Promotors

Prof. dr. ir. Nico Boon

Center for microbial ecology and technology (CMET), Department of Biochemical and Microbial Technology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.

Prof. dr. ir. Bart De Gusseme

Center for microbial ecology and technology (CMET), Department of Biochemical and Microbial Technology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.

Members of the examination committee

Prof. dr. Colin Janssen (Chairman)

GhEnToxLab, Department of applied ecology and environmental biology, Faculty of bioscience engineering, Ghent University, Ghent, Belgium

Prof. dr. Andre Skirtach (Secretary)

Nano-biotechnology, Department of molecular biotechnology, Faculty of bioscience engineering, Ghent University, Ghent, Belgium

Prof. dr. David Berry

Division of microbial ecology (DOME), Department of microbiology and ecosystem science, University of Vienna, Vienna, Austria

Dr. Aleksandra Knezev

Senior consultant biology, Het Waterlaboratorium N.V., Arnhem, The Netherlands

Prof. dr. ir. Arne Verliefde

Particle and interfacial technology (PaInT), Department of applied analytical and physical chemistry, Faculty of bioscience engineering, Ghent University, Ghent, Belgium

Dean Faculty of Bioscience Engineering

Prof. dr. ir. Marc van Meirvenne

Rector Ghent University

Prof. dr. ir. Rik Van De Walle

Single-cell optical fingerprinting for microbial community characterization

Ir. Benjamin Buysschaert

Thesis submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) in Applied Biological Sciences at Ghent University

Titel van het doctoraat in het Nederlands:
Single-cell optische fingerprinting voor de karakterizatie van microbiële gemeenschappen

Copyright © 2018
The author and promotors give the authorization to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author

Please refer to this work as:
Buysschaert, B. (2018). Single-cell optical fingerprinting for microbial community characterization, PhD thesis, Ghent University, Belgium

ISBN: 978-9-46-357076-3

This work was supported by the project grant SB-131370 of the IWT Flanders, by the IMPROVED project, subvented by The interreg V “Vlaanderen-Nederland” program, a program for transregional collaboration with financial support from the European Regional Development Fund. More info : www.grensregio.eu, by the Geconcerteerde Onderzoeksactie (GOA) of Ghent University (BOF09/GOA/005), and by the Ernest Dubois prize 2016 of the King Baudouin foundation.

Cover illustration
Arthur Lambillotte

NOTATION INDEX

A.U.	Arbitrary units
AOC	Assimilable organic carbon
ATP	Adenosine triphosphate
bp	basepairs
CARS	Coherent anti-Stokes Raman spectroscopy
CCD	Charge coupled device
cFDA	Carboxyfluorescein diacetate
cFDA-SE	Carboxyfluorescein diacetate succinimidyl ester
CFU	Colony forming unity
CPVC	Chlorinated polyvinylchloride
D_{0-2}	Hill number diversity indices
DGGE	Denaturing gradient gel electrophoresis
$DiBAC_4(3)$	Bis-(1,3-dibutylbarbituric acid)trimethine oxonol
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DOC	Dissolved organic carbon
dsDNA	Double-stranded deoxyribonucleic acid
DWDS	Drinking water distribution systems
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular polymeric substance
EU	Europe/European Union
FCM	Flow cytometer
FCS	Flow cytometry standard
FDA	Fluorescein diacetate
FL-1, FL-3, ….	Fluorescence detector of a flow cytometer, number started from the lowest wavelength

FRET	Fluorescence resonance electron transfer
FSC	Forward scatter
HE	Dihydroethidium or hydroethidine
HNA	High nucleic acid
HPC	Heterotrophic plate counts
IR	Infrared
LDA	Linear discriminant analysis
LNA	Low nucleic acid
LP	Long pass (e.g. a long pass filter)
LRV	Log reduction value
LTRS	Laser tweezers Raman spectroscopy
MF	Microfiltration
MRM	Microbial resource management
NF	Nano filtration
NR	Nile Red
OD	Optical density
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PE	Polyethylene
PI	Propidium iodide
PLS	Partial least squares
PLS-LDA	Partial least squares regression followed by linear discriminant analysis
PMT	Photo multiplier tube
PVC	Polyvinylchloride
QC	Quality control
RNA	Ribonucleic acid

RNS	Reactive nitrogen species
ROS	Reactive oxygen species
rRNA	Ribosomal ribonucleic acid
SERS	Surface enhanced Raman spectroscopy
SG	SYBR green I
SGPI	SYBR green I and propidium iodide
SIP	Stable isotope probing
SSC	Side scatter
TERS	Tip-enhanced Raman spectroscopy
TOC	Total organic carbon
UF	Ultrafiltration
UPVC	Unplasticized polyvinylchloride
UV	Ultraviolet
VBNC	Viable but non-culturable
WHO	World health organization
YE	Yeast extract

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

CHAPTER 1
INTRODUCTION

1 Microbial ecology

Microorganisms were the first living creatures on earth and its sole inhabitants for almost three billion years. They survived the most harsh conditions in the history of the earth and helped to shape the environment. Today they are still the most abundant and widespread group of living organisms and still play a vital role in our planetary ecosystem. Yet it's only in the late 17th century that they were observed for the first time with the advent of the microscope. At first, microbiology, led by Louis Pasteur and Robert Koch, focused on pure cultures and the role of bacteria in infectious diseases and fermentation. Microbial ecology itself emerged at the end of the 19th century, when Sergei Winogradsky and Martinus Beijerinck discovered the biochemical cycles such as chemolithotrophy, the carbon or the nitrogen cycle. In the first part of 20th century, microbial ecology evolved to a discipline underlining the interactions between bacteria, and with their environment in all its diversity (Bertrand *et al.*, 2015).

1.1 Microbial ecosystems

Microbial ecosystems are typically very diverse ecosystems both in structure and functionality and play a key role in nearly all biogeochemical cycles of our planet such as, the nitrogen cycle or the carbon cycle (Schmidt, 2006). These processes are only accomplished by the joint effort of microorganisms with a different functional role. The microorganisms do not act as individuals but rather as dynamically changing microbial communities (Little *et al.*, 2008, Klitgord and Segre, 2010). The diversity is thus a key aspect for the functioning of an ecosystem and is therefore central in our search to better understand microbial ecosystems for microbiological, ecological and biotechnological purposes. Microbial ecosystems can be distinguished depending on their complexity, *i.e.* natural communities, synthetic communities or pure cultures. Each of these levels can help us to better understand different aspects of how microbial communities function (**Figure 1 - 1**).

A plethora of natural microbial ecosystems have been described but only the freshwater microbial ecosystems will be discussed further. Microbial communities in fresh water typically fluctuate between 10^4-10^5 cells/mL for drinking water or ground water and 10^5-10^7 cells/mL for surface water. They play an important role in the nutrient cycle of the water bodies and can affect the overall water quality of drinking water or process water (Besmer *et al.*, 2014). The microbial community composition between different fresh water bodies can vary enormously in function of the type of water. Considering the chemical and physical differences between ground water, river water or lake water, this discrepancy is not surprising. Multiple factors shape the microbial communities, one of which is the oxygen concentration. While oxygen is abundant in air, the solubility in water is low and only in water exposed to light, phototrophic microorganisms produce oxygen. In the deeper layers, facultative bacteria consume all the oxygen to metabolize the organic matter, resulting in anoxic conditions (Salmond and Whittenbury, 1985). The nutrient concentration or trophic status also influences the microbial communities as it defines the available niches (Lindstrom, 2000). While drinking water contains roughly 2 mg/L of total organic carbon (TOC), river water can contain more than 10 mg/L TOC (Niemirycz *et al.*, 2006). The amount of organic matter again influences the available oxygen as, even in turbulent water, a high carbon load can lead to oxygen deficits due to microbial respiration (Salmond and Whittenbury, 1985). Also temperature plays a role and reports have shown that microbial communities change significantly depending on the season (Pinto *et al.*, 2014). Yet, the bacterial taxa that play the most prominent role in these ecosystems remain relatively unknown (Newton *et al.*, 2011). Our current knowledge is mainly based on the isolation and identification of the bacteria. But, as most bacteria are not able to be cultured or isolated, big gaps in our knowledge remain.

For ecological research, natural communities are very relevant though difficult to study considering their complexity and the impossibility to control the numerous external factors. Artificial communities, referred to as synthetic ecosystems, consist of an assembled microbial ecosystem created by a bottom-up approach where two or more defined microbial populations are brought together in a controlled and well-defined environment (**Figure 1 - 1**). The advantages are the lower complexity, higher controllability, and higher reproducibility (De Roy *et al.*, 2014b). However, these artificial ecosystems should not be considered as miniature versions of natural ecosystems but as a tool to test ecological theories and to study fundamental principles such as microbial interactions, nutrient cycling or the influence of external factors observed in natural communities. Synthetic ecosystems are also used to create interacting communities with specific characteristics (De Roy *et al.*, 2014b, Jessup *et al.*, 2004).

The culturing of bacteria allows us to isolate bacterial species for in-depth analysis of their physiological potential and characteristics. The isolation of bacteria is a cornerstone of microbiological research, and while pure cultures are the simplest microbial experimental set-up, they are still not fully understood (Jessup *et al.*, 2004). With the advent of technologies such as genomics, transcriptomics, proteomics and metabolomics, a lot of knowledge has been gained in the taxonomic and physiologic diversity of bacteria, which helped to understand the functionality, interactions or adaptability of these bacteria within communities (Kuypers and Jorgensen, 2007, Wagner, 2009). The combined focus on microbial functionality and activity with the in-depth analysis of genetics, has also led to new insights and a growing focus on microbial individuality (Ackermann and Schreiber, 2015). Research showed that even taxonomically homogeneous populations could be functionally heterogeneous (Elowitz *et al.*, 2002, Wagner, 2009, Ceuppens *et al.*, 2013, Orphan *et al.*, 2001, Musat *et al.*, 2008). This phenotypic diversity is a deeper level of diversity, which also shapes the microbial community in ecosystems as it is believed to be beneficial. The first advantage of phenotypic heterogeneity is that it increases the chance of a population to survive fluctuating environments. This bet-hedging strategy is an alternative strategy where microbes diversify instead of responding to environmental cues. An example of this is the occurrence of persister cells that show little growth or metabolic activity and that are therefore able to survive antibiotic treatment (Balaban *et al.*, 2013). Another advantage is the division of labor where a population divides itself in subpopulations using different pathways that are either incompatible or inefficient to combine. The cellular differentiation of cyanobacteria filaments is an example of labor division (Flores and Herrero, 2010). Several mechanisms have been reported to induce phenotypic heterogeneity. A first mechanisms is genetic modifications or mutations but, as phenotypic differentiation occurs at rates higher than any known mutational mechanism and as it is robust against the suppression of mutational mechanisms, it is not considered as the most prominent cause (Ackermann, 2015). Another mechanism is the stochasticity of gene expression (Elowitz *et al.*, 2002, Blake *et al.*, 2003) and the stochastic partitioning of molecules at cell division (Huh and Paulsson, 2011), although the latter is less important in the case of symmetrical division. Next to intrinsic mechanisms, also extrinsic mechanisms such as cell-to-cell communication (e.g. quorum sensing) and environmental fluctuations can lead to phenotypic differentiation (Ackermann, 2015) (**Figure 1 - 1**).

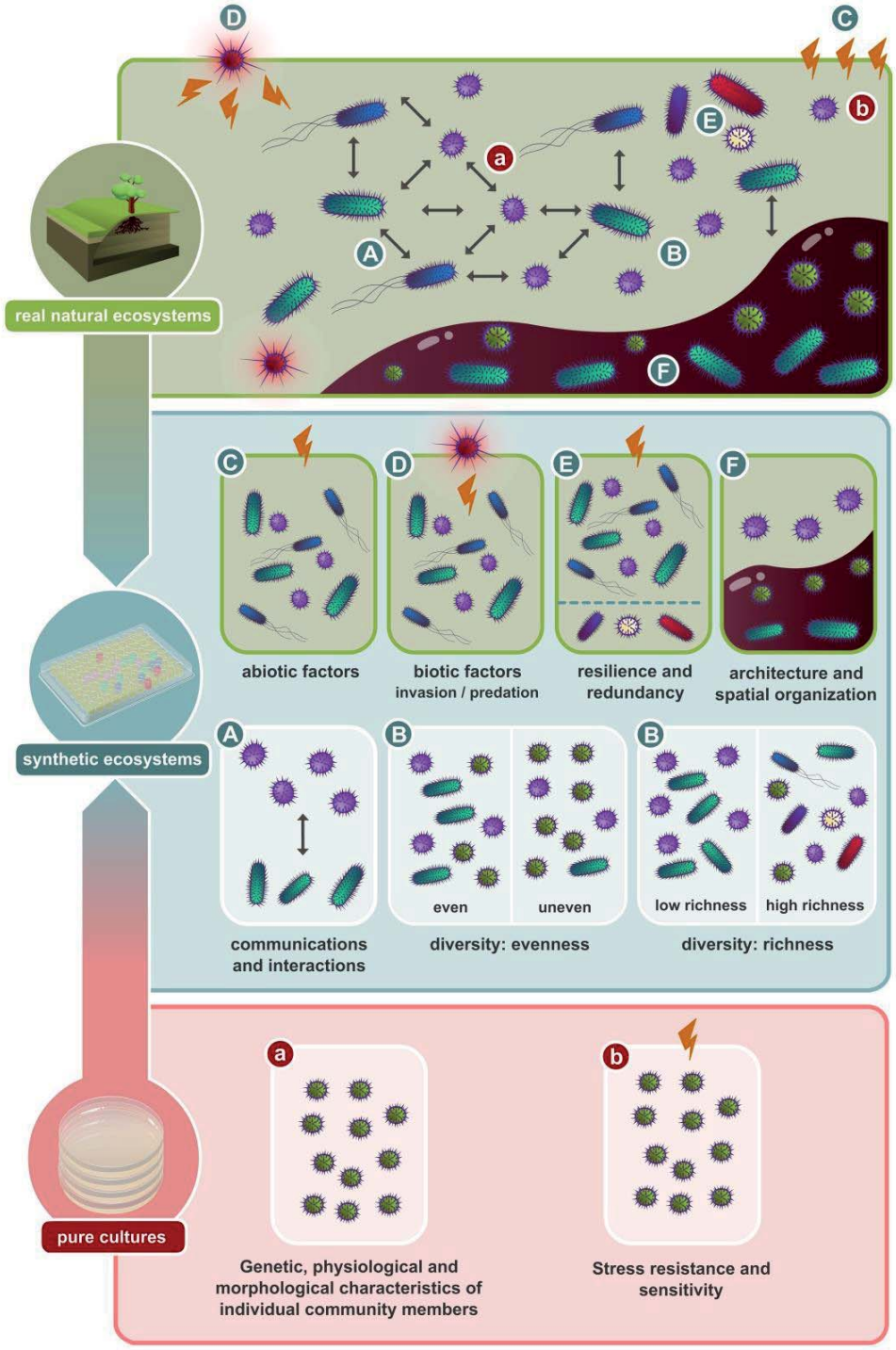


Figure 1 - 1: *The connection between the different levels of microbial ecosystems. Natural ecosystems are the most complex and are shaped by microbial interactions and communication (**A**) and by the community diversity, which is determined by both the richness and the evenness of the community (section 1.2) (**B**), abiotic factors (**C**), biotic factors (**D**), the community resilience and resistance (**E**), and its architecture and spatial organization (**F**). Synthetic ecosystems are approximations of natural ecosystems but assembled with pure cultures. Synthetic ecosystems are mainly used to study microbial ecology. Pure cultures can be considered as a community of phenotypically different subpopulations and both intrinsic (**a**) or extrinsic (**b**) mechanisms can change the phenotypic heterogeneity of a microbial population (De Roy et al., 2014a).*

1.2 Microbial resource management

Microorganisms are not only vital for earth's biogeochemical cycles but are now also crucial for our modern society because many basic food products, medication or fine chemicals are produced with microbiologically driven processes. Also wastewater treatment, bioremediation, bioleaching or bioaugmentation is facilitated by the activity of microbial communities. All these processes rely on natural microbial communities, axenic cultures or combinations thereof. It is now well-established that bacteria do not work alone but rather in communities. The way to steer biotechnological or environmental microbial processes is by steering the microbial communities as a whole. Verstraete *et al.* (2007) therefore proposed the concept of microbial resource management (MRM) that, similarly to human resource management, aims at increasing the efficiency of the microbial capital. For this, three questions about the community need to be handled. "Who is there?", "Who is doing what?", and "Who interacts with whom?". To answer these questions, different types of techniques can be used which will be discussed further (**Chapter 1, section 1.3**). These methods often yield a large amount of data and in order to deal with all this information, Marzorati *et al.* (2008) and Read *et al.* (2011) proposed MRM tools inspired by macroecological concepts. They were initially developed for gel-based molecular methods such as denaturing gradient gel electrophoresis (DGGE) but can be extended to other techniques such as sequencing or even flow cytometric fingerprinting (Props *et al.*, 2016). These MRM parameters are the richness, the evenness, and the dynamics. Depending on the method, they are calculated in a different way and the biological meaning and interpretation can differ.

The richness of an ecosystem represents the number of species present. For sequencing-based data, the interpretation is evident as the richness equals the sum of the identified taxa. A high richness indicates the high carrying capacity of the environment. Research has shown that a high taxonomic species richness helps microbial communities to resist to microbial invasion and stress (Mallon *et al.*, 2015, De Roy *et al.*, 2013). Furthermore, also community functionality improves with increasing richness, though at a decreasing rate (Bell *et al.*, 2005). In case of flow cytometry, richness is defined as the number of bins in which events are registered (**Box 1**). The relationship between cytometric richness and community functionality, resistance and resilience has yet to be reported.

The number of species alone is not sufficient to describe a community and also their abundance plays an important role. Species diversity, as defined in ecology, is therefore always a combination of species richness and species evenness. Many different indices have been proposed based on different calculations and reporting different aspect of a microbial community (Legendre and Legendre, 2012). The order-based Hill numbers are

ecologically meaningful because they are interpretable in terms of the 'effective number of species' (Hill, 1973, Props *et al.*, 2016). Depending on their order, Hill numbers take the evenness more or less into account (**Box 1**). These indices, calculated for both flow cytometry data and sequencing data, have been reported to correlate well (Props *et al.*, 2016).

The microbial community composition is continuously in flux (Shade *et al.*, 2012, Konopka *et al.*, 2007). The dynamics of a system express how much communities change over time, and is the consequence of both endogenous (e.g. microbial interactions) and exogeneous (e.g. environmental changes) factors (Konopka *et al.*, 2015, Jiao *et al.*, 2017). Its potential influence on functionality is evident, yet the system's function can be retained regardless of the changes in community composition as well (*i.e.* resilience) (Shade *et al.*, 2012). Research based on co-occurrence patterns has given valuable insights in the influence of specific factors on community dynamics (Widder *et al.*, 2016). Yet, ecological concepts underlying the dynamics are still lacking (Konopka *et al.*, 2015).

Box 1 - A phenotype, a bin, an operational phenotype, and phenotypic diversity

A phenotype is defined by the observed traits of the bacteria. These observed traits depend on the technique used to analyze the bacteria. In case of flow cytometry, the fluorescence intensity or scattered light of each cell is measured. Binning is the process of grouping continuous values into a smaller number of categories or bins. The binned value is representative of all data points within the category. The purpose of binning is to reduce the number of values and to reduce the effect of small variations in the data. In the case of flow cytometry, the continuous intensity scale can be binned for all relevant detectors or dimensions. As all cytometry data points within a bin have the same optical properties, they can be considered to be of a same operational phenotype. The density of data points in each bin corresponds to the abundance of each of these operational phenotypes. The resulting flow cytometric fingerprinting data is then similar to the output of typical sequencing based on which phenotypic diversity or MRM parameters can be calculated (**Figure 1 - 2**).

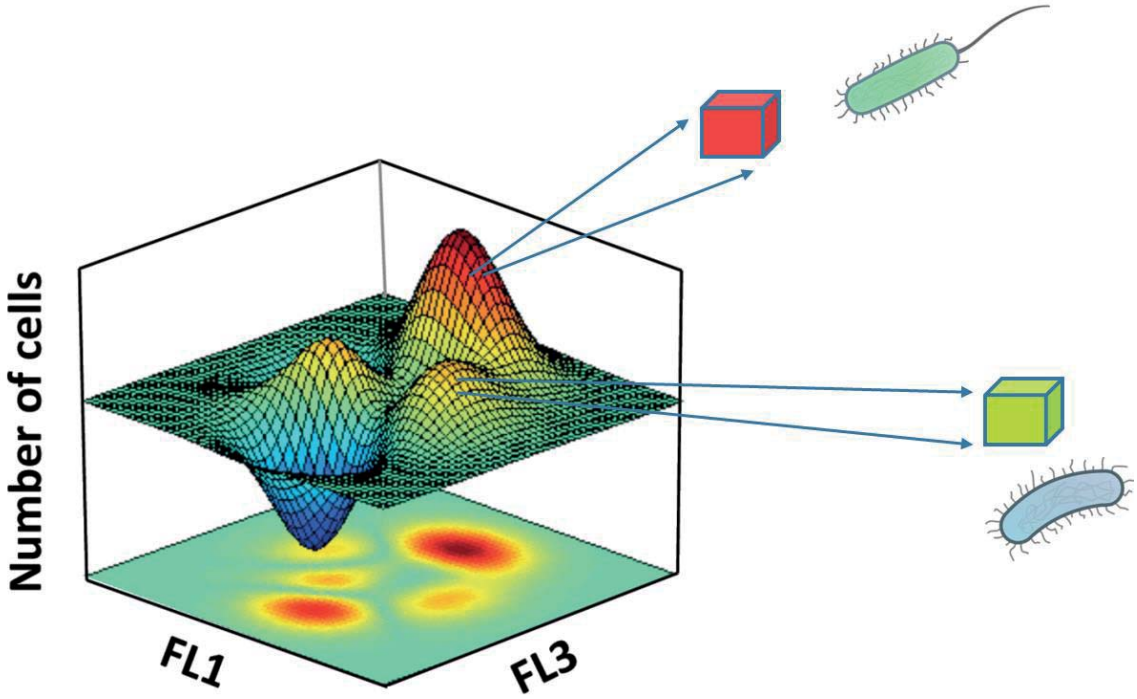


Figure 1 - 2: *Illustration of the binning process on a biparametric cytometric fingerprint shown as a three dimensional cell density landscape. The cell density is illustrated with the color in function of both fluorescence parameters FL-1 and FL-3. A grid divides the data in bins in which the cell density is calculated per bin. Because all cells in a bin have the same optical characteristics, they can be defined as operational phenotypes*

The diversity can be calculated based on the Hill number diversity indices which is represented with the following general formula (**Equation 1**) in which *a* represents to order of the equation and p_i the relative abundance of a species in an ecosystem or phenotype in an cytometry plot.

$$D_{q=a} = \left(\sum_i p_i^a\right)^{1/(1-a)} \quad (1)$$

Depending on their increasing order (q = 0 - 2), the Hill number diversity indices increasingly emphasize the evenness (**Equation 2 – 4**). For a given sample with S bins containing cells, the first order Hill number is equal to the species richness and does not take evenness into account.

$$D_{q=0} = \sum_{i=1}^{S} 1 = S \quad (2)$$

$$D_{q=1} = e^{-\sum_{i=1}^{S} p_i ln(p_i)} \quad (3)$$

$$D_{q=2} = \frac{1}{\sum_{i=1}^{S} p_i^2} \quad (4)$$

1.3 Monitoring microbial ecosystems

To further study microbial ecology or to put the concepts of microbial resource management into practice, good tools for community analysis are required. Numerous methods are available and, in general, a distinction can be made between culture-dependent and culture-independent techniques. The culture-dependent techniques are fundamental for microbiology and are still widely used for the isolation and culturing of species of interest. The tools are cheap and the methods rather easy to perform but require long incubation times and are labor intensive. Moreover, many microorganisms cannot be cultured, making their use for ecological research sometimes difficult.

Culture-independent methods can be divided further into single-cell and bulk methods. Examples of single-cell methods are flow cytometry or Raman spectroscopy, which will be discussed in **Chapter 1, section 2**. In the last decades, bulk methods such as molecular techniques gained much popularity. At first, molecular fingerprinting tools were used for community characterization such as DGGE, but the technological advances and the decreased prices have made sequencing technologies, such as 454 pyrosequencing and Illumina, a standard method. Sequencing methods are particularly useful for microbial ecology and microbial resource management as they are able to identify the microbial taxa present, as well as determine their relative abundances. However, there are a few important caveats to sequencing or other molecular approaches. A first point of attention is that they require DNA extraction, which can lead to important biases considering the heterogeneity in the microbial world and the sometimes difficult matrix in which bacteria are embedded (Starke *et al.*, 2014). In many cases, the extracted DNA also needs to be amplified which adds another bias. Universal primers are seldom truly universal because they do not amplify each DNA strand with the same efficiency (Thijs *et al.*, 2017). As a result, the relative abundances can be an over- or underestimation of the actual abundance of the taxa. The last source of bias is the comparison of 16S rRNA gene sequences with the available 16S rRNA gene databases for identification. As different databases are available also different results can be obtained for a same dataset. Also, databases are not free of errors and many depositions are of poor quality or incorrectly labeled (Ashelford *et al.*, 2005, Janda and Abbott, 2007). Due to these different sources of bias, much research is done on bioinformatics pipelines to reduce these issues though different pipelines are reported to influence the results significantly (Mysara *et al.*, 2017). For example, recently, it was proposed that the relative species abundances should be combined with absolute cell concentrations, as determined with for example flow cytometry, to calculate the absolute species abundances (Props *et al.*, 2017).

Sequencing is only answering one of the three questions of microbial resource management and is therefore insufficient to characterize microbial communities on itself. Ideally also more information on the microbial activity is necessary. For many applications, the presence or absence of one or more molecules can be used to describe the community functionality. Depending on the type of molecule or concentration different techniques can be used. Typically, a correlation between the occurrence of certain species and the required metabolic activity is made *a posteriori*. Information on how microbial communities react to external factors or how bacteria interact with each other, has been obtained this way (Brehm-Stecher and Johnson, 2004). But, this approach only considers microbial communities on the population level and in the last couple of years, emphasis on the potential importance of phenotypic heterogeneity has been published repeatedly (Delvigne and Goffin, 2014, Ackermann, 2015). This new attitude towards microbial communities calls for single-cell technologies to characterize the microorganisms, their activity and interactions at the individual level to provide an insight on how these communities work with unprecedented detail. Very promising development has been made in the field of single-cell sequencing and transcriptomics (Brehm-Stecher and Johnson, 2004). However, the method is not yet fully developed and some important issues and technical difficulty are refraining the wider application of the method. Blainey (2013) reviews the issues and perspectives of single-cell genomics and transcriptomics in the field of microbiology.

2 Alternative approaches

The increased awareness of the importance of the bacterial individuality calls for single-cell methods for microbial analysis. Ideally, fast and easy methods requiring little or no sample preparation are used. In this developing field, two promising methods among many, are flow cytometry and Raman spectroscopy.

2.1 Flow cytometry

Flow cytometry is an optical technique for analyzing individual particles, typically cells, that are suspended in a fluid. It measures and analyzes simultaneously multiple physical and (bio)chemical characteristics of the single particles as they flow in a liquid stream through an excitation light source, which is typically a laser beam (Shapiro, 2003, Hammes and Egli, 2010). The first flow cytometer was built in 1934 in Canada and comprised of a capillary mounted on a microscope for photoelectric cell counting (Moldovan, 1934). Later, Guckert and O'Konski (1947) modified the technique and their particle counter was used successfully by the United States army in the second world war to detect airborne bacterial spores (Davey

and Kell, 1996). The first modern fluorescence-based cytometer was patented in 1968 (Dittrich and Gühde, 1973) and commercialized that same year by Partec in Göttingen, Germany. Since then, flow cytometers have improved considerably due the technological advances in optics and electronics and eventually found their way in the field of microbiology in the 1980s (Wang *et al.*, 2010).

2.1.1 Working principle

A flow cytometer comprises of three parts: the fluidics, optics and electronics system (**Figure 1 - 3**). The fluidics system transports the particles in a fluid stream to the light source. By injecting the sample into a stream of sheath fluid, the flow of the sheath fluid accelerates and aligns the particles. After alignment by so-called hydrodynamic focusing, the particles are transported to the flow chamber where they pass individually through a light beam. When the particles pass the laser interrogation point, their intrinsic properties, like cell size, granularity, and morphology, lead to specific interactions with the laser light, including scattering and fluorescence. Light scattering is detected either at a small angle by a forward scatter detector (FSC), or perpendicular to the incident light by a sideward scatter detector (SSC). Fluorescence on the other hand, typically originates from the excitation of fluorochromes or from autofluorescent particles and is detected after selection with appropriate wavelength filters (Shapiro, 2003). The scattered and emitted fluorescent light is collected by a system of optical mirrors and filters to route specified wavelengths of light to different optical detectors which are all part of the optical system. The light signals are finally detected by photomultiplier tubes (PMT) and converted into electronic signals in the electronics system (Shapiro, 2003).

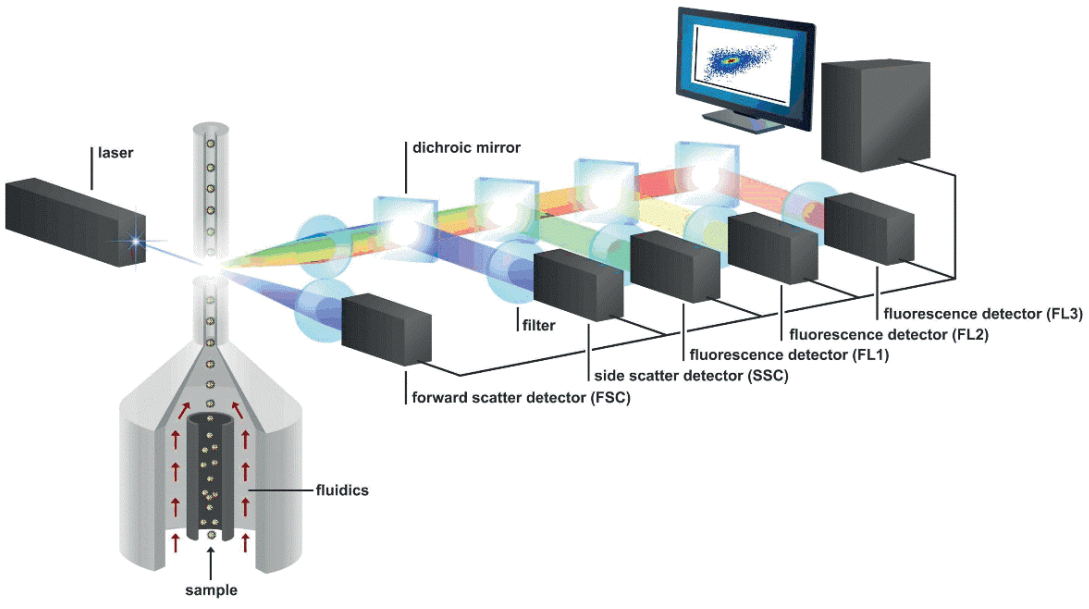


Figure 1 - 3: *Scheme of a flow cytometer (De Roy, 2014).*

The main advantage of the method is the speed of the analysis as up to 10 000 cells per second can be analyzed. Furthermore, automated sampling and integration with robotics enables an automated and high throughput analysis. Modern instruments are also very accurate with an instrument error below 5%, and sensitive with a detection as low as 100 cells/mL (Hammes and Egli, 2010). Because bacteria are small and because samples can contain a lot of abiotic particles, such as crystals and dust, which cause light scattering, it is necessary to label microbial cells with fluorescent molecules (Wang *et al.*, 2010). In general, the scattered light already provides information about basic cells characteristics (e.g. cells size or surface properties), but fluorescent probes provide additional information about specific features. A wide variety of dyes are available and depending on which dyes are used the viability, metabolic activity or intracellular pH per cell can be measured (**Chapter 2**). Multiple features can be measured simultaneously as flow cytometers are equipped with several lasers and detectors. However, when multiple fluorescent markers are used, optical cross-talk can be an issue. Cross-talk or spillover is caused by the overlapping emission spectra of the fluorochromes but can be corrected by spectral subtraction referred to as compensation. A last advantage of flow cytometry is that there is no need for a culturing step or DNA extraction which makes it very useful for environmental microbiology. Normally little sample preparation is required though some pre-treatment such as sonication can be necessary to detach bacteria from biofilms or to disintegrate flocs (Wang *et al.*, 2010, Hammes and Egli, 2010)

2.1.2 Fingerprinting

Analysis of flow cytometric data typically involves gating and feature extraction on the one hand and interpretation of the flow cytometric patterns on the other hand (Bashashati and Brinkman, 2009). In the gating process, the operator draws approximated polygons around specific areas of interest, for example around bacterial cell clusters with similar properties (Hammes and Egli, 2010). From these gates of interest, features like the percentage of cells within the identified gates, are computed for further analysis (Bashashati and Brinkman, 2009). Since gating is a highly subjective process, which is mainly based on the experience and knowledge of the operator, it is often inaccurate and prone to error (Maecker *et al.*, 2004). A solution is automated gating which classifies and clusters cells in a more objective and systematic way. In addition, automated gating is less time consuming and labor intensive than manual gating when processing a large number of samples (Bashashati and Brinkman, 2009).

