentiate between different brands of bottled water and to detect changes within a microbial community (De Roy *et al.*, 2012) - was used in a field study to evaluate when 'dirty' rinse water was again of the same quality of the 'clean' water that was used to rinse.

2. Material and methods

2.1. Rinsing and sampling

After reparation works on a main water pipe in the Ghent water distribution network, the pipe was disinfected with chlorine. This was done by adding 10 L of a sodium hypochlorite solution (15% active chlorine) to the point where the reparation works were performed. After

the addition of chlorine, the pipe was filled with water. Once full, water samples were taken every hour during 19 hours upstream of (reference water) and downstream of (rinse water) the location of the reparation works (Figure 5-1).

Prior to sampling, the tap was opened for 5 minutes to rinse the tap and to ensure the sampled water originates from the main water pipe and not the tap. Residual chlorine in the water samples was quenched by the addition of 100 mM sodium thiosulfate. Samples were stored at 4°C and analysed within 30 minutes after sampling. This procedure of sampling, chlorine neutralisation and storage is identical to the procedure applied and optimised by the water company (De Gusseme & Van de Velde, 2013).
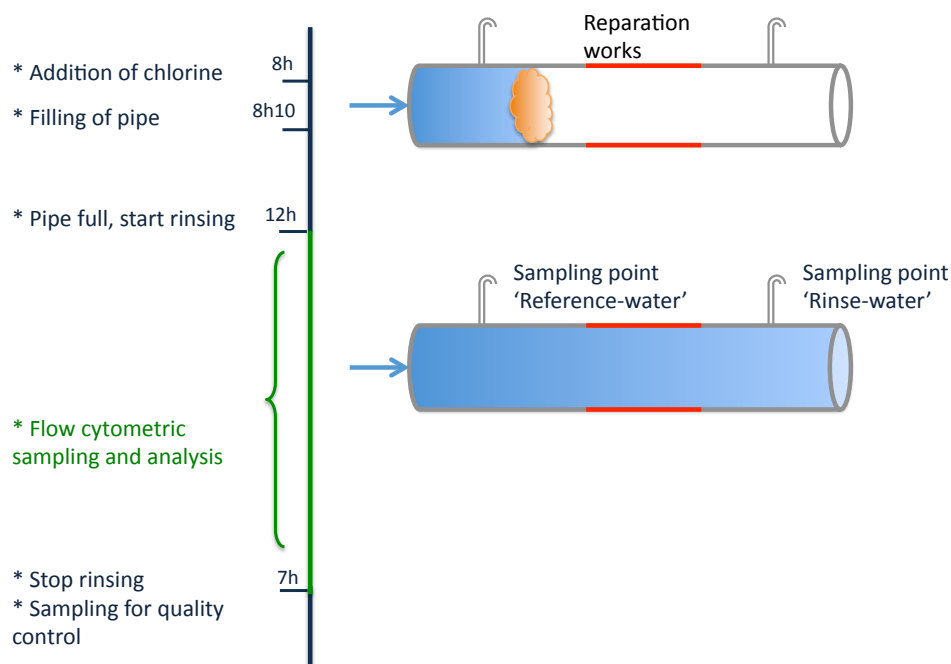
Figure 5-1. Rinsing procedure of the main water pipe after reparation works. *The pipe was filled with water after addition of chlorine (orange plug). From the moment the pipe was full, the pipe was rinsed during 19 hours. To investigate if the rinsing time (indicated in green) could be reduced, samples of the reference and rinse water were taken every hour during the rinsing process and analysed by flow cytometry.*

2.2. Flow cytometric analysis

The water samples were stained with a viability staining containing propidium iodide (PI) and SYBR Green I. The staining solution was prepared by diluting SYBR® Green I (10 000 times concentrate in DMSO, Invitrogen) 100 times in 0.22 μm filtered DMSO and adding PI (Invitrogen) to a concentration 400 μM. Water samples were stained in triplicate with 10 μL mL^{-1} staining solution and incubated for 13 minutes at 37°C in the dark (Hammes *et al.*, 2012). For each replicate, a volume of 100 μL was analysed with the Accuri C6 flow cytometer (BD

Biosciences), equipped with a 488 nm solid-state laser. Measurements were performed at high speed (66 µL min^{-1}, 22 µm core size), the threshold settings for FL1 and FSC were respectively 500 and 3000. Green and red fluorescence were collected with photomultiplier tubes using 533/30 nm (FL1) band pass and 630 nm (FL3) long pass filters, respectively. Forward (FSC) and side scatter (SSC) were detected by diodes. Milli-Q water was used as the sheath fluid.

Cell counts of damaged and intact cells were done by measuring the number of particles in each region after gating on green versus red fluorescence plots in the BD CSampler software (Van Nevel *et al.*, 2013b).

2.3. Flow cytometric fingerprinting

For flow cytometric fingerprinting (Figure 5-2), all data was exported as fcs-files and imported in the R-fingerprinting program. To create a multi-dimensional fingerprint, the most relevant parameters FL1, FL3 and SSC were selected based on the emission wavelength of SYBR® Green I and PI and the morphological characteristics of the cells.

The training dataset, initially only consisting of the data of the reference water (clean water) and rinse water (dirty water) sampled at the first time point, was pooled and used to create the initial fingerprinting model by probabilistic binning (De Roy et al., 2012). The number of bins was set at 64. For each sample in this training dataset, a fingerprint was created by determining the relative abundance of cells in each bin of the fingerprinting model. Each fingerprint was assigned to one of two groups, 'clean' or 'dirty', according to the origin of the sample. A dimension reduction was performed by principal component analysis (PCA, threshold 96%), followed by Fisher discriminant analysis (FDA) that extracts those features from the fingerprints that are most important to differentiate between the two groups. The resulting discriminants are linear combinations of the input variables. By projecting the fingerprints of new flow cytometry samples in the same discriminant space, this discriminant space can be used for classification purposes. Therefore, fingerprints were created for the rinse water of the next time point and classified using cross-validation.

Because in a real situation it is required to classify a new sample as clean or dirty when only the results of earlier time points are available, not all samples of the reference water were directly added to the training dataset. This was done sequentially according to the different time points. The complete procedure of model development, fingerprinting, dimension reduction, discrimination and classification was therefore repeated for every time point.

Flow cytometric analysis

Data extractions

Create Fingerprint model

Create Fingerprints

Create 2 groups:

'Clean'

'Dirty'

Principal component analysis (PCA)
→ Dimension reduction

Fisher discriminant analysis (FDA)
→ Find differences between groups

Classification unknown sample
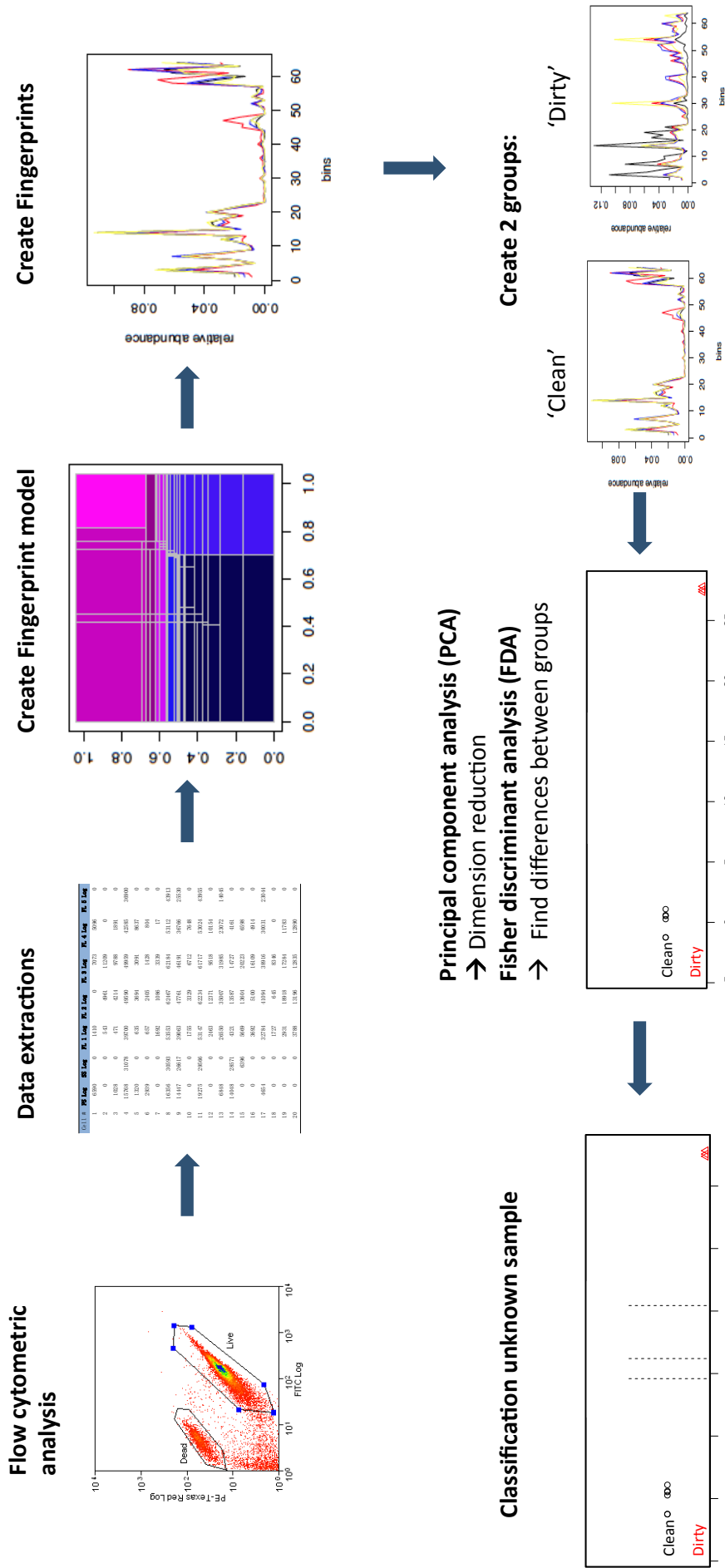
***Figure 5-2. Flow cytometric fingerprinting procedure.** After flow cytometric analysis, the data for each sample is exported as fcs-file. A training dataset is used to create the fingerprinting model, which on its turn is used to create fingerprints of the individual samples. Two groups are created within the training dataset. To differentiate between these two groups, principal component (PCA) and Fisher discriminant analysis (FDA) are performed. Fingerprints of unknown samples can be classified by performing the identical transformation as performed by PCA and FDA on the training dataset.*

3.   Results

3.1.   Cell counts

The number of damaged and intact cells in the reference and rinse water was measured after the samples were stained in triplicate with a viability staining. The number of intact cells in the reference water varied between 15 000 and 20 000 cells mL^{-1} during the first three sampling hours and reached a stable value of 15 000 cells mL^{-1} after 4 hours (Figure 5-3A). After 17 hours, there was again an increase of the number of viable cells. From the moment the pipe was full (time 0) until two hours of rinsing, the number of viable cells in the rinse water was 22-45% higher compared to the reference water, which was analysed and sampled at the same time as the rinse water. From 4 hours of rinsing on, the number of viable cell in the rinse water was lower than in the reference water.
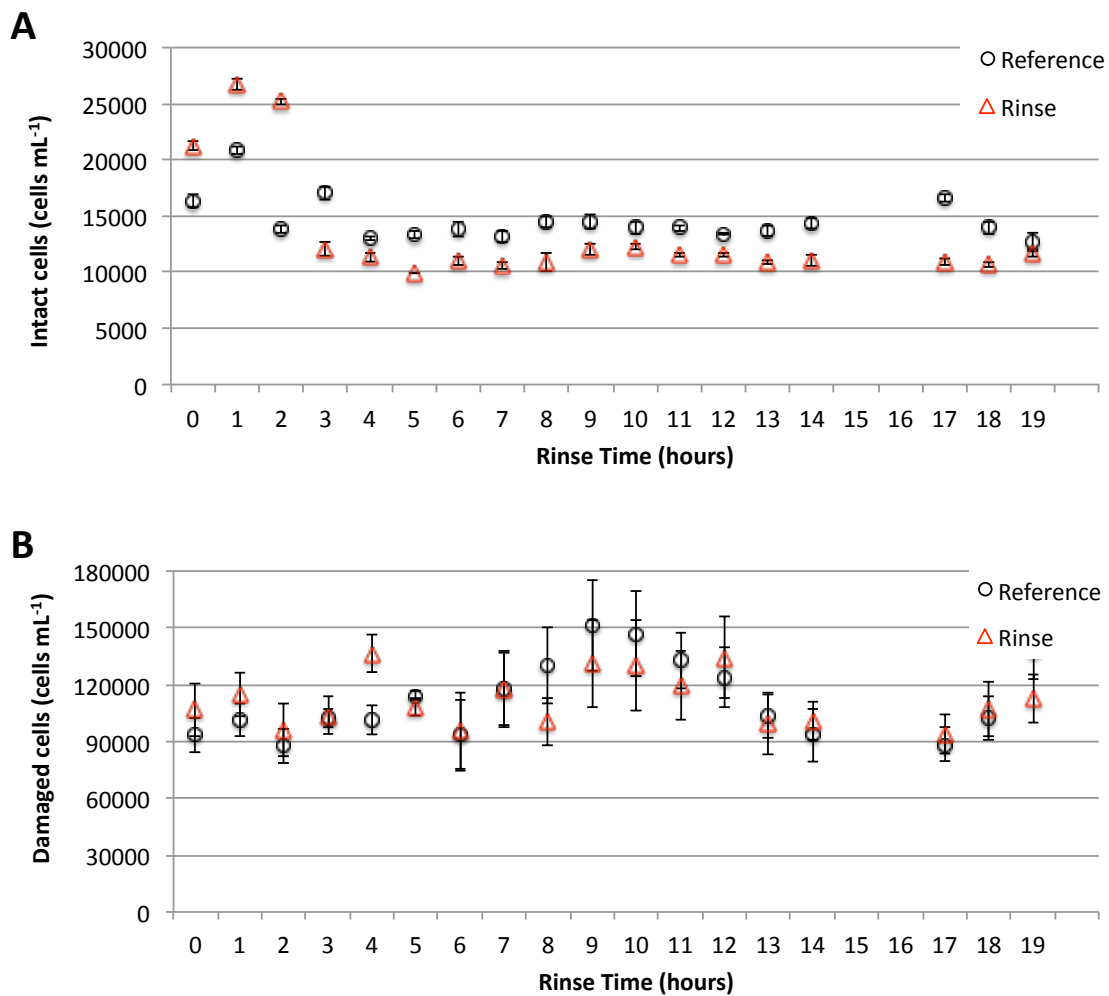
Figure 5-3. Number of intact (A) and damaged (B) cells in the reference water (black circles) and rinse water (red triangles) throughout the rinsing procedure of the water pipe (mean ± SD, n=3).

The number of damaged cells (Figure 5-3B) was the same in the reference and rinse water and was about ten times higher than the number of intact cells. It varied over time, with the lowest number at the start of the rinsing process and the highest after 9 hours rinsing.

3.2. Fingerprinting

The rinsing time required for the rinse water to become of the same quality as the reference water was determined by flow cytometric fingerprinting. The reference and rinse water at time 0, when the pipe was filled, are respectively considered as clean and dirty water and are used to create the training dataset. Every hour, the reference water was added to the 'clean' group of the training dataset. Because every hour the training dataset is getting bigger and for each time point the complete fingerprinting procedure is repeated, different discriminant values are obtained over time for the reference and dirty water.

The only discriminant, which accounts for 100% of the discrimination potential, was able to discriminate between the 'clean' water (Figure 5-4, black circles) and 'dirty' water (Figure 5-4, red triangles) at all time points. The variability within the 'clean' group did however become bigger. The rinse water that had to be classified is indicated with a dashed line. After 1 hour of rinsing, the water could not be classified as either clean or dirty based on the discriminant score. After 2 hours, the water was still qualified as dirty (indicated with a red rectangle), while it was clean after 3 hours and 4 hours (indicated with a green rectangle). From 5 hours of rinsing on (result only shown until 8 hours of rinsing), the discriminant score of the rinse water was again different from the clean water. However it remained more similar to the clean water than the dirty water, no clear conclusions can be made.

CHAPTER 5
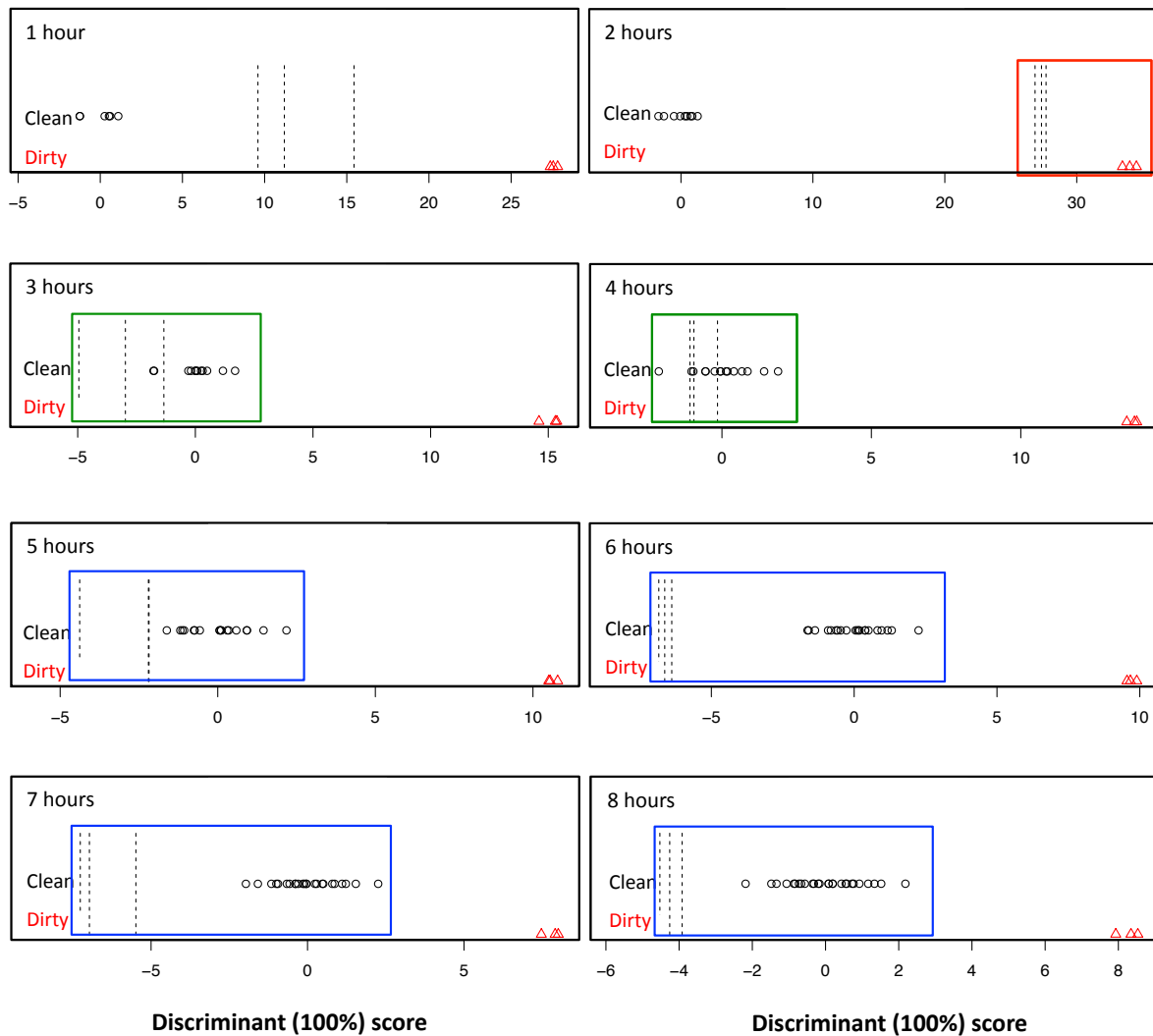
Figure 5-4. Functional discriminant analysis. *For each hour of rinsing, PCA and FDA were performed to differentiate between the reference water (clean) and initial dirty rinse water 'dirty'. The rinse water after x hours that requires classification was projected in the same discriminant space and indicated with dashed lines on the discriminant plot. Red, green and blue rectangles respectively classify the rinse water as dirty, clean or none of both. (n=3)*

4.   Discussion

To determine when rinsing water was again of the desired quality, this water was compared to water sampled earlier in the distribution network, which is known to be of good quality. As drinking water was shown before to remain stable throughout the water distribution network (Lautenschlager *et al.*, 2013), samples taken earlier in the distribution network can be compared to samples taken further on.

The number of intact cells in the rinse water was higher compared to the reference water during the first two hours of rinsing. From three hours of rinsing on, there were less intact

cells in the rinse water than in the reference water. These cell counts are a first indication that from three hours of rinsing on, the rinse water is of the desired quality to allow consumption (Hammes *et al.*, 2008). Subtle, but important differences in the community composition can however not be detected by looking only at cell numbers. Moreover, the used post-processing method is subjective and experienced personnel are needed because it requires the drawing of gates to differentiate between cells with different physiological characteristics and to remove background signals. Therefore, the fingerprinting technique that (i) takes into account the complete multivariate distribution of the flow cytometric measurements and (ii) does not require any subjective post-processing, was used as an alternative. As it was the aim to determine when the rinse water is again similar to the reference water, a training dataset consisting of two groups was created. The first group contained the reference water, which was known to have the required quality, and was therefore named 'clean'. Every hour, new reference samples were added to this group of the training dataset. A higher number of samples within the training dataset benefits the reproducibility and stability of the system and captures the variability within the microbial community of the reference water. This variability is also detectable in the cell counts. At the beginning and the end of the rinsing procedure, a higher number of intact cells was present in the reference water. These samples were respectively taken at noon and in the early morning of the next day. The variable cell count can have several causes like (i) stagnation within pipes, reservoirs or water towers (Lautenschlager *et al.*, 2010b), (ii) water of different origin, (iii) use of different water sources for the production of water and (iv) altering residual chlorine concentrations (Hwang *et al.*, 2012).

The second group in the training dataset contained the water samples that were regarded as 'dirty'. Only the first rinse water samples were added to this group, what makes this group rather small in comparison with the 'clean' group. The small number of replicates within this group impedes the classification. Therefore, the results need to be interpreted carefully.

Classification of the rinse water at different time points shows that the rinse water is similar to the reference water after three hours of rinsing. From this point on, the rinse water could be regarded as clean. After 5 hours of rinsing, the water is however again different from the reference, but still more similar to the clean water than the dirty water. This difference can also be seen in the cell numbers: after three hours, more cells are present in the reference water than in the clean water. This can be caused by attachment of the cells to the water pipes or by the residual concentrations of chlorine from the disinfection procedure at the beginning of the rinsing process.

5. Conclusion

Based on the cell counts and fingerprinting results it could be concluded that a rinsing time of three hours is sufficient. This however needs to be confirmed by the conventional plate count techniques. This would generally mean that the rinsing procedure could be reduced by more than 12 hours, what corresponds to a reduced water cost. Furthermore, the water pipes could be put back into use sooner.

Rinsing times can however vary dependent on the performed works, length of the pipes, type of pipes, etc. Therefore no fixed minimal rinsing time can be assigned. A proposed procedure is the follow up of the rinsing water with flow cytometry. When the water is predicted to be of the desired quality by the fingerprinting technique, the rinsing procedure can be stopped and samples can be taken for quality control with the compulsory conventional techniques.

It is important to mention that this fingerprinting technique is not able to replace the compulsory quality control done by conventional plating. This is still required to assure the absence of possible pathogens by showing the absence of their indicator organisms. The method can only be used to decrease the rinsing time and not the time required for quality control.

6. Acknowledgements

CHAPTER

6

GENERAL DISCUSSION AND PERSPECTIVES

GENERAL DISCUSSION AND PERSPECTIVES

Chapter partially redrafted after:

- **De Roy K**, Marzorati M, Van den Abbeele P, Van de Wiele T & Boon N (2013) Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environmental Microbiology (in press).*

1. General research outcomes

1.1. Positioning of this research

Microbial resource management or MRM has been defined as the optimal management of microbial resources in order to develop novel products and to optimise and steer microbial processes. Its final aim is to improve the environment, human health and biotechnological applications in the most sustainable way (Verstraete *et al.*, 2007; Read *et al.*, 2011). Management may occur at the level of single cells, i.e. engineering of individual microbial populations to improve their resistance to stress, to have a higher productivity or to degrade toxic compounds (Benner & Sismour, 2005). Furthermore, management may also occur at the level of the complex microbial community, which inhabits natural and anthropogenic environments and whose final functionalities often result from metabolic networking among the different members. Despite the potential of MRM, the road to translate MRM into practice is still long. Several aspects require further investigation, like the factors that influence and shape microbial communities. Synthetic microbial communities were proposed in Chapter 1 as a useful tool for the controlled study of ecological theories. In Chapter 2 and 3, synthetic microbial communities were used to investigate biological invasion and the establishment of microbial mutualism.

Another aspect that impedes MRM is the characterisation of microbial communities. Microbial community analysis techniques are required that are able to provide more information about the current status of the microbial community, like its composition, structure and physiological characteristics. Conventional techniques are generally time consuming. Flow cytometry was proposed in Chapter 1 as a fast and promising alternative. In Chapter 4, a flow cytometric fingerprinting technique was developed for the analysis of microbial communities. In Chapter 5, the fingerprinting technique developed in Chapter 4 was applied for the follow up of water quality in a water distribution network (Figure 6-1).
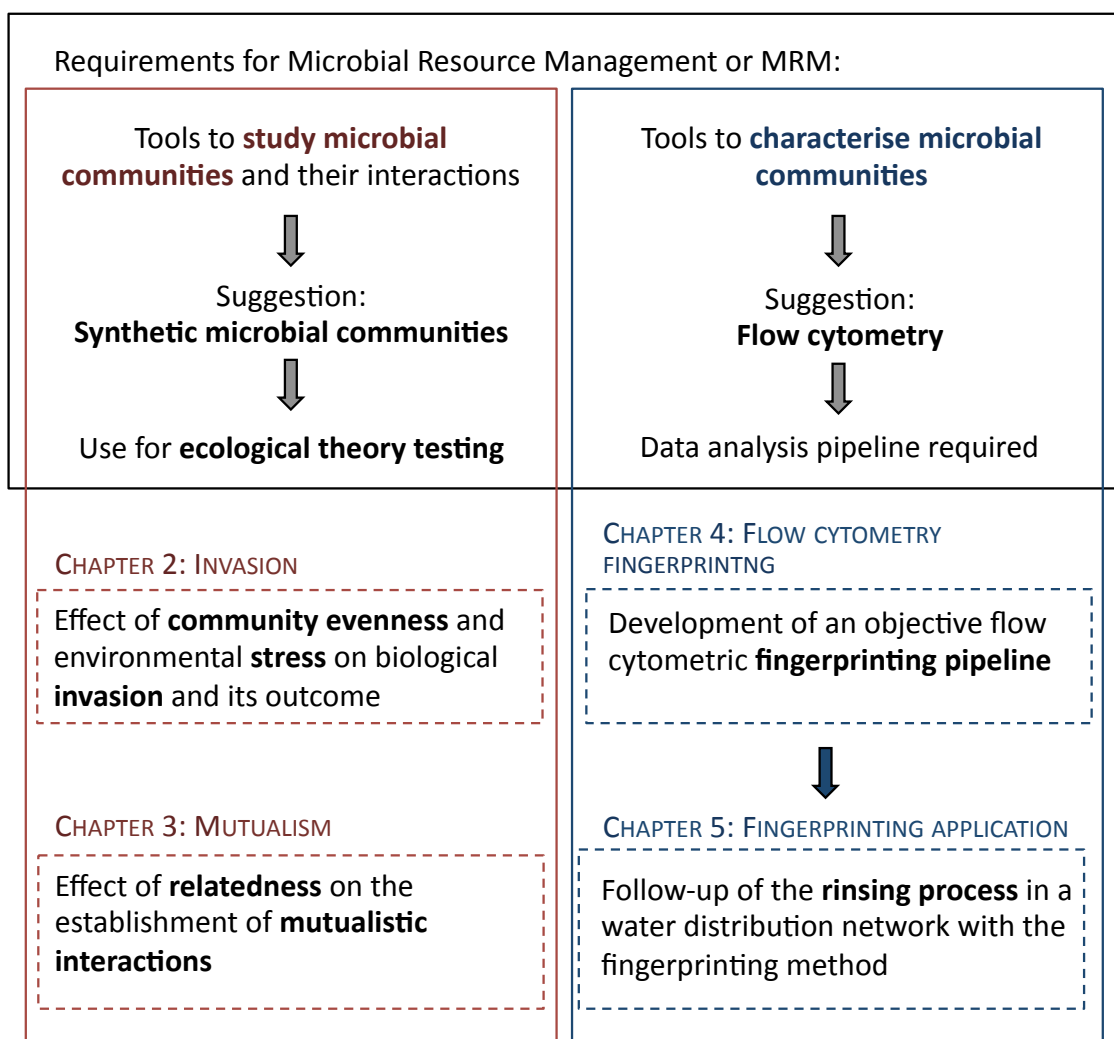
Figure 6-1.Schematic research overview for locating the different chapters.

1.2. Main research outcomes

In the first part of this dissertation, ecological theories were tested by the use of synthetic microbial communities to characterise and investigate factors that influence and affect microbial communities.

In **Chapter 2,** we evaluated the degree of invasion and the effect on the community functionality in relation to the initial community evenness under the influence of specific environmental stressors. We showed that evenness influences the level of invasion and that the introduced species can promote functionality under stress. In the absence of stress, invasion was higher in an uneven community compared to an even community and invasion had a negative effect on the community functionality. On the other hand, the evenness-invasibility relationship was neutral in the presence of stress. Under these stress conditions, the introduced species was able to maintain the functionality of uneven communities, which was lower in the absence of invasion. These results indicate that communities, having the same genetic background, in the presence of the same invader, react in a different way with respect to invasibility and functionality depending on specific environmental conditions and community evenness.

In **Chapter 3,** the effect of interspecies relatedness on the establishment of a mutualistic interaction was investigated. Synthetic microbial communities were used to test the importance of interspecies relatedness during the establishment of mutualism between previously non-interacting microorganisms. Obligatory mutualism was created by making pairwise combinations of an ampicillin resistant, tyrosine auxotrophic strain and ampicillin sensitive strains. At the initial encounter, no cooperation could be detected in any of the mixes. Closely related species were able to adapt their phenotype after longer contact times and successfully established mutualisms. While distantly related species were not able to establish mutualism and were consequently threatened with extinction.

In the second part of this dissertation a flow cytometric based approach for the fast and objective characterisation and classification of microbial communities was developed, tested and applied.

In **Chapter 4**, a flow cytometry based approach was developed for a fast and objective comparison of microbial communities based on the distribution of cellular features from single cells within these communities. The method consists of two main parts, firstly the generation of fingerprint data by flow cytometry and secondly a novel statistical pipeline for

CHAPTER 6

the analysis of flow cytometric data. The combined method was shown to be useful for the discrimination and classification of different brands of drinking water. It was also successfully applied to detect changes in the microbial community composition of drinking water caused by changing environmental factors. Generally, the method can be used as a fast fingerprinting method of microbial communities in aquatic samples and as a tool to detect shifts within these communities.

In **Chapter 5,** the applicability of the flow cytometric fingerprinting technique developed in **Chapter 4** was tested in a full-scale water distribution network. It was used for the follow up of the water quality during the rinsing process of water pipes. The conventional procedure consists of disinfection by the addition of chlorine followed by rinsing. The flow cytometric fingerprinting approach was used to determine the minimal rinsing time, this is the time needed for the rinse water to be again of the same quality as the clean and safe reference water. It was shown that the rinsing time could be drastically reduced and that flow cytometry fingerprinting is a valuable technique for the follow up of this rinsing time.

2. Synthetic microbial ecosystems

2.1. Concerns and future perspectives of synthetic microbial ecology research

The majority of synthetic ecosystems consist of only two to four species. Although being very useful to study ecological theories, the resemblance with natural ecosystems and potential for practical applications is limited. Therefore, a next step in synthetic ecology is to create synthetic ecosystems with increasing resemblance to natural ecosystem. The better a model can simulate the actual complexity of nature, the higher is its scientific value. Firstly, this can be achieved by using sophisticated experimental models that better simulate the environmental factors. An example of a sophisticated model is a high-pressure reactor to simulate the deep-sea environment (Zhang *et al.*, 2011). Secondly, synthetic ecosystems can be optimised by increasing the number of species and optimising their composition, structure and functionality. It will however be impossible to use as many species as present in a natural ecosystem. Dependent on the ecosystem, very high numbers of species are present, soil can contain up to 30 000 species (Curtis *et al.*, 2002). But most communities examined to date feature a species abundance distribution in which the majority of species is found in low abundance, while a limited number is found in high abundance (Figure 6-2) (Nemergut *et al.*, 2013). As such, synthetic ecosystems can be created focusing on the most abundant species.
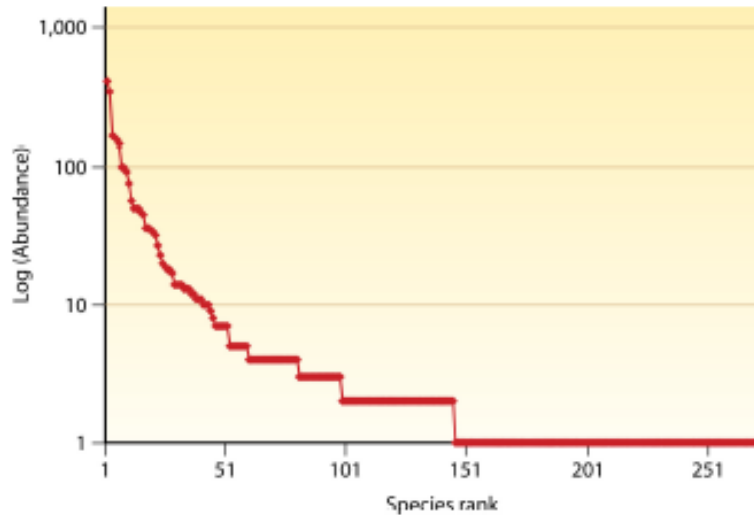
Figure 6-2. Typical rank-abundance plot. *Each point represents the abundance of one species within the community. A common feature of many microbial communities is that only a limited number of species is present in high abundance, while the majority is present in low abundance (Nemergut et al., 2013).*