Depending on the physiological state or the taxonomy of the bacteria, flow cytometric data changes in ways that cannot be quantified with traditional gates. To do so, so-called fingerprinting techniques have been developed. Fingerprinting can be done with a variety of techniques, each different in their approach and purpose. A first approach is based on the image processing of the cytometric biplots. Bombach *et al.* (2011) introduced the Dalmatian plot where an overlap image of each pair of images is created. In a first step, the operator gates the most abundant subsets of cells, and in a second step the gates are colored in black or in greyscale. Pairwise overlap images are then calculated and based on the number of identified gates, the total number of positive pixels, the average size of each gate and the relative fraction of how much each pixels accounts for in the picture the Jaccard dissimilarity between the images is calculated. As improvement to this method, the cytometric histogram image comparison (CHIC) procedure was introduced, but in contrast to the Dalmatian plot method, no initial gating is required making this method more reproducible and objective. Following this, the images are compared based on an exclusive disjunction function (XOR) and an overlap image. Both are then combined to calculate an average grey value per informative pixel which is used to calculate the dissimilarity between the samples (Koch *et al.*, 2013a). Because those methods are based on image analysis, they will not be discussed further in this introduction. For more information, the comparative study by Koch *et al.* (2014) is recommended. A second and more often used approach is based on gating or binning of cell populations. The simplest method was developed by Prest *et al.* (2013) and is mainly used for water analysis where aquatic microbial communities, stained with SYBR green I, show two populations. Based on the SYBR green's property to bind nucleic acids they are therefore called low nucleic acid bacteria (LNA) and high nucleic acid bacteria (HNA) (Lebaron *et al.*, 2001, Robertson and Button, 1989). The ratio of those both populations is supposed to be constant in a stable environment, whereas deviations are caused by changes in the microbial community. This remarkably simple method is proved to be sensitive for the detection of changes. Today, it is used to monitor drinking water quality and was integrated in the Swiss legislation (SLMB, 2012), and although flow cytometry has not been legalized in other countries for this purpose, drinking water utilities explore its use an early warning system but further development is required. A disadvantage of the method is that it only works with nucleic acid stains capable of differentiating HNA and LNA bacteria. Furthermore, the distinction between HNA and LNA bacteria is not possible for most microbial communities and for axenic lab cultures, thus precluding the universality of the method. The cytometric barcoding (CyBar) method, developed by Koch *et al.* (2013c), is based on the creation of a gating template for all samples. An ellipsoid gate is drawn for every population that appears over the course of the experiment. The normalized number of cells detected in each gate is used as a barcode characterizing the microbial community.

Similarly to the previous method, changes in the microbial community are translated in fluctuations of the cell counts per gate and subsequently the characteristic barcode. Based on this method, correlations between populations and abiotic data can be calculated. As this method relies on the gating of populations, suitable stains and instrumentation are necessary to maximize the number of detected populations. It is evident that a reduced number of populations engenders a decreased resolution and sensitivity of the method. As a result, this method is suitable for microbial communities with a high physiological and taxonomic heterogeneity who can produce a higher amount of subpopulations but it is unsuitable for less diverse communities or axenic cultures. A second downside of the method is the inherent subjectivity of gating, hindering the reproducibility of the method by independent operators. To overcome this limitation an objective and gate-free method can be implemented, as described by Rogers and Holyst (2009), where an algorithm based on all samples divides the cytometric biplots in bins with an equal number of cells. This method, referred to as probabilistic binning, creates a model applied on all samples to calculate the number of cells per bin. As a result, the biological meaning behind the binned populations is lost. This approach was successfully applied by De Roy *et al.* (2012) to discriminate different brands of bottled drinking water and to detect the impact of physicochemical perturbations on the microbial community. In addition, De Roy *et al.* (2012) improved the method by performing a discriminant analysis on the fingerprints to maximize the performance. Similarly to the method described by Koch *et al.* (2013c), the dependency of the method on the original dataset impedes its flexibility. To solve this, an equally spaced grid can be used instead of probabilistic binning as described by Van Nevel *et al.* (2016b), where we successfully showed that it was possible to use flow cytometric fingerprinting for fast monitoring of drinking water quality. However, post-processing of the data is based on a supervised discriminant analysis. For this, all the data should be analyzed simultaneously and knowledge about what has to be discriminated is required, making this approach unsuitable for exploratory and real-time analysis. To address this problem, and to make flow cytometric fingerprinting a useful tool for an early-warning system, Props *et al.* (2016) introduced the so-called phenotypic diversity indices after binning with an equally spaced grid. The convenience of reducing the flow cytometric fingerprint to a single number improves interpretation and visualization especially for time-series where the change in phenotypic diversity can be used for monitoring.

2.2 Raman spectroscopy

Raman spectroscopy is an optical technique names after the Nobel prize-winning Indian physicist Sir C. V. Raman. The method observes the vibrational, rotational, and other low frequency modes of a molecule in a cell for example, via light scattering. It is widely used in the field of chemistry to provide structural fingerprints to identify molecules. The Raman effect, *i.e.* the inelastic scattering of light, was observed for the first time in 1928 (Raman and Krishnan, 1928). But, due to the low signal intensity and the sensitivity of the method, it took until the advent of the lasers in the 1960s before Raman spectroscopy was used as a common analytical method.

2.2.1 Working principle

The periodic motion between the atoms in a molecule is referred to as molecular vibration. A number of different vibrational modes are possible and include bending and stretching of the molecular bonds. Raman spectroscopy can measure these vibrational modes directly by measuring the scattering of monochromatic light. When monochromatic light, typically from the visual part of the spectrum, interacts with the sample, the light will be partly scattered. This scattering can be split up in two parts, *i.e.* the elastic and the inelastic scatter. In both cases an incident photon is temporarily absorbed by the transition from a ground state into a virtual state of excitation and is scattered by the transition from this virtual state back to the ground state (**Figure 1 - 4**). The majority of the scattered light has the same wavelength or energy as the incoming light which is referred to as the elastic or Rayleigh scatter. In case of inelastic scatter, the scattered photons from the monochromatic light have more or less energy after interacting with the molecular vibrations. In case the photons lose energy, *i.e.* the photon wavelength increases, the scatter is called Stokes scatter. This energy loss is due to energy absorption by the molecule. In case the photons gain energy, *i.e.* the wavelength decreases, the scatter is called anti-Stokes scatter. This energy gain is due to energy dissipation from the material. Molecules that are initially in the ground state give rise to Stokes scattering, while molecules that are initially in an excited vibrational state give rise to anti-Stokes scattering. At ambient temperatures, more molecules are in their ground states, and therefore Stokes scattering is more intense than anti-Stokes scattering (Larkin, 2011).

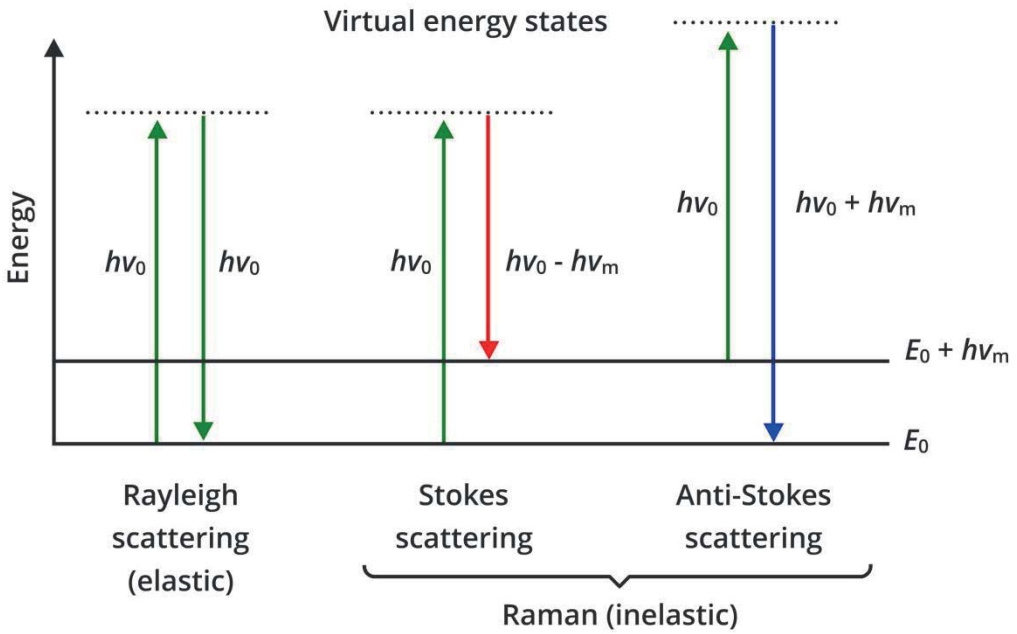


***Figure 1 - 4**: Jablonski energy diagram illustrating the different types of scattering. Horizontal lines represent different vibrational states of the molecular bonds. The incident photons from the monochromatic light have an energy level equal to* hv_0*, with* h *as Planck's constant. After exciting the electron could to a virtual state, a second photon is emitted. In case of Rayleigh or elastic scatter, the emitted photon has the same energy as the incident photon. Other types of scatter are referred to as Raman or inelastic scatter. When the photons lose energy, the scatter is called Stokes scatter. When the photons gain energy the scatter is called anti-Stokes scatter.*

This energy loss or gain is caused by the interaction of the photons with the molecular electron cloud. The vibrational states of the molecular electron cloud create electric fields and the polarizability of the molecule is the ease at which this electron cloud can be changed in shape, size or orientation in response to an external electric field. Incident photons will also generate an electromagnetic field, generated by the photon's energy state, and can polarize the electromagnetic field of the molecule it interacts with which is referred to as an induced dipole. By causing this induced dipole, the incident photons can either gain or lose some energy. As the vibrational modes of a molecule are dependent on the mass of the atoms and their geometric arrangement, the nature of their chemical bond, and their motions, Raman scattering can be used to collect this information. Thus, the structure and the properties of the molecules in a sample can be characterized. Next to that, the relative intensity of the Raman signal is also dependent on the relative concentration of the molecules making quantification of the chemical bonds possible. As polarizability is necessary for a Raman scattering, not all molecules can be detected with Raman spectroscopy. Also, the smaller the polarizability, the smaller the Raman-effect. Because of that, molecules such as H_2O have a very low Raman-activity which means that water will not interfere with Raman signals. This is an important advantage of Raman spectroscopy for the

biological sciences as water is omnipresent in biological samples. Raman scatter is an inherently weak process as only 1 out of 10^6-10^8 incident photons are Raman scattered (Petry *et al.*, 2003).

Raman spectra contain the intensity of the scatter signal as a function of the energy difference with the incident light, expressed in terms of the so-called wavenumber. **Equation 5** describes how to convert between the Raman shift expressed in wavenumbers to the spectral wavelength of the incident and emitted photons.

$$\tilde{\nu} = \frac{1}{\lambda_0} - \frac{1}{\lambda_1} \quad (5)$$

In which $\tilde{\nu}$ represent the Raman shift in wavenumbers (cm^{-1}) and λ_0 represents the wavelength of the incident photon and λ_1 the wavelength of the emitted photon. This wavenumber is directly related to the energy of the photons which is shown by **Equation 6**.

$$\tilde{\nu}_p = \frac{E}{h*c} \quad (6)$$

Where $\tilde{\nu}_p$ is the wavenumber of a photon in function of its energy *E* divided by the speed of light *c* and Planck's constant *h*.

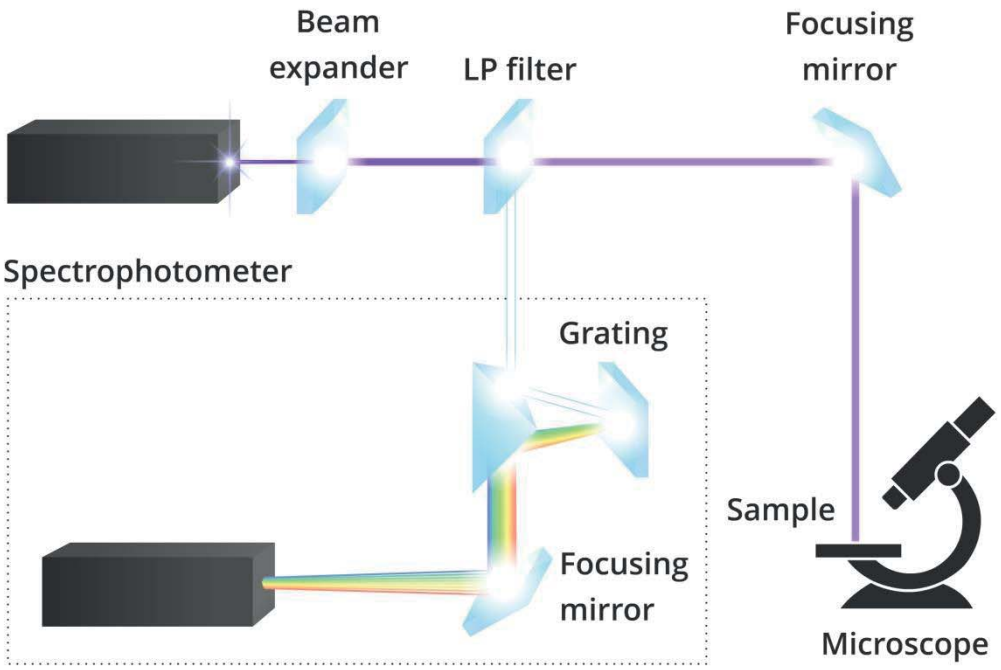


Figure 1 - 5: *Scheme of a Raman spectroscope. A monochromatic light beam is focused on a sample with a microscope. After interaction with the sample, the scattered light is collected by a lens and the Rayleigh scatter is removed by a longpass (LP) filter. The resulting light is grated to split the light in its different components and converted to an electrical signal by a detector (Butler et al., 2016).*

Box 2 - Fluorescence, Raman scattering or infrared (IR) absorption

When incident photons interact with matter, they can bring the molecules from a ground state to an excited state by transferring their energy through absorption. The lifetime of such a state is usually very short and the transition from the excited state back to the ground state can happen in different ways. For example, molecules can emit photons when transitioning from an excited singlet or triplet state to the ground state, which results in fluorescence or phosphorescence respectively. This is usually caused by high energy photons (e.g. from UV or blue lasers). When lower energy photons interact with matter, the energy can by absorbed and bring the molecules to a higher vibrational state without excitation to a higher energy level as with IR absorption. When photons are not absorbed by the molecules they can bring the molecules to a virtual state of excitation and result in scatter such as Rayleigh and Raman scatter.

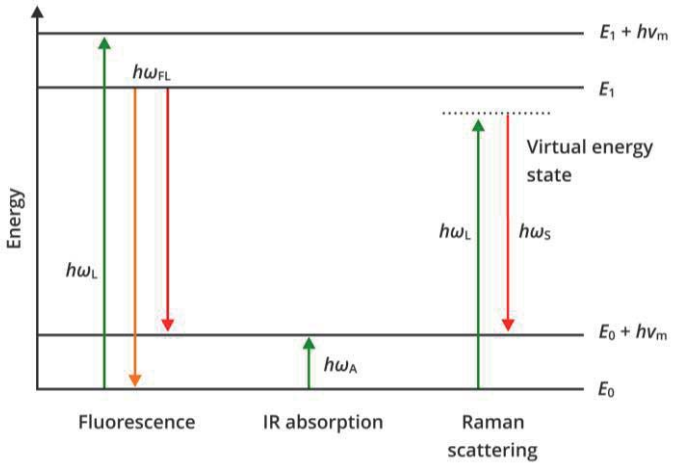


Figure 1 - 6: *Illustration of fluorescence, IR absorption, and Raman scattering with* ω_L = *laser frequency,* ω_{Fl} = *fluorescence emission,* ω_A = *IR absorption,* ω_S = *Raman scattering. Adapted from Petry et al. (2003).*

In practice, lasers with wavelengths in the visible light spectrum are used. As many biological or organic molecules can be autofluorescent with green or blue lasers, red (e.g. 633 nm) or near infra-red (785 nm) are preferred (**Box 2**). However, when increasing the wavelength of the incident light, the Raman scatter signal intensity decreases, and longer acquisition times and higher laser power may be required. For single-cell applications, the lasers are mounted on a microscope in order to focus the light on the bacteria of interest (**Figure 1 - 5**). To ensure that the incident light has only one particular wavelength, a monochromator is sometimes used before the light reaches the sample. During Illumination, the scattered light is focused with a lens and the photons resulting from the elastic scattering are removed by a Rayleigh filter. The resulting light consisting of inelastically scattered photons is then splitted according to its wavelength by a beam splitter or grating. This way, all wavelengths are

collected on a separate spot on the detector. For most applications a charge-coupled device (CCD) detector is used to convert the optical signal into an electronic signal which can then be visualized in a spectrum.

2.2.2 Applications in microbiology

Raman spectroscopy is used for a wide variety of applications but is well-established for material identification, e.g. to assess mineral composition in geology, purity of polymers or pharmaceuticals or even for chemical analysis in art and archaeology (Petry *et al.*, 2003, Vandenabeele *et al.*, 2007). Also in life sciences, Raman spectroscopy has become more popular over the years as it offers a non-invasive, non-destructive and water-insensitive method to characterize the chemical and spatial structure of tissues or cells. In the field of microbiology, the potential of Raman has been reported for species identification (Hutsebaut and Moens, 2005, Jarvis and Goodacre, 2008, Stockel *et al.*, 2016) and for determining microbial interactions by using stable isotope probing (SIP) or deuterium (Berry *et al.*, 2015, Cui *et al.*, 2017).

Raman spectroscopy suffers from some important drawback when measuring spontaneous Raman scattering. The signals are very weak as only 1 out of 10^6-10^8 incident photons are Raman scattered. Also, biological samples often have fluorescent properties which are much stronger than the Raman signals. Even very weak fluorescent signals can mask Raman scatter. To circumvent these issues, special Raman-based techniques can be used of which only a few will be discussed further here. A well-known example is the use surface enhanced Raman spectroscopy (SERS) which enhances the signal of molecules adsorbed on rough metal surfaces or nanostructures, typically silver or gold (Fleischmann *et al.*, 1974). The technique relies on a combination of chemical and electromagnetic signal enhancement. The collective excitation of the electron in the metallic nanostructures, also called the surface plasmon resonance, enhances the electromagnetic field of the molecules bound to, or in close proximity of the metal surface. The enhanced electromagnetic field subsequently induces a stronger Raman signal. The chemical enhancement relies on the charge transfer between the metal surface and the adsorbed molecules (Petry *et al.*, 2003). For this, a short sample preparation in which, for example, bacteria are coated with gold nanoparticles and the adaptation of the laser frequency for it to match the frequency of the plasmon resonance is necessary. SERS has most often been used for species discriminations (Jarvis and Goodacre, 2008). The time gain can, in combination with microfluidic devices, also lead to automation (Walter *et al.*, 2011). For subtle differences, SERS can be unreliable as only the spectra of the molecules in the vicinity of the SERS particles are amplified (Hering *et al.*, 2008). SERS also reduces the holistic character of Raman spectroscopy as only the signal of

the molecules on the surface of the bacteria are amplified (Pahlow *et al.*, 2012). Tip-enhanced Raman spectroscopy (TERS) is a variation on SERS where an almost atomically sharp gold pin is used to enhance the Raman signals. Similarly to atomic force microscopy (AFM), this pin can be used to probe surfaces and to map the Raman profiles. Budich *et al.* (2008) demonstrated how TERS could be used to map the heterogeneity of the microbial cell surface. But, to our knowledge, not much research has been done with TERS in the field of microbiology. A very promising variation on Raman spectroscopy is coherent anti-Stokes Raman spectroscopy (CARS). CARS relies on the interaction of three different lasers, a pump laser, a Stokes laser, and a probe laser. The combination of the pump laser and the Stokes laser is used to excite the molecules to a higher vibrational state (ω_{pump} - ω_{Stokes}). The probe laser then excites the molecule to a virtual state enabling anti-Stokes photons to be scattered (**Figure 1 - 7**). The resulting CARS signal is stronger and directed in comparison to spontaneous Raman scattering. The more energetic nature of the signal also makes CARS insensitive to the interference of fluorescence as it is detected on the short wavelength side of the excitation pulses (Krafft *et al.*, 2009).

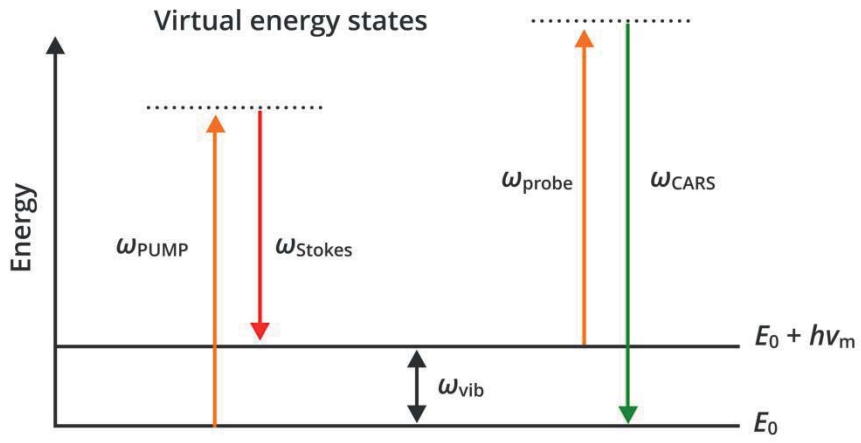


***Figure 1 - 7**: Illustration of the energy levels in coherent anti-Stokes Raman spectroscopy. The samples are probed with both a pump and a Stokes laser with frequencies ω_{pump} and ω_{Stokes}, respectively. As a result, the molecule is brought into a higher vibrational level. Subsequently, the probe laser with frequency ω_{probe} excites the molecule to a virtual state, which results in an anti-Stokes scatter with frequency ω_{CARS}. Because the CARS signal contains more energy, it is both stronger than the spontaneous Raman signals and shows a blueshift.*

The signal enhancement enables CARS images to be collected at video time rates. In contrast to SERS, CARS is non-invasive and makes it thus an ideal method for real-time imaging of cells and tissues (Cicerone, 2016). However, the suppression of background has so-far been challenging and made the application of the technique for research more difficult though solutions are now available (Arora *et al.*, 2008). The potential of CARS has been illustrated in studies of eukaryotic cells and demonstrated subcellular resolution as even

chromosomes can be imaged (Cheng *et al.*, 2002, Yue and Cheng, 2016). Similarly, Okuno *et al.* (2010) demonstrated the subcellular resolution of CARS with budding yeast. Scully *et al.* (2002) proposed and developed with his team a CARS-based approach to detect and identify bacterial spores in air (Petrov *et al.*, 2007, Pestov *et al.*, 2008) while Hong *et al.* (2016) showed that CARS could also be used to detect bacteria in complex matrices such as milk or urine.

3 Objectives and overview of the research

Microbial communities are, regardless of their complexity, central to our planet's health and many industrial processes. To make bacteria-driven processes more efficient and less prone to failures, a good understanding on how these communities work is vital. For both industrial applications or for research, a need for a fast characterization of these communities is necessary to increase this understanding. Microbial resource management has been proposed as a guidelines for steering microbial communities, and while there are many technological solutions for research, there is a lack of fast, cheap, and easy techniques for microbial community characterization. This doctoral research is based on previous work of De Roy *et al.* (2014a) and Van Nevel (2014) where flow cytometric fingerprinting is first proposed. Here we present improvements on the method, test its sensitivity, and explore potential applications. Based on this flow cytometric fingerprinting pipeline, the possibility of Raman spectroscopy as alternative method for microbial community characterization was also explored.

The first part of this thesis focuses on pure cultures as most simple type of a microbial ecosystem. Flow cytometry has often been used for cell counting of pure cultures, yet almost no research has been done to explore the possibilities of flow cytometric fingerprinting for pure cultures. Furthermore, Raman spectroscopy was also explored as method for community characterization. In the second part, flow cytometric fingerprinting was applied to monitor aquatic microbial communities to determine the most influencing factors on the fingerprints, the sensitivity, and the relevance of the method for practical applications.

PART 1: Pure cultures and cocultures

Flow cytometry relies on staining and much research has been done on dyes yet very few truly multicolor protocols have been developed. Moreover, important variations are reported between protocols and some important aspects such as stability in function of time are poorly investigated. In **Chapter 2**, different multicolor staining protocols for automated and high throughput flow cytometric analysis were compared and optimized for both a Gram positive and a Gram negative bacteria species. The purpose of this chapter was to find the most

sensitive, stable, and most biologically informative dye combination for the following research.

In **Chapter 3**, the cytometric fingerprinting toolbox was expanded by calculating the similarity between fingerprints. The sensitivity of the method was tested by comparing the fingerprints of 29 strains and species of taxonomically related Lactobacilli. Furthermore, the sensitivity of the method was calculated by using beads and bacteria mixtures. Finally, also the impact of technical issues and the importance of physiological changes due to microbial growth on the reproducibility is illustrated.

It is well-established and illustrated in previous chapter that flow cytometry fingerprints of axenic cultures change in function of their growth stage. In **Chapter 4**, a classic *E. coli* batch fermentation was monitored with flow cytometric fingerprinting to determine why the cytometric fingerprints change in function of the growth stage. For this, the cell physiology, cell density, and the fingerprints were compared with operational parameters such as respiration and metabolite production. The aim of this chapter was to further explore sensitivity of flow cytometric fingerprinting to physiological changes and to explore the usability of flow cytometry in the context of bioreactor monitoring.

The reliance of flow cytometry on fluorochromes is in some cases and advantage, but it also complicates the sample preparation and slows down the analysis. In **Chapter 5**, we tested how Raman spectroscopy, a non-invasive and non-destructive alternative for flow cytometry, could be used to detect phenotypic diversity of the model organism *E. coli*. Several data processing pipelines were compared and several fingerprinting pipelines are proposed for future research.

To assess the influence of phenotypic diversity in microbial communities, fast single-cell techniques capable of revealing the this lowest level of diversity are required. In **Chapter 6**, both flow cytometry and Raman spectroscopy were compared for the detection of phenotypic changes of two drinking water isolates. For this, microcosms were developed which allowed individual cell populations to interact while remaining physically separated.

PART 2: Environmental microbiology

As flow cytometry showed to be a valuable and sensitive tool for microbial community analysis, the application of the method on complex environmental communities was evaluated. Water distribution systems are fundamental for drinking water or process water distribution. In both cases, the water quality is important and microbial regrowth in the distribution system should be limited as much as possible. Biofilms on the piping material are an important source of microbes in the water, and in **Chapter 7**, flow cytometry was applied

for the monitoring of the biofilm and bulk water microbial community. A series of batch tests and a lab-scale flow-through experiment were set up to assess the influence of different pipe materials and water sources on the dynamics in the microbial communities assessed with flow cytometric fingerprinting. Bacteria are known to colonize biofilms and we investigated with flow cytometry if *Enterobacter amnigenus,* a known drinking water contaminant, could colonize biofilms of different pipe materials.

In the last chapter, **Chapter 8**, the sensitivity of flow cytometric fingerprinting to detect the response of a fresh water microbial community to a contamination was tested by adding elements which could be used as nutrients by the bacteria. Finally, a flow cytometer was installed in a full-scale water production plant to demonstrate the practical application of flow cytometry and cytometric fingerprinting at the different stages of a water treatment plant.

CHAPTER

2

MULTICOLOR FLOW CYTOMETRY

CHAPTER 2

MULTICOLOR FLOW CYTOMETRY

1 Abstract

Flow cytometry is a rapid and quantitative method to determine bacterial physiology. Although different stains can be used to determine the physiology, staining protocols are inconsistent and lack a general optimization approach. Very few 'true' multicolor protocols, where dyes are combined in one sample, have been developed in such a way for microbiological applications. As each dye and dye combination behaves differently within a certain combination of medium matrix, microorganism or instrument, protocols need to be tuned to obtain reproducible results. In this chapter, we review current single, double and triple stains and the different parameters that influence staining: stain kinetics, optimal stain concentration and the effect of the chelator EDTA as membrane permeabilizer. As multiwell autoloaders are now commonly used, samples are often stained at the same time but not measured simultaneously. The time difference between the analysis of samples of a multiwell assay can significantly impact the results. In the last section, we highlight the need to investigate the stability of multicolor assays to ensure correct results.

Chapter redrafted after:

Buysschaert B*, Byloos B*, Van Houdt R, Leys N, Boon N (2016). Reevaluating multicolor flow cytometry to assess microbial viability. *Applied microbiology and biotechnology*. 100:9037–9051

*Equal contribution

2 Introduction

Multicolor flow cytometry is a good approach to simultaneously estimate and assess multiple features and to characterize the physiological heterogeneity of a community in detail. However, staining bacteria is a complex interplay between dye chemistry, the target organisms and the staining conditions. For microbiological applications, the diversity of bacterial species is challenging, as even closely related organisms are known to behave very differently, making it difficult to analyze bacteria in a standardized way (Shapiro, 2000). Hence, it is important to have a reasonable amount of standardization in terms of stain concentration, used buffers, incubation time, need for permeabilization or fixation, and the necessary controls to compare different samples. In addition, a better understanding of the staining chemistry is important to estimate the reliability of a staining protocol for a specific research set-up. Unfortunately, it is exactly on those aspects that many available studies lack the necessary information and where data are poor and inconsistent. An overview of the parameters important for protocol optimization are given by Hammes *et al.* (2011). Very few 'true' multicolor protocols, where dyes are combined in one sample, have been developed in such a way for microbiological applications. A wide variety of fluorescent dyes are available for flow cytometry. Thus, very different aspects of microbial physiology can be assessed and monitored. However, it is important to understand how different dyes function and how they can be applied to draw the correct conclusions. In this chapter, we discuss the discrepancy between protocols for some popular stains and offer a way in which staining could be optimized for a specific set-up. The number of available dyes is vast and summarizing them all in detail is beyond the scope of this work as not all dyes are useful for flow cytometric fingerprinting. For more information consult Hammes *et al.* (2011), Strauber and Muller (2010), Tracy *et al.* (2010) and Shapiro (2003). Furthermore, we included novel data using red-excitable SYTO dyes combined with functional stains for double and triple staining applications.

2.1 Cell-permeant nucleic acids dyes

Nucleic acids carry the genetic information used in the development and functioning of all living organisms and viruses and could therefore be a very logical indicator of life**.** However, it's important to note that DNA can be persistent and that dead cells may still contain DNA. For example, when bacteria are killed by UV-C irradiation, lethal thymine dimers are produced that will be stained by most nucleic acid dyes (Hammes *et al.*, 2011). Nucleic acid stains can also act as good counterstains for labeling all organisms which are present, live or dead. Generally, nucleic acid stains can be divided in cell-permeant and cell-impermeant

dyes depending on their ability to pass through the cell membrane. The latter, being unable to cross intact membranes, can therefore be considered as viability-dependent dyes. Furthermore, the difference between the Gram-positive and Gram-negative bacterial cell wall structure poses additional difficulties (Berney *et al.*, 2007, Strauber and Muller, 2010, Shapiro, 2003) and permeabilization of the outer membrane of Gram-negative bacteria is often necessary to optimize staining. Either EDTA or citrate can be used for this purpose (Marie *et al.*, 1996). Both compounds permeabilize the outer membrane by chelating cations and stripping the LPS layer of the outer membrane (Chen *et al.*, 2004).These cations can also be deleterious as they decrease binding efficiency of certain dyes such as DAPI or Hoechst 33342 (Marie *et al.*, 1996). Here, we will focus on EDTA as previous results from Marie *et al.* (1996) showed that EDTA gave similar to those with citrate.

2.2 Functional dyes

Cell-impermeant nucleic acid stains such as propidium iodides (PI) are not able to cross membranes because of their size and charge, and are therefore used as an indicator of membrane permeabilization (Berney *et al.*, 2007). Since membrane integrity is vital to keep the intracellular environment stable, membrane damage can be an indication of cell death (Hammes *et al.*, 2011). However, it is known that contact time, incubation temperature and stain concentration are crucial factors for proper staining, which emphasizes the importance of standardization (Van Nevel *et al.*, 2013, Hammes *et al.*, 2012). Furthermore, a recent study on *E. coli* suggested that porins and periplasmic transporters induced by substrate limitation facilitate PI entry into cells and that staining efficiency is influenced by the physiological state (Brognaux *et al.*, 2014). Shi *et al.* (2007) showed that more bacteria were stained during early exponential phase than during the early lag phase. This indicates that PI cannot be used as viability estimator and suggests the necessity for multiparameter viability determination. Besides PI, other dyes with a similar mode of action are available such as SYTOX dyes, the TOTO and TO-PRO family of dyes (Shapiro, 2003).

Another aspect of viability is maintaining the cell's membrane potential. All healthy microbial cells need to keep their membrane potential, which is produced through a functional electron transport chain. The membrane potential also powers processes such as ATP synthesis and solute-ion transport. If the membrane potential decreases, the cell will be unable to transport essential molecules eventually leading to cell death. Membrane potential should be considered as a more conservative measurement of viability compared to membrane permeabilization because of the link between membrane potential and cell activity (Hammes *et al.*, 2011). $DiBAC_4(3)$, also known as bis oxonol (BOX), is mostly used to evaluate

membrane potential. It enters depolarized cells, because of its anionic structure and binds non-specifically to intracellular proteins (Muller and Nebe-von-Caron, 2010). In contrast, the cationic Rhodamine 123 (Rh123) only accumulates in cells with active membrane potential (Diaper *et al.*, 1992). Since this stain can be actively pumped out by certain cells it has a limited use in standardized protocols (Tracy *et al.*, 2010). Alternative dyes are 3,3'-dihexyloxacarbocyanine iodide (Di-OC_6(3)), 3,30- diethyloxacarbocyanine (Di-OC_2(3)) and 3,30-dipropylthiadicarbocyanine (DiSC_3(5)) (Shapiro, 2003). Various protocols exist for the use of DiBAC_4(3) with final stain concentrations, incubation time and temperatures ranging from 0.24 µM to 29 µM (Herrera *et al.*, 2002, Nielsen *et al.*, 2009), 2 to 20 minutes (Lopez-Amoros *et al.*, 1995, Comas and Vives-Rego, 1997, Rezaeinejad and Ivanov, 2011), and room temperature to 40°C (Rezaeinejad and Ivanov, 2011, Linhova *et al.*, 2012), respectively. **Table 2 - 1** provides an overview of staining protocols and their references.