Studies using synthetic ecosystems with a high number of species have mostly been restricted to short-term experiments, due to stability issues of synthetic communities. It was theoretically shown by ecological models that some species and specific mixtures of agonistic and mutualistic interactions between species are necessary to obtain a stable ecosystem (Boyd, 2012; Mougi & Kondoh, 2012). The integration of such models in microbial ecology would be of high value. Research with synthetic microbial ecosystems created an enormous amount of complementary data, in addition, the genomes of numerous microorganisms have been sequenced. By combining these data and information, *in silico* models making use of 'digital microorganisms' can be created and used for the construction of synthetic ecosystems with desired characteristics (Figure 6-3) (Yedid *et al.*, 2009). Furthermore, these models can be used to predict an ecosystem's behaviour like stability, resistance and functionality. The problem with many ecological models is the lack of validation and overparameterisation. Therefore, we argue to use real ecosystem, *in vivo* models, sophisticated *in vitro* models or synthetic ecosystems for the validation of *in silico* theoretical models and correct for possible overparameterisation. But also to use real ecosystems to check the relevance of synthetic ecosystems, since numerous important factors could be missed. Only when this is done, models can really contribute to the understanding, prediction and management of ecosystems.
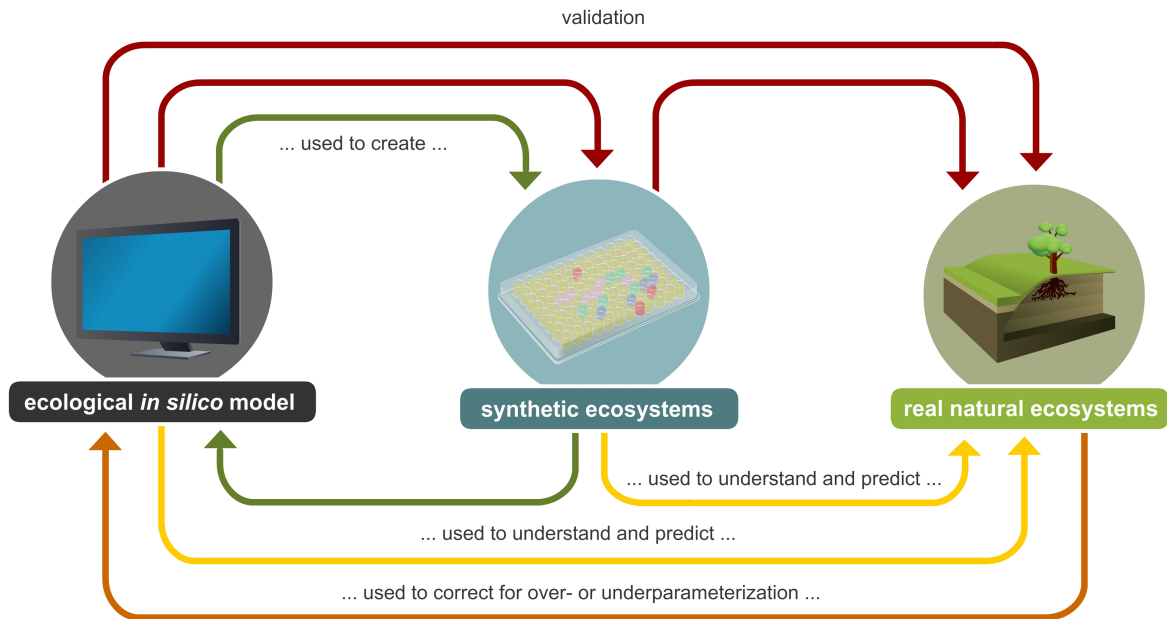
Figure 6-3. The future of synthetic ecosystem research. *Research with synthetic ecosystems drastically increased the knowledge on microbial ecosystems. All this information could be used to create in silico models that can predict an ecosystem's behaviour. After validation and correction for possible over- or underparameterisation, these models could be used to understand, predict, manage and create ecosystems.*

2.2.   The applied use of synthetic microbial communities

Synthetic microbial communities were used in the first part of this thesis for testing different ecological theories. The use of these synthetic communities is however not restricted to research. Synthetic ecology has been shown numerous times to be important in the development of specific applications, representing a good balance in terms of complexity, relevance and manageability.

Synthetic communities can be used, for instance, to recycle waste products. The European Space Agency (ESA) designed MELiSSA (Micro-Ecological Life Support System Alternative), a bioregenerative life support system for the complete recycling of gas, liquid and solid wastes during long distance space exploration (Fulget *et al.*, 1999; Hendrickx *et al.*, 2006b). In MELiSSA, cyanobacteria and plants were use as food sources. As both cyanobacteria and plants preferentially take up nitrogen as nitrate, the ammonium-enriched liquid waste derived from human activities needs to be nitrified to nitrate to create the most optimal recycling system. Therefore, ammonia is oxidized to nitrite by ammonia-oxidizing bacteria (i.e. genera *Nitrosomonas*, *Nitrosococcus*, *Nitrosospira*, *Nitrosolobus*, and *Nitrosovibrio,* currently a coculture of *Nitrosomonas europaea* ATCC 19178 and *Nitrobacter*

winogradskyi ATCC 25391 is used) and then nitrite to nitrate by nitrite oxidizers (i.e. genera *Nitrobacter*, *Nitrococcus*, and *Nitrospira*). Considering that MELiSSA has been designed for space exploration, the stability of the system is a key aspect in order to assure long-term functionality. In this respect, the choice of a synthetic community should assure both a functional and compositional stability as the environment is well defined and the required metabolic conversions are not complex. In fact, according to Pimm (1984), the more the functionality of one species depends on the activity of another species, the fewer species will be necessary to maintain ecosystem stability. Moreover, as the loss of a species would lead to the disruption of the whole ecosystem, the designed synthetic community should be also resilient to perturbation (Pimm, 1984).

Synthetic communities also play a key role in the industrial fermentation and production of chemical compounds. In industrial bioethanol production, most ethanol is produced by the fermentation of glucose or sucrose from corn, sugar cane or beets. Because this competes with food production, alternative sources of sugar are investigated, such as lignocellulosic biomass. Glucose and xylose are the two dominant sugars. But current approaches are inefficient, since no native microorganisms can convert all sugars into ethanol at high yield. Therefore co-cultures of strains that have a high yield for different sugars are used (Chen, 2011). Patle and Lal (2007), showed that a very simple community composed of *Zymomonas mobilis* and *Candida tropicalis* was able to transform enzymatically hydrolysed lignocellulosic biomass into ethanol with a yield of 97.7%. Mixed-culture fermentation from lignocellulosic biomass for ethanol production can increase ethanol yield and production rate and reduce process cost.

Synthetic microbial communities consisting of *Ketogulonicigenium vulgare* and *Bacillus megaterium* have been used in industry to produce 2-keto-gulonic acid (2-KGA), the precursor of vitamin C (Ma *et al.*, 2011). By means of quantitative systems biology analysis, it was shown that the cell lysis of *B. megaterium* provided key elements necessary for *K. vulgare* to grow better and produce more 2-KGA, as compared to the production as a pure strain.

Another field of application for synthetic communities is the bioremediation of contaminated areas. This approach often relies on the addition of microorganisms with the metabolic potential to degrade a specific contaminant, i.e. bioaugmentation. Given the high complexity of some contaminants, bioaugmentation of single strains may not be sufficient to achieve a good 'removal efficiency', as demonstrated in the case of the pesticide linuron (Dejonghe *et al.*, 2003). *Variovorax* sp. strain WDL1 could degrade linuron using it as C, N and energy

source. Conversely, *Delftia acidovorans* WDL34 and *Pseudomonas* sp. strain WDL5 were not able to use linuron but only some intermediate of its degradation. When these strains were mixed in a synthetic community, the rate of linuron degradation improved due to the synergistic interaction of the strain WDL1 with the other bacteria. A similar case is represented by the degradation of 4-chlorosalicylate (4-CS). This compound can only be degraded if *Pseudomonas reinekei* (MT1), *Wautersiella falsenii* (MT2), *Achromobacter spanius* (MT3) and *Pseudomonas veronii* (MT4) work together (Pawelczyk *et al.*, 2008).

A final example is the application of synthetic microbial communities as a safe alternative for human faecal transplants. Because the human gut contains a dense (10^{13}-10^{14} microbial cells) and diverse microbial community (Eckburg *et al.*, 2005), consisting of several hundreds of microbial species, severe disturbances of this ecosystem are unlikely to be resolved by the administration of a single probiotic strain. Indeed, recurrent *Clostridium difficile*-associated diarrhoea (Khoruts *et al.*, 2010; Guo *et al.*, 2012), which is thought to result from persistent disruption of the commensal gut microbiota, was cured upon transplantation of a complex faecal microbiota derived from a healthy human donor (Shahinas *et al.*, 2012). This approach is however only applied in severe cases given the high complexity of a human faecal sample, which is inherently associated with a certain risk for transmitting disease. As a result, there is a large potential for synthetic ecology to mix a well-characterised and safe set of gut microorganisms. Petrof *et al.* (2013) synthesized a synthetic microbiota consisting of 33 individual microbial species and indeed demonstrated the potential of such synthetic microbiota in the eradication of *Clostridium difficile* infections. Such approaches may result in a replacement of commonly used antibiotics.

All the cases described in this section demonstrate the potential for synthetic communities in practical applications.

3.   Cooperation and interactions in microbial communities

Chapter 2 and 3 both deal with the microbial interactions at play in a microbial community and how these interactions are crucial for the survival and functionality of a community. In both chapters, the interactions were investigated at a different level of complexity. In chapter 2 microbial interactions were investigated on community level using synthetic communities with a high richness and different levels of initial evenness. While the development of mutualistic interactions was studied on species level using co-incubations of only two strains in chapter 3.

3.1. Satellite populations supporting the core population

'Invasion', which is the introduction of non-native species to a community, is generally considered as negative. In this thesis, the outcome of microbial invasion was shown to be greatly dependent on the environmental conditions and community structure. The invader was able to improve an ecosystem's functionality under stress, while the functionality was strongly impeded in the absence of the invader. As such, the invader had a supportive role within the stressed community. In a way, the invader helped the community to maintain its functionality. This supportive effect was also reported in an activated sludge bioreactor: a bioaugmented reactor could better protect the native community its structure and function compared to not bioaugmented reactors (Boon *et al.*, 2003). The mechanism of these supportive interactions is however still unclear and lot of new questions are raised: How does the invader improve the functionality? Are specific microbial interactions involved? Does every functional community require a supporting population? etc. These questions could be further investigated by the use of synthetic communities to create a 'collaborome' that consists of a functional 'core population' and a supporting 'satellite population'. By (i) using fully characterised species, (ii) gradually increasing the number of species within the 'collaborome', (iii) analysing the performance of the communities in the presence and absence of stress (e.g. growth and functionality) and (iv) performing proteomics, metabolomics and transcriptomics, it will be possible to fully map the microbial community and the interactions involved in this community. Transcriptomics and proteomics allow the identification of crucial genes and proteins that are being up- or down-regulated under specific conditions. Metabolomics allows the identification of small molecules that can serve as signal or communication molecule and that possibly alter the gene expression and consequently the phenotype of the participating strains. Like that, a better insight into the invasion process and its effect on the microbial community could be obtained.

3.2. Invasion in disparate ecosystems

Next to the supporting role that an invader can have on a microbial community, it was also shown that the composition of the native community is a key factor that determines invasion success. This information can be used to perform MRM in environments where invasion should be avoided or where it should be enhanced. In different environments, such as drinking water and the human gut, invasion of bacteria, such as pathogens, should be avoided. According to the results obtained in this thesis, this would mean that the microbial

communities should have a high evenness. This was also confirmed by a recent study in which the invasion in sewage and drinking water was compared. In sewage water, which was shown to have a higher evenness than drinking water, the invasion of a pathogen was less than in drinking water (Van Nevel *et al.*, 2013a). Situations in which invasion is desirable are e.g. bioaugmentation of polluted ecosystems with bacteria that are able to degrade the toxic compounds or the supplementation of probiotics to the human gut via food to have a health benefit. According to our findings, invasion can be improved in these ecosystems by stressing it and giving it a more dominant community structure.

3.3. Establishment of mutualistic interactions

In chapter 3, phylogenetic relatedness was shown to be very important in the development of mutualistic interactions. Apparently, distantly related species prefer to die than to cooperate. However, this is probably not a matter of 'preferring' but rather a matter of 'having the ability' to cooperate. Distantly related species are possibly not able to communicate with each other in a proper way or do not have the ability to react properly to the signals. This is however not yet clearly understood and the answers given in this chapter raises again some questions: Why do some species cooperate and others not? How do species cooperate? Is it possible to manage and steer the communication and collaboration between microorganisms? How do species adapt their phenotype? etc. To answer all these questions and unravel the underlying mechanisms of microbial interactions, in depth studies are required. Performing co-incubations of two species followed by in depth analysis of the phenotype and genotype can provide more information. Metabolomics will allow the identification of signal molecules such as alarmones, which are hypothised in this work to be important in the establishment of mutualism and which are shown to be an important warning system for bacteria (Rowbury, 2001). The effect of these alarmones on gene expression of closely and distantly related species and the importance in the establishment of mutualistic interactions could be tested and the knowledge used for microbial resource management. This will also provide an explanation why closely related species are better cooperators than distantly related species. Once some of the important factors in the establishment of interactions are identified, this information can be used to enhance microbial interaction. Or to do the opposite, quench interactions. Furthermore, the information can be used to construct functionally stable microbial consortia with desired characteristics (Kerner *et al.*, 2012).

4.   Flow cytometric fingerprinting

4.1.   General considerations

The flow cytometric fingerprinting technique was developed as a diagnostic tool for the characterisation of microbial communities to allow MRM. This method was developed and optimised using drinking water. This liquid matrix allows easy flow cytometric analysis without the requirements of any sample pre-treatment. Furthermore, lot of information is available on this 'easy environment' (Hammes & Egli, 2010; Van Nevel *et al.*, 2013a). Therefore, drinking water was a suited environment and good showcase for development of a diagnostic tool.

On the other hand, drinking water is a very stable environment and does not require MRM, unlike many other environments such as wastewater in a wastewater treatment plant, soil, etc. (Verstraete *et al.*, 2012). The applicability of the flow cytometric fingerprinting technique for these 'difficult environments' still needs to be investigated. These environments are regarded as difficult due to the limitations of flow cytometry. Flow cytometry requires a liquid matrix with suspended cells and without particles larger than 50 μm, which can cause clogging of the system. Pre-treatment steps for the extraction and purification of cells are therefore essential for soil, sediments, plant, biofilms, etc. (Amalfitano & Fazi, 2008). All these pre-treatment steps need optimisation and standardisation to allow the comparison of samples by the flow cytometric fingerprinting techniques. Therefore, it will be challenging to use and apply flow cytometry and the fingerprinting technique in the different ecosystems.

4.2.   Comparison with alternative flow cytometric fingerprinting techniques

4.2.1.   Alternative fingerprinting techniques

Since the development of our flow cytometric fingerprinting technique (De Roy *et al.*, 2012), two alternatives have been proposed by Koch and coworkers (Koch *et al.*, 2013a; Koch *et al.*, 2013b; Koch *et al.*, 2013c). All these methods have clear advantages and disadvantages, which are summarised here:

The first method of Koch *et al.* (2013c) creates a fingerprint based on the number of cells in cell clusters, which are defined by gates. The gates comprise mixed cell compositions and are regarded as subcommunities. These fingerprints were correlated with abiotic parameters. This way, they were able to identify activity hot spots in microbial communities and detect

functional rather than phylogenetic subcommunities. The main disadvantage of this method is the subjectivity of the gating procedure, which is minimized by the use of a gating template. The position of the gates is dependent on the type of instrument, its settings, the type of staining and the staining procedure. Furthermore, a flow cytometric pattern is different for every environment. Therefore, new gating templates need to be created once one of these parameters is changed. The most important advantage of this fingerprinting method is the ability to link the flow cytometry pattern to the functionality of the microbial community. By correlating the fingerprints with abiotic data, they are able to make a prediction on the functional outcome of microbial communities in a specific environment. This is very interesting with the eye on microbial resource management.

In the second method they have developed, they tackled the gating problem. They did this by exporting the flow cytometric dot plots to image analysis software that compares the plots by comparing single pixel values (Koch *et al.*, 2013a). This procedure is called "Cytometric Histogram Image Comparison" or CHIC. This is an easy and straightforward method that is person independent. The main problem of this method is the loss of very informative data. Because the method is image-based there is no more information on cell counts and fluorescence intensities. The main advantage compared to our method is that no model is needed to create the fingerprints.

4.2.2. Limitations of the developed fingerprinting technique

The main limitation of the flow cytometric fingerprinting technique is the importance of the training dataset, which is a complete database that contains all samples of an experiment or a part of this database. The training dataset is used in the initial step of the fingerprinting technique: the probabilistic binning algorithm pools the samples of the training dataset to create the fingerprinting model. This training dataset needs to represent an ecosystem and the expected variability within this ecosystem and therefore requires a sufficient amount of samples. Only when this is the case, a reliable model can be made that generates fingerprints, which are sensitive enough to detect minor changes in a microbial community or differences between microbial communities. For samples originating from a different environment or stained with another dye, a new model needs to be created. Therefore, a new way of creating the fingerprinting model that does not require a training dataset is proposed in the future perspectives of this chapter.

A second part of the fingerprinting procedure in which the database is of importance is the dimension reduction and discrimination. The linear projection of the variables that is performed by PCA and FDA is data driven. This means that a different dataset will lead to different projections (explaining the differences in discriminant values between the different time points in Chapter 5). Therefore, the dataset and the number of samples within each treatment group needs to be sufficiently big, so small changes within the database only have a small influence on the PCA and FDA. This is however strongly dependent on the research question of the experimental setup. A sufficiently big database is also required to allow good classification of unknown samples.