Table 2 - 1 : *Overview of concentration, incubation time and temperature used in staining protocols for different dyes and organisms compared with our findings.*

Dye	Conc. [µM]	Incubation Time [min]	Incubation Temp. [°C]	Organism studied	Reference
DiBAC_4(3)	5	22-40	37	*C. metallidurans CH34, L. brevis LMG 18022*	This study
	1	4	37	*Bififobacterium adolescentis*	Amor *et al.* (2002)
	1	2	RT[a]	*E. coli*	Comas and Vives-Rego (1997)
	19.4	10	RT	*E. coli*, *P. aeruginosa, S. aureus*	Jepras *et al.* (1995)
	0.24	10	RT	*P. fluorescens, Pythium ultimum, Rhizoctonia solani*	Nielsen *et al.* (2009)
	0.5	15	37	*S. macedonicus*	Papadimitriou *et al.* (2007)
	1	20	37	*E. coli*	Rezaeinejad and Ivanov (2011)
	0.48	20	40	*L. delbrueckii* subsp. *bulgaricus*	Rault *et al.* (2008)
	1	2	RT	*E. coli*, *S.typhimurium*	Lopez-Amoros *et al.* (1995)
	1.94	7	RT	*C. pasteurianum*, *C. beijerinckii*	Linhova *et al.* (2012)
	29	10	RT	*E. coli*	Herrera *et al.* (2002)
	10	10	20	*E. coli*	Berney *et al.* (2009)
	3.87	10	RT	*Pichia pastoris*	Hyka *et al.* (2010)
cFDA	10	30-40	37	*C. metallidurans CH34, L. brevis LMG 18022*	This study
	50	10	30	*Lactobacillus plantarum*	Bunthof and Abee (2002)
	50	10	30	*Lactococcus lactis*	Bunthof *et al.* (1999)
	10	10-20-30	RT-30-40	Lake water bacteria	Porter *et al.* (1995)

	10	10	37	Lactic acid bacteria	Chen *et al.* (2012)
	10	30	37	*Bifidobacterium adolescentis*	Amor *et al.* (2002)
	2E-04	10	40	*L. delbrueckii* subsp. *bulgaricus*	Rault *et al.* (2008)
	0.002	10	40	*L. delbrueckii* subsp. *bulgaricus*	Rault *et al.* (2009)
	25	45	37	*S. macedonicus*	Papadimitriou *et al.* (2007)
	10.8	10	37	*Pichia pastoris*	Hyka *et al.* (2010)
	50	60	30	*B. cereus* endospores	Cronin and Wilkinson (2008a)
	10	30	37	activated sludge bacteria	Forster *et al.* (2002)
	21.7	10	RT	*C. pasteurianum*, *C. beijerinckii*	Linhova *et al.* (2012)
	10	15	35	*Aeromonas hydrophila, B. subtilis, E. coli, P. aeruginosa, S. epidermidis* and bacteria from environmental waters	Hoefel *et al.* (2003)
	21.7	30	37	*E. coli*, *P. aeruginosa, S. aureus*	Jepras *et al.* (1995)
	5	30	30	*L. lactis*	Hansen *et al.* (2015)
cFDA-SE	0.2	38-40	37	*C. metallidurans CH34, L. brevis LMG 18022*	This study
	0.045	10	40	*L. delbrueckii* subsp. *bulgaricus*	Rault *et al.* (2008)
	100	150	25-37	Strains from aquatic environment	Fuller *et al.* (2000)
	10	30	35	*Aeromonas hydrophila, B. subtilis, E .coli, P. aeruginosa, S. epidermidis* and bacteria from environmental waters	Hoefel *et al.* (2003)
	11	20	37	*S. pyogenes*	Hytonen *et al.* (2006)
	1	15	30	*Bifodobacterium*, *Lactobacillus*	Lahtinen *et al.* (2006)
	50	20	37	*L. casei*	Lee *et al.* (2004)
	10	30	37	*B. licheniformis*	Hornbaek *et al.* (2002)
	35	10	RT	*Lactobacillus plantarum*	Fitzgerald *et al.* (2004)
	1	10	30	*L. lactis*	Breeuwer *et al.* (1996)
	44.8	10	40	*Oenococcus oeni*	Bouix and Ghorbal (2015)
	50	30	30	*L. lactis*	Hansen *et al.* (2015)
	0.5	4	RT	*Cronobacter spp.*	Arku *et al.* (2011)
HE	5	40	37	*C. metallidurans CH34, L. brevis LMG 18022*	This study
	31.7	10	RT	*E. coli*	Herrera *et al.* (2002)

	10	0	RT	*Cupriavidus, shewanella, E. coli, deinococcus*	Baatout *et al.* (2005)
	0.07	4	RT	*Cronobacter spp.*	(Arku *et al.*, 2011)
Nile Red	0.13	10-40	37	*C. metallidurans CH34, L. brevis LMG 18022*	This study
	10,100	30	RT	*Synechocystis spp, E. coli*	Tyo *et al.* (2006)
	157	10-15	RT	*P. auroginosa*	Vidal-Mas *et al.* (2001)
	3.1	30	RT	*Escherichia coli*	Herrera *et al.* (2002)
	94.2	10	RT	*Ralstonia eutropha*	Gorenflo *et al.* (1999)
	60	10	RT	environmental bacteria	Koch *et al.* (2013c)
	0.126	30	RT	*Methylobacterium rhodesianum*	Ackermann *et al.* (1995)

[a] *RT: Room Temperature*

Besides membranes, other aspects of the cell can be used to assess viability or functionality. All bacteria possess housekeeping enzymes such as esterases or dehydrogenases that are linked to the respiratory activity of metabolically active cells. The inactivity of these enzymes indicates metabolic inactivity, but not necessarily cell death as these enzymes can still be active even after cell death has occurred. Measurement of cellular enzymatic activity is useful although the mentioned limitations needs to be considered. Generally, dyes used to monitor enzymatic activity are cleaved upon uptake in the cell, leading to the production of a fluorescent signal (Shapiro, 2003). Again, a wide variety of dyes that target different enzymatic activities are available. A popular dye is 5-cyano-2,3-ditolyl tetrazolium chloride (CTC), which is reduced by dehydrogenases to fluorescent membrane-impermeant formazan (Lopez-Amoros *et al.*, 1997). Since respiratory activity is linked to the maintenance of the membrane potential, CTC reduction and $DiBAC_4(3)$ diffusion are complementary (Hammes *et al.*, 2011). Fluorescein diacetate (FDA), another popular dye, is cleaved by esterases to release the fluorescent fluorescein. Since fluorescein easily leaks from cells, FDA modifications have been developed such as carboxyfluorescein diacetate (cFDA), with better retention kinetics, and modifications of cFDA to further reduce leakage such as carboxyfluorescein diacetate acetoxymethyl ester (cFDA-AM), 20,70-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein-AM (BCECF-AM), calcein-AM and carboxyfluorescein diacetate succinimidyl ester (cFDA-SE). Evidently to avoid leakage, cell permeabilization (e.g. by EDTA or citrate) to improve staining is not recommended with those dyes. Furthermore, permeabilization of the membranes can affect viability. An interesting feature of these dyes is that the fluorescence emission intensity of fluorescein depends on the pH, thereby giving additional information about cell metabolism. With a maximal emission at pH 9 and minimal emission at pH 5, changes around the neutral pH can be detected (Johnson and Spence,

2010). Again, many different protocols for cFDA and cFDA-SE have been reported (**Table 2 - 1)**.

Oxidative stress can also be assessed by flow cytometry. Reactive oxygen and nitrogen species (ROS and RNS) such as superoxide anion radical ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2) and the hydroxyl radical ($HO^{\bullet}$) are naturally occurring by-products of respiration and oxidation. To protect themselves from those toxic compounds, aerobic organisms use enzymes like superoxide dismutase (SOD) or non-enzymatic anti-oxidants like glutathione (GSH) to control the level of ROS. Environmental oxidizers, such as UV irradiation or chlorination can increase intracellular ROS levels leading to increased oxidative stress and eventually to cell death. To measure the increase of ROS, dihydroethidium (hydroethidine; HE) can be used. The oxidation of HE results in the formation of ethidium, a fluorescent compound that intercalates DNA (ex./em. 520/610 nm) (Munzel *et al.*, 2002). Research has shown that not only $O_2^{\bullet-}$ but also cytochrome c and other reactive oxygen and nitrogen species can oxidize HE (Tarpey *et al.*, 2004). The relative reactivity of the different components are $ONOO^-$ > $Fe(II)/H_2O_2$ (*i.e.* $HO^{\bullet}$) > $O_2^{\bullet-}$ > H_2O_2 (Murrant and Reid, 2001) and show that HE provides an indication of both ROS and RNS production (Gomes *et al.*, 2005). Literature on the use of HE for assessing oxidative stress in bacteria by flow cytometry is scarce and staining protocols used concentrations between 0.07 and 31 µM (Arku *et al.*, 2011, Herrera *et al.*, 2002) and incubation times between zero and ten minutes at room temperature (Baatout *et al.*, 2005, Herrera *et al.*, 2002) (**Table 2 - 1**).

Lipid composition is an interesting parameter to investigate cell physiology as it reveals more about the cells energy storage. Lipid stains can roughly be divided into two groups: lipid analogues and lipophilic organic molecules. The BODIPY-labeled fatty acid analogues are often used for mammalian cells and microscopy, their use in flow cytometry applications is rather rare (Benincasa *et al.*, 2009, Papadimitriou *et al.*, 2007). An example of the second class is nile red (NR), which binds selectively to non-polar lipid droplets inside cells (Johnson and Spence, 2010) and can be used to detect the presence of storage lipids (PHA/PHB) in spectrophotometry (Greenspan and Fowler, 1985) and flow cytometry (Tyo *et al.*, 2006, Vidal-Mas *et al.*, 2001, Ackermann *et al.*, 1995, Gorenflo *et al.*, 1999, Herrera *et al.*, 2002). Tyo *et al.* (2006) showed that ionic strength of the dilution buffer influenced staining efficiency and recommended to use deionized water as dilution buffer instead of physiological saline solution (0.9% NaCl) to improve signal to noise ratio. The authors also mentioned the need for membrane permeabilization for specific bacterial species, which reduced viability. Interestingly, they also showed that the optimal concentration of NR is species dependent, potentially because of differences in PHA contents. Protocols vary with

concentrations between 3 and 100 µM (Tyo *et al.*, 2006, Herrera *et al.*, 2002) and incubation times ranging from 10 minutes to 30 minutes at room temperature (Gorenflo *et al.*, 1999, Tyo *et al.*, 2006) (**Table 2 - 1**).

2.3 Stain combinations

Combining different stains offers the possibility to simultaneously assess different physiological states of bacteria within a population, thereby improving understanding of bacterial behavior within a specific condition (Nebe-von-Caron *et al.*, 2000, Nielsen *et al.*, 2009, Rezaeinejad and Ivanov, 2011). Several studies used and described double and triple staining protocols in order to determine different functional properties of a bacterial community (Hewitt *et al.*, 1999, Johnson and Spence, 2010). A widely-used stain combination is available in the commercialized Live/Dead BacLight kit (Thermo Fisher Scientific), which uses a combination of SYTO 9 and PI to distinguish intact 'live' cells from permeabilized 'dead' cells. This kit has been used in numerous studies (Vriezen *et al.*, 2012, Mah *et al.*, 2003, Leys *et al.*, 2009, Lawrence *et al.*, 1998, Dalwai *et al.*, 2006, Alonso *et al.*, 2002). **Table 2 - 2** gives an overview of genuine double and triple staining protocols and their applications. Apart from these combinations, many multicolor protocols have been optimized in which dyes are separately added to different technical replicates of the same sample. This is different from a genuine multicolor set-up in which dyes are added to the sample simultaneously.

To combine different dyes, it is important to choose dyes that possess the right spectral properties, to determine the incubation time, incubation conditions and dye concentration for each fluorescent probe separately, and to assess possible interference. When combining stains, one of the most common issues is overspill which is a consequence of the spectral characteristics of the dyes. To resolve overspill, other dye combinations can be made or compensation can be applied. A second type of interference, also related to the spectral characteristics of the dyes, is FRET (fluorescence resonance electron transfer). In this case the emission of one dye (donor) is absorbed by a second dye (acceptor) in close proximity. Consequently, the fluorescence intensity of the donor decreases (quenching) and the fluorescence intensity of the acceptor increases (Horvath *et al.*, 2005). Besides FRET, also the matrix can cause fluorescence quenching. Each dye combination behaves differently within a certain combination of medium matrix, microorganism or instrument and needs to be tuned to determine possible compensations and obtain reproducible results (Hyka *et al.*, 2010, Tracy *et al.*, 2010).

Table 2 - 2: *Overview of double and triple combination of stains, set-up of combinations and subject in which they were used.*

Stains	Set-up	Subject	Reference
Hexidium iodide/cFDA	Double stain	Physiological characterization of *Escherichia coli* during fermentations	Hewitt *et al.* (1999)
cFDA/PI/$DiBAC_4(3)$	Double cFDA/PI Single $DiBAC_4(3)$	Bile salt stress on bifidobacteria population	Amor *et al.* (2002)
cFDA/PI	Double stain	Viability assessment of *Lactococcus lactis* exposed to different stresses	Bunthof *et al.* (1999)
cFDA/PI	Double stain	Gastro-intestinal stress on probiotic bacteria	Chen *et al.* (2012)
Rh123/ $DiBAC_4(3)$/PI/ SYTO13/SYTO17	Double PI/SYTO13 Single other stains	Effect of gramicidin, formaldehyde, and surfactants on *Escherichia coli*	Comas and Vives-Rego (1997)
SYTO9/PI/cFDA/ Hoechst3342/resazurine/ SYTOX green	Double SYTO9/PI Single other stains	Growth phase-related changes in *Bacillus cereus*	Cronin and Wilkinson (2008b)
Hexidium iodide/cFDA	Double stain	Gram staining of bacteria in activated sludge	Forster *et al.* (2002)
PI/$DIOC_6$	Double stain	Starvation response in *Bacillus licheniformis*	da Silva *et al.* (2005)
Rh123/$DiBAC_4(3)$/PI	Double Rh123/PI and $DiBAC_4(3)$/PI	Starvation survival in seawater of *Escherichia coli* and *Salmonella typhimurium*	Lopez-Amoros *et al.* (1995)
Thiazoleorange(TO)/PI/ $DiBAC_4(3)$	Double TO/PI Single $DiBAC_4(3)$	Characterization of bacteria in microbial fuel cells	Matos and Lopes da Silva (2013)
SYBR green/EB/ $DiBAC_4(3)$/PI	Double SYBR green/PI Triple SYBR green/EB/$DiBAC_4(3)$	*Pseudomonas fluorescens* DR54 biocontrol	Nielsen *et al.* (2009)
PI/cFDA/ $DiBAC_4(3)$	Double PI/cFDA Single $DiBAC_4(3)$	Acid stress response of *Streptococcus macedonicus*	Papadimitriou *et al.* (2007)
CTC/fluorescent-antibody/DAPI	Double CTC/fluorescent-antibody Single DAPI	Enumeration of enterohemorrhagic *Escherichia coli* O157:H7	Pyle *et al.* (1995)
cFDA/cFDA-SE/PI/ $DiBAC_4(3)$	Double cFDA/PI Single $DiBAC_4(3)$ and cFDA-SE	Fermentation of *Lactobacillus delbrueckii*	Rault *et al.* (2008)
cFDA/PI	Double	Viability of *Escherichia coli* O157:H7 in river water	Tanaka *et al.* (2000)
cFDA/PI/CTC	Double cFDA/PI Single CTC	Effect of pollution on the activity of microbial communities in river water	Yamaguchi and Nasu (1997)

2.4 Stability

An advantage of flow cytometry is the possibility to analyze samples in a high throughput manner. The use of multiwell autoloaders has become very common and samples are often stained and incubated at the same time. However, this can lead to discrepancies as samples are not measured simultaneously. In that time frame, biological changes such as aggregation or physiological and chemical changes such as bleaching, dye extrusion or intrusion can occur and significantly alter the results and subsequently affect the reliability and reproducibility (Hyka *et al.*, 2010). Dye stability is therefore an important factor to consider, particularly when measuring in high throughput screening mode. Only Hammes *et al.* (2012) and Van Nevel *et al.* (2013) discussed this issue for the double SYBR green and propidium iodide staining. To assess the stability of the dye, Van Nevel *et al.* (2013) aliquoted a known sample in a multiwell plate after staining with their optimized staining protocol and monitored the mean fluorescence intensity and cell concentration of the detected populations. This way the gradual changes in the results could be detected. They showed that SYBR green I staining is stable for at least 74 minutes, making this stain suitable for multiwell plates. On the other hand, they demonstrated that the combination of SYBR green and propidium iodide is less stable and that, albeit the number of cells in each population remained stable, the fluorescence intensity changed over time. Thus, care should be taken when many samples are analyzed in batch and when a fixed gating template (Prest *et al.*, 2013) or flow cytometric fingerprinting (De Roy *et al.*, 2012) is used to analyze the data. The authors also clearly mention that the dye stability is different when other dye concentrations or samples are tested and illustrate the usability of such tests to develop high throughput assays.

3 Materials and methods

3.1 Preparation of bacteria cultures

Two bacterial strains were used. The Gram-positive *Lactobacillus brevis* LMG 18022 and the Gram-negative *Cupriavidus metallidurans* CH34. Both were grown overnight at 28°C in aerobic conditions in liquid media, MRS broth (Oxoid) and LB broth (Oxoid), respectively. As example of a mixed microbial culture, bottled Evian water was used. Bacterial cultures were diluted in phosphate buffered saline (PBS) solution and adjusted to a concentration around 10^5 cells/mL prior analysis based on flow cytometric cell enumeration (Prest *et al.*, 2013).

3.2 Flow cytometry and used stains

Bacteria were analyzed on a Accuri C6 (BD Biosciences) with a blue (488 nm, 20 mW) and red (640 nm, 14.7 mW) laser. Standard optical filters were used and included FL-1 (530/30 nm), FL-2 (585/40 nm) and FL-3 (670 LP) for the blue laser and FL-4 (675/25 nm) for the red laser. The tested dyes include $DiBAC_4(3)$ (Sigma Aldrich), cFDA (Sigma Aldrich) and cFDA-SE (Sigma Aldrich) and were detected with FL-1. The dyes dihydroethidium (Sigma Aldrich) and Nile red (Sigma Aldrich) were detected with FL-3 and the red-excitable SYTO dyes 17, 59-64 (Thermo Fisher) were detected with FL-4. A daily quality control with fluorescent beads (05-4008, Sysmex) and a cleaning cycle were performed before the experiments to assess both the accuracy (bead count and position) and the cleanliness of the machine. Samples were analyzed in the Accuri C6 software (version 1.0.264.21) unless stated otherwise.

3.3 Assessment of staining kinetics

Staining kinetics were assessed by staining 1.5 mL of a 10^5 cells/mL bacterial suspension (as described above) and measuring continuously for maximum 40 min (or one hour for HE as stabilization of the signal took longer) using the instrument's "unlimited run" function. The fluidics speed was set at low flow rates (16 µL/min) to keep a low number of events per second in all channels (< 2000). During acquisition, the sample was incubated at 37°C. For the red-excitable SYTO dyes, time-gates were made of one minute using the *flowCore* package v1.38.1 in R. For each gate the average fluorescence intensity of the cell population was extracted. Cell concentrations were not extracted in this experiment as they were not considered reliable for the given set-up.

3.4 Assessment of optimal conditions for the single stains

The concentration, incubation time and the effect of EDTA were tested on all SYTO stains and on both bacteria. Cell suspensions were prepared as described above. Three dye concentrations (5 µM, 0.5 µM and 0.05 µM) were tested at different incubation times (0 min, 15 min and 30 min) both with and without EDTA. The final concentration of EDTA was 5 mM. To determine if dyes could be used as benchmark, cell concentrations with the SYTO dyes were compared to the standard SYBR green I staining as described by Prest *et al.* (2013). For all functional dyes, one single concentration was used based on literature and titration. The optimal stain concentration was chosen based on maximized signal to background distinction (**Appendix Figure 2 - 1**). The final concentrations chosen were 5 µM, 1 µM, 0.2 µM, 10 µM and 0.3 µM for dihydroethidium, $DiBAC_4(3)$, cFDA-SE, cFDA and Nile Red respectively. For all SYTO stains a sample filtered over a 0.22 µm filter (Merck Millipore) was

used as a negative control. A sample was heat-killed by incubation at 100°C for 30 min to serve as positive control sample for $DiBAC_4(3)$ and as a negative control sample for cFDA and cFDA-SE. An oxidized sample was prepared by adding 10 µl of 30% H_2O_2 (Sigma Aldrich) to 500 µl of sample. After 30 min of incubation, the suspension was centrifuged and washed three times with 0.22 µm-filtered PBS to remove the H_2O_2 as to avoid bleaching of the stains. This sample was used as positive control for dihyroethidium staining.

3.5 Assessment of triple staining

Following triple stains were tested $DiBAC_4(3)$/ dihydroethidium/ SYTO 60, $DiBAC_4(3)$/ Nile red/ SYTO 60, cFDA/ dihydroethidium/ SYTO 60, cFDA/ Nile Red/ SYTO 60, cFDA-SE/ dihydroethidium/ SYTO 60, cFDA-SE/ Nile Red/ SYTO 60. For each triple stain, all double stain combinations were tested and all single stains as well. The stains were tested on both bacteria (prepared as described above) and on all controls described previously. Finally, a mixture of different samples was made to have a positive and negative signal for each stain (**Appendix Table 2 - 1**). All samples were incubated at 37°C for 30 min. After flow cytometry read out, a gating protocol was established. For the double stains with SYTO 60, first both FL-1 positive and FL-3 positive cells were gated versus FL-4. Then both positive and negative FL-1 and FL-3 populations were gated in the opposite channel (*e.g.* the FL-3 positive population was gated in the FL-1 channel) to establish the necessary compensation. After the right gates and compensation were established, 4 populations could be distinguished with the triple stains: $FL\text{-}1^+/FL\text{-}3^+$, $FL\text{-}1^-/FL\text{-}3^+$, $FL\text{-}1^+/FL\text{-}3^-$, $FL\text{-}1^-/FL\text{-}3^-$. The gating protocol is further illustrated in **Figure 2 - 1**.

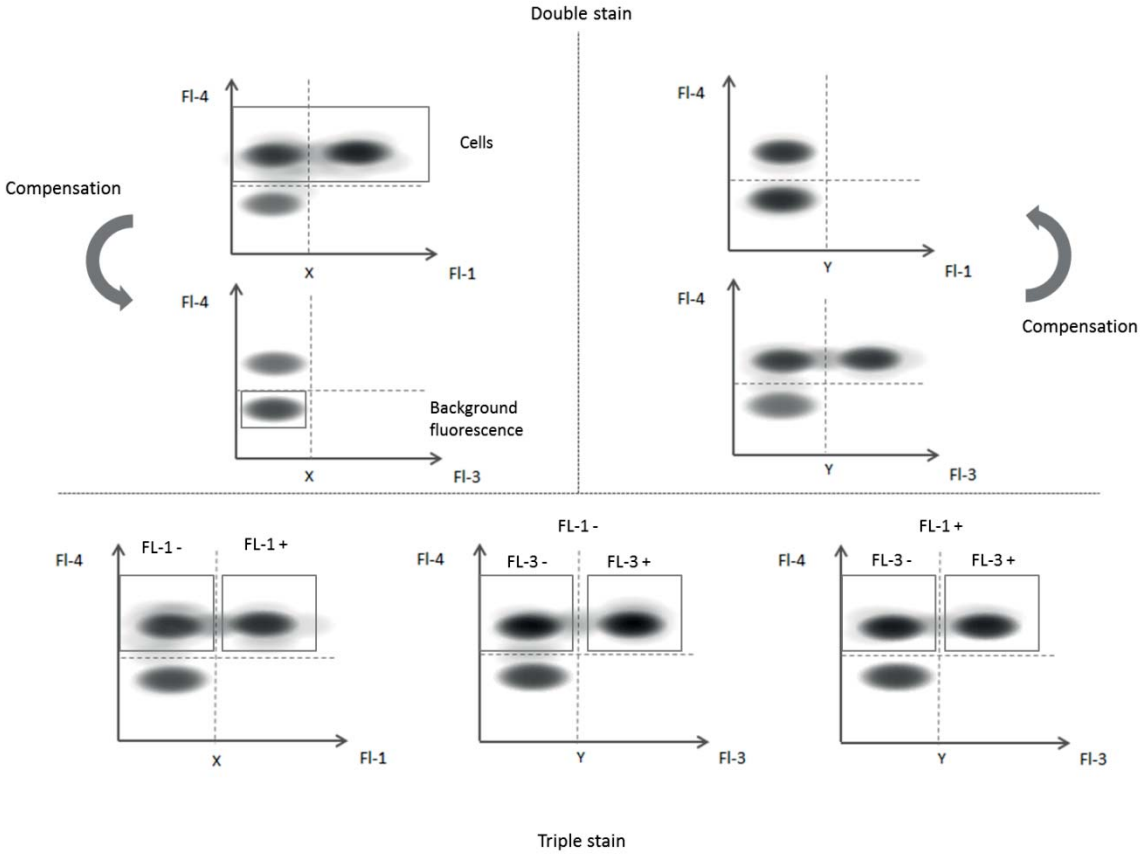


Figure 2 - 1: *Illustration of the gating protocol used to discriminate the four populations when using triple stains. The red-excitable nucleic acid stain SYTO 60 was used as a counterstain to differentiate cells from background. For the double stains with SYTO 60, first both FL-1 positive (**upper left**) and FL-3 positive cells were gated versus FL-4 (**upper right**). Then both positive and negative FL-1 and FL-3 populations were gated in the opposite channel (e.g. the FL-3 positive population was gated in the FL-1 channel) to establish the necessary compensation. After the right gates and compensation were established, four populations could be distinguished with the triple stains: FL-1^{+}/FL-3^{+}, FL-1^{-}/FL-3^{+}, FL-1^{+}/FL-3^{-},and FL-1^{-}/FL-3^{-}.*

3.6 Stability

A mixed microbial community with different cell populations (*e.g.* dead, live or oxidized were prepared of either *L. brevis* LMG 18022 or *C. metallidurans* CH34 and stained with any of the triple combinations mentioned in the previous section. The samples were then subsampled and loaded in a 96-well plate and measured sequentially with the flow cytometer after 30 min of incubation at 37°C. Cell concentrations were then determined by applying the gating protocol mentioned in the previous section allowing to look at the stability of these concentrations over the time needed to process the complete 96-well plate (about 1 h 30 min).

4 Results and Discussion

To further scrutinize the usability of SYTO dyes, we investigated the possible application of the red-excitable SYTO dyes 17, 59, 60, 61, 62, 63 and 64 for both the Gram-negative bacterium *Cupriavidus metallidurans* CH34 and the Gram-positive bacterium *Lactobacillus brevis* LMG 18022. First, we tested the staining kinetics by measuring the change in fluorescence over time. Results indicated that dye uptake was immediate for the Gram-positive bacterium while an incubation period of approximately 15 minutes was required for the Gram-negative bacterium to reach maximum fluorescence intensity (**Appendix Figure 2 - 2** and **Appendix Figure 2 - 3**). For the Gram-positive population, fluorescence intensities were stable, indicating time-independent uptake. For the Gram-negative population, fluorescence intensities increased over time showing time-dependent uptake. After uptake, the fluorescence signal remained stable during the entire measurement period (30 minutes). This difference in uptake between Gram-positive and Gram-negative bacteria is likely due to differences in cell membrane composition and is also observed for blue-excitable SYTO dyes Lebaron *et al.* (1998). A second test was performed where different dye concentrations and the effect of EDTA were assessed for all dyes. This test was performed because cell concentrations can vary depending on the used dye concentrations and incubation time regardless of the fluorescence intensity of the cell population. Thus, different dye concentrations and the effect of EDTA were assessed at different time points during incubation (0 min, 15 min and 30 min) for all red-excitable SYTO dyes. As a benchmark, cell concentrations were compared to the standard SYBR green I staining as described by Prest *et al.* (2013). Results of this experiment for SYTO 60 and the SYBR green bench mark are shown in **Figure 2 - 2**, the results of the other dyes are available in the publication of Buysschaert *et al.* (2016). A stain was assessed as good, if cell concentrations did not deviate more than 10% from the SYBR green benchmark. Our results showed that a final stain concentration of 0.5 µM is preferable for all red-excitable SYTO dyes and that an incubation period between 15 to 30 minutes, depending on the organism, is sufficient to obtain a reliable estimate of cell numbers (data not shown). Only SYTO 64 was unable to stain bacteria under these conditions, as no cells could be detected with the flow cytometer (data not shown). This optimal stain concentration is comparable to the concentration found by Comas and Vives-Rego (1997) who tested SYTO 17 at 1 µM on *Escherichia coli*. However, our protocol relied on a shorter incubation time of 15 minutes instead of 60 minutes. The difference in incubation temperature could facilitate dye intrusion by diffusion and could explain the difference (Johnson and Spence, 2010). As cell concentrations remained stable after the minimal incubation time it indicates that the dyes are not pumped out of the cells. In case no bleaching occurs, longer incubation times can also be used. The

addition of 5 µM EDTA did not improve staining efficiency nor signal intensity, except when lower dye concentrations (0.05 µM) were used for the Gram-negative bacterium.

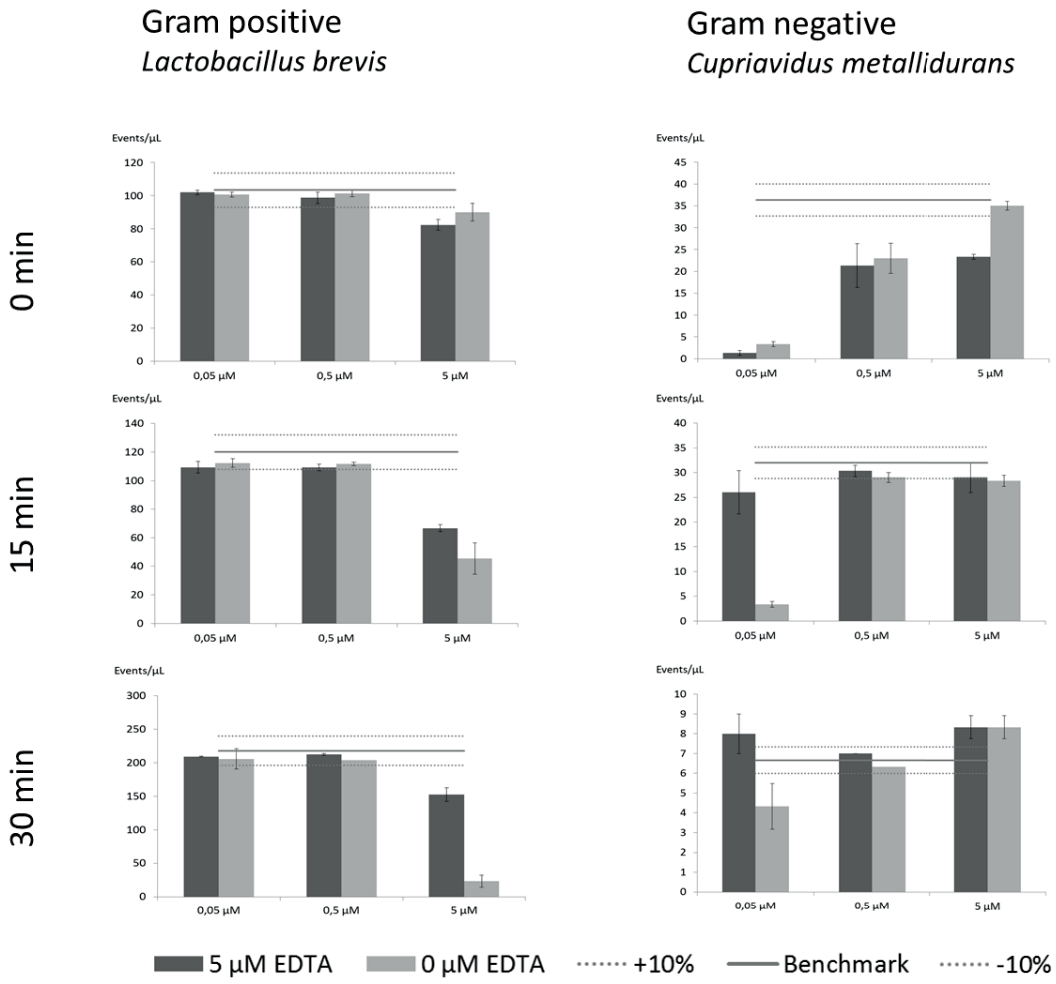


Figure 2 - 2: *Optimization of the red-excitable dye SYTO 60 for both the Gram-positive* Lactobacillus brevis *LMG 18022 (**left**) and Gram-negative* Cupriavidus metallidurans *CH34 (**right**) on three different time points: 0 min (**top**), 15 min (**middle**) and 30 min (**bottom**). Staining was performed with three different stain concentrations (5 µM, 0.5 µM and 0.05 µM) with and without 5 µM EDTA. All samples were measured in triplicate. Cell concentrations are expressed as events/µL and should be compared to the results obtained with SYBR green I staining as benchmark. A maximum 10% deviation on the SYBR green I results was accepted.*

For the different functional dyes, staining kinetics were assessed in the same way as the nucleic acid dyes. Most dyes showed a time-dependent dye uptake but after fluorescence intensity maximized, the fluorescence intensity remained stable until the end of the measurement (40 min). In contrast to the other functional dyes tested, HE showed a time-independent uptake in the Gram-positive population (**Appendix Figure 2 - 4**). For $DiBAC_4(3)$, the minimal incubation time was 22 minutes regardless of the type of bacteria. For the other functional dyes, a difference could be noticed between the Gram-positive and Gram-negative

bacteria. Incubation times for the Gram-positive *L. brevis* LMG 18022 are 21, 38, 24, and 10 minutes for cFDA, cFDA-SE, HE, and NR, respectively. For the Gram-negative *C. metallidurans* CH34, the incubation times are 30, 26, 30, and 5 minutes for cFDA, cFDA-SE, HE, and NR, respectively (**Appendix Figure 2 - 4** and **Appendix Figure 2 - 5**). The optimal stain concentration was chosen based on maximized signal to background distinction (**Appendix Figure 2 - 1**). A concentration of 1, 10, 1, 5, and 0.63 µM was chosen for $DiBAC_4(3)$, cFDA, cFDA-SE, HE, and NR, respectively. The use of EDTA was evaluated for all functional stains, except for $DiBAC_4(3)$ as EDTA can alter membrane permeability and thus impact membrane potential. Results showed an important increase in background fluorescence because of leakage for cFDA and cFDA-SE. For HE and NR, the use of EDTA did not improve staining significantly (data not shown). The addition of EDTA with any functional dye is therefore not recommended as it will impact the measurement. The fluorescence stability and the similar incubation conditions enable the combination of different dyes with seemingly incompatible staining protocols despite the different minimal incubation times.