4.3. Future perspectives

4.3.1. Optimising flow cytometric fingerprinting

As discussed above, the different fingerprinting techniques used to characterise and compare microbial communities have distinct advantages and disadvantages. Therefore, an 'optimal fingerprinting technique' that is based on our own fingerprinting technique and that combines the advantages of all three methods is proposed here.

The first step in our original fingerprinting procedure was the development of the fingerprinting model using the probabilistic binning algorithm. For this, a training dataset that represents the ecosystem and in which all samples were analysed and stained in an identical way, was necessary. For each ecosystem and each type of staining, a new training dataset and fingerprinting model had to be created. Therefore, it would be interesting to create a model that is (i) independent of the ecosystem or type of staining and (ii) for which no training dataset needs to be created. Instead of using the probabilistic binning algorithm, we propose using an n-dimensional grid with bins that all have the same size and shape. In case two parameters are of interest, the model would be a 2D-grid in which the bins are squares of equal size, for three parameters this is a 3D-grid with cubes, etc. And thus, the number of bins will depend on the number of parameters of interest. As the method should be able to detect subtle changes or differences in fluorescence intensity, FSC or SSC, this number will probably be much larger then the number of bins we conventionally work with.

In the next step, the fingerprint will be created. This will be done as previously described, by determining the number of cells in each region. Also the statistical pipeline used to look for differences between microbial communities does not require changes. This pipeline is a

combination of FDA and PCA and is a supervised technique for which the samples are first ascribed to groups. Subsequently, the method looks for the most important features to differentiate groups and it allows classification of unknown samples to one of the groups. Although the method does not require a training dataset to create the fingerprinting model, for classification of unknown samples there is of course still need for a training dataset. The bigger this dataset, the more variability is included and the more reliable the classification will be. The statistical pipeline also allows going back to the initial flow cytometry dimensions and locate those bins that are most important to differentiate groups and to see how microbial communities change.

As this method requires to first define groups, some basic things like a simple comparison between a few samples can become too complicated. Therefore we propose to create an additional statistical pipeline to perform a simple comparison between fingerprints, that is also able to localise the differences.

As mentioned earlier, correlating fingerprints with abiotic and biotic factors, like done by Koch *et al.* (2013c), might be very interesting in view of MRM. It enables the identification of activity hotspots in microbial communities, but might also allow the detection of specific kinds of environmental stress, invasion, interactions, etc. Factors that have been shown in Chapter 2 and 3 to be import for microbial communities' functionality. Therefore it is useful to create a deeper understanding of the correlation between flow cytometric fingerprints and different abiotic and biotic factors.

4.3.2. Flow cytometry based MRM parameters

As an alternative to linking the fingerprints to abiotic and biotic factors, the flow cytometry data and fingerprints can also be used to develop "flow cytometric MRM parameters". The conventional MRM parameters, being range-weighted richness (Rr), dynamics (Dy) and functional organisation (Fo), were developed by Marzorati *et al.* (2008). They were introduced as a conceptual interpretation of molecular fingerprinting patterns like DGGE, TGGE and T-RFLP. The parameters provide an ecological and predictive value for the analysis of the structure and diversity of microbial communities in a certain environment. The MRM parameters have been proven useful in different environments, including drinking water (Lautenschlager *et al.*, 2010a), microbial fuel cells (Aelterman *et al.*, 2008), wastewater treatment (Vlaeminck *et al.*, 2009) and the human gut (Possemiers *et al.*, 2010; Van den

Abbeele *et al.*, 2010). Therefore it is useful to redefine the MRM parameters based on flow cytometry data. This idea was first raised by Wang *et al.* (2010) and was applied a first time in the doctoral dissertation of Koch (2013). There, the MRM parameters were defined after the flow cytometric data was converted using the CHIC-procedure and are based on pixel values. Here we propose similar MRM parameters based on our fingerprinting technique and how to interpret these parameters.
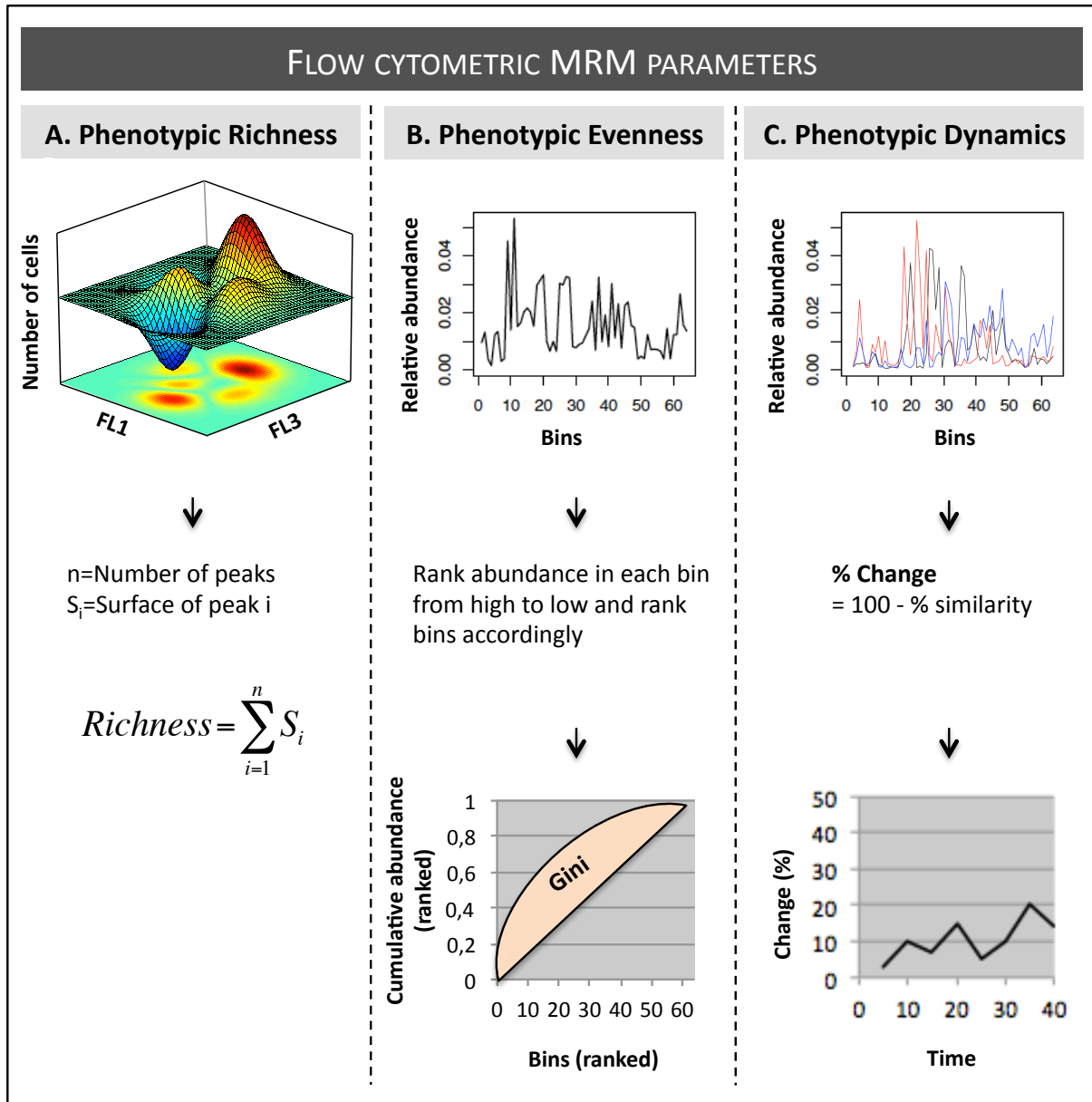
Figure 6-4. New MRM parameters based on flow cytometric data and fingerprints.

Phenotypic richness

The range-weighted richness, or species richness in more general, is conventionally calculated based on the number of bands in a DGGE-profile. In a DGGE, sequences are separated

according to the GC-content of the DNA. Theoretically, one band corresponds to one species and the number of bands is used as a measure of species richness. In flow cytometry, microorganisms with similar physiological and morphological characteristics will create signals with similar fluorescence intensity, FSC and SSC. They will be positioned closely together in dot plots and create peaks in 2D histograms (Figure 6-4A). The more peaks are present, the more microorganisms with clearly different characteristics are present. A peak with a big surface area contains more species with slightly different characteristics, while small peaks all contain very similar microorganisms. Therefore, a combination of peak surface area (S_i) and the number of peaks (n) is proposed as a measure of phenotypic richness:

$$Phenotypic\ richness = \sum_{i=1}^{n} S_i$$

The surface area within this equation is used when the phenotypic richness is based on only two parameters, like FL1 and FL3, as shown in Figure 6-4A. In case of three or more parameters, this surface area will become a volume or hypervolume and the equation needs to be changed accordingly.

It is important to mention that this richness does not correspond to the species richness, but to the phenotypic and thus the physiological and morphological richness of a microbial community. This is because identical strains can create a different pattern if they are under stress, dying, dividing, metabolically active or inactive, etc.

Phenotypic evenness

The functional organisation (Fo), lately more frequently referred to as community organisation (Co) or community evenness (Read *et al.*, 2011), is visualised by Lorenz evenness curves. It describes the difference between the relative abundance of different species and is a measure for the organisation of the community. For DGGE, the Lorenz curves are based on the relative band intensities. The bands are ranked from high to low intensity and the cumulative relative band intensities form the y-axis, while the cumulative normalized bands are used as the x-axis. The surface between the Lorenz curve and the diagonal is know as the Gini coefficient and describes the evenness. A high evenness corresponds to a low Gini coefficient, while a low evenness has a high Gini coefficient.

For flow cytometry, the Lorenz curves can be drawn in a very similar way, by ranking the bins from the fingerprint from high to low relative cell abundance (Figure 6-4B). The

cumulative relative abundance forms the y-axis and the ordered bins the x-axis. The Gini coefficient is calculated as before.

The phenotypic evenness is a measure of the difference in relative abundance of microorganisms with specific morphological and physiological characteristics. If the Gini is high, the community is dominated by microorganisms with specific phenotypic characteristics, while microbial communities with a low Gini have a wide variety of cells.

Phenotypic dynamics

The dynamics determines how much a microbial community changes over time and is visualised by a moving window analysis, which shows the percentage of change (= 100 - %similarity) between two consecutive time points. When applied to DGGE, it refers to the number of species that are detected to be of significance in a given habitat at a certain time point (Marzorati *et al.*, 2008; Read *et al.*, 2011).

For flow cytometry, the percentage of change can be determined by comparing two fingerprints (Figure 6-4C). It shows how much the microbial community phenotypically changes over time. This can have multiple causes: (i) some microorganisms may become more dominant, while others are disappearing, (ii) they can become active or inactive, (iii) they can start growing, (iv) change morphologically, etc. As such, the dynamics is a good way to follow microbial communities over time.

When these MRM parameters are created using the original fingerprinting model, the fingerprints and thus the MRM parameters are dependent on the training dataset that is used. When this dataset is sufficiently big, this will only have a minor influence on the fingerprints. Furthermore, the same model will be used for all samples, what makes the importance of the training dataset limited for the flow cytometry based MRM parameters.

When the new proposed fingerprinting model is used, no training dataset is required for making the fingerprint model and thus the fingerprints and the flow cytometry based MRM parameters.

4.3.3. Integration of flow cytometry with established microbiological tools

Flow cytometry is a very fast method to study microbial communities and is for this reason interesting for fast screening of microbial communities. It has been shown multiple times that, with the correct staining, flow cytometry can give more information on the phenotypic status

of bacterial cells. With the development of the flow cytometry based MRM parameters and the fingerprinting technique, flow cytometry could even help understanding microbial communities and their behaviour. But giving an in depth interpretation of flow cytometry data remains difficult. It is for example not possible to easily identify species, follow gene expression or analyse microbial interactions. Established microbiological and molecular tools, such as sequencing, plating, DGGE etc. have the ability to provide all this information. Therefore, flow cytometry should not be regarded as a stand-alone-tool, but should be used in combination with conventional and well-understood techniques.

By analysing microbial communities by flow cytometry, changes within the microbial community can be easily detected, but not explained. Transcriptomics, proteomics and metabolomics can give a better insight into this. They provide information on the taxonomic and functional diversity, the population structure, the presence of genes as well as their expression levels and levels of translation into proteins (VerBerkmoes *et al.*, 2009).

5.   Conclusions

In this thesis, synthetic ecosystems were used to test different ecological theories to improve our knowledge on microbial communities and the processes involved within these communities. Factors like community composition, stress and invasion were shown to influence the behaviour, resilience and functionality of a microbial community. A flow cytometric fingerprinting technique was developed to analyse and compare microbial communities. New MRM parameters based on this fingerprint technique were proposed to quickly determine the current status of a microbial community. By combining the knowledge gained on microbial communities and the flow cytometric fingerprinting method to analyse the current status of a microbial community, it will become possible to predict the behaviour of a microbial community, localise possible problems and respond in a suitable way, and thus perform MRM.

However we shortened the road to MRM, we are not there yet. There is room for improvement for both the synthetic ecosystems and the flow cytometric fingerprinting technique. For synthetic ecosystem research, we propose to work in parallel with *in silico* and *in vivo* methods to be able to model and predict the behaviour of microbial communities. We also recommend doing this by using the fingerprinting technique in parallel with other established techniques. For flow cytometric fingerprinting, we opt for optimising the fingerprinting technique, to further develop the flow cytometry based MRM parameters and

to apply these MRM parameters to different environments to see how they correlate with stress resistance, functionality, etc. Eventually, the combination of the knowledge on microbial communities and the technique that quickly provides information on the current status of a microbial community will allow us to perform MRM.

CHAPTER 6

ABSTRACT

Microorganisms contribute to nearly all biogeochemical cycles on earth and play a crucial role in biotechnological processes such as wastewater treatment, production of food and pharmaceuticals, digestion, cleaning of chemically contaminated sites and production of bioenergy. Their genetic and physiological diversity result in an enormous metabolic potential, which makes them essential for life on earth. Most of the processes they are involved in are accomplished by joint effort of microorganisms with different functional roles. They influence each other's behaviour and possibly alter the biochemical phenotypes of the participating strains. Understanding the factors that shape and influence these microbial ecosystems is essential from a microbiological, ecological and biotechnological point of view. It will allow us to predict the response of microbial communities to perturbations and environmental changes. Eventually, this knowledge will make it possible to manage, steer and optimise microbial processes, an approach that is better known as "Microbial Resource Management" or MRM. Its final aim is to improve the environment, human health and biotechnological applications in the most sustainable way. Despite the potential of MRM, the road to translate MRM into practice is still long. Several aspects require further investigation, like the establishment of microbial interactions, resistance to stress and perturbations, etc. Furthermore, there is need for a fast and objective technique that provides more information on the current status of the microbial community, like its composition, structure and physiological characteristics.

In Chapter 1, a theory driven approach was proposed to study the aspects that influence microbial communities. This theory driven approach encompasses the generation of ecological theories based on existing observational data and its verification using quantitative research. A deliberate choice of the experimental setup, methodology and microbial model systems is indispensable for optimal hypothesis testing. Synthetic microbial ecosystems with intermediate complexity and high controllability were shown to be very useful for this. These synthetic communities were used in Chapter 2 to investigate a first ecological phenomenon: biological invasion. Invasion is widely studied in the animal and plant kingdom, however conclusions on the outcome of this process mainly originate from observations in systems that

leave a large number of experimental variables uncontrolled. These confounding factors and the reciprocal interactions lead to opposite conclusions on the role of specific parameters. In this dissertation, a fully controlled system, consisting of assembled bacterial communities, was used to investigate how community evenness and salinity stress influence invasion of non-native species in a microbial community and the effect this invader has on the microbial community. In the absence of stress, invasion was higher in uneven communities compared to even communities and resulted in a small decrease of the community functionality. In the presence of salinity stress, community evenness had no effect on invasion. Under these stress conditions but in the absence of invasion, uneven communities had a lower functionality compared to even communities. The invader on the other hand, which did not contribute to the community functionality, was able to maintain the functionality at a maximum and thus supported the community. These results indicate that communities, having the same genetic background, in the presence of the same invader, react in a different way with respect to invasibility and functionality depending on specific environmental conditions and community evenness.

The second ecological theory tested in this thesis was the role of species relatedness on the establishment of mutualistic interactions. Relatedness is regarded as one of the key reasons for microorganisms to perform an altruistic behaviour that is costly to the actor but beneficial to the recipients. However, its importance during the establishment of mutualistic interactions, which are beneficial to both the actor and recipient, remains unclear. Here, synthetic communities, consisting of two species, were used to create obligatory mutualism. Pairwise combinations of previously not interacting strains were made. One was an ampicillin resistant, tyrosine auxotrophic strain, the other one an ampicillin sensitive strain. Ten ampicillin sensitive strains were selected based on their relatedness to the auxotroph. At the initial encounter, no cooperation could be detected in any of the mixes. After three transfers, closely related species were able to adapt their phenotype and successfully established mutualism, while distantly related species were not able to establish mutualism and were consequently threatened with extinction.

The results observed in Chapter 2 and 3 bring us one step closer to a complete understanding of microbial communities and its management. Furthermore, synthetic microbial ecology has been proven to be a valuable tool for the study of microbial communities and the interactions involved within these communities.