To test the feasibility of 'true' triple staining protocols, stain combinations were chosen on the basis of spectral characteristics. A green fluorescent stain ($DiBAC_4(3)$, cFDA ,and cFDA-SE) was combined with an orange fluorescent stain (HE, NR) and a red-fluorescent counterstain (SYTO 60). Several important parameters such as distinguishable populations, total cell concentration, interference between stains and overspill in other channels were evaluated and a gating protocol was established. This has been further illustrated in **Figure 2 - 1**. Four different *C. metallidurans* or *L. brevis* suspensions were made (*i.e.* a heat-killed, a peroxide-exposed, a stationary phase culture, and a mixed population), which were subsequently stained either with a single stain or all combinations of double and triple stains. This approach was necessary to determine the appropriate compensation, thresholds, and gating. For all triple stain combinations, four populations could be identified with flow cytometry. The threshold and the number of events detected were always affected when stains were combined because of the increased background to signal ratio as a result of spectral overlap and compensation. In addition, compensation was necessary for all green and orange fluorescent dyes, as all caused overspill in the other fluorescent channels. $DiBAC_4(3)$ caused the most spillover and a compensation of more than 100% was necessary making this dye unsuitable in combination with NR and HE. Both cFDA and cFDA-SE showed to be more suitable for triple staining. Combinations with NR required a 25% and 15% compensation in combination with cFDA and cFDA-SE, respectively. This is slightly higher than for HE, which required a 9% and 13% compensation in combination with cFDA and cFDA-SE, respectively. SYTO 60 did not require compensation and allowed to distinguish cells from the background

as well as correcting for false positive and false negative events. Because of this, any combination of a functional dye was possible with SYTO 60, resulting in more reliable population counts.

We also investigated the stability of the previously described triple staining protocols. A synthetic microbial community with different cell populations (*e.g.* dead, live or oxidized cells) was aliquoted in a 96-well plate, stained and incubated as previously described. Stability was assessed by comparing the number of cells in the pre-established gates of each sample. As all samples were measured consecutively and for a fixed time (one minute), the stability in time could be determined. For few triple stains a satisfying degree of stability (10% deviation in the cell counts per gate) for the different populations was found based on the analysis of a 96-well plate (**Appendix Table 2 - 2**). While $DiBAC_4(3)$ is difficult to combine with both NR and HE, only combinations with cFDA and cFDA-SE are considered. As expected, both carboxyfluorescein stains impair the stability of triple stains due to the previously mentioned leakage and a maximum stability of 15 minutes was measured, making these triple stain combinations unsuitable for high throughput screening tests requiring more than 15 min analysis time. Similar to the results found by Van Nevel *et al.* (2013), stability can be affected or even improved when other dye combinations or other dye concentrations are used.

5 Conclusions and perspectives

Since the introduction of flow cytometry for research, it has been used to analyze cell populations through high-throughput single-cell analysis. Numerous staining protocols have been developed for many applications. However, knowledge about the influence of the different methodological factors on the measurement, and its subsequent interpretation, is still lacking. This is important as they impact the reliability and reproducibility of a staining protocol. Few genuine multicolor protocols have been developed for microbiology and even fewer give an explanation why a certain concentration, incubation time, and temperature were chosen or how stable the added fluorochromes were. Furthermore, few results have been published regarding the performance of multicolor protocols in multiwell assays.

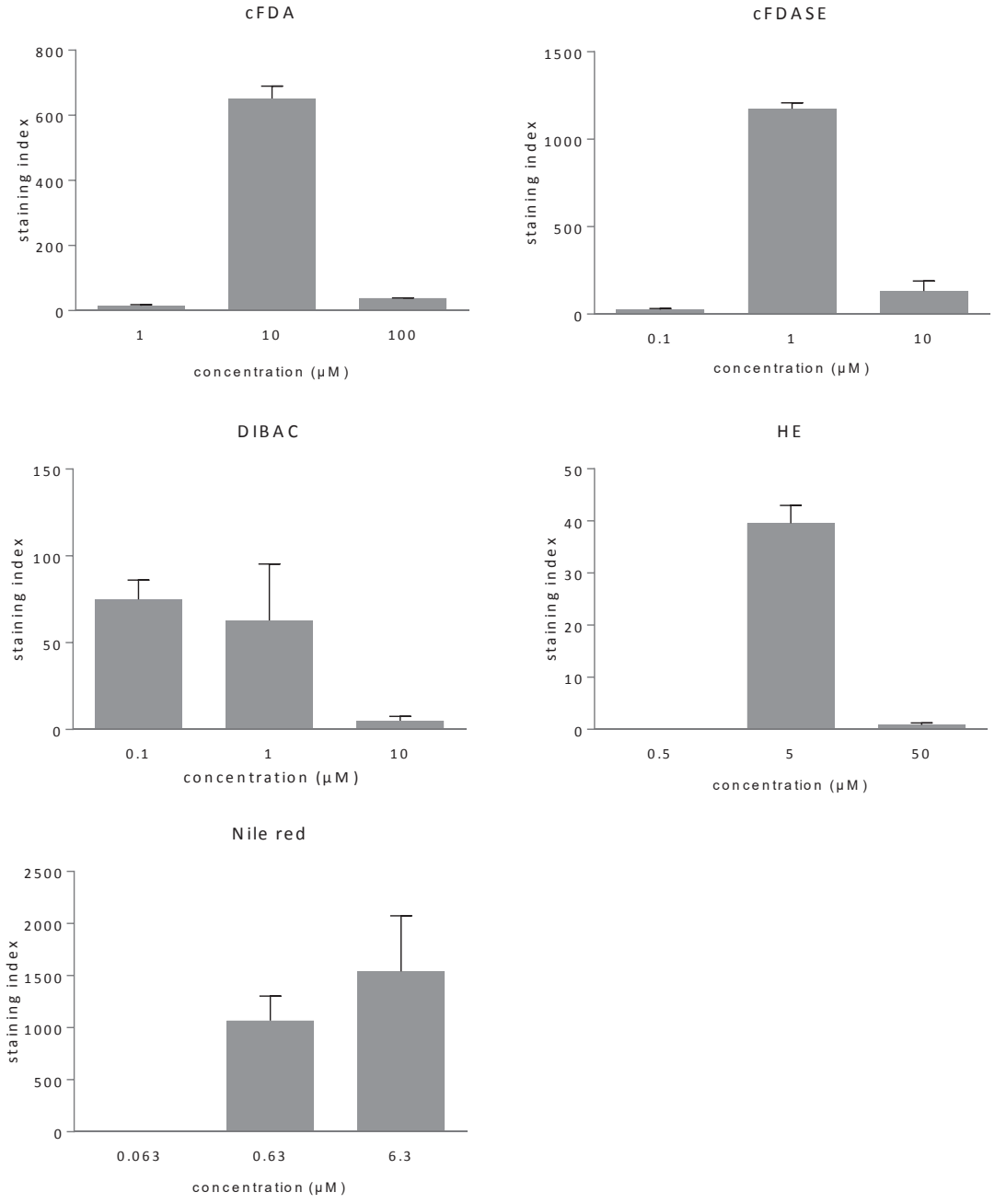
In this chapter, we demonstrated how red-excitable SYTO dyes can be integrated as a counterstain for multicolor protocols. Moreover, we tested some functional dyes and showed that, like the SYTO dyes, the efficiency differed between the tested organisms, confirming that optimization is necessary for accurate functional measurements. For all tested dyes we showed that their fluorescence intensity remained stable after fluorescence intensity maximized, offering the possibility to combine dyes with different minimal incubation times. However, we found that the combined dyes influence the results and that appropriate

controls and compensation are crucial for a correct analysis. $DiBAC_4(3)$ was difficult to combine as it produced a lot of background fluorescence and spillover while cFDA and cFDA-SE were more suitable for combinations. Being excited by a red laser, SYTO 60 did not create spillover and was thus easily combined with other dyes. In general, the addition of a nucleic acid stain improved the results as more background fluorescence could be filtered out. Another parameter that influences stain performance is the dye stability during analysis, as the last sample of a batch analysis will not be analyzed at the same time as the first sample and may have undergone changes in biology (*e.g.* aggregation, sedimentation behavior, physiological changes) as well as in staining chemistry (*e.g.* bleaching or leakage). Both literature and our preliminary results clearly show that stability is important to ensure correct results. It can be concluded that the simultaneous discrimination of certain physiological states is possible when protocols and staining conditions are optimized and appropriate compensation is set. The addition of a red excited nucleic acid dye as counterstain reduces the background to signal ratio and improves the separation between the positive and negative populations.

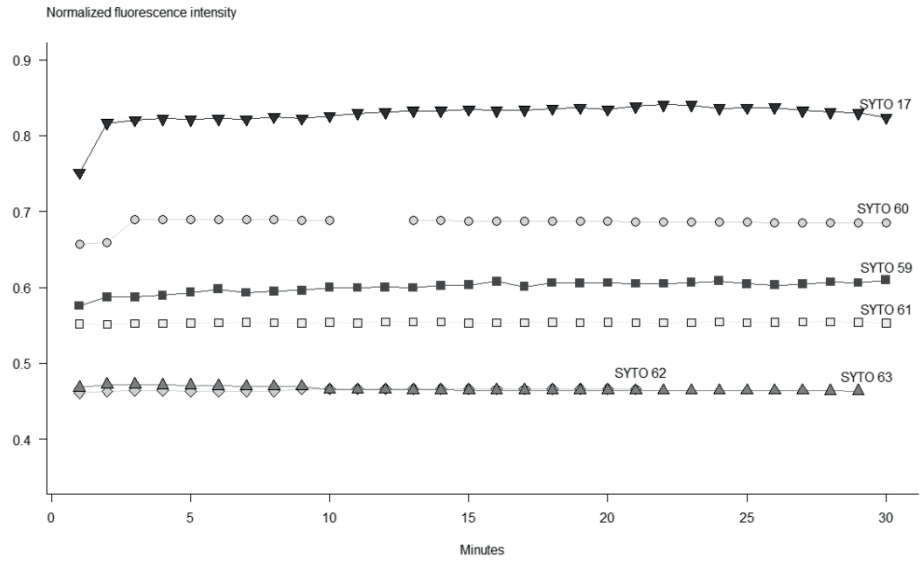
6 Acknowledgements and author contributions

This work was supported by the project grant SB-131370 of the Agency for Innovation by Science and Technology (IWT Flanders) (Benjamin Buysschaert), the European Space Agency (ESA-PRODEX), Belgian Science Policy (Belspo) through the E-GEM/BIOROCK project (Bo Byloos) and the Inter-University At- traction Pole (IUAP) “µ-manager” funded by the Belgian Science Policy (BE, 305 P7/25). The authors want to thank An-Sofie Lerno for the help in the lab. B. Byloos assessed the functional dyes while B. Buysschaert assessed the red-nucleic acid dyes. Combinations were tested by both authors. R.V.H., N.L. and N.B. helped with data interpretation, scientific guidance and preparation of the manuscript

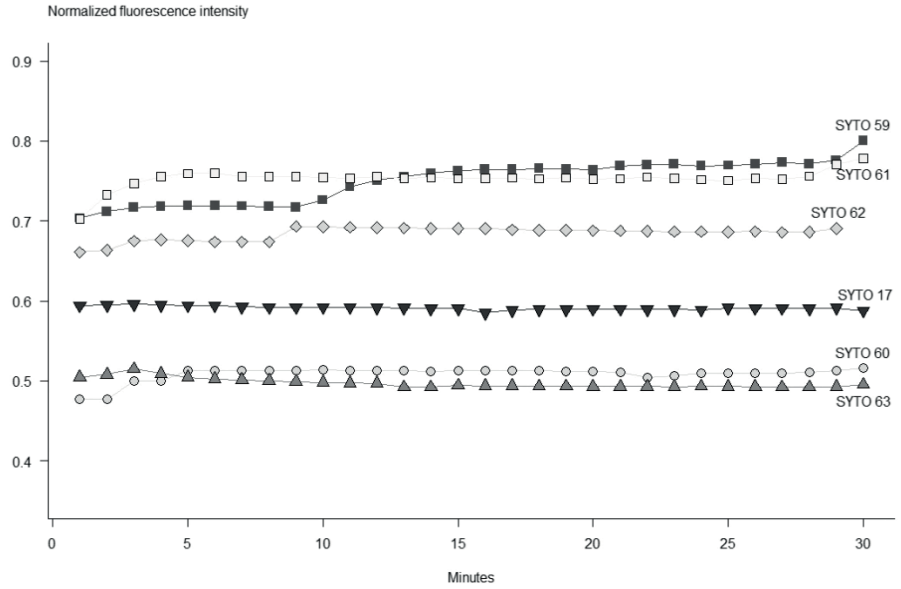
7 Appendix – Supplementary information for Chapter 2



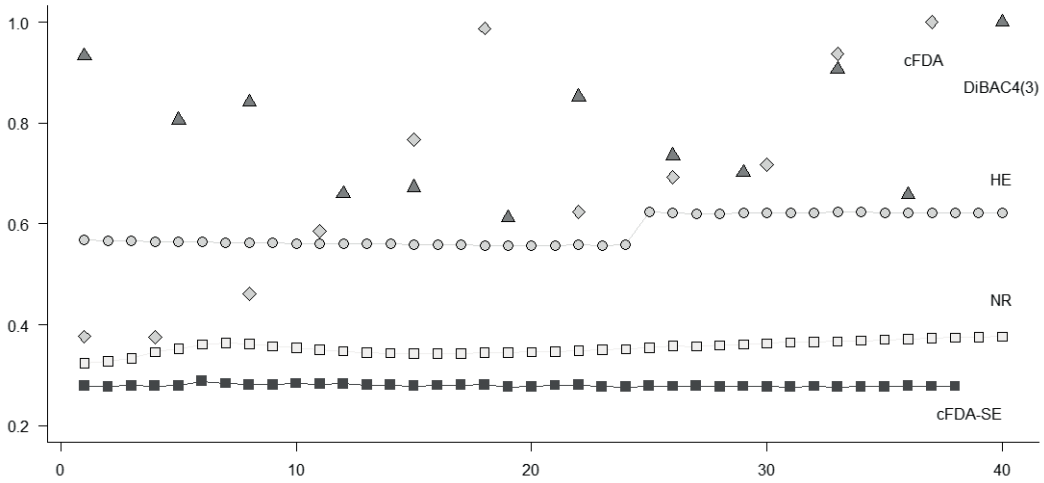
***Appendix Figure 2 - 1**: Staining index of the different functional stains tested. Staining index was defined by Maecker et al. (2004) and is defined as the relative difference between the background and the fluorescent population. Lower concentrations showed the same mean fluorescence positive cell population but higher variance and lesser cells detected while the higher concentration shifted both background and the positive cell population higher and induced more background signal.*



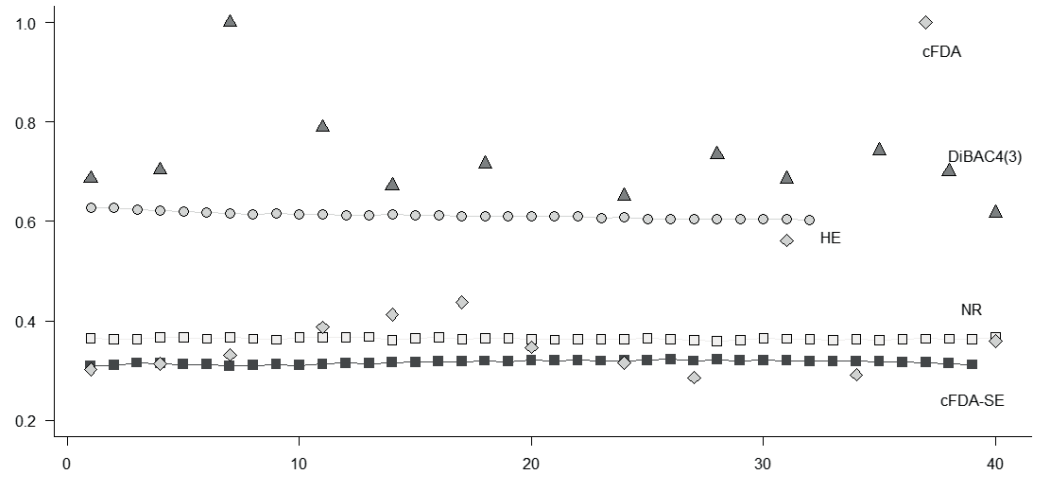
Appendix Figure 2 - 2: *Evolution of fluorescence intensity over time during incubation at 37°C for the different SYTO dyes with the Gram-positive bacteria,* Lactobacillus brevis *LMG 18022 cell suspension of 10^5 cells/ml. For each of the SYTO stains 0.5 µM was used. SYTO 64 was not included as no cells could be detected at any point in time. Intensity was normalized. The first time point (t=0) was not included due to poor quality of the data. The last time points of SYTO 62 and a few time points of SYTO 60 were also removed due to improper acquisition. The cell concentration was not calculated for this experiment as results can be unreliable for this set-up.*



Appendix Figure 2 - 3*: Evolution of fluorescence intensity over time during incubation at 37°C for the different SYTO dyes with the Gram-negative bacteria* Cupriavidus metallidurans *strain CH34 cell suspension of 10^5 cells/ml. For each of the SYTO stains 0.5 µM was used. SYTO 64 was not included as no cells could be detected at any point in time. Intensity was normalized. The first time point (t=0) was not included due to poor quality of the data. The last time point of SYTO 62 was also removed due to improper acquisition. The cell concentration was not calculated for this experiment as results can be unreliable for this set-up.*



Appendix Figure 2 - 4*: Evolution of fluorescence intensity over time during incubation at 37°C for the different functional dyes with the Gram-positive bacteria,* Lactobacillus brevis *LMG 18022 cell suspension of 10^5 cells/ml. Concentrations used are final concentrations 5 µM, 1 µM, 0.2 µM, 10 µM and 0.3 µM for HE, $DiBAC_4(3)$, cFDA-SE, cFDA, and NR respectively. Intensity was normalized. The first time point (t=0) was not included due to poor quality of the data. The cell concentration was not calculated for this experiment as results can be unreliable for this set-up.*



Appendix Figure 2 - 5*: Evolution of fluorescence intensity over time during incubation at 37°C for the different functional dyes with the Gram-negative bacteria* Cupriavidus metallidurans *strain CH34 cell suspension of 10^5 cells/ml. Concentrations used are final concentrations 5 µM, 1 µM, 0.2 µM, 10 µM and 0.3 µM for HE, $DiBAC_4(3)$, cFDA-SE, cFDA, and NR respectively. Intensity was normalized. The first time point (t=0) was not included due to poor quality of the data. The cell concentration was not calculated for this experiment as results can be unreliable for this set-up.*

Appendix Table 2 - 1: *overview of the samples used as positive and negative controls. To make ensure a positive and a negative population for all dyes when triple stains were applied, a mix of different samples was made.*

	Positive control	**Negative control**
cFDA	Live cellular suspension	Heat-killed sample
cFDA-SE	Live cellular suspension	Heat-killed sample
$DiBAC_4(3)$	Heat-killed sample	Live cellular suspension
NR	Stationary phase *C. metallidurans* culture	Exponential phase *L. brevis*
HE	Peroxidised sample	Live cellular suspension
SYTO 60	Cellular suspension	0.22 µm-filtered sample

Appendix Table 2 - 2*: Stability of the different triple stains expressed in minutes. Stability was considered sufficient when cell counts in the predefined gates of the detected populations did not deviate more that 10% from the first sample. Gating was of the populations was performed as described previously. Triple stains with $DiBAC_4(3)$ were not considered because of a high spillover.*

Triple Stain	**Stability [min]**
cFDA/NR/SYTO 60	13
cFDA/HE/SYTO 60	3
cFDA-SE/NR/SYTO 60	15
cFDA-SE/HE/SYTO 60	13

CHAPTER

3

FLOW CYTOMETRIC FINGERPRINTING FOR MICROBIAL STRAIN DISCRIMINATION AND PHYSIOLOGICAL CHARACTERIZATION

CHAPTER

3

FLOW CYTOMETRIC FINGERPRINTING FOR MICROBIAL STRAIN DISCRIMINATION AND PHYSIOLOGICAL CHARACTERIZATION

1 Abstract

The analysis of microbial populations is fundamental, not only for developing a deeper understanding of microbial communities but also for their engineering in biotechnological applications. Many methods have been developed to study their characteristics and over the last few decades, molecular analysis tools, such as DNA sequencing, have been used with considerable success to identify the composition of microbial populations. Recently, flow cytometric fingerprinting is emerging as a promising and powerful method to analyze bacterial populations. So far, these methods have primarily been used to observe shifts in the composition of microbial communities of natural samples. In this chapter, we apply a flow cytometric fingerprinting method to discriminate among 29 *Lactobacillus* strains. Our results indicate that it is possible to discriminate among 27 *Lactobacillus* strains by staining with SYBR green I and that the discriminatory power can be increased by combined SYBR green I and propidium iodide staining. Furthermore, we illustrate the impact of physiological changes on the fingerprinting method by demonstrating how flow cytometric fingerprinting can discriminate the different growth phases of a microbial culture. The sensitivity of the method is assessed by its ability to detect changes in the relative abundance of a mix of polystyrene beads down to 1.2%. When a mix of bacteria was used, the sensitivity was as between 1.2% and 5%. The presented data demonstrate that flow cytometric fingerprinting is a sensitive and reproducible technique with the potential to be applied as a method for the dereplication of bacterial isolates.

Chapter redrafted after:

Buysschaert,B., Kerckhof, F.-M., Vandamme, P., De Baets, B. & Boon, N. (2017). Flow cytometric fingerprinting for strain discrimination. *Cytometry part A.*

2 Introduction

The application of flow cytometry in microbiology became popular in the last few decades due to the improved resolution of the instrumentation (Shapiro, 2003). The initial reason for its success was the possibility to count planktonic bacteria. The high throughput of the technique increases the accuracy of the measurement while decreasing the analysis time, making flow cytometry the method of choice for bacterial enumeration (Muller and Nebe-von-Caron, 2010). A second reason for the success is the possibility to simultaneously assess different physiological parameters by means of multicolor staining assays (Hammes *et al.*, 2011, Tracy *et al.*, 2010, Strauber and Muller, 2010). As a consequence, specific cellular reactions can be monitored and quantified with single-cell resolution, rendering flow cytometry a very appealing technique to study microorganisms (Van Nevel *et al.*, 2013). Not only has much research been done on characterizing the microbial physiology of single species by means of flow cytometry, also environmental (community) microbiology and microbial ecology gained much insight in population dynamics thanks to flow cytometry (Wang *et al.*, 2010, Porter *et al.*, 1997, Vives-Rego *et al.*, 2000, Prest *et al.*, 2014). While usage and applications increased, challenges became apparent and experiments with different dyes demonstrated that the heterogeneity in microbial life influences the staining kinetics (Lebaron *et al.*, 1998, Buysschaert *et al.*, 2016, Strauber and Muller, 2010). This resulted in the need to optimize staining protocols for each type of bacterium or bacterial community (Muller *et al.*, 1995). Furthermore, it has been shown that environmental factors and the microbial growth phase influence the outcome of flow cytometric analyses (Shi *et al.*, 2007, Brognaux *et al.*, 2014). Although this could hinder conventional data analysis based on gating where fixed regions of interest are determined based on operator experience and control samples, it can also be seen as an opportunity to detect changes related to both the organism itself and its environment. One way to exploit this sensitivity is by flow cytometric multidimensional fingerprinting, a method developed in the last few years (Props *et al.*, 2016, De Roy *et al.*, 2012, Koch *et al.*, 2013a, Koch *et al.*, 2013c, Prest *et al.*, 2013, Bombach *et al.*, 2011, Van Nevel *et al.*, 2016b).

Flow cytometric fingerprinting algorithms have been presented in various ways, each different in their approach and purpose. A first approach is based on image processing of the flow cytometric density plots. Examples of these kind of methods are the Dalmatian plot method and the cytometric histogram image comparison (CHIC) method (Bombach *et al.*, 2011, Koch *et al.*, 2013a). In this case pairwise comparisons of the images are used as a tool to differentiate microbial communities. For more information on these methods we refer the reader to the comparative study by Koch *et al.* (2014). A second approach is based on either

gating or binning of flow cytometry cell populations. This approach relies on the hypothesis that phenotypic or genotypic changes in the microbial communities are reflected by changes in event counts in the different gates or bins in a multidimensional space. Recently, methods based on this approach have been published by Prest *et al.* (2013) and Koch *et al.* (2013b). The method developed by Prest *et al.* (2013) is optimized for drinking water and its usefulness is illustrated by the integration of flow cytometry for monitoring drinking water in the Swiss legislation (SLMB, 2012). The method named CyBar, developed by Koch *et al.* (2013b) on the other hand was used to monitor the microbial community of a biogas reactor. However, both methods have two main disadvantages: Firstly, they rely on gating which is subjective and dependent on the expertise of the operator. Consequently, the reproducibility of these methods is decreased. Secondly, as the gates are based on the microbial community of a specific environment or set-up, the universality of the method is constrained. An objective and gate-free method can be implemented to overcome these limitations as described by Van Nevel *et al.* (2016b), (Van Nevel *et al.*, 2016a). There we showed that cytometric fingerprinting is a viable strategy for the fast monitoring of drinking water quality. The data post-processing is based on a supervised discriminant analysis to separate different types of samples. This method requires a representative and labeled dataset to compute discriminant functions for further classification, rendering this approach unsuitable for exploratory and real-time analysis. To summarize, although all previously described methods demonstrate that flow cytometric fingerprinting is a sensitive method to assess the dynamics of microbial populations in communities, they can be either limited to specific applications, subjective, time consuming, complex or they require a complete dataset and knowledge of the observed system to be implemented. The combination of these factors has made it difficult to implement flow cytometric fingerprinting as an exploratory tool. To address this problem, Props *et al.* (2016) introduced the so-called phenotypic diversity index. The convenience of reducing the flow cytometric fingerprint to a single number, referred to as the phenotypic fingerprints, improves interpretation and visualization, especially for time-series data. For time-independent data, however, the phenotypic diversity, expressed in arbitrary units, is less suitable as it does not explain the relationship between samples and lacks the resolution to discriminate among similar samples. In the current chapter, we describe a clustering-based unsupervised algorithm with flow cytometric fingerprinting, complementary to the fingerprinting method of Props *et al.* (2016), for dereplication of microbial isolates.

To assess the suitability of our unsupervised flow cytometric fingerprinting method, we compared 29 bacterial strains encompassing closely as well as more distantly related *Lactobacillus* species and demonstrated that it was possible to discriminate 27 strains from one another solely based on their flow cytometric fingerprint. Discrimination could be

improved depending on the dyes used and by excluding background fluorescence. As flow cytometric fingerprints are known to change in function of the microbial physiology, the effect of the growth phase on the fingerprinting was also assessed and our results show that the different growth stages could be distinguished using our fingerprinting strategy, which illustrates that the growth stage directly impacts the fingerprints. Furthermore, we determined the sensitivity of the method with polystyrene beads and by mixing two bacteria strains. We demonstrated that a comparative shift of 1.2 to 5 % of events is sufficient to be detected by our fingerprinting method. Ultimately, we assessed the reproducibility of the method as well.

3 Materials and methods

3.1 Flow cytometric fingerprinting for taxonomic differentiation

Bacterial strains. 29 *Lactobacillus* strains representing eight distinct species were selected for this experiment. We included multiple strains of several species to determine to which extent flow cytometric fingerprinting permitted to distinguish among strains representing the same species. In addition, four of the species selected, *i.e. Lactobacillus casei, Lactobacillus paracasei, Lactobacillus rhamnosus* and *Lactobacillus zeae,* all belong to the so-called *Lactobacillus casei* species cluster, and are therefore taxonomically and phylogenetically closely related (Pot *et al.*, 2014). The remaining species studied (*i.e. Lactobacillus acidophilus, Lactobacillus brevis, Lactobacillus farciminis* and *Lactobacillus salivarius*) each represent an additional *Lactobacillus* species cluster (Pot *et al.*, 2014) (**Appendix Figure 3 - 1**). The strains analyzed included *L. acidophilus* LMG 8151, LMG 9433^{T} (*^{T}, type strain), LMG 11428 and LMG 11430; *L. brevis* LMG 6906^{T}, LMG 7761, LMG 11774, LMG 11993 and LMG 18022; *L. casei* LMG 6904^{T}; *L. farciminis* LMG 9200^{T}, R-42629 and R-46564; *L. paracasei* LMG 9191^{T} (*L. paracasei* subsp. *tolerans*), LMG 10774, LMG 13087^{T} (*L. paracasei* subsp. *paracasei*), LMG 13729 and R-21695; *L. rhamnosus* R-32689, LMG 6400^{T}, LMG 10775, LMG 12166 and LMG 18030; *L. salivarius* LMG 9476^{T} (*Lactobacillus salivarius* subsp. *salicinius*), LMG 9477 (*Lactobacillus salivarius* subsp. *salivarius*), LMG 14476, LMG 14477 and LMG 22873; and *L. zeae* LMG 17315^{T} (**Table 3 - 1**). All strains originated from the BCCM/LMG collection or from the research collection of the Laboratory of microbiology, faculty of science, UGent, (LM-UGent).

Table 3 - 1: *Overview of the different species and their cultivation conditions.*

Species Name	Strain Number	Condition	Temperature [°C]
L. acidophilus	8151	Anaerobic	37
L. acidophilus	9433^T	Anaerobic	37
L. acidophilus	11428	Anaerobic	37
L. acidophilus	11430	Anaerobic	37
L. brevis	6906^T	Aerobic	28
L. brevis	7761	Aerobic	28
L. brevis	11774	Aerobic	28
L. brevis	11993	Aerobic	28
L. brevis	18022	Aerobic	28
L. casei	6904^T	Aerobic	28
L. farciminis	9200^T	Aerobic	28
L. farciminis	R-42629	Aerobic	28
L. farciminis	R-46564	Aerobic	28
L. paracasei	9191^T	Aerobic	28
L. paracasei	10774	Aerobic	28
L. paracasei	13087^T	Aerobic	28
L. paracasei	13729	Aerobic	28
L. paracasei	R-21695	Aerobic	28
L. rhamnosus	R-32689	Aerobic	37
L. rhamnosus	6400^T	Anaerobic	37
L. rhamnosus	10775	Anaerobic	37
L. rhamnosus	12166	Anaerobic	37
L. rhamnosus	18030	Anaerobic	37
L. salivarius	9476^T	Aerobic	28
L. salivarius	9477	Aerobic	28
L. salivarius	14476	Aerobic	28
L. salivarius	14477	Aerobic	28
L. salivarius	22873	Aerobic	28
L. zeae	17315	Aerobic	28

Cultivation method. The lactobacilli were cultivated on MRS agar plates (Oxoid). After three days of incubation, colonies were picked and transferred to liquid MRS broth (Oxoid) in quadruplicate. After exactly 24 hours of growth, the strains were transferred to fresh medium in a 2% (v/v) ratio and samples were taken after exactly 24 hours of growth. Anaerobic strains (LMG 8151, LMG 9433^T, LMG 11428, LMG 11430, LMG 6400^T, LMG 10775, LMG

12166, LMG 18030) were incubated at 37°C. The other strains were incubated at 28°C except R-32689 which was also incubated at 37°C (**Table 3 - 1**). The incubation time was chosen to increase reproducibility but did not ensure that all strains were in the same growth stage after 24 hours.

Staining protocol. Two different stains were applied: SYBR Green I (SG) and a combination of SYBR green I and propidium iodide (SGPI) as a viability indicator. PI (20 mM in dimethyl sulfoxide (DMSO), Invitrogen) and SYBR Green I (10 000X concentrate in DMSO, Invitrogen) were diluted 50 and 100 times respectively in 0.22 µm-filtered DMSO (IC Millex, Merck). In either case, samples were stained with 10 µL/mL staining solution (Prest *et al.*, 2013). While SG stains all cells regardless of their phenotypic state, PI only enters cells with a damaged or permeabilized membrane, thus differentiating intact versus putative dead or damaged cells.

Flow cytometry. All samples were measured with a benchtop Accuri C6 cytometer (BD Biosciences). The stability of the instrument was controlled daily using 3 µm calibration beads (05-4018, Sysmex-Partec) and the instrument was calibrated according to the manufacturers standard. The blue laser (488 nm) was used for the excitation of the stains. The filters for the (fixed gain) photomultiplier detectors used during the measurements were 533 nm with a bandpass of 30 nm for the green fluorescence (FL-1) and 670 nm longpass filter for the red fluorescence (FL-3). The threshold was set on the 533/30 nm (FL-1) detector at the arbitrary unit of 500. The reproducibility of the method was tested with a FACSVerse cytometer (BD Biosciences). The performance of the instrument was monitored daily and the instrument was calibrated with the CS&T calibration beads (BD Biosciences). The blue laser (488 nm) was used for the excitation of the stains. The optical filters used were 527 nm with a bandpass of 32 nm for the green fluorescence and 700 nm with a bandpass of 54 nm for the red fluorescence.

3.2 Flow cytometric fingerprinting for physiological differentiation

To assess the effect of the growth phase on the fingerprinting method, a batch experiment was performed where *L. paracasei* LMG 10774 was cultivated in MRS medium (Oxoid). Biological triplicates of the culture were cultivated in flasks shaken at 200 rpm and incubated at 28°C during the entire experiment. Inoculation concentration was approximately 10^7 cells/mL. Samples were taken from the flasks with an OnCyt© staining robot (Oncyt, Switzerland) coupled to the Accuri C6 flow cytometer as described by Besmer *et al.* (2014). In short, every 18 minutes the autosampler took a sample from one of the flasks and diluted it 1000 times with 0.22 µm-filtered water. Subsequently, the samples were automatically

stained with SYBR green I in the same concentration as described above and incubated at 37°C for 13 minutes. The growth curves were made with the cell concentration as determined by flow cytometry. A universal gate was used to remove most of the background as described below.

3.3 Sensitivity analysis and reproducibility

Two types of fluorescent polystyrene beads were selected for this test (0.88 µm and 1.34 µm Nano Polystyrene and Fluorescent Size Standard, Spherotech). The beads were analyzed on the Accuri C6 and excited by the blue laser. As the beads had different sizes and a different fluorescence intensity, a distinct fingerprint could be made based on both the scatter signals and the fluorescence signals (530/30 nm and 670 LP). To determine the sensitivity, mixes of different ratios were made (*i.e.* 0, 1, 5, 10, 20, 30, 40, 50 percent for both bead types). To that end, the beads were first diluted to a concentration of approximately 10^5 beads/mL in 0.22 µm-filtered water. Both suspensions with an equal concentration of beads were then mixed in different ratios. The mixtures were analyzed in triplicate. The ratio of beads was calculated based on the bead counts with the Accuri C6 Csampler software (version 1.0.264.21, BD Biosciences) as the measured count deviated from the theoretical values. Similarly, *Lactobacillus brevis* LMG 18022 and *Lactobacillus paracasei* LMG 9191^T were mixed in different ratios. Both strains could easily be discriminated but their cell populations show some overlap in the fluorescence and scatter channels (**Appendix Figure 3 – 2**). Prior to mixing, stocks of equal concentration (10^5 cells/mL) were made for both strains. Concentrations were determined by flow cytometric cell counting (Van Nevel *et al.*, 2013) after standardized culturing, ensuring that all samples contained the same number of cells. The majority of the background fluorescence was removed prior analysis using the universal gate described below. To assess reproducibility of the dereplication method, the experiment was repeated three times for LMG 9191^T, LMG 9200^T, LMG 9477 and LMG 18022. For each repetition, quadruplicates of the strains were cultivated, stained with SGPI and analyzed similarly to the other datasets.

3.4 Data analysis

Unless mentioned otherwise, all data was extracted from the proprietary Accuri C6 Csampler software version 1.0.264.21 or FACSuite software version 1.0.4 (BD Biosciences, Belgium) in the flow cytometry standard (FCS 3.0) format and subsequently imported into R version 3.4.0 (R Core Team, 2015) through the functionality offered by the *flowCore* package v1.42.2 (B. Ellis *et al.*). Data was first log transformed and subsequently normalized by dividing all

values by the maximum fluorescence intensity signal. No compensation was applied. Gating to reduce the background was performed in R studio using the *flowCore* package on both SG and SGPI stained samples. A 0.22 µm-filtered control was used to determine the position of the background and a heat-killed sample was used to determine the position of the permeabilized cell population (Berney *et al.*, 2007). Based on this, a universal gate was constructed to remove as much background as possible without removing the permeabilized cell population although the distinction between PI positive cells and the background can be difficult for some samples. Additionally, a stained sample of the dilution buffer was used to assess the quality of the dilution buffer and of the stain (**Figure 3 - 1**). The data quality was evaluated and improperly acquired data was cleaned using the *flowAI* package v1.4.3 (Monaco *et al.*, 2016) in order to remove anomalies in the data related to changes in flow rate, unstable signal or outliers in the lower limit of the dynamic range. Samples which failed the quality control (QC) were removed from the dataset. When the microbial growth was monitored, 2000 events per sample were randomly selected to avoid an influence of the different number of detected events during the different growth stages on the fingerprint. This was done after background was removed with the universal gate. Next, a single-step discretization ('binning') and Gaussian bivariate density estimation was performed on the selected parameters (green and red fluorescence, FSC-H and SSC-H) using the *KernSmooth* package (Wand, 2015). A binning grid of 128 x 128 was fixed for each bivariate density estimation (Props *et al.*, 2016). All bivariate density estimations were concatenated to a one-dimensional feature vector which we refer to as the fingerprint. Subsequently, the dissimilarity of the fingerprints was calculated using the quantitative Jaccard distance measure (Ružička index) as implemented in the function *vegdist* from the *vegan* package v2.4.3 (Oksanen *et al.*, 2016). The Jaccard distance matrices were visualized as dendrograms based on agglomerative clustering with Ward's minimum variance method (Ward.D2 from the *hclust* package) as linkage from the stats package (R Core Team, 2015). Uncertainty of the clustering was evaluated with the *pvclust* package v2.0.0 (Suzuki and Shimodaira, 2015). The latter package provides two types of *p*-values: AU (Approximately Unbiased) *p*-value and BP (Bootstrap Probability) value. The AU *p*-value, which is computed by multiscale bootstrap resampling, is a better approximation of the unbiased *p*-value than the BP value computed by normal bootstrap resampling. The number of permutations was determined based on the standard error of the calculated *p*-values. For all clustering dendrograms, 1000 permutations were sufficient and resulted in a maximum error of approximately 0.02. For *p*-values >95% the hypothesis that the cluster does not exist is rejected at a 5% significance level. Dendrograms were visualized with the *iToL* software version 3.5.3 (Letunic and Bork, 2016).

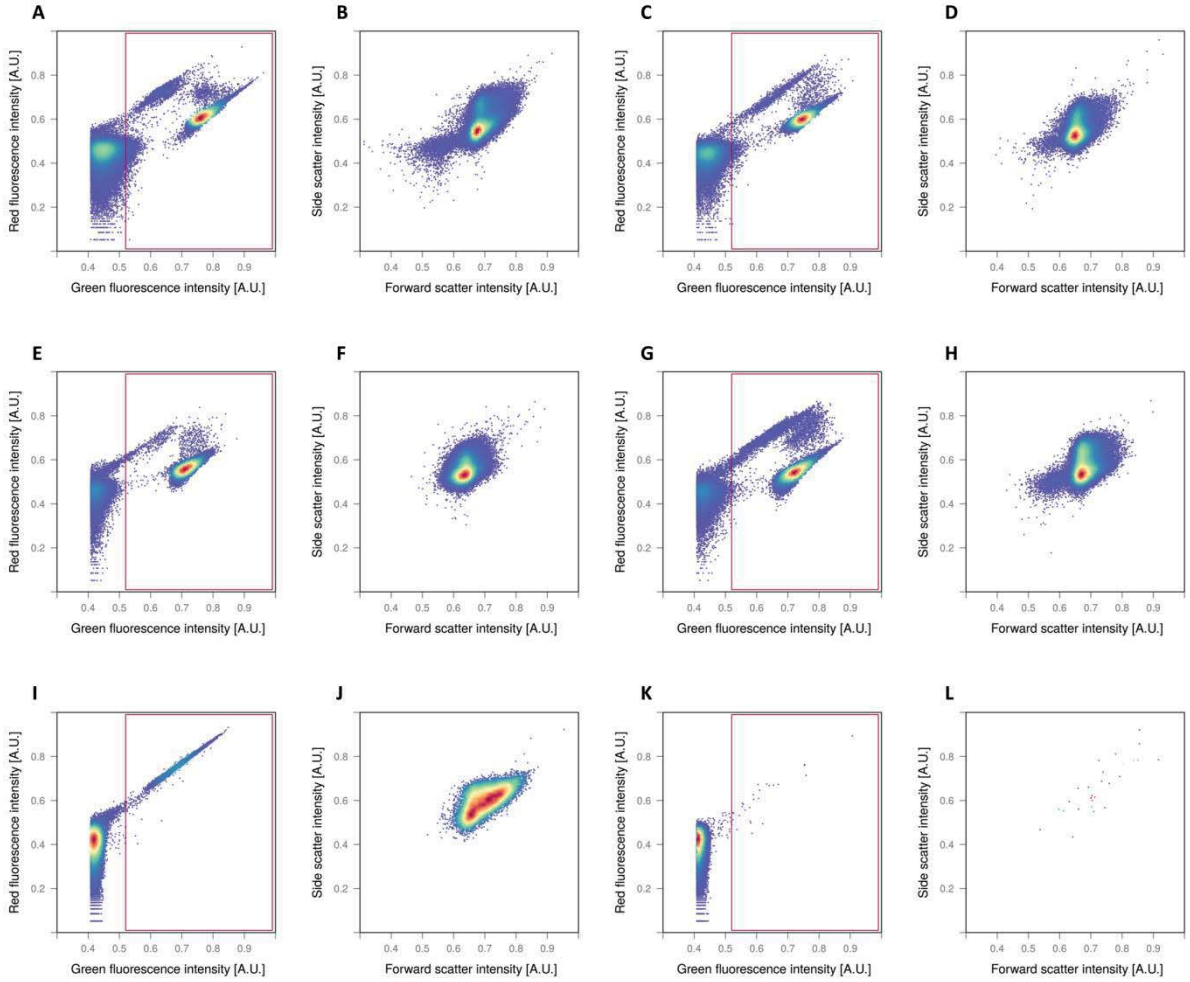


Figure 3 - 1: *Illustration of representative fingerprints of different* Lactobacillus *spp. strains and the controls used to construct the gate. Fingerprints of LMG 9200^T (**A,B**), LMG 9477 (**C,D**), LMG 7761 (**E,F**) and LMG 14476 (**G,H**) were obtained by staining the bacteria with SGPI. A universal gate, shown as a red rectangle, was constructed to remove as much background as possible. The gate was based on a heat killed sample (**I,J**) and a 0.22 µm-filtered sample (**K,L**). The fingerprint plots based on the scatter signals only show the data after background was removed with the universal gate (**B,D,F,H,J,L**).*

4 Results and discussion

Flow cytometric fingerprinting has been established as a useful tool to monitor microbial communities in the past decade (Koch, 2013, Prest *et al.*, 2013, Van Nevel *et al.*, 2016a, Props *et al.*, 2016, Van Nevel *et al.*, 2016b). In this chapter we explore the possibility to use flow cytometry to analyze axenic cultures based on the division of the cytometric data in bins by an equally spaced grid as described by Van Nevel *et al.* (2016b) and Props *et al.* (2016) to create fingerprints. The Jaccard distance measure between these fingerprints was subsequently calculated to cluster the data based on their similarity. Thus, our method can be considered as an extension of those methods as they are founded on a common binning

approach. Using binning instead of gating reduces operator dependency and increases interoperability. Another advantage is that, in theory, it can be used for any kind of sample or stain in contrast with the methods developed by Prest *et al.* (2013) and Koch *et al.* (2013a) as it does not rely on gating of specific subpopulations. In comparison to other binning approaches such as the flowFP method (De Roy *et al.*, 2012, Rogers and Holyst, 2009) this algorithm does not require a training dataset to construct a binning model. Regarding the subsequent treatment of the fingerprints, Van Nevel *et al.* (2016b) relied on supervised discriminant analysis to look for the differences between samples and a reference. Despite the better performance of supervised methods, a disadvantage is that it requires labeled data which is, in many cases, not available. Props *et al.* (2016) showed that it is also possible to calculate diversity indices on the binned data and that the dynamics of these diversity indices is correlated to the dynamics in community composition as determined by amplicon sequencing of the 16S rRNA gene. This approach does however not describe the similarity between samples. Calculation of the phenotypic diversity on our dataset showed to be difficult to interpret as not every strain yielded a different diversity whereas no relationship between the samples is established (**Appendix Figure 3 - 3**). The CHIC method (Koch *et al.*, 2013a) does establish this relationship by pairwise image comparison. The downside of the method is the workflow where data must be pre-processed first. Then images must be created and saved in the cytometry software. Subsequently, the images are imported and analyzed in ImageJ. The output of the image analysis is then exported from imageJ and imported in R for final statistical analysis. The use of different software platforms makes this workflow impractical, and as it relies on image comparison, multiple images should be generated per sample to analyze the multivariate flow cyometry data making it also a laborious analysis. Our initial choice for the Jaccard distance measure to calculate the similarity was based on its frequent use in ecological studies. To validate our choice, a comparison was made with other distance measures (**Appendix Figure 3 - 4**). The results showed that also other distance measures could be used successfully.

4.1 Flow cytometric fingerprinting for taxonomic differentiation

A wide variety of dyes are available for flow cytometry to assess different aspects of microbial cells (Buysschaert *et al.*, 2016). In addition, most dyes exhibit different spectral characteristics, changing the flow cytometric data and as a consequence, the cytometric fingerprint. To assess the effect of the stain that was used we tested both SYBR green I (SG) and the combination of SYBR green I and propidium iodide (SGPI) on all biological quadruplicates of the 29 *Lactobacillus* strains. Fingerprints of both the scatter signals and the fluorescence signals (530/30 nm and 670 LP) were made for each sample, and sample

dissimilarities were calculated based on these flow cytometry data. In both datasets, improperly acquired samples due to technical failures were found. These samples show abnormal results or an unsteady fluorescence intensity over time. For the SG stained dataset these are LMG 9191^T rep1 and rep3, LMG 11428 rep4, LMG 10774 rep3, LMG 12166 rep3, LMG 18030 rep2, LMG 9476 rep1, LMG 14476 rep1 and LMG 22873 rep1. For the SGPI stained dataset LMG 9191^T rep1 and 3 and LMG 11428 rep4 were acquired improperly. The results were clustered and visualized in dendrograms.

The dendrogram based on fingerprints obtained with SG staining without gating indicated that 5 out of 116 fingerprints were misclassified and that two pairs of bacteria could not be discriminated (**Appendix Figure 3 - 5**). When SGPI was used as a stain, the results improved to 3 out of 116 fingerprints that were misclassified and one pair of strains that could not be discriminated (**Appendix Figure 3 - 6**). Although some samples experienced some issues during acquisition, not all of these were misclassified (**Appendix Figure 3 - 5** and **Appendix Figure 3 - 6**). The successful discrimination of the different strains based on SG or SGPI, both nucleic acid dyes, is in part related to the difference in genome sizes that may vary from over three million base pairs (bp) in the *L. casei* species to around two million bp for *L. acidophilus.* Also between strains of the same species, important differences in genome size can be noticed (Wassenaar and Lukjancenko, 2014). PI is generally used as a viability indicator, therefore the extra information necessary to improve the results is related to the viability of the bacteria. It is important to note that not all strains had the same growth rate, thus not all strains were in stationary phase at the time of analysis. Consequently, the number of permeabilized or dead cells, which is related to the growth rate, differed between the strains as well. The combination of dyes can improve clustering performance, if they reveal relevant information for discrimination and when used in a reproducible and standardized way. Fingerprinting techniques ideally rely on ungated data as each form of gating might alter the results and decrease the universality of the method. However, background fluorescence is often present and might alter the results. To assess the impact of background fluorescence on the clustering, the background was removed using a universal gate (**Figure 3 - 1**) both for samples stained with SG and SGPI and the cell populations were processed in the same way as the ungated samples. When both background and cells were used, there were five misclassifications and two pairs that could not be discriminated for the SG stained samples. When only cells were considered, the misclassification decreased to four misclassified fingerprints and one indiscriminate pair (**Appendix Figure 3 - 7**). Clustering of the samples stained with SGPI showed no improvement after removing the background fluorescence. Based on this, we argue that the background did not worsen our results markedly.

Flow cytometry is a sensitive technique and several problems can occur during acquisition such as an unstable flow due to clogs or an unstable fluorescence signal due to cell settling or improper staining. These irregularities can influence the fingerprints and subsequently the results. By using an automated algorithm to detect and remove anomalies, the data could be cleaned from these anomalies (Monaco *et al.*, 2016). In case of the gated SG stained dataset, samples LMG 9191^T rep3, LMG 11428 rep4, LMG 6906 rep3 and LMG 6906 rep4 were removed due to an insufficient number of cells causing an error during quality control. Among the gated SGPI stained dataset, samples LMG 9191^T rep3, LMG 6906 rep3 and 4, LMG 14477 rep1 and LMG 11428 rep4 were removed. In both cases, sample LMG 11428 rep4 was also misclassified. Results improved after automated cleaning to 3 misclassified fingerprints out of 112 and one indiscriminate pair for SG stained samples (**Appendix Figure 3 - 8**) and 2 misclassified fingerprints out of 111 and one indiscriminate pair for the SGPI stained samples (**Figure 3 - 2**). Some samples were still misclassified after quality control and background removal. Results suggest the misclassification is related to the staining as the mean fluorescence intensity of the cell populations showed to be noticeably different from the other replicates. Based on this, we conclude that misclassification can be caused by background fluorescence, invalid data, improper staining or a combination thereof.

As clustering algorithms tend to form clusters even in homogeneous datasets, it is important to validate the clustering. Thus it cannot be assumed that every cluster is meaningful (Hennig, 2007). Using bootstrapping analysis, we calculated the stability of the clusters and proved that the clustering was stable (**Appendix Figure 3 - 9** and **Appendix Figure 3 - 10**). Phylogenetic relatedness between the strains could not be observed in the flow cytometric fingerprinting dendrograms. Nevertheless, when cells were stained with SG and no background was removed or data was cleaned, one of the four main branches contained almost exclusively strains grown in anaerobic conditions (**Appendix Figure 3 - 5**).

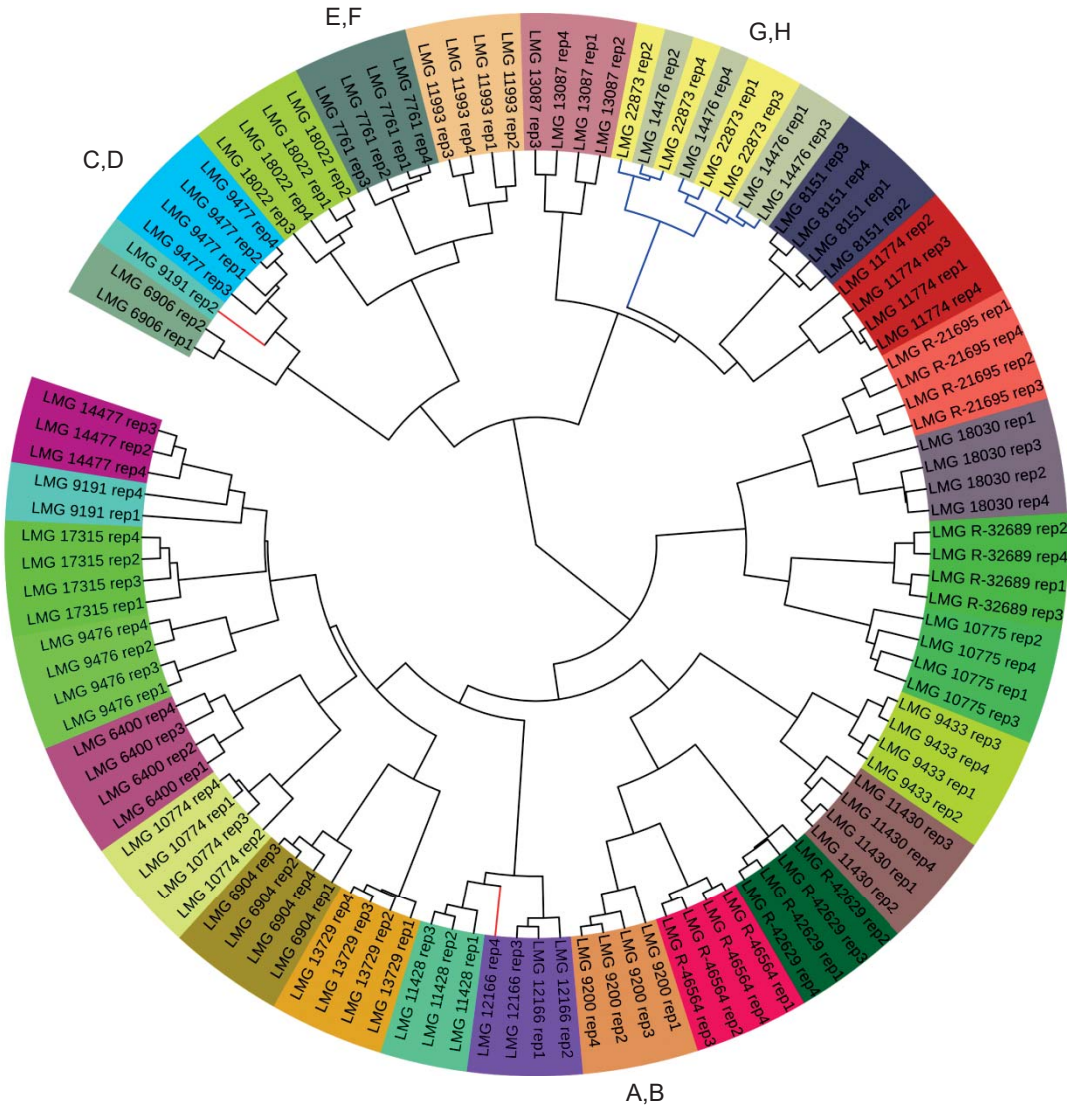


Figure 3 - 2: *Dendrogram of the flow cytometric fingerprints of the 29* Lactobacillus *strains and species stained with SGPI. All species and strains were cultivated in biological quadruplicates. After acquisition, flow cytometry data was gated to remove the background and cleaned from anomalies. Subsequently, the multidimensional space was binned and the density based Jaccard distance measures were calculated for all samples and visualized in a dendrogram constructed with the Ward hierarchical clustering method. The different strains are indicated with a different colored label. Clades highlighted in blue indicate strains that could not be discriminated. Nodes highlighted in red show the misclassified fingerprints. Fingerprints of LMG 9200^T (**A,B**), LMG 9477 (**C,D**), LMG 7761 (**E,F**) and LMG 14476 (**G,H**) are shown in **Figure 3 – 1**.*

4.2 Flow cytometric fingerprinting for physiological differentiation

As flow cytometric fingerprints of bacteria vary in function of incubation time, the capability of our method to discriminate different growth phases from one another was assessed. To this end, a batch culture of *L. paracasei* LMG 10774 was monitored during its different growth phases. The growth curves of the replicates were obtained based on the flow cytometric cell

counts and are shown in **Figure 3 - 3** (above). The growth curve can be divided in three sections: the lag phase from the start to approximately 12 hours of growth, the exponential phase between 12 hours and 36 hours of growth and the early stationary phase that starts approximately at 36 hours of growth. Fingerprints of the samples were compared using the density-based Jaccard distance measure and the result of the clustering is visualized in **Figure 3 - 3** (below). The clustering shows four branches dividing the data according to the time of incubation. A comparison between the growth curve and the dendrogram shows that the lag phase, the early exponential phase, the late exponential phase and the early stationary phase can be discriminated from one another. Despite the good separation between the lag and early exponential phase, a less precise distinction could be made between the late exponential phase and the early stationary phase. A possible explanation for this result is the average change in morphology and physiology of the bacteria during growth. While cells tend to be larger in the lag phase to prepare for division, the cells in the exponential phase show higher activity and contain more nucleic acids as they are actively dividing. Cells in the stationary phase decrease in size and change their metabolism as nutrients deplete and waste increases. The possibility to discriminate among different growth stages, which could be labeled as physiological fingerprinting, has been reported in literature (Steen and Boye, 1980, Boye and Lobner-Olesen, 1991) and suggests that these results are independent of the microorganism tested. Our findings are consistent with this literature as the authors also based their conclusions on nucleic acid stains. On the other hand, our results also show an important overlap between the late exponential phase and the stationary phase. This is partially related to the variability inherent to the data which is reflected in the variability of the cell concentrations but could also be related to the fact that the transition between the two phases is not a clear transition. It is important to add that the concentration of cells could not exert an effect on the clustering as for all samples 2000 cells were randomly sampled after gating from the FCS files for subsequent fingerprinting. Although samples from the first six hours did not contain 2000 cells due to the low cell concentration, this is not reflected in the clustering of the data. A second important conclusion that can be drawn from these results is that the growth stage at which bacteria are sampled for phylogenetic fingerprinting influences the outcome and the reproducibility of the results. To overcome this issue, a standardized cultivation protocol is necessary to improve reproducibility. Such a protocol should ensure that isolates are analyzed in the same growth stage for every measurement. As illustrated by our results, even relatively small time differences can be important.

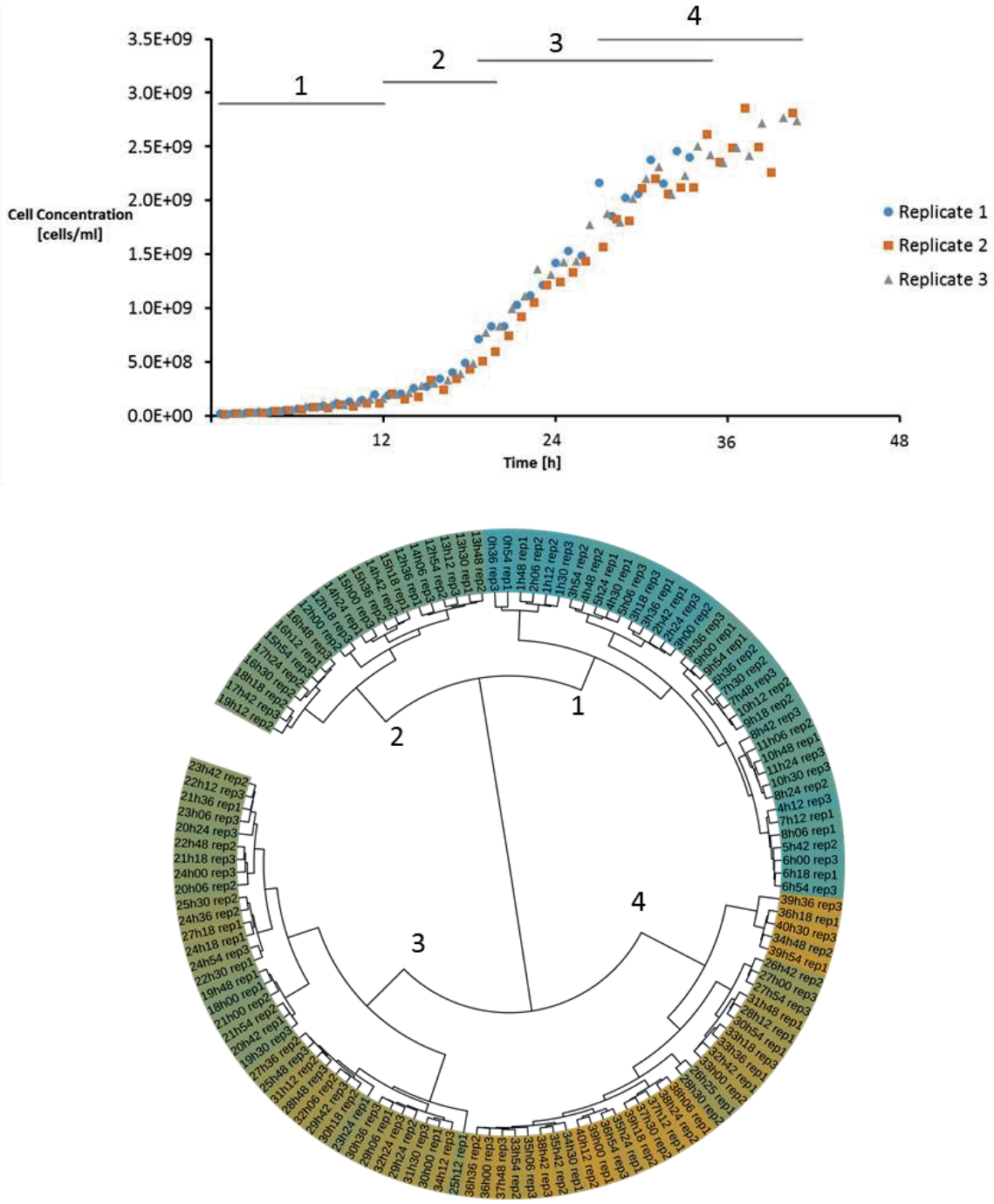


Figure 3 - 3: *Growth curve of* L. paracasei *LMG 10774 as determined by flow cytometric cell concentration (**above**) and the subsequent clustering of the flow cytometric fingerprints (**below**). The growth curve was constructed based on the cell concentration detected by flow cytometry and the SG staining. Three biological replicates were monitored for this experiment. The growth curve shows the different stages of the microbial growth with the lag phase until approximately 12 hours of growth, the stationary phase starting after approximately 36 hours of growth and the exponential phase in between. Clustering of the data after gating and cleaning, shows four clusters coinciding with the lag phase, the early exponential phase, the late exponential phase and the stationary phase. Samples are labeled according to the time sampling after inoculation and colored in function of this time. The timespan of the four clusters is plotted on top of the growth curve to visualize the overlap between the clusters.*

4.3 Sensitivity analysis and reproducibility

To assess the sensitivity of the method, fluorescent beads of 0.88 µm and 1.34 µm were mixed in different ratios. Both bead populations could easily be discriminated based on their fluorescent or scatter signals and showed no overlap in any of the channels used for fingerprinting. Prior to mixing, dilutions of both bead solutions were made to ensure that all samples contained a similar concentration of beads. The standard deviation on the detected concentration was 4%. The results are shown in **Figure 3 - 4** where the ratios are expressed in percentage of detected 0.88 µm beads. The clustering shows to easily discriminate samples with a difference of approximately 10% in bead populations and shows some larger clusters with comparable ratios. The sensitivity of the method can be determined by looking more closely at the lower or higher concentrations of 0.88 µm beads. When taking the average concentration of 0.88 µm beads into account in both end of the dilution spectrum, it can be concluded that a difference of 1.2% of events with different optical characteristics in all parameters used for fingerprinting can be detected. This illustrates that even very small changes have an impact on the results. It should be noted that this threshold was based on two types of beads with distinct fingerprints in all channels considered in this analysis. Consequently, it should be considered as the maximal sensitivity of the method. However, both the 5.6% and the 67.8% samples showed to be misclassified and lower mean fluorescence intensity was noticed of the 1.34 µm beads for the 67.8% sample. For the 5.6% sample no important difference in the fluorescence intensity of the bead populations could be detected.

By replicating the sensitivity analysis with mixtures of LMG 9191^T and LMG 18022, we show that also for bacteria a high sensitivity can be obtained down to a theoretical mix of 1% (**Appendix Figure 3 - 2**). However, as both microbial populations showed some overlap, it is hard to calculate the actual fraction of each strain to take technical errors into account. Because of this finding, we conclude that for these two strains, a sensitivity of at least 5% can be reached. Although this figure is higher than what we found for a mixture of beads, it is expected as the sensitivity should decrease with an increasing overlap between cytometric cell populations. Prest *et al.* (2013) show that their method detects changes of 3% of the relative cell concentration, whereas our method exhibited a comparable but better performance with a sensitivity of 1.2%.

Next to biological variation, the instruments can influence the results of the fingerprint as they vary in the type of lasers, detectors or optical filters. Instrumental drift can also compromise the reproducibility. Although this effect is small on the short term, this could be of higher importance after in the long term. This technical bias can be monitored with fluorescent

beads as daily quality control. To avoid any issues over time it could be useful to spike samples with beads and use the beads to correct the data for drift. Our results show that the method is reproducible for three independent repetitions of four different strains. The measurements were performed on an instrument with the possibility to recalibrate the optics to avoid the effect of instrumental drift. The reproducibility of the method was tested by repeating the analysis three times with LMG 9191^T, LMG 9477, LMG 9200^T and LMG 18022 as pure cultures stained with SGPI. Results after gating and cleaning the data show that the method is reproducible (**Figure 3 - 5**).