The second part of this dissertation focused on the development of a new methodology for the fast characterisation and comparison of microbial communities. In Chapter 1, flow cytometry

was shown to be a promising tool for the substitution of conventional time consuming techniques. However, some issues, like the subjective handling of the flow cytometry results, still had to be tackled before flow cytometry could be used for community analysis. Therefore, an objective method to quickly and automatically create and compare fingerprints of microbial communities was developed in Chapter 4. The method consists of two main parts. Firstly fingerprints are created out of the flow cytometry data. These fingerprints are based on cellular features of the single cells. In a second step, the fingerprints are analysed with a tailored statistical data analysis pipeline. This method was first applied on different brands of bottled water and on water that was treated with different kinds of environmental stress. Based on the similarities and differences between the microbial fingerprints, we were able to discriminate the different brands of bottled water, classify unknown samples and detect shift of microbial communities caused by changing physico-chemical conditions. This opened opportunities for the real-time follow up of microbiological processes like wastewater treatment, fermentation processes and the production and distribution of drinking water. One of these opportunities was further investigated by using the fingerprinting method for the follow up of the rinsing procedure in a full-scale water distribution network (Chapter 5). After reparation works, the water pipes are conventionally rinsed with an amount of water equal to three times the volume of the water pipe, which is a long and expensive procedure. By the follow up of the microbial community of this rinsing water, it was investigated if the rinsing procedure could be shortened. Therefore, the fingerprints of the rinsing water were compared every hour with the fingerprints of the reference water. After three hours of rinsing, no difference could be detected between these two fingerprints, and the rinsing water could be regarded as clean and safe. As such, the rinsing time could be drastically reduced by real-time flow cytometry in combination with the fingerprinting pipeline. Furthermore, it confirms the potential of flow cytometry for the characterisation of microbial communities.

In conclusion, new tools were developed and used to study and characterise microbial communities with the eye on bringing MRM into practice. A newly developed flow cytometric fingerprinting tool was proven useful for the characterisation of microbial communities, while synthetic microbial ecosystems were used to study the factors that influence and shape microbial communities. Environmental stress, community evenness, invasion and relatedness were shown to partly determine the fate of microbial communities.

ABSTRACT

SAMENVATTING

Micro-organismen dragen bij tot bijna alle biochemische kringlopen op aarde en spelen een cruciale rol in biotechnologische processen zoals afvalwaterzuivering, productie van voedingsstoffen en farmaceutica, spijsvertering, reiniging van chemisch verontreinigde sites en productie van bio-energie. Hun genetische en fysiologische diversiteit resulteert in een enorme metabolische potentieel die hen essentieel maken voor het leven op aarde. De meeste processen waarin ze betrokken zijn komen tot stond door een gezamenlijke inspanning van micro-organismen met verschillende functies. Zij beïnvloeden elkaars gedrag en wijzigen mogelijk het elkaars fenotype. Het begrijpen van de factoren die deze microbiële gemeenschappen vormen en beïnvloeden is daarom belangrijk vanuit zowel een microbiologische, ecologisch als biotechnologisch oogpunt. Deze kennis zal ons toelaten om de reactie van microbiële gemeenschappen op verstoringen en omgevingsveranderingen in te schatten. Dit zal het uiteindelijk mogelijk maken om microbiële processen te managen, sturen en optimaliseren. Deze aanpak is beter bekend als "Microbial Resource Management" of MRM en heeft als uiteindelijk doel om op de meest duurzame manier het milieu, de gezondheid van de mens en biotechnologische toepassingen te verbeteren. Ondanks het potentieel van MRM is de weg naar de omzetting in de praktijk nog lang. Verschillende aspecten dienen verder onderzocht te worden, zoals de ontwikkeling van microbiële interacties, resistentie tegen stress en verstoringen. Bovendien is er nood aan een snelle en objectieve techniek die informatie verschaft over de actuele toestand van de microbiële gemeenschap, zoals zijn samenstelling, structuur en fysiologische eigenschappen.

In hoofdstuk 1 werd een theorie-gedreven aanpak voorgesteld voor het bestuderen van de aspecten die microbiële gemeenschappen beïnvloeden. Deze theorie-gedreven aanpak houdt in data ecologische theorieën geformuleerd worden op basis van waargenomen data die vervolgens geverifieerd worden door het uitvoeren van kwantitatief onderzoek. Een weloverwogen keuze van de experimentele setup, methodologie en de microbiële model systemen is noodzakelijk voor het optimaal testen van hypotheses. Het werd aangetoond dat synthetische microbiële gemeenschappen met een intermediaire complexiteit en een hoge controleerbaarheid hiervoor heel nuttig kunnen zijn. Deze synthetische gemeenschappen

werden in hoofdstuk 2 gebruikt voor het onderzoeken van een eerste ecologisch fenomeen: biologische invasie. Invasie is sterk bestudeerd in het rijkdom van plant en dier, maar de conclusies omtrent de effecten van invasie zijn meestal afkomstig van observaties in systemen die een groot aantal variabelen ongecontroleerd laten. Deze verstorende factoren leiden vaak tot tegengestelde conclusies omtrent de rol van specifieke parameters. Daarom werd in deze thesis gebruik gemaakt van een volledig gecontroleerd systeem bestaande uit synthetische microbiële gemeenschappen om te onderzoeken hoe de gelijkheid of evenness van een microbiële gemeenschap en zoutstress de invasie van uitheemse stammen in deze microbiële gemeenschap beïnvloedt en wat het effect ervan is op de microbiële gemeenschap. In afwezigheid van stress was de invasie groter in ongelijke gemeenschappen vergeleken met gelijke gemeenschappen en resulteerde de invasie in een kleine daling van de functionaliteit. Gelijkheid had in de aanwezigheid van stress geen effect op invasie. Onder deze stress omstandigheden, maar in de afwezigheid van invasie, hadden ongelijke gemeenschappen een verlaagde functionaliteit vergeleken met gelijke gemeenschappen. De invader, die niet bijdroeg tot de functionaliteit, was echter in staat om de functionaliteit van ongelijke gemeenschappen de onderhouden en maximaal te houden en had dus een ondersteunende rol voor de gemeenschap. Deze resultaten tonen aan dat gemeenschappen die initieel dezelfde genetische achtergrond hebben, in de aanwezigheid van dezelfde invader, op een andere manier reageren op het vlak van invasibiliteit en functionaliteit afhankelijk van de omgevingsomstandigheden en de gelijkheid van de gemeenschap.

De tweede ecologische theorie die getest werd in deze thesis is de rol van verwantschap tussen stammen bij de ontwikkeling van mutualistische interacties. Verwantschap wordt beschouwd als een van de belangrijkste redenen voor micro-organismen voor het uitvoeren van een altruïstische daad die kostelijk is voor zichzelf, maar voordelig voor anderen. Het belang bij de ontwikkeling van mutualistiche interacties die voordelig zijn voor zowel de uitvoerder als de ontvanger is echter niet gekend. Om dit te onderzoeken werd in deze thesis synthetische gemeenschappen met twee stammen ontwikkeld met verplicht mutualisme. Hiervoor werden gepaarde combinaties gemaakt van een ampicilline resistente, tyrosine auxotrofe stam en telkens één van de tien ampicilline gevoelige stam. Deze ampicilline gevoelige stammen werden gekozen op basis van hun verwantschap met de auxotroof. Bij de initiële ontmoeting kon bij geen enkele van de paren een samenwerking worden waargenomen. Na drie overentingen waren de nauw verwante stammen echter in staat om hun fenotype aan te passen en werd er succesvol een mutualistische interactie ontwikkeld. Terwijl

ver verwant stammen niet in staat waren om deze mutualistische interactie te ontwikkelen, waardoor de stammen binnen deze combinaties met uitsterven werden bedreigd.

De resultaten uit hoofdstuk 2 en 3 brengen ons een stap dichter bij het managen en volledig begrijpen van microbiële gemeenschappen. Bovendien werd aangetoond dat synthetische microbiële ecologie een waardevolle methode is voor het bestuderen van microbiële gemeenschappen en de interacties die er een rol in spelen.

In het tweede deel van deze thesis werd er gefocust op de ontwikkeling van een nieuwe methodologie voor de snelle karakterisatie en het vergelijken van microbiële gemeenschappen. In hoofdstuk 1 werd flow cytometrie voorgesteld als een veelbelovende techniek voor het vervangen van tijdrovende conventionele technieken. Maar hiervoor dienden eerst enkele problemen, zoals de subjectieve behandeling van de flow cytometrie resultaten, verholpen te worden om het gebruik van flow cytometrie mogelijk te maken voor de analyse van microbiële gemeenschappen. Daarom werd in hoofdstuk 4 een objectieve methode ontwikkeld om snel en automatisch fingerprints van microbiële gemeenschappen te maken en vergelijken. Deze methode bestaat uit twee delen. Eerst worden fingerprints gemaakt uit flow cytometrie data. Deze fingerprints zijn gebaseerd op de cellulaire eigenschappen van individuele cellen. In een tweede stap worden deze fingerprints geanalyseerd met een nieuw ontwikkelde statistische data analyse pipeline. Deze fingerprinting methode werd eerst toegepast op verschillende merken flessenwater en water dat behandeld werd met verschillende soorten stress. Op basis van de gelijkenissen en verschillen tussen de microbiële fingerprints was het mogelijk om de verschillende merken flessenwater van elkaar te onderscheiden, onbekende stalen te classificeren en verschuivingen binnen een microbiële gemeenschap veroorzaakt door veranderde fysico-chemische omstandigheden vast te stellen. Dit opende mogelijkheden om de methode te gebruiken voor het real-time opvolgen van microbiële processen zoals afvalwater zuivering, fermentaties en de productie en distributie van drinkwater. Één van deze mogelijkheden werd verder onderzocht door de fingerprinting techniek te gebruiken voor het opvolgen van het spoelproces van waterleidingen in een full-scale water distributie netwerk (hoofdstuk 5). Na reparaties aan waterleidingen worden deze normaal gespoeld met een hoeveelheid water gelijk aan drie maal het volume van de waterleiding. Dit is een tijdrovende en dure procedure, daarom werd onderzocht of dit spoelproces kan worden ingekort door het opvolgen van de microbiële gemeenschap met de fingerprinting methode. De fingerprints van het spoelwater werden elk uur vergeleken met deze van het propere referentiewater. Na drie uur spoelen konden geen verschillen meer worden waargenomen tussen deze twee fingerprints en kon het

spoelwater aanzien worden als proper en veilig. Deze resultaten tonen aan dat het spoelproces van waterleidingen drastisch kan worden ingekort door real-time flow cytometrie te gebruiken in combinatie met de fingerprinting methode. Bovendien bevestigt dit het potentieel van flow cytometrie voor de karakterisatie van microbiële gemeenschappen.

In conclusie, nieuwe methodes voor het bestuderen en karakteriseren van microbiële gemeenschappen werden ontwikkeld en toegepast met als doel om MRM in de praktijk te brengen. Het werd aangetoond dat de nieuw ontwikkelde flow cytometrische fingerprinting methode nuttig is voor de karakterisatie van microbiële gemeenschappen. Synthetische microbiële ecosystemen werden gebruik voor het bestuderen van de factoren die microbiële gemeenschappen vormen en beïnvloeden. Omgevingsstress, de gelijkheid van de gemeenschap, invasie en verwantschap zijn hierbij belangrijke factoren die het lot van microbiële gemeenschappen bepalen.

REFERENCES

REFERENCES

AELTERMAN P, VERSICHELE M, MARZORATI M, BOON N & VERSTRAETE W (2008) Loading rate and external resistance control the electricity generation of microbial fuel cells with different three-dimensional anodes. *Bioresource Technology* **99**: 8895-8902.

ALLEN B & NOWAK MA (2013) Cooperation and the Fate of Microbial Societies. *Plos Biology* **11**.

ALLEN MJ, EDBERG SC & REASONER DJ (2004) Heterotrophic plate count bacteria - what is their significance in drinking water? *International Journal of Food Microbiology* **92**: 265-274.

AMALFITANO S & FAZI S (2008) Recovery and quantification of bacterial cells associated with streambed sediments. *Journal of Microbiological Methods* **75**: 237-243.

AMANN R (2000) Who is out there? Microbial Aspects of Biodiversity. *Systematic and Applied Microbiology* **23**: 1-8.

AMANN RI (1995) In situ identification of micro-organisms by whole cell hybridization with rRNA-targeted nucleic acid probes. *Molecular Microbial Ecology Manual*. Akkermans ADL, Elsas JD & Bruijn FJ (eds): Springer Netherlands, pp. 331-345.

AMANN RI, LUDWIG W & SCHLEIFER KH (1995) Phylogenetic identification and in-situ detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**: 143-169.

ANDONIAN K, HIERRO JL, KHETSURIANI L, BECERRA P, JANOYAN G, VILLARREAL D, CAVIERES L, FOX LR & CALLAWAY RM (2011) Range-Expanding Populations of a Globally Introduced Weed Experience Negative Plant-Soil Feedbacks. *Plos One* **6**: e20117.

ANDORRA I, MONTEIRO M, ESTEVE-ZARZOSO B, ALBERGARIA H & MAS A (2011) Analysis and direct quantification of Saccharomyces cerevisiae and Hanseniaspora guilliermondii populations during alcoholic fermentation by fluorescence in situ hybridization, flow cytometry and quantitative PCR. *Food Microbiology* **28**: 1483-1491.

ANDREOTE FD, AZEVEDO JL & ARAUJO WL (2009) Assessing the diversity of bacterial communities associated with plants. *Brazilian Journal of Microbiology* **40**: 417-432.

ATKINSON A, DONEV A & TOBIAS R (2007) *Optimum Experimental Designs, with SAS*. Oxford University Press, USA.

BARTRAM J, COTRUVO J, EXNER M, FRICKER C & GLASMACHER A (2003) *Heterotrophic Plate Counts and Drinking-water Safety*. IWA Publishing on behalf of the World Health Organization, London.

BASHASHATI A & BRINKMAN RR (2009) A Survey Of Flow Cytometry Data Analysis Methods. *Advances in Bioinformatics* **2009**: 584603.

BECKER S, BOGER P, OEHLMANN R & ERNST A (2000) PCR bias in ecological analysis: A case study for quantitative Taq nuclease assays in analyses of microbial communities. *Applied and Environmental Microbiology* **66**: 4945-4953.

BELL T, NEWMAN JA, SILVERMAN BW, TURNER SL & LILLEY AK (2005) The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157-1160.

BENNER SA & SISMOUR AM (2005) Synthetic biology. *Nature Reviews Genetics* **6**: 533-543.

BERNEY M, HAMMES F, BOSSHARD F, WEILENMANN HU & EGLI T (2007) Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight kit in combination with flow cytometry. *Applied and Environmental Microbiology* **73**: 3283-3290.

BERNEY M, VITAL M, HULSHOFF I, WEILENMANN HU, EGLI T & HAMMES F (2008) Rapid, cultivation-independent assessment of microbial viability in drinking water. *Water Research* **42**: 4010-4018.

BISSETT A, BROWN MV, SICILIANO SD & THRALL PH (2013) Microbial community responses to anthropogenically induced environmental change: towards a systems approach. *Ecology Letters* **16**: 128-139.

BOON N, DE WINDT W, VERSTRAETE W & TOP EM (2002) Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. *Fems Microbiology Ecology* **39**: 101-112.

BOON N, TOP EM, VERSTRAETE W & SICILIANO SD (2003) Bioaugmentation as a tool to protect the structure and function of an activated-sludge microbial community against a 3-chloroaniline shock load. *Applied and Environmental Microbiology* **69**: 1511-1520.

BOYD IL (2012) The Art of Ecological Modeling. *Science* **337**: 306-307.

BRENNER K & ARNOLD FH (2011) Self-Organization, Layered Structure, and Aggregation Enhance Persistence of a Synthetic Biofilm Consortium. *Plos One* **6**: e16791.

BRUNO JF, STACHOWICZ JJ & BERTNESS MD (2003) Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution* **18**: 119-125.

BULL JJ & HARCOMBE WR (2009) Population Dynamics Constrain the Cooperative Evolution of Cross-Feeding. *Plos One* **4**: e4115.

BURKOVSKI A & KRAMER R (2002) Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Applied Microbiology and Biotechnology* **58**: 265-274.

CADE BS & NOON BR (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1**: 412-420.

CHAPIN FS, ZAVALETA ES, EVINER VT, NAYLOR RL, VITOUSEK PM, REYNOLDS HL, HOOPER DU, LAVOREL S, SALA OE, HOBBIE SE, MACK MC & DIAZ S (2000) Consequences of changing biodiversity. *Nature* **405**: 234-242.

CHAVEZ-BEJAR MI, BAEZ-VIVEROS JL, MARTINEZ A, BOLIVAR F & GOSSET G (2012) Biotechnological production of L-tyrosine and derived compounds. *Process Biochemistry* **47**: 1017-1026.

CHEN YL (2011) Development and application of co-culture for ethanol production by co-fermentation of glucose and xylose: a systematic review. *Journal of Industrial Microbiology & Biotechnology* **38**: 581-597.

CHUANG JS, RIVOIRE O & LEIBLER S (2010) Cooperation and Hamilton's rule in a simple synthetic microbial system. *Molecular Systems Biology* **6**: 398.

COUTO JA & HOGG T (1999) Evaluation of a commercial fluorochromic system for the rapid detection and estimation of wine lactic acid bacteria by DEFT. *Letters in Applied Microbiology* **28**: 23-26.

CURTIS TP & SLOAN WT (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Current Opinion in Microbiology* **7**: 221-226.

CURTIS TP, SLOAN WT & SCANNELL JW (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 10494-10499.

DAHLLÖF I (2002) Molecular community analysis of microbial diversity. *Current Opinion in Biotechnology* **13**: 213-217.

DAIMS H, LUCKER S & WAGNER M (2006) daime, a novel image analysis program for microbial ecology and biofilm research. *Environmental Microbiology* **8**: 200-213.

DAVIS M, CHEW MK, HOBBS RJ, LUGO AE, EWEL JJ, VERMEIJ GJ, BROWN JH, ROSENZWEIG ML, GARDENER MR, CARROLL SP, THOMPSON K, PICKETT STA, STROMBERG JC, DEL

TREDICI P, SUDING KN, EHRENFELD JG, GRIME JP, MASCARO J & BRIGGS JC (2011) Don't judge species on their origins. *Nature* **474**: 153-154.

DAVIS MA (2009) *Invasion Biology*. Oxford University Press, UK.

DE GUSSEME B & VAN DE VELDE F (2013) Rinsing procedure of the water distribution network (TMVW). (personal communication)

DE ROY K, CLEMENT L, THAS O, WANG Y & BOON N (2012) Flow cytometry for fast microbial community fingerprinting. *Water Research* **46**: 907-919.

DE ROY K, MARZORATI M, NEGRONI A, THAS O, BALLOI A, FAVA F, VERSTRAETE W, DAFFONCHIO D & BOON N (2013) Environmental conditions and community evenness determine the outcome of biological invasion. *Nature Communications* **4**: 1383.

DEJONGHE W, BERTELOOT E, GORIS J, BOON N, CRUL K, MAERTENS S, HOFTE M, DE VOS P, VERSTRAETE W & TOP EM (2003) Synergistic degradation of linuron by a bacterial consortium and isolation of a single linuron-degrading Variovorax strain. *Applied and Environmental Microbiology* **69**: 1532-1541.

DEWETTINCK T, HULSBOSCH W, VAN HEGE K, TOP EM & VERSTRAETE W (2001) Molecular fingerprinting of bacterial populations in groundwater and bottled mineral water. *Applied Microbiology and Biotechnology* **57**: 412-418.

DIAZ M, HERRERO M, GARCIA LA & QUIROS C (2010) Application of flow cytometry to industrial microbial bioprocesses. *Biochemical Engineering Journal* **48**: 385-407.

DING LL, ZHOU QX, WANG L & ZHANG QR (2011) Dynamics of bacterial community structure in a full-scale wastewater treatment plant with anoxic-oxic configuration using 16S rDNA PCR-DGGE fingerprints. *African Journal of Biotechnology* **10**: 589-600.

DOROSHENKO V, AIRICH L, VITUSHKINA M, KOLOKOLOVA A, LIVSHITS V & MASHKO S (2007) YddG from Escherichia coli promotes export of aromatic amino acids. *Fems Microbiology Letters* **275**: 312-318.

DU J, SHAO ZY & ZHAO HM (2011) Engineering microbial factories for synthesis of value-added products. *Journal of Industrial Microbiology & Biotechnology* **38**: 873-890.

DUFOUR A (2003) *Assessing microbial safety of drinking water: improving approaches and methods*. IWA Publishing on behalf of the World Health Organization and the Organisation for Economic Co-operation and Development, London.

DUGATKIN LA, PERLIN M, LUCAS JS & ATLAS R (2005) Group-beneficial traits, frequency-dependent selection and genotypic diversity: an antibiotic resistance paradigm. *Proceedings of the Royal Society B-Biological Sciences* **272**: 79-83.

DUNHAM MJ (2007) Synthetic ecology: A model system for cooperation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1741-1742.

DUNSTAN PK & JOHNSON CR (2006) Linking richness, community variability, and invasion resistance with patch size. *Ecology* **87**: 2842-2850.

ECKBURG PB, BIK EM, BERNSTEIN CN, PURDOM E, DETHLEFSEN L, SARGENT M, GILL SR, NELSON KE & RELMAN DA (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635-1638.

ELSNER JB, KOSSIN JP & JAGGER TH (2008) The increasing intensity of the strongest tropical cyclones. *Nature* **455**: 92-95.

EMERY SM & GROSS KL (2007) Dominant species identity, not community evenness, regulates invasion in experimental grassland plant communities. *Ecology* **88**: 954-964.

ENDY D (2005) Foundations for engineering biology. *Nature* **438**: 449-453.

ESTRELA S, TRISOS CH & BROWN SP (2012) From Metabolism to Ecology: Cross-Feeding Interactions Shape the Balance between Polymicrobial Conflict and Mutualism. *American Naturalist* **180**: 566-576.

FAUST K & RAES J (2012) Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**: 538-550.

FINK KA & WILSON SD (2011) Bromus inermis invasion of a native grassland: diversity and resource reduction. *Botany-Botanique* **89**: 157-164.

FORSTER S, LAPPIN-SCOTT HM, SNAPE JR & PORTER J (2003) Rains, drains and active strains: towards online assessment of wastewater bacterial communities. *Journal of Microbiological Methods* **55**: 859-864.

FRANK SA (1994) Genetics of mutualism - the evolution of altruism between species. *Journal of Theoretical Biology* **170**: 393-400.

FULGET N, POUGHON L, RICHALET J & LASSEUR C (1999) MELISSA: Global control strategy of the artificial ecosystem by using first principles models of the compartments. *Life Sciences: Artificial Ecosystems*, Vol. 24. Brechignac F & MacElroy R (eds), pp. 397-405.

GOOD P (2005) *Permutation, parametric and bootstrap tests of hypotheses*. Springer, New York.

GRAVEL D, BELL T, BARBERA C, BOUVIER T, POMMIER T, VENAIL P & MOUQUET N (2011) Experimental niche evolution alters the strength of the diversity-productivity relationship. *Nature* **469**: 89-94.

GREGORI G, CITTERIO S, GHIANI A, LABRA M, SGORBATI S, BROWN S & DENIS M (2001) Resolution of viable and membrane-compromised bacteria in freshwater and marine waters based on analytical flow cytometry and nucleic acid double staining. *Applied and Environmental Microbiology* **67**: 4662-4670.

GRIFFIN AS, WEST SA & BUCKLING A (2004) Cooperation and competition in pathogenic bacteria. *Nature* **430**: 1024-1027.

GUENTHER S, KOCH C, HUEBSCHMANN T, ROESKE I, MUELLER RA, BLEY T, HARMS H & MUELLER S (2012) Correlation of Community Dynamics and Process Parameters As a Tool for the Prediction of the Stability of Wastewater Treatment. *Environmental Science & Technology* **46**: 84-92.

GUO B, HARSTALL C, LOUIE T, VAN ZANTEN SV & DIELEMAN LA (2012) Systematic review: faecal transplantation for the treatment of Clostridium difficile-associated disease. *Alimentary Pharmacology & Therapeutics* **35**: 865-875.

GUO F & ZHANG T (2013) Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Applied Microbiology and Biotechnology* **97**: 4607-4616.

HAMILTON WD (1963) Evolution of altruistic behavior *American Naturalist* **97**: 354-356.

HAMILTON WD (1964) Genetical evolution of social behaviour. *Journal of Theoretical Biology* **7**: 1-16.

HAMMES F & EGLI T (2010) Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. *Analytical and Bioanalytical Chemistry* **397**: 1083-1095.

HAMMES F, BERGER C, KOSTER O & EGLI T (2010) Assessing biological stability of drinking water without disinfectant residuals in a full-scale water supply system. *Journal of Water Supply Research and Technology-Aqua* **59**: 31-40.

HAMMES F, BERNEY M, WANG YY, VITAL M, KOSTER O & EGLI T (2008) Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Research* **42**: 269-277.

HAMMES F, BROGER T, WEILENMANN H-U, VITAL M, HELBING J, BOSSHART U, HUBER P, ODERMATT RP & SONNLEITNER B (2012) Development and laboratory-scale testing of a fully automated online flow cytometer for drinking water analysis. *Cytometry Part A* **81A**: 508-516.

HENDRICKX B, JUNCA H, VOSAHLOVA J, LINDNER A, RUEGG I, BUCHELI-WITSCHEL M, FABER F, EGLI T, MAU M, SCHLOMANN M, BRENNEROVA M, BRENNER V, PIEPER DH, TOP EM, DEJONGHE W, BASTIAENS L & SPRINGAEL D (2006a) Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: Distribution of the

genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *Journal of Microbiological Methods* **64**: 250-265.

HENDRICKX L, DE WEVER H, HERMANS V, MASTROLEO F, MORIN N, WILMOTTE A, JANSSEN P & MERGEAY M (2006b) Microbial ecology of the closed artificial ecosystem MELiSSA (Micro-Ecological Life Support System Alternative): Reinventing and compartmentalizing the Earth's food and oxygen regeneration system for long-haul space exploration missions. *Research in Microbiology* **157**: 77-86.

HEYLEN K, GEVERS D, VANPARYS B, WITTEBOLLE L, GEETS J, BOON N & DE VOS P (2006) The incidence of nirS and nirK and their genetic heterogeneity in cultivated denitrifiers. *Environmental Microbiology* **8**: 2012-2021.

HIBBING ME, FUQUA C, PARSEK MR & PETERSON SB (2010) Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology* **8**: 15-25.

HOEFEL D, GROOBY WL, MONIS PT, ANDREWS S & SAINT CP (2003) Enumeration of water-borne bacteria using viability assays and flow cytometry: a comparison to culture-based techniques. *Journal of Microbiological Methods* **55**: 585-597.

HOEFEL D, MONIS PT, GROOBY WL, ANDREWS S & SAINT CP (2005) Profiling bacterial survival through a water treatment process and subsequent distribution system. *Journal of Applied Microbiology* **99**: 175-186.

HORNER-DEVINE MC, CARNEY KM & BOHANNAN BJM (2004) An ecological perspective on bacterial biodiversity. *Proceedings of the Royal Society B-Biological Sciences* **271**: 113-122.

HOSODA K, SUZUKI S, YAMAUCHI Y, SHIROGUCHI Y, KASHIWAGI A, ONO N, MORI K & YOMO T (2011) Cooperative Adaptation to Establishment of a Synthetic Bacterial Mutualism. *Plos One* **6**: e17105.

HUIBREGTSE IL, ZAAT SA, KAPSENBERG ML, DA SILVA MAS, PEPPELENBOSCH MP, VAN DEVENTER SJH & BRAAT H (2012) Genetically Modified Lactococcus lactis for Delivery of Human Interleukin-10 to Dendritic Cells. *Gastroenterology Research and Practice* **2012**: 639291.

HUNTER PR (1993) The Microbiology Of Bottled Natural Mineral Waters - A Review. *Journal of Applied Bacteriology* **74**: 345-352.

HUTTER KJ (2002) Flow cytometry - A new tool for direct control of fermentation processes. *Journal of the Institute of Brewing* **108**: 48-51.

HWANG C, LING F, ANDERSEN GL, LECHEVALLIER MW & LIU W-T (2012) Microbial Community Dynamics of an Urban Drinking Water Distribution System Subjected to Phases of Chloramination and Chlorination Treatments. *Applied and Environmental Microbiology* **78**: 7856-7865.

JESSUP CM, FORDE SE & BOHANNAN BJM (2005) Microbial experimental systems in ecology. *Advances in Ecological Research, Population Dynamics and Laboratory Ecology*, Vol. 37. Desharnais RA (ed). London: Elsevier Academic Press Ltd, pp. 273-307.

JIANG L & MORIN PJ (2004) Productivity gradients cause positive diversity-invasibility relationships in microbial communities. *Ecology Letters* **7**: 1047-1057.

KENNEDY TA, NAEEM S, HOWE KM, KNOPS JMH, TILMAN D & REICH P (2002) Biodiversity as a barrier to ecological invasion. *Nature* **417**: 636-638.

KERNER A, PARK J, WILLIAMS A & LIN XN (2012) A Programmable Escherichia coli Consortium via Tunable Symbiosis. *Plos One* **7**: e34032.

KHORUTS A, DICKSVED J, JANSSON JK & SADOWSKY MJ (2010) Changes in the Composition of the Human Fecal Microbiome After Bacteriotherapy for Recurrent Clostridium difficile-associated Diarrhea. *Journal of Clinical Gastroenterology* **44**: 354-360.

KIM HJ, BOEDICKER JQ, CHOI JW & ISMAGILOV RF (2008) Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 18188-18193.

KLITGORD N & SEGRE D (2010) Environments that Induce Synthetic Microbial Ecosystems. *Plos Computational Biology* **6**: e1001002.

KNEITEL JM & PERRAULT D (2006) Disturbance-induced changes in community composition increase species invasion success. *Community Ecology* **7**: 245-252.

KOCH C (2013) Monitoring microbial community dynamics with flow cytometry. Thesis, University of Leipzig, Germany.

KOCH C, FETZER I, HARMS H & MUELLER S (2013a) CHIC - An automated approach for the detection of dynamic variations in complex microbial communities. *Cytometry Part A* **83A**: 561-567.

KOCH C, FETZER I, SCHMIDT T, HARMS H & MUELLER S (2013b) Monitoring Functions in Managed Microbial Systems by Cytometric Bar Coding. *Environmental Science & Technology* **47**: 1753-1760.

KOCH C, GUENTHER S, DESTA AF, HUEBSCHMANN T & MUELLER S (2013c) Cytometric fingerprinting for analyzing microbial intracommunity structure variation and identifying subcommunity function. *Nature Protocols* **8**: 190-202.

KOENKER R (2005) *Quantile regression*. Cambridge University Press, New York.

KUCNEROWICZ F, VANDEWERF R & VERSTRAETE W (1982) Estimation of microbial ATP in waters and soils. *Antonie Van Leeuwenhoek Journal of Microbiology* **48**: 522-523.

KÜMMERLI R, JIRICNY N, CLARKE LS, WEST SA & GRIFFIN AS (2009) Phenotypic plasticity of a cooperative behaviour in bacteria. *Journal of Evolutionary Biology* **22**: 589-598.

LAMBERTINI M, LEAPE J, MARTON-LEFEVRE J, MITTER-MEIER RA, ROSE M, ROBINSON JG, STUART SN, WALDMAN B & GENOVESI P (2011) Invasives: A Major Conservation Threat. *Science* **333**: 404-405.

LAUTENSCHLAGER K, BOON N, WANG Y, EGLI T & HAMMES F (2010a) Overnight stagnation of drinking water in household taps induces microbial growth and changes in community composition. *Water Research* **44**: 4868-4877.

LAUTENSCHLAGER K, BOON N, WANG YY, EGLI T & HAMMES F (2010b) Overnight stagnation of drinking water in household taps induces microbial growth and changes in community composition. *Water Research* **44**: 4868-4877.

LAUTENSCHLAGER K, HWANG C, LIU W-T, BOON N, KOESTER O, VROUWENVELDER H, EGLI T & HAMMES F (2013) A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. *Water Research* **47**: 3015-3025.

LEE SK & KEASLING JD (2005) A propionate-inducible expression system for enteric bacteria. *Applied and Environmental Microbiology* **71**: 6856-6862.

LEONARD E, NIELSEN D, SOLOMON K & PRATHER KJ (2008) Engineering microbes with synthetic biology frameworks. *Trends in Biotechnology* **26**: 674-681.

LI CS, CHIA WC & CHEN PS (2007) Fluorochrome and flow cytometry to monitor microorganisms in treated hospital wastewater. *Journal of Environmental Science and Health Part a-Toxic/Hazardous Substances & Environmental Engineering* **42**: 195-203.

LI JT, DUAN HN, LI SP, KUANG JL, ZENG Y & SHU WS (2010) Cadmium pollution triggers a positive biodiversity-productivity relationship: evidence from a laboratory microcosm experiment. *Journal of Applied Ecology* **47**: 890-898.

LITTLE AEF, ROBINSON CJ, PETERSON SB, RAFFA KE & HANDELSMAN J (2008) Rules of Engagement: Interspecies Interactions that Regulate Microbial Communities. *Annual Review of Microbiology* **62**: 375-401.

LO K, HAHNE F, BRINKMAN RR & GOTTARDO R (2009) flowClust: a Bioconductor package for automated gating of flow cytometry data. *Bmc Bioinformatics* **10**: 145.

LOPEZ-GARCIA P & MOREIRA D (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Research in Microbiology* **159**: 67-73.

LOREAU M, NAEEM S, INCHAUSTI P, BENGTSSON J, GRIME JP, HECTOR A, HOOPER DU, HUSTON MA, RAFFAELLI D, SCHMID B, TILMAN D & WARDLE DA (2001) Ecology -

Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294**: 804-808.

LORENZ MO (1905) Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association* **9**: 209-219.

LUTA G (2011) On Extensions of k-Means Clustering for Automated Gating of Flow Cytometry Data. *Cytometry Part A* **79A**: 3-5.

LUTKE-EVERSLOH T & STEPHANOPOULOS G (2007) L-tyrosine production by deregulated strains of Escherichia coli. *Applied Microbiology and Biotechnology* **75**: 103-110.

MA Q, ZHOU J, ZHANG WW, MENG XX, SUN JW & YUAN YJ (2011) Integrated Proteomic and Metabolomic Analysis of an Artificial Microbial Community for Two-Step Production of Vitamin C. *Plos One* **6**: e26108.

MAECKER HT, RINFRET A, D'SOUZA P, DARDEN J, ROIG E, LANDRY C, HAYES P, BIRUNGI J, ANZALA O, GARCIA M, HARARI A, FRANK I, BAYDO R, BAKER M, HOLBROOK J, OTTINGER J, LAMOREAUX L, EPLING CL, SINCLAIR E, SUNI MA, PUNT K, CALAROTA S, EL-BAHI S, ALTER G, MAILA H, KUTA E, COX J, GRAY C, ALTFELD M, NOUGAREDE N, BOYER J, TUSSEY L, TOBERY T, BREDT B, ROEDERER M, KOUP R, MAINO VC, WEINHOLD K, PANTALEO G, GILMOUR J, HORTON H & SEKALY RP (2005) Standardization of cytokine flow cytometry assays. *Bmc Immunology* **6**: 13.