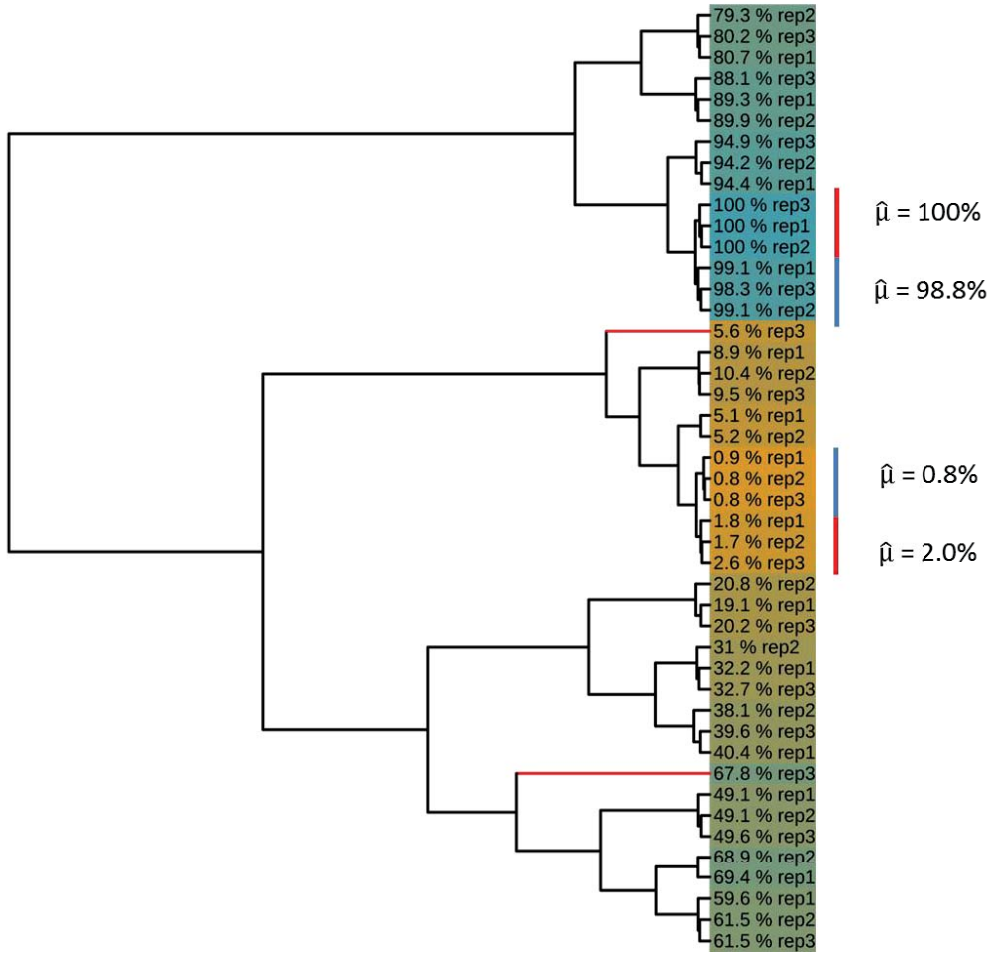


Figure 3 - 4: *Hierarchical clustering of the mixtures of 0.88 µm and 1.34 µm fluorescent polystyrene beads after cleaning the data with the* flowAI *package. Samples were prepared and measured in triplicates. Labels indicate the measured percentage of 0.88 µm beads and label colors are based the theoretical percentage of 0.88 µm beads. Red nodes indicate incorrectly clustered samples. For the highest and lowest percentages of both populations, the average percentage of 0.88 µm beads was calculated to estimate the maximal sensitivity.*

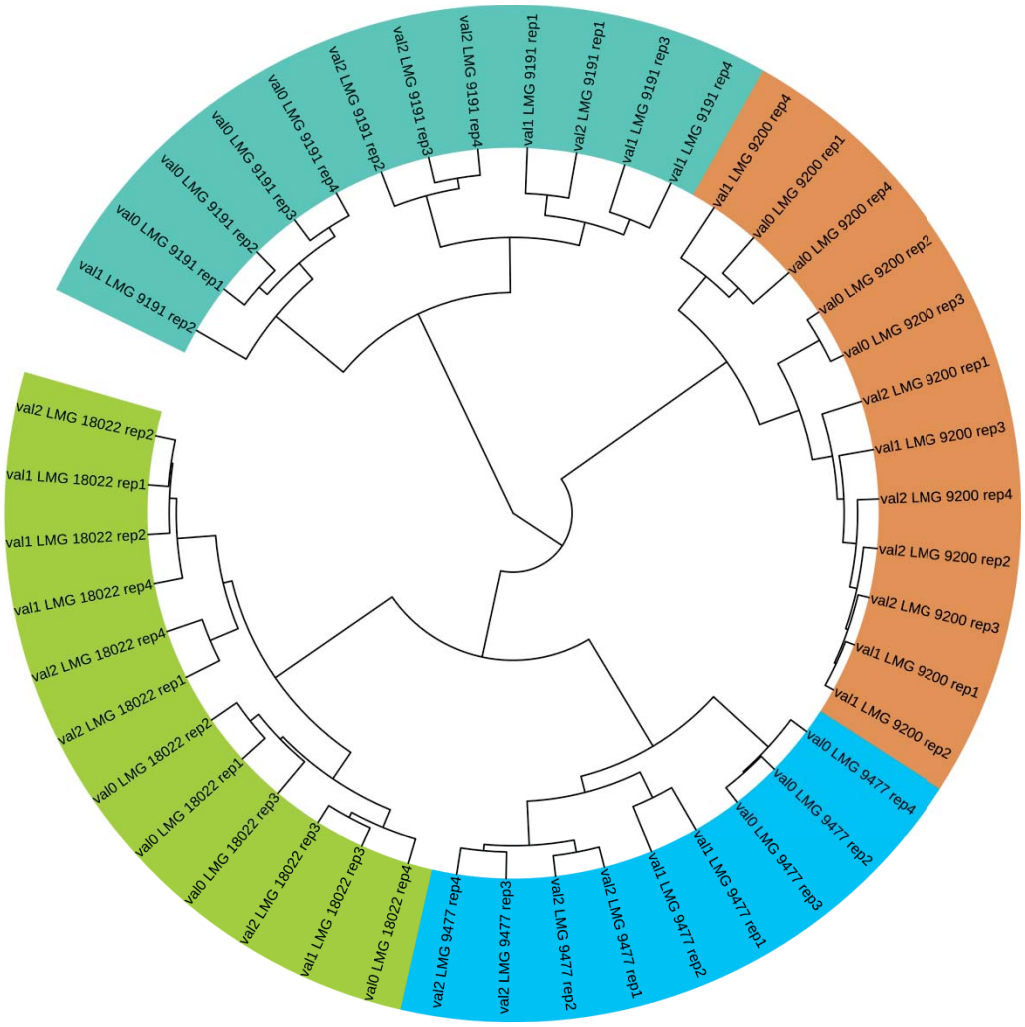


Figure 3 - 5: *illustration of the reproducibility of the method assessed by three independent repetitions of the method. To this end, four strains were selected and cultivated and analyzed as described above. Prior clustering, data was gated and cleaned as described above. Strains are indicated with a different color.*

Flow cytometric fingerprinting can be divided into what could be named phylogenetic cytometric fingerprinting on one hand and physiological cytometric fingerprinting on the other hand. The former is demonstrated by the experiment with the different strains and shows that pure cultures have characteristic fingerprints which can be used to discriminate taxa from one another. A possible application of this method in the context of environmental microbiology is the dereplication of microbial isolates before molecular characterization to improve isolation efficiency and to increase throughput (Dieckmann *et al.*, 2005). Today, characterization of isolates can also be performed in high throughput using MALDI-TOF for peptide fingerprinting of a culture (De Bruyne *et al.*, 2011). Similar to our proposed method, this approach requires a standardized cultivation protocol to ensure reproducibility. In this respect, both methods are comparable to one another but advantages of flow cytometry are the single-cell resolution and the possibility to assess different aspects of the cells by using

specific dyes. Physiological fingerprinting, on the other hand, is an interesting way to characterize the microbial physiology and could be used to better understand and estimate the impact of microbial physiology on its functionality. While physiological characterization using different dyes is well established (Strauber and Muller, 2010, Hammes *et al.*, 2011) and has been shown useful in biotechnology (da Silva *et al.*, 2012, Hewitt *et al.*, 2000), they often rely on complex multicolor assays and gating strategies. Fingerprinting could help facilitate the analysis and interpretation of this data.

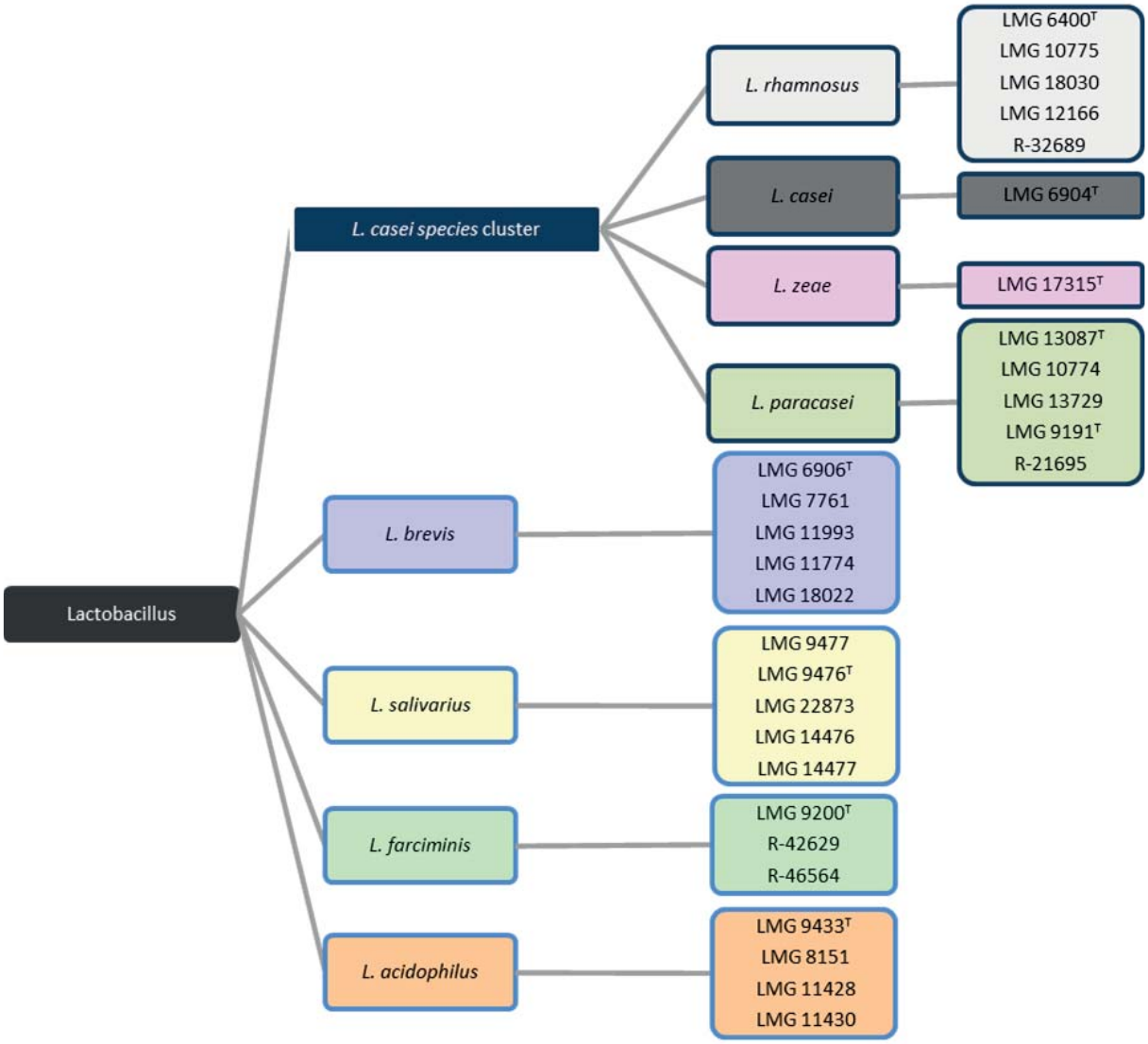
5 Conclusions and perspectives

In the current chapter we describe a new approach for flow cytometric fingerprinting that shares the same binning algorithm as described by Van Nevel *et al.* (2016a) and Props *et al.* (2016) and should be considered as a complementary method. Whereas the approach of Props *et al.* (2016) is more suitable for monitoring of time-series data, our method is better suited as comparative analysis tool for samples that are not necessarily time-dependent. In contrast to the method of Van Nevel *et al.* (2016a), our method is not supervised and is thus not reliant on labeled data. We successfully demonstrate that it is possible to discriminate bacterial strains and species with flow cytometry using only SG as nucleic acid dye. We also demonstrate that an alternative choice of stains might improve the discriminatory power. In a second experiment, we showed how fingerprinting could discriminate the different growth stages and illustrate that it may affect the reproducibility of the fingerprinting method. Finally, we show that a 1.2% change in the pattern can already be detected with our method based on a test with beads. When biological samples are used, with more similar spectral characteristics than beads, we show that changes between 1% and 5% of the events are detected. Because of this sensitivity and because of biological and mechanical variability, it is important to work with a standardized growth protocol to improve reproducibility. To conclude, we discussed the possibility to use flow cytometric fingerprinting as a fast dereplication tool to screen microbial isolates as well as the use of flow cytometric fingerprinting for physiological characterization of axenic microbial cultures through time. These proposed strategies increase the realm of possibilities for bacterial flow cytometry, especially for axenic cultures.

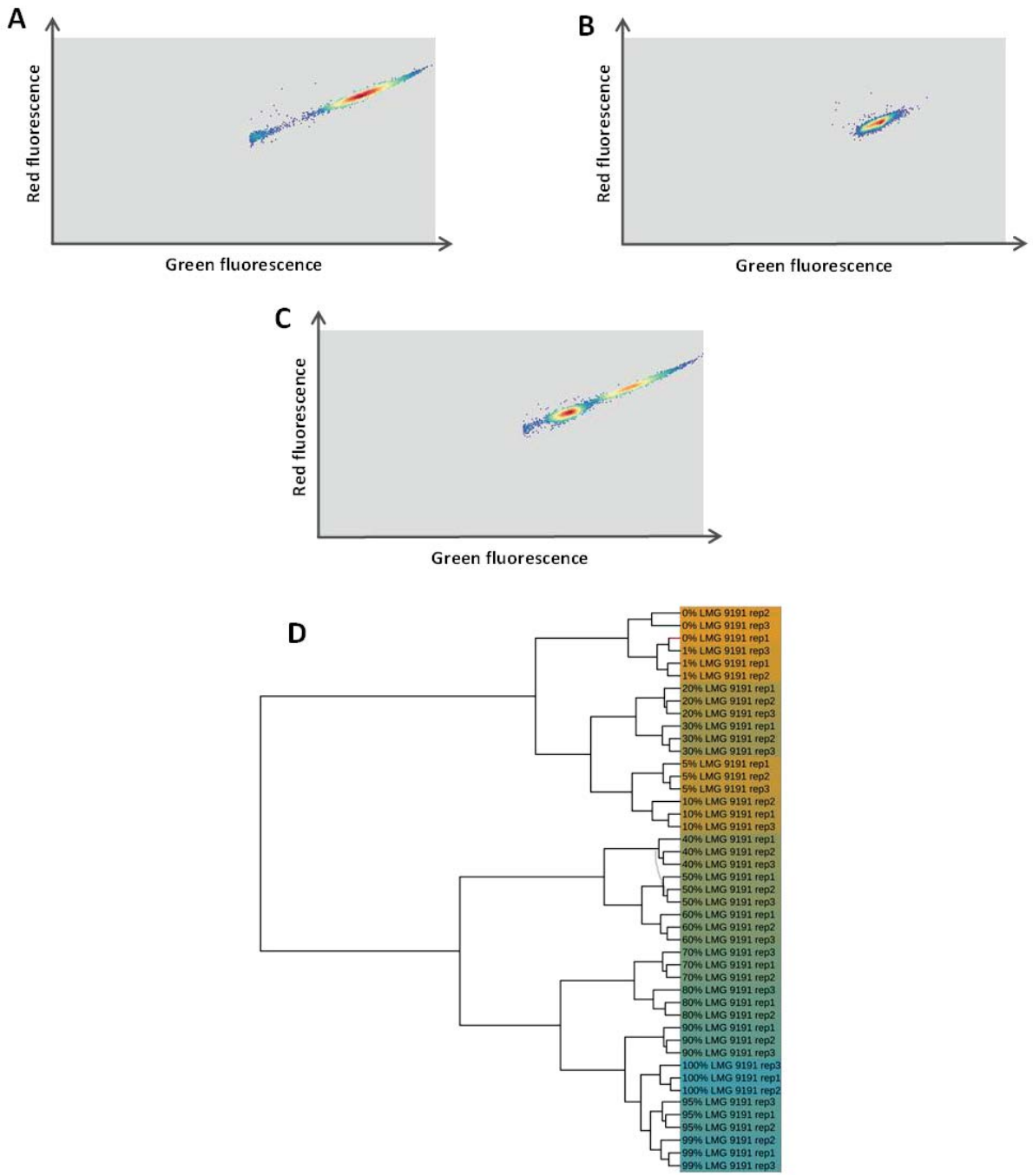
6 Acknowledgements

This work was supported by the project grant SB-131370 of the IWT Flanders and by the IMPROVED project, subvented by The interreg V “Vlaanderen-Nederland” program, a program for transregional collaboration with financial support from the European Regional Development Fund. More info : www.grensregio.eu (in Dutch). F.M. Kerckhof was supported by the GOA “sustainable methanotrophs” (BOF09/GOA005) and the Inter-University Attraction Pole (IUAP) “μ-manager” (BELSPO, P7/25). The authors want to thank Jana De Bodt and Tom Bellon for the help in the lab, the Department of Microbiology for providing us the cultures and Charlotte De Rudder and Jeet Varia for critically reading the manuscript.

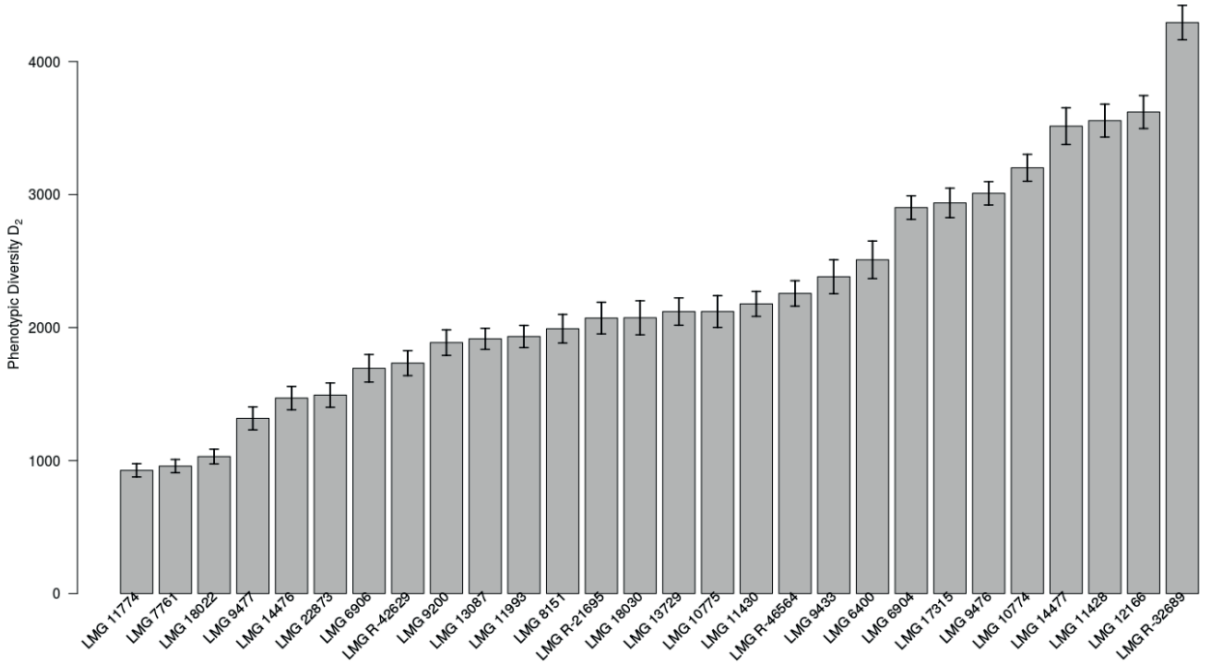
7 Appendix – Supplementary information for chapter 3



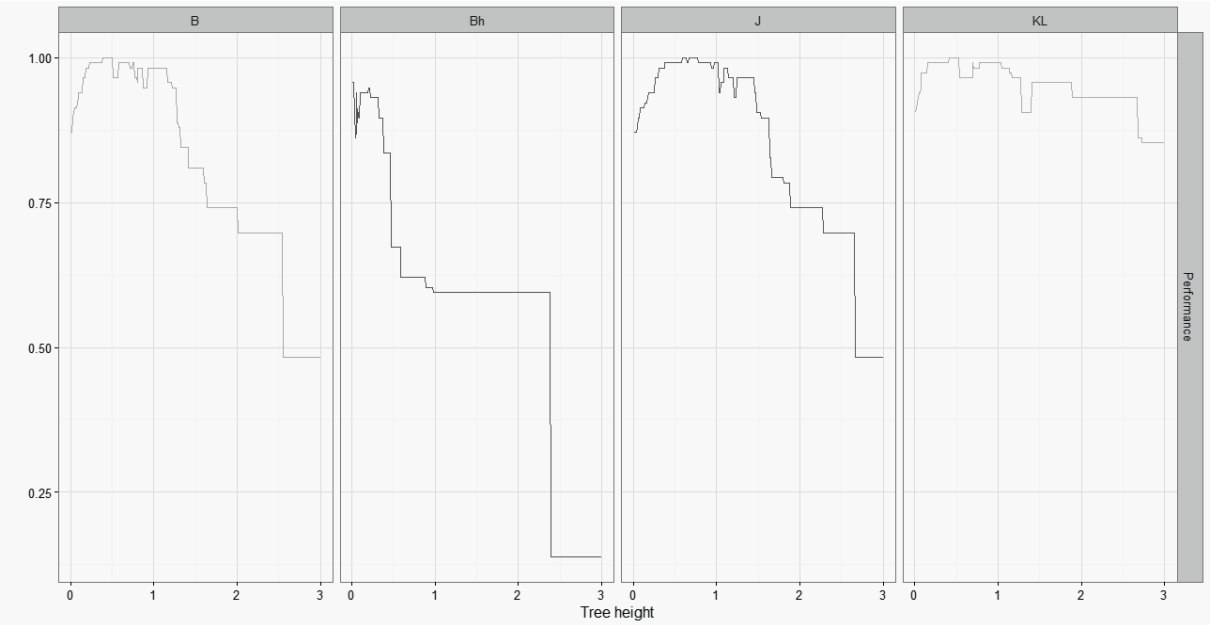
Appendix Figure 3 - 1: *Phylogenetic relationship of the different* Lactobacillus *species and strains as described by Pot et al. (2014).*



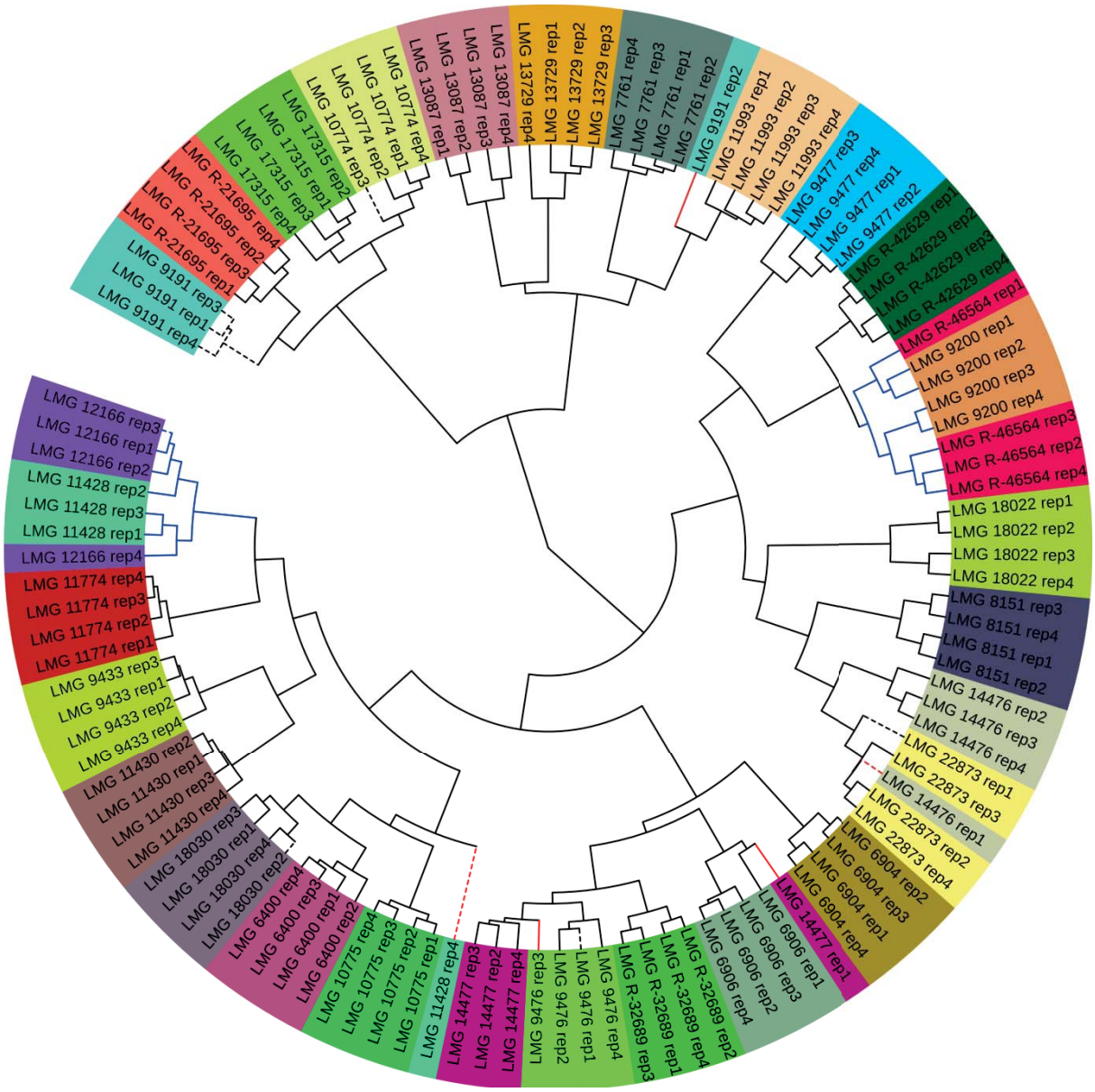
Appendix Figure 3 - 2: *Illustration of the fingerprints of both LMG 9191^T (**A**), LMG 18022 (**B**) and a mixture of both (**C**). The resulting dendrogram of the different mixes expressed as theoretical percentage of LMG 9191^T (**D**). As both cell populations overlap partially, it was not possible to recalculate the measured percentage of LMG 9191^T. Data was gated to remove the background and cleaned as described above. All samples were prepared and measured as technical triplicates. One sample is misclassified (red node) and visual assessment attributes this difference to an effect of the staining.*



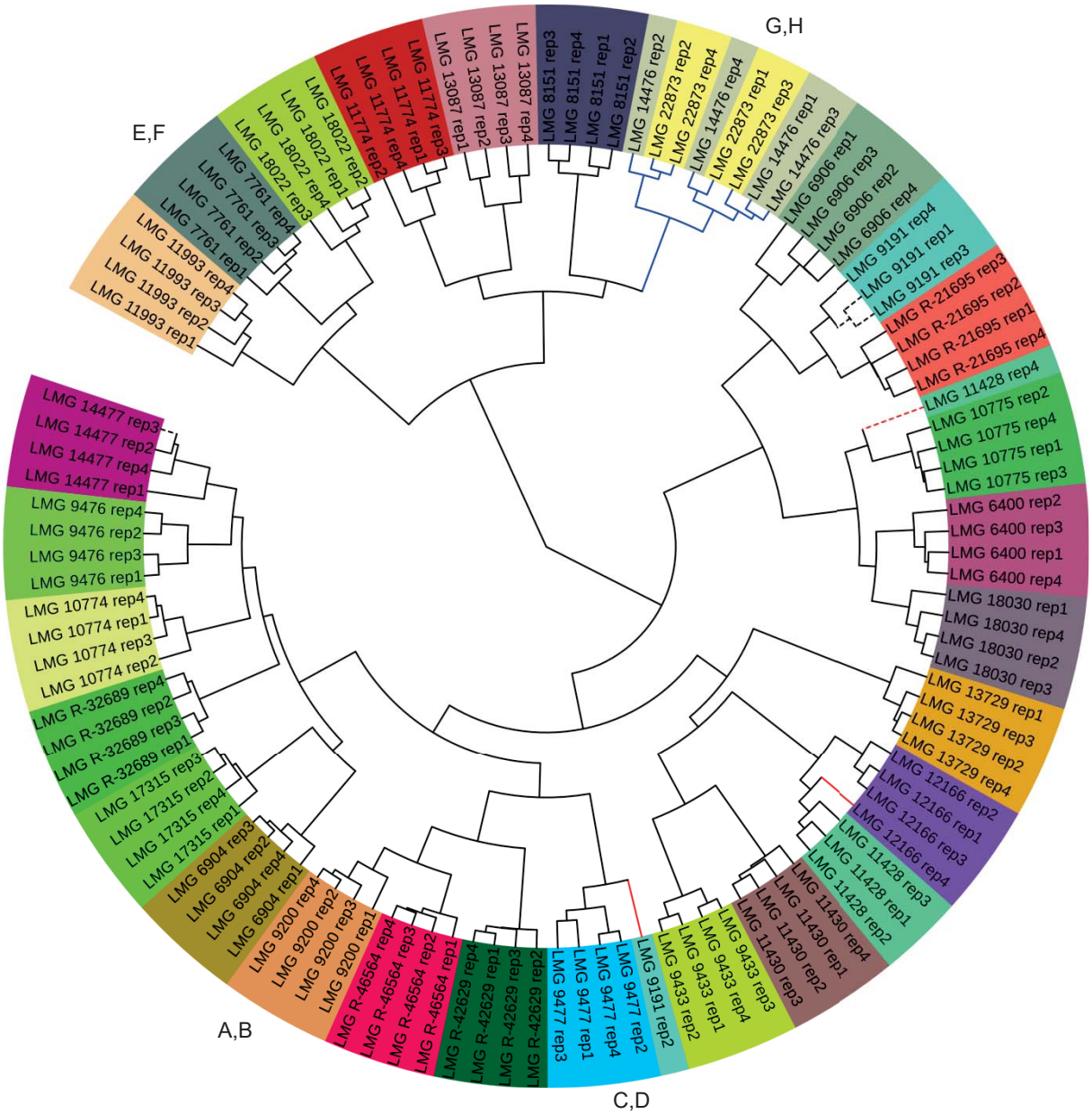
Appendix Figure 3 - 3: *Hill number diversity index D_2 as described by Props et al. (2016) for the 29 Lactobacillus strains stained with SGPI after background removal and cleaning of the data. All samples were measured in biological quadruplicates.*



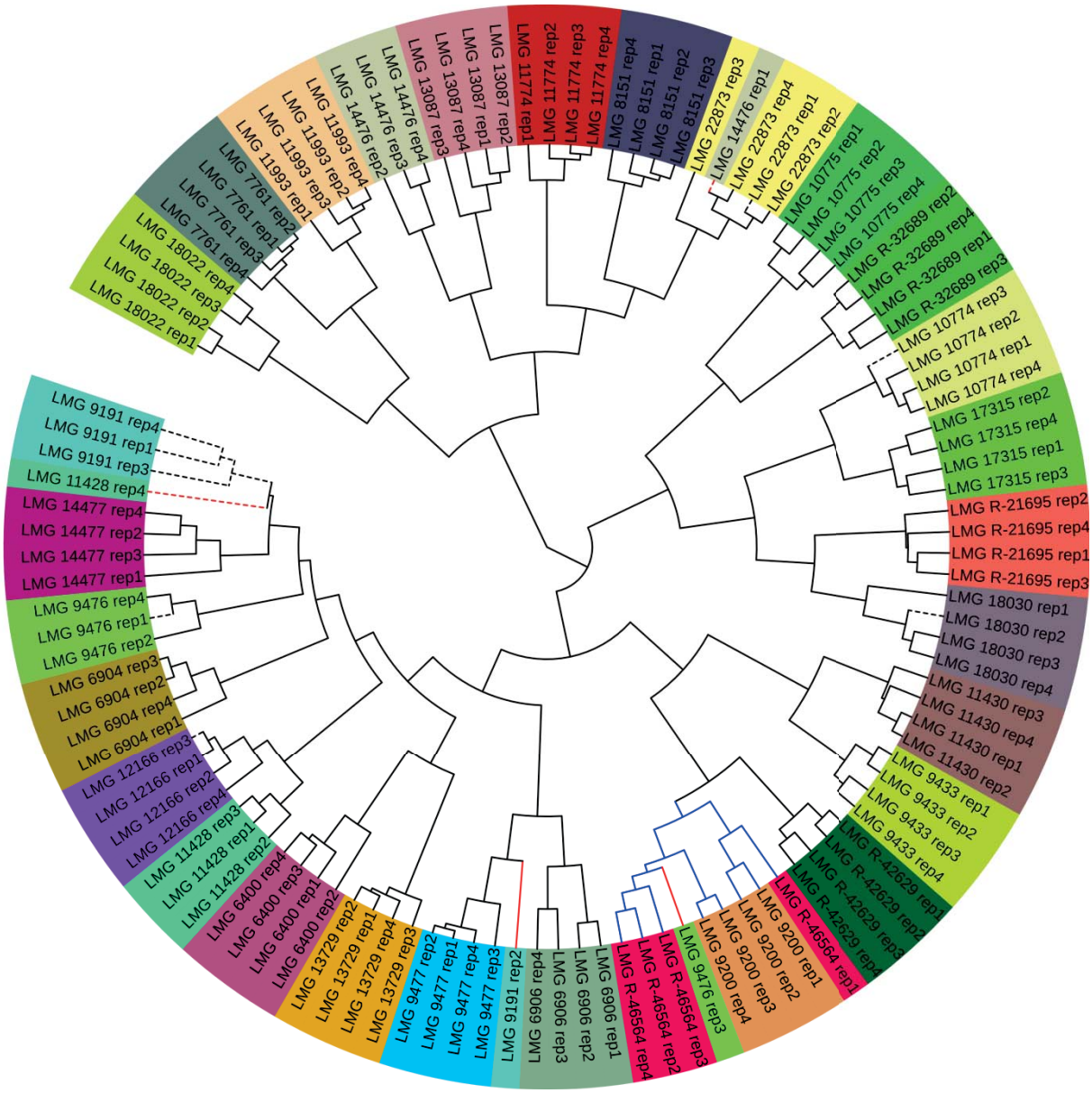
Appendix Figure 3 - 4: *Comparison of different distance measures for the SG stained dataset. Both ecological diversity distance measures (Bray-Curtis [**B**] and Jaccard-Ružička [**J**]) as probability distribution distance measures (Bhattacharyya [**Bh**] and Kullback-Liebler [**KL**]) were compared. Both the Bray Curtis and the Jaccard-Ružička distance were calculated based on the* vegan *package v2.3.4 (Oksanen et al., 2016). The Kullback-Liebler distance was calculated with the* flexmix *package v2.3.13 (Grun and Leisch, 2008). Because the Kullback-Liebler distance measure is an asymmetrical measure, the Kullback-Liebler distance measure was calculated as 1/2(KL(p,p')+KL(p',p)) to make it symmetrical. This is also known as the Jeffrey divergence. To conclude, the Bhattacharyya distance measure was calculated with the* fpc *package v2.1.10. To compare the methods, the performance was calculated as the normalized number of pairs of consecutive samples that are either of the same strain and cluster together or that are of different strains and do not cluster together. This was done at multiple heights to see the overall evolution of the performance. All methods except Bhattacharyya show to reach the maximum performance of 1.*



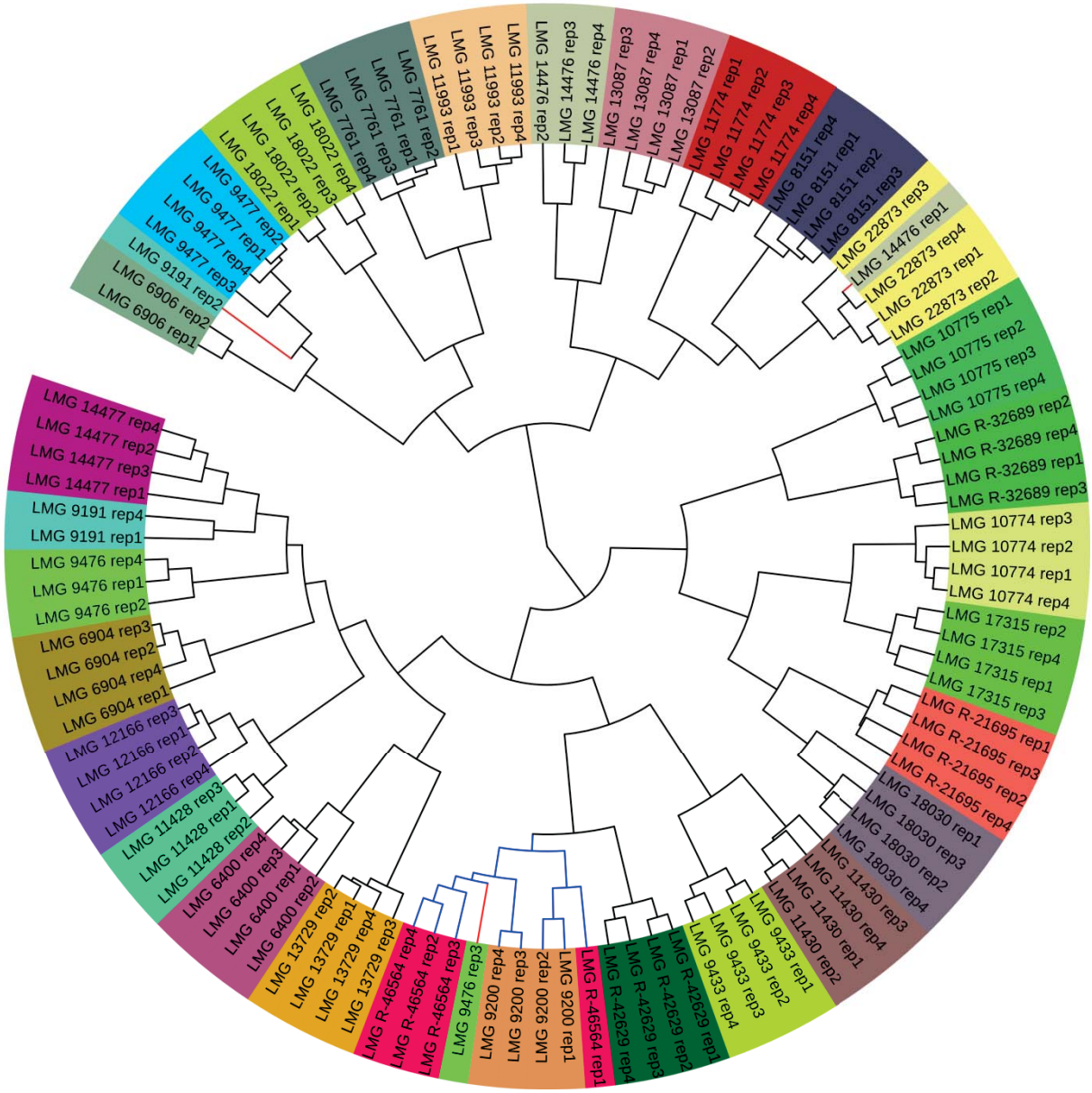
Appendix Figure 3 - 5: *Dendrogram of the flow cytometric fingerprints of the 29* Lactobacillus *strains and species stained with SG. All species and strains were cultivated in biological quadruplicates. After acquisition, the multidimensional space was binned and the density based Jaccard distance measures were calculated for all samples and visualized in a dendrogram constructed with the Ward hierarchical clustering method. Background was not removed and no quality assessment was performed on the data. The different strains are indicated with a different colored label. Clades highlighted in blue indicate strains that could not be discriminated. Nodes highlighted in red show the misclassified fingerprints. Dashed lines indicate samples that showed anomalies in their acquisition after visual inspection of the green fluorescence in function of time.*



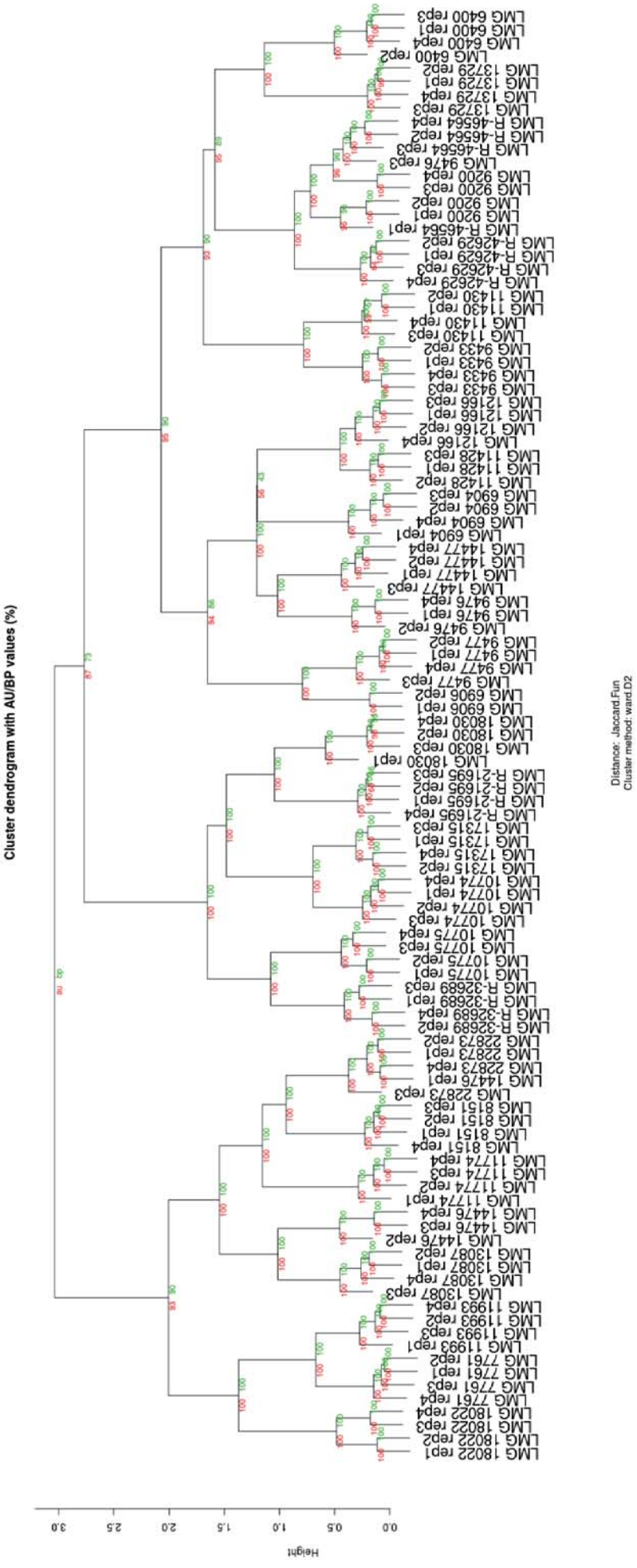
Appendix Figure 3 - 6: *Dendrogram of the flow cytometric fingerprints of the 29* Lactobacillus *strains and species stained with SGPI. All species and strains were cultivated in biological quadruplicates. After acquisition, the multidimensional space was binned and the density based Jaccard distance measures were calculated for all samples and visualized in a dendrogram constructed with the Ward hierarchical clustering method. Background was not removed with a gate and no quality assessment was performed on the data. The different strains are indicated with a different colored label. Clades highlighted in blue indicate strains that could not be discriminated. Nodes highlighted in red show the misclassified fingerprints. Dashed lines indicate samples showed anomalies in their acquisition after visual inspection of the green fluorescence in function of time. Fingerprints of LMG 9200^T (**A,B**), LMG 9477 (**C,D**), LMG 7761 (**E,F**) and LMG 14476 (**G,H**) are shown in **Figure 3 – 1**.*



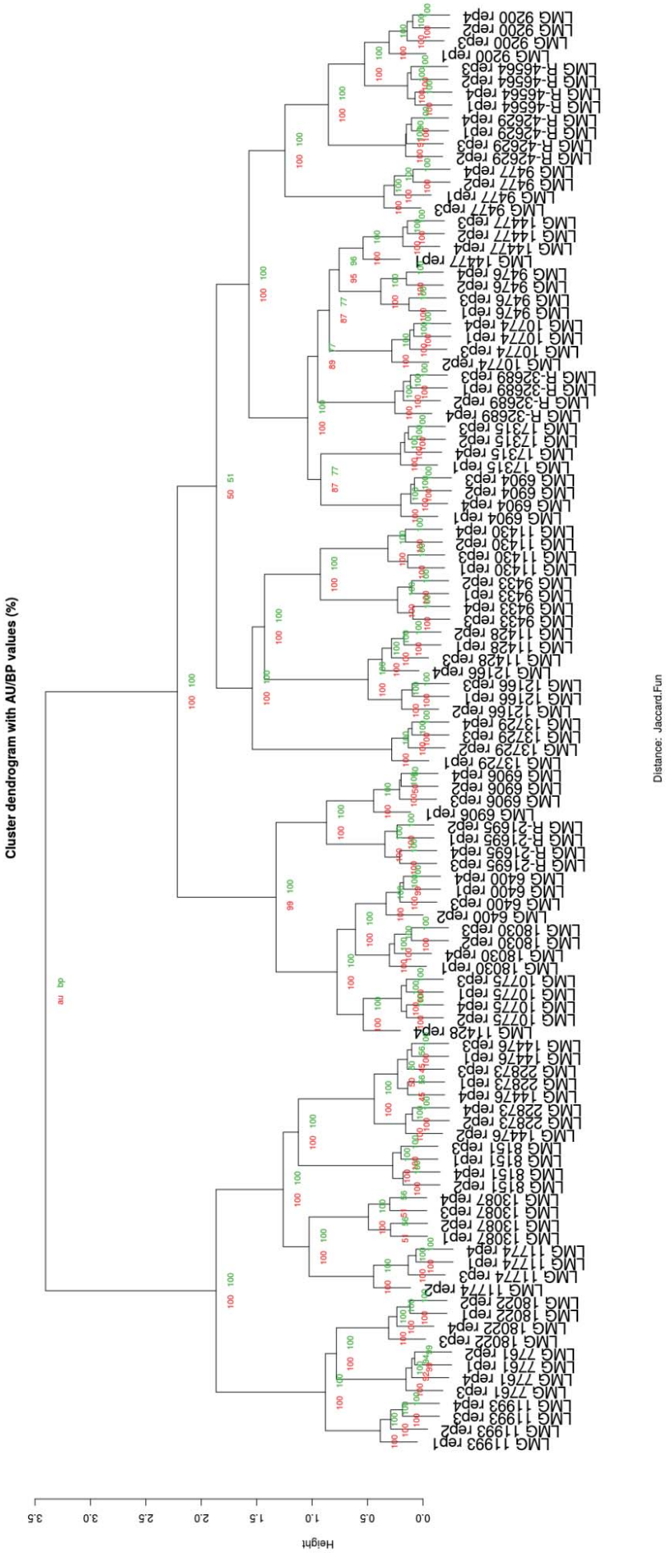
Appendix Figure 3 - 7: *Dendrogram of the flow cytometric fingerprints of the 29* Lactobacillus *strains and species stained with SG. All species and strains were cultivated in biological quadruplicates. After acquisition, background was removed by using a universal gate. Next, the multidimensional space was binned and the density based Jaccard distance measures were calculated for all samples and visualized in a dendrogram constructed with the Ward hierarchical clustering method. No quality assessment was performed on the data. The different strains are indicated with a different colored label. Clades highlighted in blue indicate strains that could not be discriminated. Nodes highlighted in red show the misclassified fingerprints. Dashed lines indicate samples showed anomalies in their acquisition after visual inspection of the green fluorescence in function of time.*



Appendix Figure 3 - 8: *Dendrogram of the flow cytometric fingerprints of the 29* Lactobacillus *strains and species stained with SG. All species and strains were cultivated in biological quadruplicates. After acquisition, flow cytometry data was gated to remove the background and cleaned from anomalies. Subsequently, the multidimensional space was binned and the density based Jaccard distance measures were calculated for all samples and visualized in a dendrogram constructed with the Ward hierarchical clustering method. The different strains are indicated with a different colored label. Clades highlighted in blue indicate strains that could not be discriminated. Nodes highlighted in red show the misclassified fingerprints.*



***Appendix Figure 3 - 9**: Validation of the clustering with the pvclust package of the* Lactobacillus *strains stained with SYBR green I. All samples were measured in biological quadruplicates. First, all samples were gated to reduce the background. Following this, data was cleaned to remove anomalies and the Jaccard distance measures were clustered according to the Ward algorithm. Both the AU (Approximately Unbiased)* p-value *in red and the BP (Bootstrap Probability)* p-values *in green were calculated for the resulting dendrogram. The number of bootstraps was set to 1000 for a small standard error on the calculated* p-values. P-values > 95 *indicate that the clusters are supported by the data.*



Appendix Figure 3 - 10: *Validation of the clustering with the pvclust package of the* Lactobacillus *strains stained with SYBR green I and propidium iodide. All samples were measured in biological quadruplicates. First, all samples were gated to reduce the background. Following this, data was cleaned to remove anomalies and the Jaccard distance measures were clustered according to the Ward algorithm. Both the AU (Approximately Unbiased)* p*-value in red and the BP (Bootstrap Probability)* p*-values in green were calculated for the resulting dendrogram. The number of bootstraps was set to 1000 for a small standard error on the calculated* p*-values. P-values > 95 indicate that the clusters are supported by the data.*

CHAPTER 4

SINGLE-CELL CHARACTERIZATION FOR BIOREACTOR PERFORMANCE

CHAPTER

4

SINGLE-CELL CHARACTERIZATION FOR FERMENTATION BIOREACTOR PERFORMANCE

1 Abstract

Fermentation bioreactors are challenging systems to manage because of the complex biological processes they mediate and the sensitive and sometimes capricious nature of the biology. Monitoring the microbial populations is therefore indispensable for good operational practices to detect possible problems in an earlier stage or to help determine the most important process parameters influencing activity and stability. Most microbial monitoring techniques used today are either indirect or bulk methods, providing only partial information about the microbial populations and potentially masking cell-to-cell heterogeneity. Flow cytometry is a fast, multiparametric, and cheap single-cell method which can be used to characterize isogenic bacterial populations and quantify their phenotypic heterogeneity. In this chapter, we applied flow cytometry to assess the phenotypic heterogeneity of an *E. coli* during a batch fermentation and we establish the relationship between the changes in phenotypic diversity and the substrate depletion thus demonstrating the usability of flow cytometry as an early-warning system for bioreactor monitoring.

Chapter redrafted after:

Buysschaert, B., Props, R., De Mey, M., Boon, N. Using phenotypic diversity for microbial bioreactor monitoring. *In preparation*

2 Introduction

Microbial fermentation is a key technology for our modern society and is used on an industrial scale for the production of a great number of products. At the heart of the fermentation are the microorganisms comprising typically one, but sometimes different, microbial species (Harrison, 1978). Regardless of phylogenetic diversity, fermentation processes are mediated by microbial communities consisting of various (isogenic) subpopulations (Nebe-von-Caron *et al.*, 2000). A detailed understanding of the population dynamics in response to changing conditions is therefore necessary to improve control of the bioprocesses and to understand the apparent capricious nature of the biology (Lidstrom and Konopka, 2010, Verstraete *et al.*, 2007). Monitoring is a vital aspect of the fermentation technology but the majority of the parameters measured are related to the process operation itself while only few parameters are directly related to the microbial populations (Pohlscheidt *et al.*, 2009). Yet, the detection of changing dynamics in an early stage can significantly contribute in preventing issues in reactor performance.

Many methods are available for biomass monitoring but they all provide only limited insight and are hampered by practical constraints. A first approach is to quantify the biomass through, for example, cell dry weight or volumetric measurements. However, these quick and easy methods are imprecise and do not provide information about the cellular physiology (Pohlscheidt *et al.*, 2009). Alternative approaches are based on optics such as optical density (OD) or microscopy measurements. Despite that OD measurements are faster and more accurate than the cell dry weight, they remain a coarse estimation of the cell density, especially at lower concentrations, and do not provide additional information on cellular physiology (Lewis *et al.*, 2014). Microscopy on the other hand is a powerful tool and can, with the use of dyes, reveal more about microbial physiology. Its main drawbacks are the sample preparation and the low throughput. Moreover, the precision and accuracy to count bacteria decreases if the output is based on less than 400 counted cells (Bolter *et al.*, 2002). One of the most interesting methods developed in the last few years is the capacitance measurement to quantify the living cells in suspension. It has already been applied to a large variety of fermentations and cultures (Knabben *et al.*, 2011) yet it does not provide physiological information apart from cell integrity. As none of the aforementioned methods directly measure cellular activity, other methods should be considered. In the case of dissolved oxygen (DO) and exhaust-gas composition analysis, the relation between activity and oxygen consumption of carbon dioxide production has been established (Salmond and Whittenbury, 1985). All aforementioned methods are already well-established and have proven to be useful for bioreactor monitoring, yet they all share some important

shortcomings. Firstly, most of these methods are indirect and do not measure the bacteria. Some parameters such as the CO_2 concentration are a good indicator of microbial activity but fail to provide a deeper insight in the microbiology necessary to anticipate on how the microorganisms are reacting (Muller *et al.*, 2010). Microorganisms can be analyzed with molecular tools such as RNA analysis but these methods are generally time-consuming, costly, and require specialist skills making them only suitable for research purposes. Secondly, all these methods, with the exception of microscopy, are bulk methods returning only a single summary statistic per microbial population measured. This approach may mask subpopulations and the underlying dynamics in the microbial population (Delvigne and Goffin, 2014). As they are undetected, no relationship can be established between these dynamics and the reactor output or operational parameters. Ideally, a single-cell method capable of measuring different physiological parameters should be used to gain better insight in the microbial population dynamics.

Several techniques are available to analyze microorganisms and their behavior on a single-cell level, but only a few are apt to monitor bioreactors as the method in question needs to be fast and cheap to enable a high measuring frequency. Flow cytometry offers a more complete option as it is able to measure thousands of cells and multiple physiological parameters of interest by means of specific fluorochromes. Furthermore, it is possible to automate the process in a reproducible way (Hammes *et al.*, 2012, Besmer *et al.*, 2014, Brognaux *et al.*, 2013) (**Chapter 8**). However, the main bottleneck of flow cytometry is the complex and subjective data analysis which has restricted its use to research purposes (Hammes and Egli, 2010). The complexity of the data analysis is caused by both the quantity of and the nature of the data. Whereas using one fluorochrome poses no difficulty for the interpretation, a combination of fluorochromes makes the data analysis complicated as a good understanding of fluorochrome kinetics and optical properties are necessary (Buysschaert *et al.*, 2016, Roederer, 2002). Furthermore, it is known that the flow cytometric data changes noticeably depending on both the growth conditions and the growth stage (Muller, 2007, Ambriz-Avina *et al.*, 2014, Wang *et al.*, 2010, Buysschaert *et al.*, 2017). Yet, it is difficult to take these small changes into account with the traditional gating-based data analysis. To process the high amount of data and to quantify the subtle changes in the cytometric data, so-called flow cytometric fingerprinting can be used. The purpose of cytometric fingerprinting is to process the data in an objective, automated, reproducible way, and to condense the obtained information in an uncomplicated result. Fingerprinting is not a new approach and the usefulness of this approach has been published in the fields of environmental microbiology and microbial ecology (Koch *et al.*, 2013a, Koch *et al.*, 2013c, Koch *et al.*, 2014, Prest *et al.*, 2013, Props *et al.*, 2016, Buysschaert *et al.*, 2017) but not yet

in the context of bioreactor monitoring. The method described by Props *et al.* (2016) translates the flow cytometric plot in a single summary statistic (*i.e.* the phenotypic diversity metric) and is the most promising method for bioreactor monitoring because of its automated pipeline and because of its suitability for longitudinal data.

In this chapter, we demonstrate that flow cytometric fingerprinting can be used to characterize and track the phenotypic heterogeneity based on the nucleic acid profile in a microbial bioreactor. For this, we monitored twice a batch fermentation of *E. coli* grown in a minimal medium containing glucose as carbon source. Flow cytometric measurements were performed with SYBR green I (SG) as nucleic acid stain on one hand and a combination of SYBR green I and propidium iodide (SGPI) as membrane integrity stain on the other hand to assess the effect of dyes and the additional physiological information they provide. Results showed a good relationship between the fingerprints and the respiration profile and demonstrate that flow cytometric fingerprinting could be used to rapidly detect issues with both dyes.

3 Materials and methods

3.1 Preparation of bacteria cultures

Escherichia coli MG1655 (ATCC 47076) was obtained from the American Type Culture Collection (ATTC). Lysogeny broth (LB) was used to recover the strain from cryovials. LB was composed of 1 % tryptone-peptone (Difco), 0.5 % yeast extract (Difco) and 1 % sodium chloride (VWR). LB agar plates contain the same components as LB with the addition of 1 % agar. For growth experiments in flasks a defined medium was used. This defined medium contained 18 µM $FeCl_2 \cdot 4H_2O$ (Merck), 26 µM $CaCl_2$ (Merck), 10.2 µM $MnCl_2 \cdot 4H_2O$ (Merck), 2.2 µM $CuCl_2 \cdot 2H_2O$ (Sigma), 2.1 µM $CoCl_2 \cdot 6H_2O$ (Merck), 6.9 µM $ZnCl_2$ (Merck), 0.4 µM H_3BO_4 (Merck), 40.3 µM $Na_2EDTA \cdot 2H_2O$ (Fluka), 3 µM thiamine·HCl (Sigma), 0.4 µM $Na_2MoO_4 \cdot 2H_2O$ (Fluka), 2.7 µM SeO_2 (Sigma), 37.4 mM NH_4Cl (Merck), 37.8 mM $(NH_4)_2SO_4$ (Merck), 22 mM KH_2PO_4 (Acros), 42 mM KH_2PO_4 (Acros), 40 mM MOPS (Sigma), 2 mM $MgSO_4 \cdot 7H_2O$ (Fluka), 8.6 mM NaCl (VWR), and 83.3 mM glucose·H_2O (Cargill). Trace element solution and molybdate solution were sterilized with a bottle top filter (Corning PTFE filter, 0.22 µm). Carbon source and $MgSO_4$ were dissolved in 200 ml H_2O and autoclaved separately to avoid Maillard reaction. All other components (except trace element and molybdate solution) were dissolved in 800 mL H_2O and set to pH 7.0 with 1 M K_2HPO_4 (Acros) solution. The sterile trace elements and molybdate solution were added after

autoclaving. Growth experiments in bioreactors were performed in defined medium without MOPS buffer.

A preculture from one colony on a LB-plate in 5 mL LB was grown during 8 h at 37°C on an orbital shaker (LS-X AppliTek orbital shaker) at 200 rpm. From the preculture, 2 mL was transferred to 100 mL of MM-flask medium in a 0.5 L shake flask and incubated overnight (16 h) at 37°C on an orbital shaker at 250 rpm. The inoculum was injected into the vessel with a sterile syringe through an inoculation port sealed with a septum and was 5% of the bioreactor working volume.

3.2 Batch Fermentations

Two batch fermentations were carried out in a 2 L Biostat B culture vessel (Sartorius-BBI Systems) with a working volume of 1.5 L. Temperature (37°C), pH (7.0), stirring rate (600 rpm), and airflow rate (1.5 L/min) were controlled by the Biostat B control unit. For maintaining the pH at 7.0, 0.05 M H_2SO_4 (VWR), and 2 M KOH (Sigma) were used. Gas that exits the bioreactor passed through an exhaust cooler (Thermostat DC1 equipped with cooler K20, Haake) set at 4°C. O_2 and CO_2 concentration in the exhaust gas were measured by a URAS 10E off-gas analyzer from Hartmann and Braun. The maximum sampling frequency was one sample every 30 min for Batch 1 and every 15 min for batch 2. To avoid foaming, a solution of 10% silicone antifoaming agent (BDH 331512K, VWR) was added to the culture vessel. All parameters described above were monitored using MFCS/win software (Sartorius AG).

In its interior, the bioreactor contains a harvest pipe which exits the vessel through a sampling port. The harvest sampling pipe consists of a Bio-Rad HPLC tubing (Bio-Rad). Outside the vessel, the sampling pipe is connected to a Masterflex 16 tubing (Cole Parmer) that creates a circuit back to the vessel and that includes a harvest port with a sterile glass sampling vial. The system has been designed to obtain a low retention time of the culture broth in the tubing. Such a sampling system is required to have a reliable method for the rapid assessment of cell physiology. This system is referred to as “rapid sampling loop”. During batch experiments, a sample for OD_{600} and extracellular measurements was taken each hour using the rapid sampling loop and the cold stainless bead sampling method. For both batches, samples were taken every 30 minutes. Around the time of metabolic switch, samples were taken every 15 minutes for batch 1 and every 5 minutes for batch 2.

Organic acids were determined by high performance liquid chromatography (HPLC) on a Varian Prostar HPLC system (Varian), using an Aminex HPX-87H column (Bio-Rad)

equipped with a one centimeter reversed phase precolumn, with 5 mM H_2SO_4 (0.6 mL/min) as mobile phase and heated at 65°C. Detection was done by a dual-wave UV−vis (210 and 265 nm) detector (Varian Prostar 325) and a differential refractive index detector (MERCK LaChrom L-7490, Merck).

3.3 Flow Cytometry

Samples were analyzed on an Accuri C6 (BD Biosciences) with a blue (488 nm) and red (640 nm) laser. Standard optical filters were used and included FL-1 (530/30 nm), FL-2 (585/40 nm) and FL-3 (670 LP) for the blue laser and FL-4 (675/25 nm) for the red laser. An optimized staining protocol was used from Van Nevel *et al.* (2013). Samples were diluted in with physiological solution filtered over a 0.22 µm syringe filter (Merck) to a concentration of approximately 10^6 cells/mL for a more precise cell counts. Two independent staining procedures were used to measure both total and intact cells. Bacteria were stained with 10 µL/mL of SYBR Green I (SG, Invitrogen, 100x diluted in DMSO from stock) for total cell counting. For intact cell counts, Propidium Iodide (PI, Invitrogen, final concentration 4 µM) is added together with SG. The samples were then incubated for 13 minutes at 37°C to optimize the staining. All measurements were performed in triplicate.

All data was extracted from the proprietary Accuri C6 *CSampler* software version 1.0.264.21 in the flow cytometry standard (FCS 3.0) format and subsequently imported into R v3.4.0 (R Core Team, 2015) through the functionality offered by the *flowCore* package v1.42.2 (B. Ellis *et al.*). Data was first log transformed and then normalized by dividing all values by the maximum fluorescence intensity signal. No compensation was applied. Gating to reduce the background was performed in R studio using the *flowCore* package on both SG and SGPI stained samples. A 0.22 µm-filtered control was used to determine the position of the background and a heat-killed sample was used to determine the position of the permeabilized cell population (Berney *et al.*, 2007). Based on this, a universal gate was constructed to remove as much background as possible without removing the permeabilized cell population although the distinction between PI positive cells and the background can be difficult for some samples (**Appendix Figure 4 - 1**). Additionally, a stained sample of the dilution buffer was used to assess the quality of the dilution buffer and of the stain. The data quality was evaluated and improperly acquired data was cleaned using the *flowAI* package v1.4.3 (Monaco *et al.*, 2016) in order to remove anomalies in the data related to changes in flow rate, unstable signal or outliers in the lower limit of the dynamic range. Samples which failed the QC were removed from the dataset. To avoid a bias due to different cell numbers, 10 000 events per sample were randomly selected. This was done after background was

removed with the universal gate. Next, a single-step discretization ('binning') and Gaussian bivariate density estimation was performed on the selected parameters (green and red fluorescence, FSC-H and SSC-H) using the *KernSmooth* package v2.23.15 (Wand, 2015). An equally spaced grid (binning grid) of 128 x 128 was fixed for each bivariate density estimation using the *flowFDA* package v1.0. All bivariate density estimations were concatenated to a one-dimensional feature vector which we refer to as the fingerprint. Subsequently, phenotypic alpha diversity was calculated according to the publication of Props *et al.* (2016) where Hill diversity indices are applied to describe the diversity of conceptual phenotypes within and between samples. The code is available at https://github.com/rprops/Phenoflow_package. To test if any of the operational parameters were correlated to the phenotypic diversity, the Spearman's rank correlation coefficient was calculated by means of the package *Hmisc* v4.0.3 (Harrell, 2017).

4 Results and discussion

To explore the added value of flow cytometric fingerprinting for bioreactor monitoring, an *E. coli* batch fermentation with glucose as carbon source was set up. The fermentation was monitored using conventional parameters such as respiration (CO_2 production) alongside with flow cytometry. The impact of two different dyes and several data analysis parameters on the flow cytometry results were compared in this chapter. Two independent fermentation runs were conducted within a period of ten days.

4.1 Conventional parameters

The fastest conventional method for monitoring microbial activity during a batch fermentation is online CO_2 measurements. Results of the online measurements showed that the microbial respiration increases from an atmospheric concentration of 0.04% to 3.01% CO_2 after 6 h 10 min for batch 1, and to 2.95% CO_2 after 6 h 28 min for batch 2, after which respiration (CO_2 production) decreases to 0.58% and 0.61% CO_2 in 15 min for batch 1 and 2 respectively (**Figure 4 - 1**). CO_2 production increases again for both fermentations until 10 h 4 min for batch 1 and 10 h 28 min for batch 2 to decrease again afterwards and decrease steadily from around 0.3% to 0.08% CO_2 for the rest of the reactor run. The first increase of CO_2 production coincides with the increase of acetate and lactate, both by-products of the glucose respiration.

The first decrease is due to the depletion of glucose and a reduction of the microbial growth rate as can be seen from the cell concentrations (**Figure 4 - 2**). The second increase in respiration coincides with the decrease in acetate and, to a lesser extent, lactate

concentrations until the concentration of acetate is almost 0 g/L and the concentration of lactate 0.5 g/L. The change in substrate concentration and the fluctuation of the microbial activity indicate that the *E. coli* population first consumes the glucose to produce acetate and lactate as by-products. When glucose is depleted, the bacteria consume acetate and lactate as carbon source instead. This behavior of *E. coli* has been documented extensively before and demonstrates that the fermentation runs as expected (Salmond and Whittenbury, 1985).

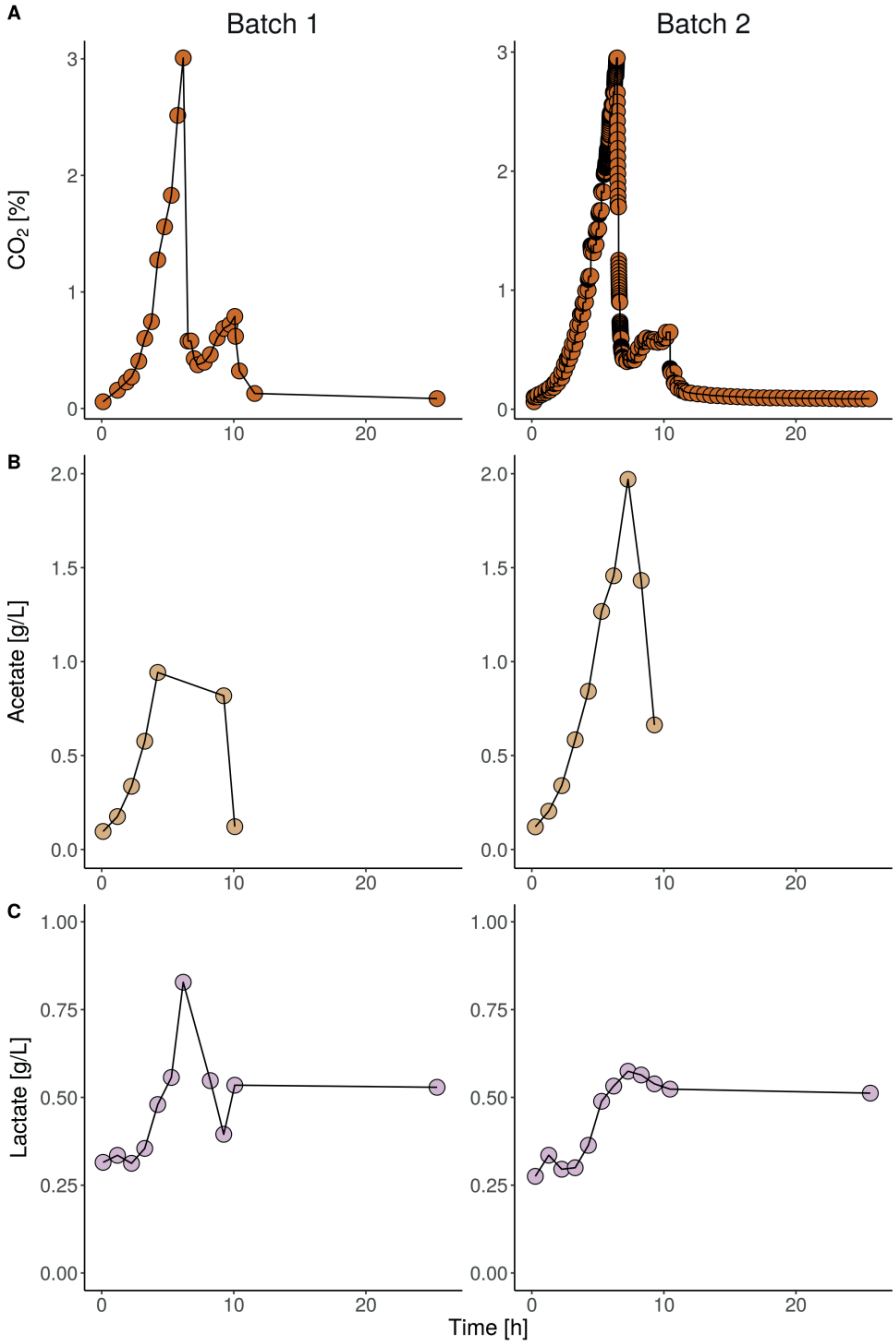


Figure 4 - 1: *The CO_2 concentration expressed as percent of the total air volume for batch 1 and batch 2 (**A**). The acetate concentration in the reactor for batch 1 and batch 2 (**B**) and lactate concentration for batch 1 and 2 (**C**) in function of time. Both compounds are expressed in g/L sample and were measured with HPLC. The gas composition was measured online with an off-gas analyzer.*

4.2 Cell density

Flow cytometry enables the distinction between intact (SG stained) and permeabilized (SG and PI stained) cells. The results of both batches show a consistent pattern with an increase and peak in the percentage of permeabilized cells before the first decrease of the CO_2 concentration at 5 h 45 min and 6 h 23 min for batch 1 and 2 respectively, and a peak in the percentage of permeabilized cells after the second decrease of the CO_2 concentration at 10 h 11 min and 10 h 53 min for batch 1 and 2 respectively (**Figure 4 - 2**). The appearance of the first increase of permeabilized bacteria appears 25 min and 5 min before the first decrease in the CO_2 concentration, and the second increase of permeabilized bacteria, 7 min and 25 min after the second decrease in CO_2 concentration for batch 1 and 2 respectively.