MANEFIELD M & TURNER SL (2002) Quorum sensing in context: out of molecular biology and into microbial ecology. *Microbiology-Sgm* **148**: 3762-3764.

MANUEL J, SELIN C, FERNANDO WGD & DE KIEVIT T (2012) Stringent response mutants of Pseudomonas chlororaphis PA23 exhibit enhanced antifungal activity against Sclerotinia sclerotiorum in vitro. *Microbiology-Sgm* **158**: 207-216.

MARZORATI M, WITTEBOLLE L, BOON N, DAFFONCHIO D & VERSTRAETE W (2008) How to get more out of molecular fingerprints: practical tools for microbial ecology. *Environmental Microbiology* **10**: 1571-1581.

MARZORATI M, MAIGNIEN L, VERHELST A, LUTA G, SINNOTT R, KERCKHOF FM, BOON N, VAN DE WIELE T & POSSEMIERS S (2013) Barcoded pyrosequencing analysis of the microbial community in a simulator of the human gastrointestinal tract showed a colon region-specific microbiota modulation for two plant-derived polysaccharide blends. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **103**: 409-420.

MCARTHUR JV (2006) *Microbial ecology: an evolutionary approach*. Elsevier Academic Press, USA.

MERTENS B, BOON N & VERSTRAETE W (2005) Stereospecific effect of hexachlorocyclohexane on activity and structure of soil methanotrophic communities. *Environmental Microbiology* **7**: 660-669.

MIETHKE M, WESTERS H, BLOM E-J, KUIPERS OP & MARAHIEL MA (2006) Iron starvation triggers the stringent response and induces amino acid biosynthesis for bacillibactin production in Bacillus subtilis. *Journal of Bacteriology* **188**: 8655-8657.

MONTGOMERY H & DYMOCK JF (1961) Determination of nitrite in water. *Analyst* **86**: 414-416.

MORTON SC, ZHANG Y & EDWARDS MA (2005) Implications of nutrient release from iron metal for microbial regrowth in water distribution systems. *Water Research* **39**: 2883-2892.

MOUGI A & KONDOH M (2012) Diversity of Interaction Types and Ecological Community Stability. *Science* **337**: 349-351.

MUYZER G, DEWAAL EC & UITTERLINDEN AG (1993) Profiling of complex microbial-populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16S ribosomal-RNA. *Applied and Environmental Microbiology* **59**: 695-700.

NAEEM S (2009) Gini in the bottle. *Nature* **458**: 579-580.

NEMERGUT DR, SCHMIDT SK, FUKAMI T, O'NEILL SP, BILINSKI TM, STANISH LF, KNELMAN JE, DARCY JL, LYNCH RC, WICKEY P & FERRENBERG S (2013) Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* **77**: 342-356.

PARKHILL J (2013) What has high-throughput sequencing ever done for us? *Nature Reviews Microbiology* **11**: 664-665.

PATLE S & LAL B (2007) Ethanol production from hydrolysed agricultural wastes using mixed culture of Zymomonas mobilis and Candida tropicalis. *Biotechnology Letters* **29**: 1839-1843.

PAWELCZYK S, ABRAHAM WR, HARMS H & MULLER S (2008) Community-based degradation of 4-chorosalicylate tracked on the single cell level. *Journal of Microbiological Methods* **75**: 117-126.

PERSKY NS, FERULLO DJ, COOPER DL, MOORE HR & LOVETT ST (2009) The ObgE/CgtA GTPase influences the stringent response to amino acid starvation in Escherichia coli. *Molecular Microbiology* **73**: 253-266.

PETROF E, GLOOR G, VANNER S, WEESE S, CARTER D, DAIGNEAULT M, BROWN E, SCHROETER K & ALLEN-VERCOE E (2013) Stool substitute transplant therapy for the eradication of Clostridium difficile infection: 'RePOOPulating' the gut. *Microbiome* **1**: 3.

PETROSINO J, CANTU C & PALZKILL T (1998) beta-Lactamases: Protein evolution in real time. *Trends in Microbiology* **6**: 323-327.

PIMM SL (1984) The complexity and stability of ecosystems *Nature* **307**: 321-326.

PINDER AC, PURDY PW, POULTER SAG & CLARK DC (1990) Validation of flow cytometry for rapid enumeration of bacterial concentrations in pure cultures. *Journal of Applied Bacteriology* **69**: 92-100.

PORTER J & PICKUP RW (2000) Nucleic acid-based fluorescent probes in microbial ecology: application of flow cytometry. *Journal of Microbiological Methods* **42**: 75-79.

POSSEMIERS S, MARZORATI M, VERSTRAETE W & VAN DE WIELE T (2010) Bacteria and chocolate: A successful combination for probiotic delivery. *International Journal of Food Microbiology* **141**: 97-103.

PRAKASH O, SHOUCHE Y, JANGID K & KOSTKA JE (2013) Microbial cultivation and the role of microbial resource centers in the omics era. *Applied Microbiology and Biotechnology* **97**: 51-62.

PROSSER JI, BOHANNAN BJM, CURTIS TP, ELLIS RJ, FIRESTONE MK, FRECKLETON RP, GREEN JL, GREEN LE, KILLHAM K, LENNON JJ, OSBORN AM, SOLAN M, VAN DER GAST CJ & YOUNG JPW (2007) Essay - The role of ecological theory in microbial ecology. *Nature Reviews Microbiology* **5**: 384-392.

RAES J & BORK P (2008) Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology* **6**: 693-699.

READ S, MARZORATI M, GUIMARAES BCM & BOON N (2011) Microbial Resource Management revisited: successful parameters and new concepts. *Applied Microbiology and Biotechnology* **90**: 861-871.

REZNICK D, BRYANT MJ & BASHEY F (2002) r- and K-selection revisited: The role of population regulation in life-history evolution. *Ecology* **83**: 1509-1520.

RIIS SB, PEDERSEN HM, SORENSEN NK & JAKOBSEN M (1995) Flow cytometry and acidification power test as rapid techniques for determination of the activity of starter cultures of *Lactobacillus delbrueckii* ssp. *bulgaricus*. *Food Microbiology* **12**: 245-250.

ROGERS WT & HOLYST HA (2009) FlowFP: A Bioconductor Package for Fingerprinting Flow Cytometric Data. *Advances in Bioinformatics* **2009**: 193947.

ROGERS WT, MOSER AR, HOLYST HA, BANTLY A, MOHLER ER, SCANGAS G & MOORE JS (2008) Cytometric Fingerprinting: Quantitative characterization of multivariate distributions. *Cytometry Part A* **73A**: 430-441.

ROUT ME & CALLAWAY RM (2009) An Invasive Plant Paradox. *Science* **324**: 734-735.

ROWBURY RJ (2001) Extracellular sensing components and extracellular induction component alarmones give early warning against stress in Escherichia coli. *Advances in Microbial Physiology, Vol 44*, Vol. 44. Poole RK (ed), pp. 215-257.

RUSZCZYNSKA A, SZTEYN J & WISZNIEWSKA-LASZCZYCH A (2007) Flow cytometry as a method for the evaluation of raw material, product and process in the dairy industry. *Polish Journal of Veterinary Sciences* **10**: 119-121.

SALEEM M, FETZER I, HARMS H & CHATZINOTAS A (2013) Diversity of protists and bacteria determines predation performance and stability. *ISME J* **7**: 1912-1921

SALEEM M, FETZER I, DORMANN CF, HARMS H & CHATZINOTAS A (2012) Predator richness increases the effect of prey diversity on prey yield. *Nature Communications* **3**: 1305.

SANCHEZ A & GORE J (2013) Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations. *Plos Biology* **11**: e1001547.

SANTOS CNS, XIAO W & STEPHANOPOULOS G (2012) Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 13538-13543.

SCHINK B (2002) Synergistic interactions in the microbial world. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **81**: 257-261.

SCHMIDT TM (2006) The maturing of microbial ecology. *International Microbiology* **9**: 217-223.

SEURINCK S, DEFOIRDT T, VERSTRAETE W & SICILIANO SD (2005) Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environmental Microbiology* **7**: 249-259.

SHAHINAS D, SILVERMAN M, SITTLER T, CHIU C, KIM P, ALLEN-VERCOE E, WEESE S, WONG A, LOW DE & PILLAI DR (2012) Toward an Understanding of Changes in Diversity Associated with Fecal Microbiome Transplantation Based on 16S rRNA Gene Deep Sequencing. *Mbio* **3**: e00338-00312.

SHAPIRO HM (1994) *Practical Flow Cytometry*. John Wiley and Sons, New Jersey.

SHARMA S, SZELE Z, SCHILLING R, MUNCH JC & SCHLOTER M (2006) Influence of freeze-thaw stress on the structure and function of microbial communities and denitrifying populations in soil. *Applied and Environmental Microbiology* **72**: 2148-2154.

SHOU WY, RAM S & VILAR JMG (2007) Synthetic cooperation in engineered yeast populations. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1877-1882.

SIMBERLOFF D, SOUZA L, NUNEZ MA, BARRIOS-GARCIA MN & BUNN W (2012) The natives are restless, but not often and mostly when disturbed. *Ecology* **93**: 598-607.

SIMBERLOFF D, MARTIN J-L, GENOVESI P, MARIS V, WARDLE DA, ARONSON J, COURCHAMP F, GALIL B, GARCÌA-BERTHOU E, PASCAL M, PYSEK P, SOUSA R, TABACCHI E & VILÀ M (2013) Impacts of biological invasions: what's what and the way forward. *Trends in ecology & evolution* **28**: 58-66.

SONG C, LUONG T-D, KONG TF, NGUYEN N-T & ASUNDI AK (2011) Disposable flow cytometer with high efficiency in particle counting and sizing using an optofluidic lens. *Optics Letters* **36**: 657-659.

SOUSA R, MORAIS P, DIAS E & ANTUNES C (2011) Biological invasions and ecosystem functioning: time to merge. *Biological Invasions* **13**: 1055-1058.

STACHOWICZ JJ, WHITLATCH RB & OSMAN RW (1999) Species diversity and invasion resistance in a marine ecosystem. *Science* **286**: 1577-1579.

STEIDLER L, HANS W, SCHOTTE L, NEIRYNCK S, OBERMEIER F, FALK W, FIERS W & REMAUT E (2000) Treatment of murine colitis by Lactococcus lactis secreting interleukin-10. *Science* **289**: 1352-1355.

STERNBERG C, CHRISTENSEN BB, JOHANSEN T, NIELSEN AT, ANDERSEN JB, GIVSKOV M & MOLIN S (1999) Distribution of bacterial growth activity in flow-chamber biofilms. *Applied and Environmental Microbiology* **65**: 4108-4117.

STOLYAR S, VAN DIEN S, HILLESLAND KL, PINEL N, LIE TJ, LEIGH JA & STAHL DA (2007) Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology* **3**: 92.

TEMPERTON B & GIOVANNONI SJ (2012) Metagenomics: microbial diversity through a scratched lens. *Current Opinion in Microbiology* **15**: 605-612.

TILMAN D, KNOPS J, WEDIN D, REICH P, RITCHIE M & SIEMANN E (1997) The influence of functional diversity and composition on ecosystem processes. *Science* **277**: 1300-1302.

TOLKER-NIELSEN T & MOLIN S (2000) Spatial organization of microbial biofilm communities. *Microbial Ecology* **40**: 75-84.

TRACY BF & SANDERSON MA (2004) Forage productivity, species evenness and weed invasion in pasture communities. *Agriculture Ecosystems & Environment* **102**: 175-183.

VAN DEN ABBEELE P, GROOTAERT C, MARZORATI M, POSSEMIERS S, VERSTRAETE W, GERARD P, RABOT S, BRUNEAU A, EL AIDY S, DERRIEN M, ZOETENDAL E, KLEEREBEZEM M, SMIDT H & VAN DE WIELE T (2010) Microbial Community Development in a Dynamic Gut Model Is Reproducible, Colon Region Specific, and Selective for Bacteroidetes and Clostridium Cluster IX. *Applied and Environmental Microbiology* **76**: 5237-5246.

VAN NEVEL S, DE ROY K & BOON N (2013a) Bacterial invasion potential in water is determined by nutrient availability and the indigenous community. *Fems Microbiology Ecology* **85**: 593-603.

VAN NEVEL S, KOETZSCH S, WEILENMANN H-U, BOON N & HAMMES F (2013b) Routine bacterial analysis with automated flow cytometry. *Journal of Microbiological Methods* **94**: 73-76.

VERBERKMOES NC, DENEF VJ, HETTICH RL & BANFIELD JF (2009) SYSTEMS BIOLOGY Functional analysis of natural microbial consortia using community proteomics. *Nature Reviews Microbiology* **7**: 196-205.

VERSTRAETE W, BOON N, VAN DE WIELE T & VLAEMINCK SE (2012) Special Issue: Microbial Resource Management preface. *Microbial Biotechnology* **5**: 305-306.

VERSTRAETE W, WITTELBOLLE L, HEYLEN K, VANPARYS B, DE VOS P, VAN DE WIELE T & BOON N (2007) Microbial resource management: The road to go for environmental biotechnology. *Engineering in Life Sciences* **7**: 117-126.

VIVES-REGO J, LEBARON P & NEBE-VON CARON G (2000) Current and future applications of flow cytometry in aquatic microbiology. *Fems Microbiology Reviews* **24**: 429-448.

VLAEMINCK SE, TERADA A, SMETS BF, VAN DER LINDEN D, BOON N, VERSTRAETE W & CARBALLA M (2009) Nitrogen Removal from Digested Black Water by One-Stage Partial Nitritation and Anammox. *Environmental Science & Technology* **43**: 5035-5041.

WAEGEMAN H & SOETAERT W (2011) Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering. *Journal of Industrial Microbiology & Biotechnology* **38**: 1891-1910.

WANG Y, HAMMES F, BOON N, CHAMI M & EGLI T (2009) Isolation and characterization of low nucleic acid (LNA)-content bacteria. *Isme Journal* **3**: 889-902.

WANG Y, HAMMES F, DE ROY K, VERSTRAETE W & BOON N (2010) Past, present and future applications of flow cytometry in aquatic microbiology. *Trends in Biotechnology* **28**: 416-424.

WEST SA & BUCKLING A (2003) Cooperation, virulence and siderophore production in bacterial parasites. *Proceedings of the Royal Society B-Biological Sciences* **270**: 37-44.

WEST SA, GRIFFIN AS & GARDNER A (2007a) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology* **20**: 415-432.

WEST SA, GRIFFIN AS, GARDNER A & DIGGLE SP (2006) Social evolution theory for microorganisms. *Nature Reviews Microbiology* **4**: 597-607.

WEST SA, DIGGLE SP, BUCKLING A, GARDNER A & GRIFFINS AS (2007b) The social lives of microbes. *Annual Review of Ecology Evolution and Systematics* **38**: 53-77.

WHO (2011) *Guidelines for Drinking-water Quality*. World Health Organization, Geneva.

WILSEY BJ & POLLEY HW (2002) Reductions in grassland species evenness increase dicot seedling invasion and spittle bug infestation. *Ecology Letters* **5**: 676-684.

WINTERMUTE EH & SILVER PA (2010) Emergent cooperation in microbial metabolism. *Molecular Systems Biology* **6**: 407.

WINTERMUTE EH & SILVER PA (2011) Dynamics in the mixed microbial concourse. *Genes & Development* **24**: 2603-2614.

WITTEBOLLE L, BOON N, VANPARYS B, HEYLEN K, DE VOS P & VERSTRAETE W (2005) Failure of the ammonia oxidation process in two pharmaceutical wastewater treatment plants is linked to shifts in the bacterial communities. *Journal of Applied Microbiology* **99**: 997-1006.

WITTEBOLLE L, MARZORATI M, CLEMENT L, BALLOI A, DAFFONCHIO D, HEYLEN K, DE VOS P, VERSTRAETE W & BOON N (2009) Initial community evenness favours functionality under selective stress. *Nature* **458**: 623-626.

WU XL & CONRAD R (2001) Functional and structural response of a cellulose-degrading methanogenic microbial community to multiple aeration stress at two different temperatures. *Environmental Microbiology* **3**: 355-362.

YEDID G, OFRIA CA & LENSKI RE (2009) Selective Press Extinctions, but Not Random Pulse Extinctions, Cause Delayed Ecological Recovery in Communities of Digital Organisms. *American Naturalist* **173**: E139-E154.

YU TC, SMITH RP & YOU LC (2012) Engineering microbial systems to explore ecological and evolutionary dynamics. *Current Opinion in Biotechnology* **23**: 791-797.

ZENGLER K & PALSSON BO (2012) A road map for the development of community systems (CoSy) biology. *Nature Reviews Microbiology* **10**: 366-372.

ZHANG T & FANG HHP (2000) Digitization of DGGE (denaturing gradient gel electrophoresis) profile and cluster analysis of microbial communities. *Biotechnology Letters* **22**: 399-405.

ZHANG Y & EDWARDS M (2010) Nutrients and metals effects on nitrification in drinking water systems. *Journal American Water Works Association* **102**: 56-66.

ZHANG Y, ARENDS JBA, VAN DE WIELE T & BOON N (2011) Bioreactor technology in marine microbiology: From design to future application. *Biotechnology Advances* **29**: 312-321.

ZHAO R, NATARAJAN A & SRIENC F (1999) A flow injection flow cytometry system for on-line monitoring of bioreactors. *Biotechnology and Bioengineering* **62**: 609-617.

CURRICULUM VITAE

Personalia

Full name	Karen De Roy
Date of birth	December 4th, 1986
Place of birth	Rumst
Nationality	Belgian
Address	Komijnstraat 15/202
	9000 Gent
Phone	+32 477 55 82 24
E-mail	Karen.DeRoy@ugent.be
	Karenderoy@hotmail.com

Education

1998 – 2004	Virgo Sapiensinstituut, Londerzeel, Belgium
	Secondary education Science-Mathematics (8 hours)
2004 – 2009	Ghent University, Faculty of Bioscience Engineering, Gent, Belgium
	Bioscience engineer - Cell & Gene Biotechnology
2008 – 2009	Master thesis: Microbially produced nitric oxide in aquatic ecosystems
	Promotor: Prof. dr. ir. Nico Boon
	Co-promotor: Prof. dr. ir. Peter Bossier
	Performed at: Laboratory of Microbial Ecology and Technology (LabMET)

Professional activities during PhD

2009 – 2013	PhD research on the FWO-project "Managing ecosystem functioning by steering microbial community structure and dynamics" grant no. G.0700.10N
2009 - 2013	Scientific collaborator at Laboratory of Microbial Ecology and Technology (LabMET).

- Tutor of 2 master thesis students
- Teaching activities: Supervisor and responsible of practical exercises of the course "Environmental microbiology"
- Reviewer for
  - o Water science and Technology
  - o Cytometry

o Journal of Applied Microbiology
o Water Research
o Environmental Science and Technology

Publications in international peer-reviewed journals

Wang Y, Hammes F, **De Roy K,** Verstraete W & Boon N (2010) Past, present and future applications of flow cytometry in aquatic microbiology. *Trends in Biotechnology* **28**: 416-424. (IF=9,644)

Evens R, De Schamphelaere KaC, Balcaen L, Wang Y, **De Roy K**, Resano M, Del Rosario Florez M, Van Der Meeren P, Boon N, Vanhaecke F & Janssen CR (2011) Liposomes as an alternative delivery system for investigating dietary metal toxicity to Daphnia magna. *Aquatic Toxicology* **105**: 661-668. (IF=3,761)

De Roy K, Clement L, Thas O, Wang Y & Boon N (2012) Flow cytometry for fast microbial community fingerprinting. *Water Research* **46**: 907-919. (IF=4,655)

Evens R, De Schamphelaere KaC, Balcaen L, Wang Y, **De Roy K**, Resano M, Florez M, Boon N, Vanhaecke F & Janssen CR (2012) The use of liposomes to differentiate between the effects of nickel accumulation and altered food quality in Daphnia magna exposed to dietary nickel. *Aquatic Toxicology* **109**: 80-89. (IF=3,730)

De Roy K, Marzorati M, Negroni A, Thas O, Balloi A, Fava F, Verstraete W, Daffonchio D & Boon N (2013) Environmental conditions and community evenness determine the outcome of biological invasion. *Nature Communications* **4**. (IF=10,015)

Gozdereliler E, Boon N, Aamand J, **De Roy K**, Granitsiotis MS, Albrechtsen H-J & Sorensen SR (2013) Comparing Metabolic Functionalities, Community Structures, and Dynamics of Herbicide-Degrading Communities Cultivated with Different Substrate Concentrations. *Applied and Environmental Microbiology* **79**: 367-375. (IF=3,678)

Van Acker H, Sass A, Bazzini S, **De Roy K,** Udine C, Messiaen T, Riccardi G, Boon N, Nelis HJ, Mahenthiralingam E & Coenye T (2013) Biofilm-grown *Burkholderia cepacia* complex cells survive antibiotic treatment by avoiding production of reactive oxygen species. *Plos One* **8**: e58943. (IF=3,730)

Van Nevel S, **De Roy K** & Boon N (2013) Bacterial invasion potential in water is determined by nutrient availability and the indigenous community. *Fems Microbiology Ecology* **85**: 593-603. (IF=3,563)

De Roy K, Marzorati M, Van den Abbeele P, Van de Wiele T & Boon N. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environmental microbiology (In press)*. (IF=5,756)

Ho A, **De Roy K**, Thas O, De Neve J, Hoefman S, Vandamme P, Heylen K & Boon N. The more, the merrier: heterotroph richness stimulates methanotrophic activity. *(Submitted to the ISME journal)*

Clement L, **De Roy K,** Boon N & Thas O. flowFDA: a Bioconductor package for functional data analysis of flow cytometry data. *(Submitted to Bioinformatics)*

De Roy K, Thas O, Marzorati M, Vilchez R and Boon N. Interspecies relatedness determines the success of microbial mutualistic interactions. (*In preparation for Nature Communications)*

Active participations in (inter)national scientific conferences

De Roy K, Clement L, Thas O, Wang Y. & Boon N. Flow cytometry community fingerprinting to detect quickly stress in drinking water systems. Proceedings of the First International Conference on Microbial Diversity 2011 – Environmental Stress and Adaptation, Milan, Italy October 26-28 2011. ***Oral*** *presentation*

De Roy K, Clement L, Thas O, Wang Y. & Boon N. Flow cytometry for fast microbial community fingerprinting. 17th symposium on applied biological sciences, Leuven, Belgium, February 10, 2012. ***Poster*** *presentation.*

De Roy K, Marzorati M, Negroni A, Thas O, Balloi A, Fava F, Verstraete W, Daffonchio D & Boon N. Environmental conditions and community evenness determine the outcome of biological invasion. 5th Congress of European Microbiologists, Leipzig, Germany, July 21-25, 2013. ***Poster*** *presentation.*

Contributions to participations in (inter)national scientific conferences

Clement L, **De Roy K**, Thas O, Wang Y. & Boon N. Flow cytometry for fast microbial community fingerprinting. 23rd Annual Conference of the International Environmetrics Society, Anchorag, Alaska, USA, June 10-14, 2013. ***Invited oral*** *presentation.*

Van Acker H, Sass A, **De Roy K**, Boon N, Mahenthiralingam E, Nelis HJ & Coenye T. Persistence in B*urkholderia cenocepacia* j2315 biofilms. BSM-meeting "Life, Death and Survival of Micro-organisms", Brussels, Belgium, November 16, 2011. ***Oral*** *presentation*

Van Acker H, Sass A, **De Roy K**, Boon N, Mahenthiralingam E, Nelis HJ & Coenye T. Transcriptome analysis of persister cells in *Burkholderia cepacia* complex biofilms. 15th International *Burkholderia cepacia* Working Group Meeting, Prague, Czech Republic, April 13-16, 2011. ***Oral*** *presentation*

Van Nevel S, **De Roy K** & Boon N. Bacterial invasion potential in water is determined by nutrient availability and the indigenous community. 3rd BeNeLux Young Water Professionals Conference, Belval, Luxembourg, October 2-4, 2013. ***Oral** presentation*

Boon N & **De Roy K**. Synthetic microbial communities: high throughput models to test new ecological hypothesis, at the Département de Protéomique et de Microbiologie (PROTMIC) de la Faculté des Sciences de l'Université de Mons, Mons, Belgium, August 31, 2012. ***Invited oral** presentation.*

Boon N, **De Roy K**, Clement L, Thas O, Wang Y. Flow cytometry community fingerprinting to detect quickly stress in drinking water systems" at the Annual Meeting of the German Society for Cytometry, Bonn, Germany, October 10-12, 2012. ***Keynote** lecture*

Boon N & **De Roy K.** Microbial Resource Management: from high throughput models to pilot reactors. Swiss Federal Institute for Environmental Science and Technology (EAWAG), Department of Microbiology and Molecular Ecotoxicology, Dübendorf, Switzerland, December 18, 2012. ***Invited oral** presentation*

Boon N, **De Roy K,** Marzorati M, Negroni A, Thas O, Balloi A, Fava F, Verstraete W & Daffonchio D. The role of evenness and invading populations to preserve microbial community functionality. Royal Netherlands Institute for Sea Research, Yerseke, The Netherlands, February 28, 2013. ***Invited oral** presentation*

Boon N & **De Roy K.** Microbial Resource Management: from high throughput models to pilot reactors. Taida College, Nankai University, Tianjin, China, April 18, 2013. ***Invited oral** presentation*

Boon N, **De Roy K**, Clement L, Thas O, Wang Y**.** μ-workshop on flow cytometry community fingerprinting. School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, April 22, 2013. ***Invited oral** presentation*

Boon N & **De Roy K.** Microbial Resource Management: from high throughput models to pilot reactors. School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, April 24, 2013. ***Invited oral** presentation*

Boon N & **De Roy K.** Microbial Resource Management – new tools and explorations. College of Environment and Resources, Jilin University, Changchun, China, April 25, 2013**.** ***Invited oral** presentation*

Evens R, De Schamphelaere KAC, Wang Y, **De Roy K,** Balcaen L, Vanhaecke F, Boon N & Janssen CR. The use of liposomes for dietary toxicity studies. SETAC Europe 20th Annual Meeting, Seville, Spain, 23-27 May, 2010. ***Poster** presentation*

Gözdereliler E, **De Roy K,** Aamand J, Boon N & Sørensen S. Molecular and physiological fingerprinting of herbicide-degrading microbial communities in response to different

herbicide concentrations. Nordic Environmental Nucleotide Network (NENUN) Workshop, Reykjavik, Iceland, September 8-10, 2011. ***Oral*** *presentation*

Gözdereliler E, **De Roy K**, Aamand J, Boon N & Sørensen S. Comparison of low nucleic acid (LNA) and high nucleic acid (HNA)-content bacterial communities enriched from a groundwater aquifer using herbicide 2-methyl-4-chlorophenoxyacetic acid (MCPA). 2011 Symposium of The Danish Microbiological Society, Denmark, November 7, 2011. ***Oral*** *presentation*

Gözdereliler E, Boon N, Aamand J, **De Roy K**, Granitsiotis M, Albrechtsen HJ & Sørensen SR. Enrichment and characterization of high- and low-affinity herbicide-degrading bacteria from a groundwater aquifer. 14th International Symposium on Microbial Ecology (ISME 14), Copenhagen, Denmark, August 19-24, 2012. ***Poster*** *presentation*

Participation to international workshops

3rd workshop on the biological stability of drinking water, May 25, 2012, Delft, the Netherlands.

CV

DANKWOORD

Eindelijk toegekomen aan de laatste en waarschijnlijk meest gelezen bladzijden van mijn boekske. Hoewel enkel mijn naam vooraan op de cover staat, is dit geen werk van mij alleen, in tegendeel! Daarom wil ik deze laatste bladzijden heel graag gebruiken om enkele mensen van harte te bedanken.

All members of the examination committee, thank you for your time and effort for critically evaluating my PhD manuscript: Prof. Monica Höfte, Prof. Hauke Harms, Prof. David Gillan, Dr. Tom Defoirdt, Prof. Daniele Daffonchio, Prof. Colin Janssen and Prof. Els Van Damme. I think all your comments greatly improved my final manuscript.

Dan wil ik de twee belangrijkste mensen in heel dit doctoraat bedanken, de twee meest complementaire promotoren die een doctoraatsstudent zich kan voorstellen bedanken, Prof. Boon en Prof. Thas. Nico, bij jou kon ik altijd terecht voor alles wat met microbiële ecologie te maken had en je hebt me meteen de liefde voor de flow cytometer doorgegeven. Olivier, jij stond dan weer paraat voor alles omtrent statistiek. Ik moet toegeven, mijn liefde voor de statistiek heeft misschien toch iets langer op zich laten wachten. Het was altijd heel aangenaam om met jullie samen te werken, de vele uurtjes die we samen in het vergaderlokaal doorbrachten om de data steeds opnieuw en opnieuw te bekijken, om elk woord in het 'nature paper' te wikken en te wegen, ... En als ik het over dat artikel heb, dan kom ik automatisch terecht bij Massimo en Andrea. Massimo, I had the honour to work further on your previous research and to use all your experience. The meetings we had were always relaxed and fun, but intensive and productive at the same time. Andrea, I want to thank you so much for your help during the three tough months of the invasion experiments. We made a good team: you reading the mixes composition, me pipetting like crazy. In line of the same experiments, I would also like to thank Prof. Daffonchio, Prof. Verstraete, Prof. Fava and Annalisa for their contributions.

Verder zou ik graag Lieven Clement en Yingying Wang enorm willen bedanken. Samen kwamen we op het idee van de flow cytometrie fingerprinting en hebben we dat verder uitgewerkt. Lieven, ik blijf nog altijd versteld staan van uw ongelofelijke kennis en briljante ideeën. Opvolging voor het fingerprinting verhaal is ondertussen verzekerd: Benjamin, heel

veel succes en plezier. Voor alles wat met flow cytometrie te maken heeft, wil ik ook Sam Van Nevel bedanken die me na enkele 'eenzame flow cytometrie jaren' gezelschap kwam houden aan de flow cytometer. Ons labo in de watertoren met panoramisch zicht over heel Gent zal me altijd bij blijven.
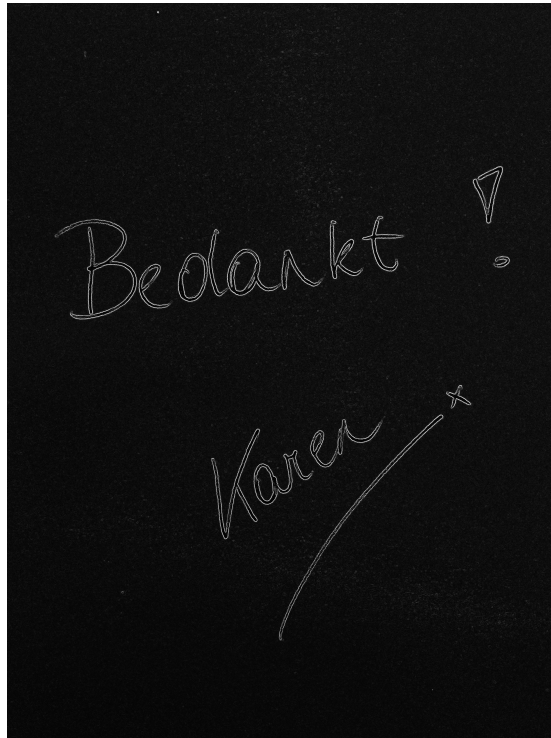
Naast de mensen waar ik heel intensief mee heb samen gewerkt, wil ik ook iedereen van het technische personeel (Siska, Tim, Greet, Mike, Jana, René en Robin) en secretariaat (Christine, Regine en Annelies) bedanken. Geen vraag was te kort of te lang voor jullie en jullie stonden altijd meteen paraat om te helpen. We zeggen het misschien niet voldoende, maar alles wat jullie doen en gedaan hebben, wordt enorm geapprecieerd. Tim, jou wil ik ook nog eens expliciet bedanken voor al de mooie figuurtjes in mijn boekje.

En dan kom ik toe aan al de mensen die ik wil bedanken om het hele doctoraat-gebeuren zo aangenaam te maken. Als kersvers doctoraatsstudentje kwam ik terecht in de bureau beneden. Ik heb heel veel leuke momenten beleefd met iedereen die over de jaren heen in onze bureau heeft gezeten: Pieter VdA, Pieter V, Beatriz, Jan, Carlos, Hannele, Varvara, Irene, Jianyun, Ramiro, Kim, Floor en Sofie. De wijnclusters, weddenschappen (die uiteraard allemaal door mij gewonnen werden), etentjes, de véééle koffiepauzes, ... zal ik niet snel vergeten.

Toen ik aan het bedenken was wie ik zeker niet mocht vergeten te bedanken en waarom, kwam ik er al snel achter dat ik de meeste moet bedanken om me te vergezellen tijdens mijn pauzekes door de dag heen: de koffietjes, de lunch, vieruurtje, een pintje na het werk, verschillende pintjes na de werkweek, ... Voor de koffietjes: Nikki, Jan A, Sam, Eva en uiteraard de bureaugenootjes. De lunch: Annelies, Jo, Jan, Eva, Jana, Marlies, Eline, Rosemarie, Pieter, David, Joan, Liesje, de rotonde, ... De vieruurtjes: Alexander. De pintjes en spelletjes darts na het werk: Sam. Bedankt iedereen om er te zijn voor zowel de leuke als de moeilijke momenten. Ik heb heel wat goede vrienden over gehouden aan mijn LabMET-avontuur!! En voor de vele gezellige drinks, spring walks, culturele en team building activiteiten, winterweekends met sauna, ...: bedankt LabMET!!!

Verder wil ik nog al mijn vrienden, zowel uit Gent als Londerzeel bedanken voor de vele aangename spelletjesavonden, etentjes, drinks, ... Samengevat, bedankt voor alles wat me mijn doctoraat even deed vergeten.

And last but no least: mijn familie. Mama en Papa, Sofie, Katrien en Sam, Pieter en Desi, en uiteraard, Axl en kleine ??? : bedankt voor alles wat jullie voor mij hebben gedaan (en vooral, wat jullie nog voor mij gaan doen ;)). Voor de zorgeloze weekends, de taxi-dienst, de auto-uitleen-dienst, de shop-uitstapjes, de goede zorgen, ... Kortom, bedankt voor alles! Zonder jullie was er zeker geen doctoraat geweest! Dit doctoraat is voor een groot deel ook jullie werk!

2 februari 2014