Discrepancies between both batches can be explained by the different sampling frequencies for off-gas analysis and flow cytometry. While samples for flow cytometry were taken every 15 min for batch 1, samples were taken every 5 min for batch 2 for the first CO_2 peaks. Afterwards, samples were taken every 30 minutes for both batches. Regardless of the percentage of permeabilized cells, an increase and subsequent decrease over time is noticeable in both batches at comparable time points. The cell concentration shows, regardless of cell integrity, the typical characteristics of a sigmoidal growth curve (**Figure 4 - 2a**). A potential cause for the increase in permeabilized cells is the substrate depletion leading to cell death, which is determined by cell permeability for this experiment (Hewitt *et al.*, 2000). However, to determine if cells are actually dead, other techniques should be used aside from PI uptake because membrane permeability could be induced by other factors which do not always lead to cell death (Amor *et al.*, 2002, Shi *et al.*, 2007) . Studies using molecular methods showed that, in the event of substrate depletion, *E. coli* cells can express more transmembrane transport proteins to allow the scarce nutrients to enter the cells (Egli, 2010, Wick *et al.*, 2001). These membrane modifications and enhanced permeability can increase the PI uptake (Brognaux *et al.*, 2014). The speed and timing at which the cell permeability of our observations changed, agrees with this hypothesis.

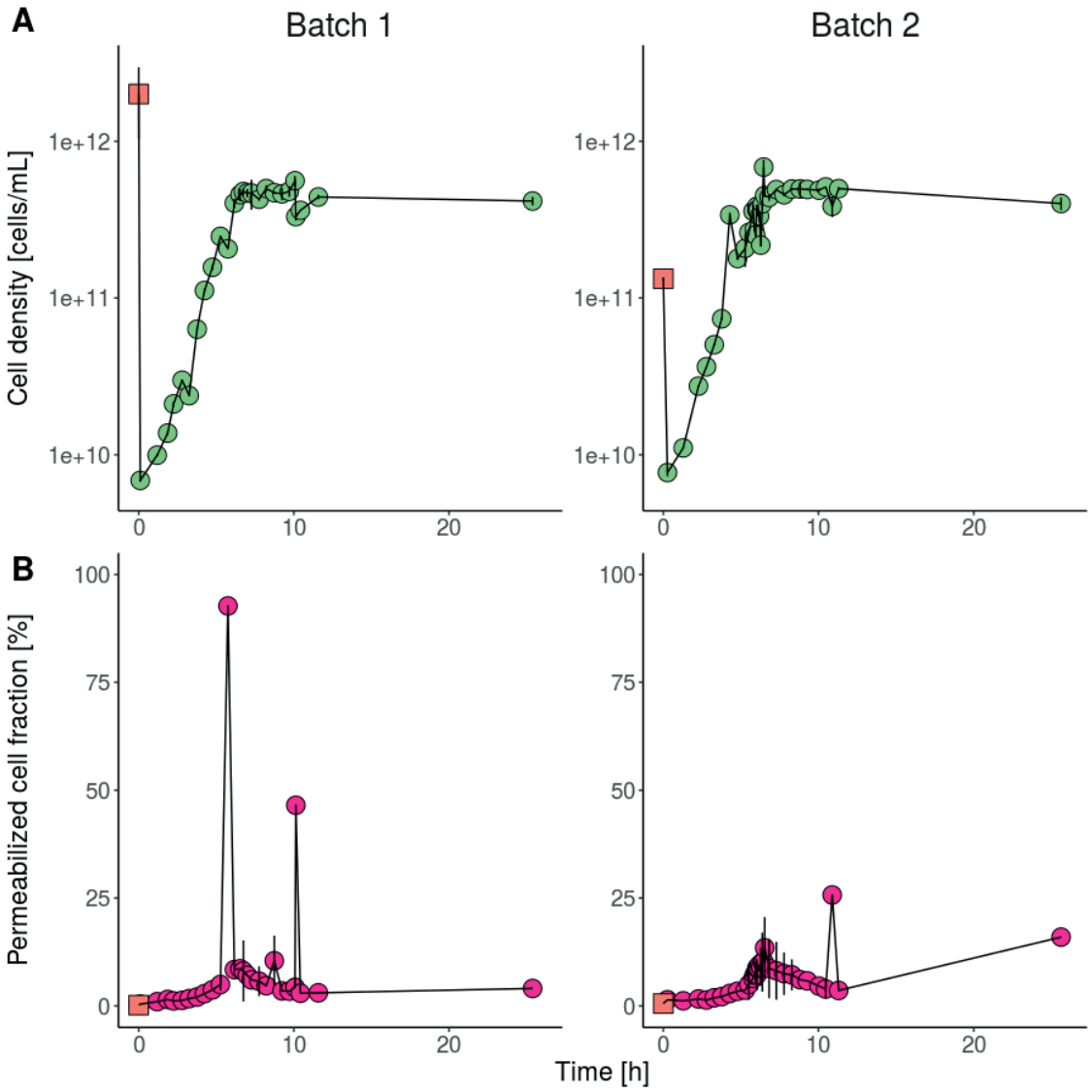


Figure 4 - 2: *Cell concentration expressed as cells/mL (**green, A**) and the fraction of permeabilized cells expressed in percentage (**red, B**) in function of time. The data for both batch 1 (left) and batch 2 (right) are presented here. Cell concentrations were determined with flow cytometry by means of gating. The standard deviation of the cell concentrations was calculated with the triplicate measurements. The first sample represents the inoculum.*

4.3 Flow cytometric fingerprinting

Besides cell concentration and the distinction between intact and permeabilized cells, fingerprinting algorithms can be used to extract more information from the flow cytometry data. Various ways of fingerprinting are available but for time-series data, the method described by Props et al. (2016) is promising. With this method, three Hill number diversity indices can be calculated, *i.e.* D_0, D_1, and D_2. D_0 describes the total number of bins – multidimensional density plot discretizations – containing cells, regardless of the concentration of cells per bin. D_1 and D_2 also take the relative density of the number of cells

per bin into account. In case of D_2, more weight is given to those bins containing many cells in comparison to D_1. As all cells in one bin are supposedly similar in their optical characteristics, they can therefore be categorized as being from a same conceptual phenotype. The data shows that, just before the respiration decreases for the first time, the phenotypic diversity D_2 increases noticeably in both batches and for both SGPI (**Figure 4 - 3**) and SG (**Appendix Figure 4 - 2**). This observation is similar for all three indices (data not shown). For batch 1, a first maximum is reached 25 minutes before the CO_2 concentration reaches its maximum and at the same time as the permeabilized cell population peaks. For batch 2 the first local maximum of three consecutive samples is reached 21 minutes before the CO_2 concentration reaches a maximum.

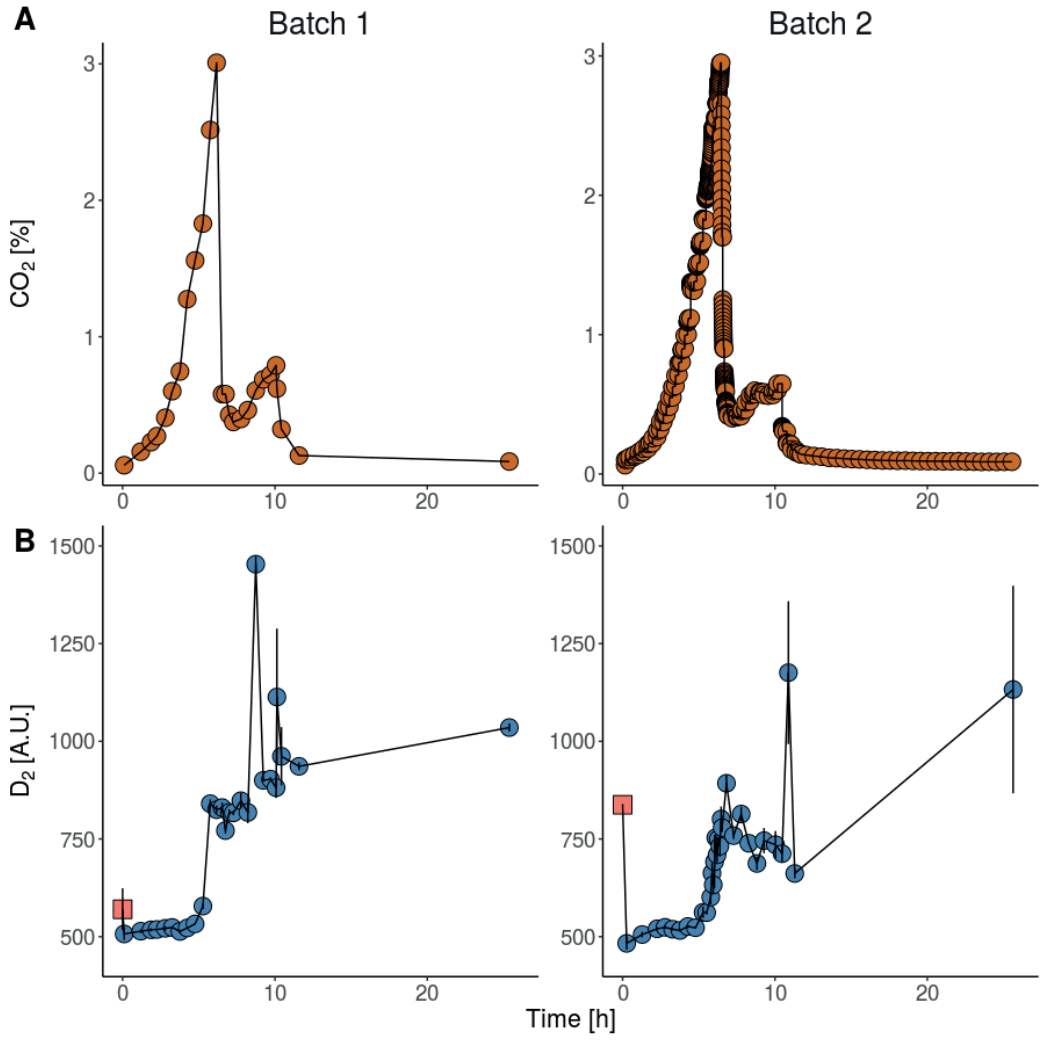


***Figure 4 - 3**: Respiration is illustrated by the concentration of CO_2 expressed as percentage of air volume in function of the time since the start of fermentation for batch 1 (left) and batch 2 (right) (**A**). Hill number diversity index D_2 describing phenotypic diversity based on the flow cytometry data after SGPI staining is expressed in arbitrary units (**B**). The diversity indices are estimated averages based on bootstrapping (n=100) and the standard error of the triplicates was added. Prior analysis, background fluorescence was removed by gating. The first sample represents the inoculum.*

An explanation for these results could be the appearance of a permeabilized cell population after the first decrease in respiration. However, a comparison of the results after staining with SGPI (**Figure 4 - 3**) or SG (**Appendix Figure 4 - 2**) indicate that both dyes exhibit similar trends. As SG is not affected by damaged cells, it is improbable that the permeabilized cell population causes the change in phenotypic heterogeneity. Both dyes have the property of binding to nucleic acids and a change in the nucleic acid composition is therefore a more likely explanation. This hypothesis is supported by contrasting the density plots of the peak samples with the density plots of the samples with low phenotypic diversity. For both batches stained with either SGPI (**Figure 4 - 4**) or SG (**Appendix Figure 4 - 3**), the mean fluorescence intensity shifts down, suggesting changes in the nucleic acid profile of the cell population. The substrate depletion, coinciding with the decrease of the respiration, induces a metabolic switch which slows down growth. Due to this diauxic shift, a fraction of the cell population enters a non-growing state while another fraction will metabolize the remaining nutrients (Kotte *et al.*, 2014). Taymaz-Nikerel *et al.* (2010) also showed that the biomass composition of an *E. coli* cultured in minimal medium changes with a changed growth rate. The decrease of growth induces a decrease in fluorescence intensity as less dsDNA will be available in the cells and metabolizing cells also contain relatively more RNA which yields weaker fluorescent signals upon binding with SYBR green I (Solopova *et al.*, 2014, Johnson and Spence, 2010). Thus, the increased diversity D_2 suggests that, after substrate depletion, the *E. coli* cell population shows a higher phenotypic diversity.

Several mechanisms have been reported to induce phenotypic heterogeneity. A first mechanisms is related to the genetic differences but, as phenotypic differentiation occurs at rates higher than any known mutational mechanism and as it is robust against the suppression of mutational mechanisms, it is not considered as the most prominent cause (Ackermann, 2015). Another explanation is the stochasticity of gene expression (Elowitz *et al.*, 2002, Blake *et al.*, 2003) and stochastic partitioning of molecules at cell division (Huh and Paulsson, 2011) although the latter is less important in the case of symmetrical division. Next to intrinsic mechanisms also extrinsic mechanisms such as cell-to-cell communication (*e.g.* quorum sensing) and environmental fluctuations can lead to phenotypic differentiation. The response to the latter is known as a bet-hedging strategy of bacteria to survive in challenging environments. Scientific studies reported this before based on observations of the expression of fluorescent markers (Acar *et al.*, 2008, Solopova *et al.*, 2014) and a well-known example is the occurrence of slow growing persister cell subpopulations that are resistant to antibiotics in contrast to the antibiotic sensitive cells growing at normal rates (Balaban *et al.*, 2004, Keren *et al.*, 2004). Kotte *et al.* (2014) reported how *E. coli* cells diversified in two distinct phenotypes based on their growth rate after glucose depletion. Depending on the substrate

concentration, a larger cell population enters a non-growing state with low metabolic activity thus explaining the decreased fluorescence intensity. However, Kotte *et al.* (2014) reported that the phenotypic diversification is a response to the metabolic switch while our results show that the diversification occurs before the metabolic switch indicated by the CO_2 concentration.

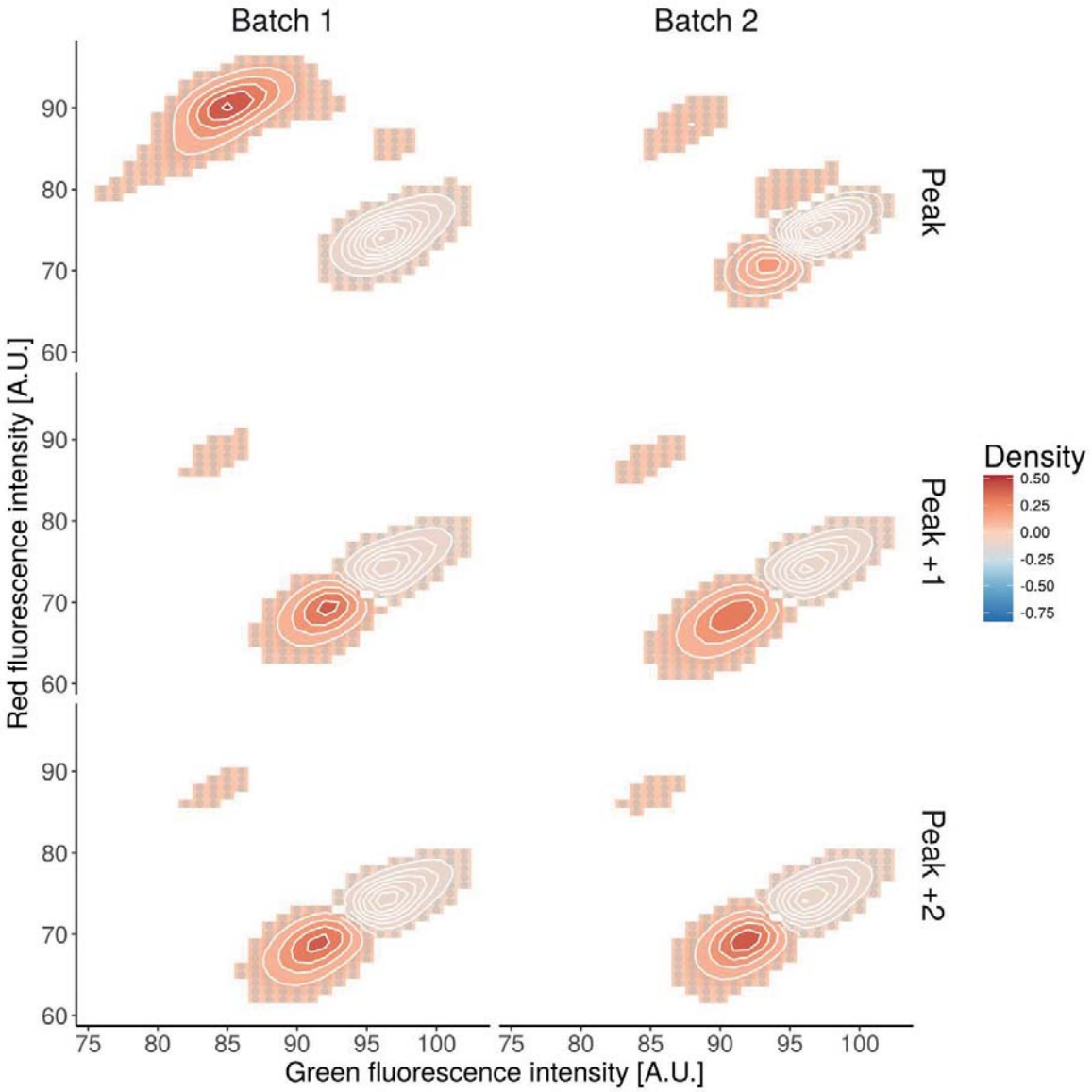


***Figure 4 - 4**: Contrast of the density plots of the peak samples with the density plots of the samples with low phenotypic diversity for both batch 1 (**left**) and batch 2 (**right**). First, only the peak sample is considered (**above**) but then the next sample (**peak +1**) and the sample after that is contrasted (**peak +2**). The red color indicates a region with higher density in the samples with high phenotypic diversity D_2 in contrast to the samples with low phenotypic diversity D_2. Inversely, the regions in blue indicate regions of lower density in the samples with high phenotypic diversity D_2 in contrast to the samples with low phenotypic diversity D_2. Samples were stained with SGPI and background was removed prior analysis. If the difference between the two communities is lower than 0.02, no contrast value is shown on the graphs, which causes the appearance of different clusters.*

4.4 Correlation with operational parameters

In order to understand the relationship between phenotypic diversity changes and operational parameters, the correlation between the Hill number diversity indices and the different operational parameters was calculated with a Spearman's rank correlation coefficient. Furthermore, the correlations coefficients can show which dye or gating strategy is best. A comparison of the resulting coefficients for the different batches, dyes, and gating strategies, show that the gated data correlates better with parameters such as optical density or acetate and lactate concentrations (**Table 4 - 1**). As all these parameters are related to the microbiology, the correlation was expected. Gating the data reduces the noise created by the background fluorescence, thus improving the correlation with the biological parameters. This result also confirms our previous findings (**Chapter 3**). Also, no important difference can be seen between SG and SGPI stained data or the different diversity indices. No clear positive or negative correlation was found between the CO_2 or O_2 concentration. It is important to note that the correlations between flow cytometry and CO_2 and O_2 concentrations should be interpreted with caution as the dataset was incomplete for many data-points. As a consequence, the pairwise correlations coefficients were not always calculated on the same number of data-points, potentially over- or underestimating the correlations.

Using flow cytometry to monitor bioreactors is a promising line of research and several attempts have been made in the past (Brognaux *et al.*, 2013). As illustrated in this chapter, one of the clear advantages of flow cytometry for bioreactor monitoring is the possibility to count cells and characterize physiology *e.g.* viability (Hewitt *et al.*, 2000, Nebe-von-Caron *et al.*, 2000, Delvigne and Goffin, 2014). For other physiological traits, a plethora of fluorescent dyes is available making the possibilities almost limitless. Because of the complex nature of the staining kinetics and the data output, different pipelines of fingerprinting were developed to facilitate the data analysis and interpretation. However, few have been made to use these fingerprinting techniques to monitor bioreactors. Here we show that more complementary information to cell viability and operational parameters can be extracted from the flow cytometry complementary data. Despite the promising results, flow cytometry still has some shortcomings that should be resolved before the method could be used as a proper monitoring tool. A first issue is related to the possibility to measure online. Some self-made prototypes have been developed for research (Brognaux *et al.*, 2013, Besmer *et al.*, 2014) yet no such features are available on commercially available cytometers. Some third-party companies offer automated online sampling technology but this increases the price of the instrument. A second issue is related to the dye chemistry. As bacteria are small, fluorochromes are necessary to visualize them. Despite the advantages of assessing specific features with selective fluorochromes, this approach limits the measured biological

information as only labeled features are measured. Moreover, almost all dyes necessitate an incubation time which decreases the rapidity of the method. For example, SG and SGPI require a minimum incubation time of 13 minutes while our results show that cell permeability or phenotypic diversity can change markedly in a less than 15 minutes. Research on faster staining protocols or alternative dyes capable of revealing phenotypic heterogeneity is therefore needed. Alternatively, the applicability of a label-free approach such as, for example, Raman spectroscopy or FT-IR should be investigated.

Table 4 - 1: *Spearman's rank correlation coefficients between the Hill diversity indices calculated from the flow cytometry data as described by Props et al. (2016) and operational parameters. The numbers were calculated for both batches, both stain types and with and without background removal.*

			CO_2	pO_2	Temperature	Lactate	Acetate	OD_{600}
			SG					
Batch 1	Ungated	D_0	0.24	0.07	0.14	0.56	0.56	-0.16
		D_1	0.41	0.26	0.05	0.59	0.66	-0.21
		D_2	0.40	0.02	0.42	**0.70**	0.58	0.42
	Gated	D_0	0.27	0.20	0.61	0.66	0.45	**0.83**
		D_1	0.14	0.32	0.52	0.65	0.41	**0.81**
		D_2	0.12	0.28	0.51	0.63	0.44	**0.78**
Batch 2	Ungated	D_0	0.18	0.28	0.20	-0.07	0.16	0.02
		D_1	0.35	0.29	0.30	-0.05	-0.20	0.24
		D_2	0.41	0.26	0.26	0.26	0.05	0.29
	Gated	D_0	0.01	0.32	0.56	0.61	**0.73**	**0.98**
		D_1	0.13	0.16	0.51	**0.88**	**0.79**	**0.98**
		D_2	0.16	0.08	0.53	**0.95**	**0.82**	**0.86**
			SGPI					
Batch 1	Ungated	D_0	0.40	0.14	0.29	0.68	0.38	0.01
		D_1	0.33	0.03	0.33	0.73	0.50	0.08
		D_2	0.15	0.25	0.56	0.65	0.39	0.58
	Gated	D_0	0.15	0.35	**0.70**	0.67	0.48	**0.91**
		D_1	0.11	0.43	0.64	0.69	0.43	**0.93**
		D_2	0.05	0.43	0.63	0.62	0.41	**0.81**
Batch 2	Ungated	D_0	0.54	0.60	0.35	-0.18	0.23	-0.40
		D_1	0.21	0.01	0.32	0.22	0.24	0.26
		D_2	0.10	0.11	0.25	0.59	0.61	0.43
	Gated	D_0	0.15	0.41	**0.75**	0.61	**0.77**	0.52
		D_1	0.10	0.19	0.52	**0.83**	**0.83**	**0.98**
		D_2	0.17	0.07	0.55	**0.87**	**0.87**	**0.93**

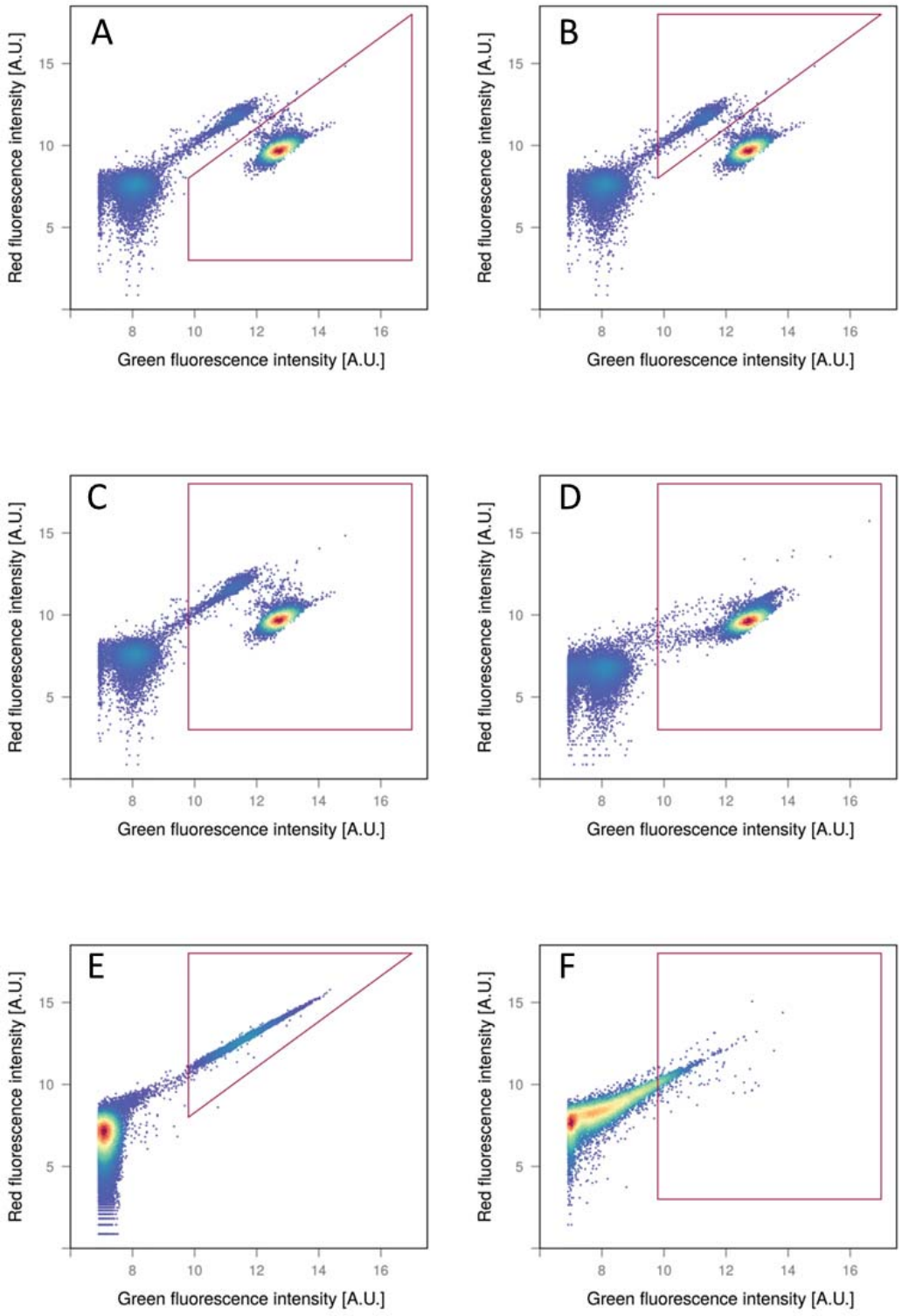
5 Conclusions and perspectives

The possibility of flow cytometry to determine both cell density and cell physiology in high throughput makes it the method of choice to monitor the microbial population in bioreactors. In this chapter we demonstrate that much more information can be extracted from these measurements by flow cytometric fingerprinting. Our results show that the operational phenotypic diversity increases approximately 20 minutes before CO_2 concentration decreases, making flow cytometry faster than exhaust-gas analysis. Based on literature and other research, we conclude that the increased cytometric diversity is caused by a phenotypic diversification as a consequence of substrate depletion. The emergence of metabolically different subpopulations with different nucleic acid profiles explains why fingerprinting with only one nucleic acid dye, SYBR green, is capable of detecting the population dynamics. Our findings open possibilities in the context of fundamental research, but also in the context of industrial bioprocesses, in order to detect disturbances in reactor performance in an early stage and to assist in steering operational parameters. We made a first attempt to correlate the phenotypic diversity indices with operational parameters and found good correlation with the biological parameters. A more in-depth research with variable operational conditions could help to gain a better insight in the relationship between reactor operations and phenotypic diversity. Moreover, also the relationship between yield and phenotypic heterogeneity should be established under different conditions.

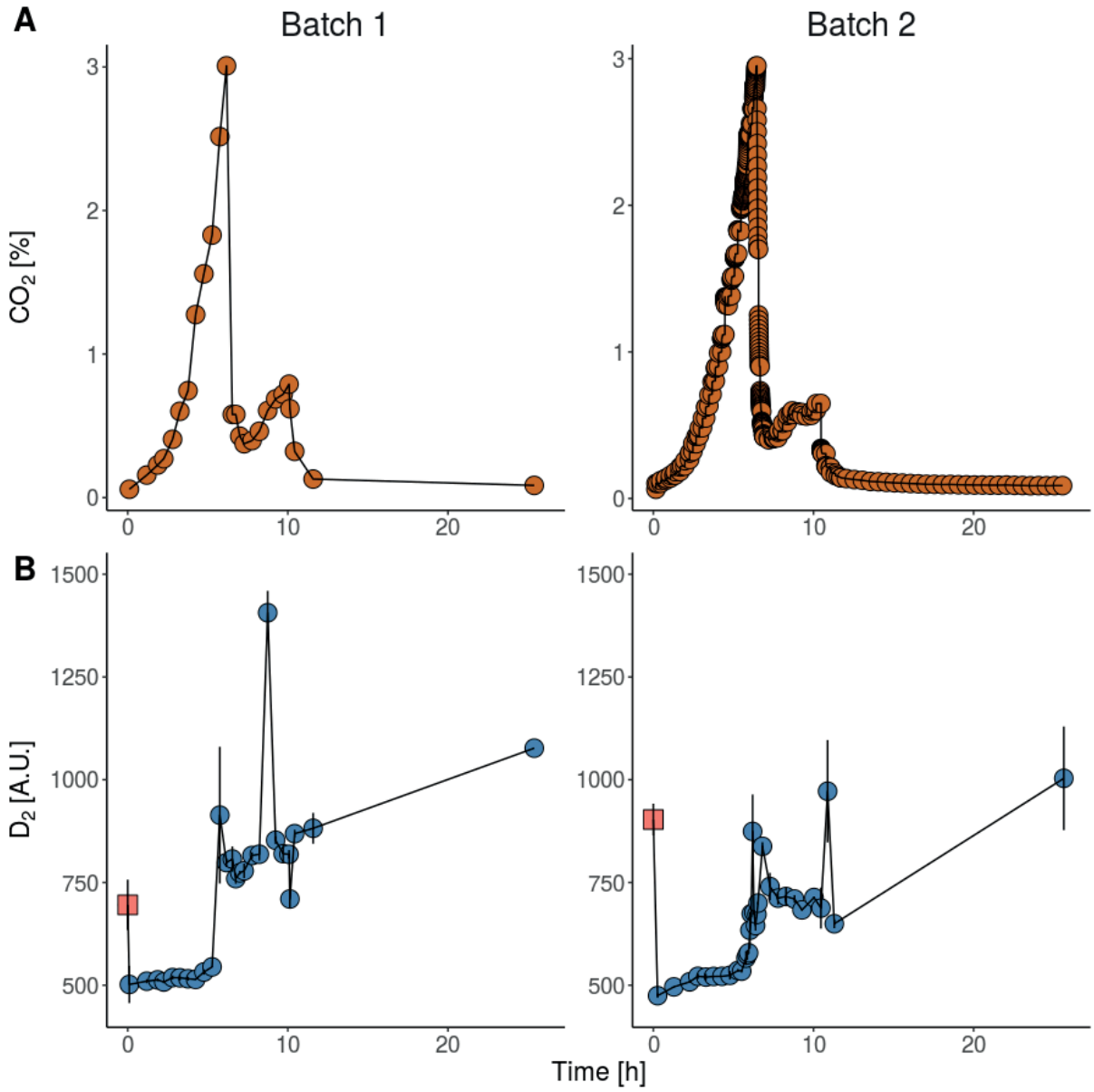
6 Acknowledgements

This work was supported by the project grant SB-131370 of the IWT Flanders, the Geconcerteerde Onderzoeksactie (GOA) of Ghent University (BOF09/GOA/005), and the IMPROVED project, subvented by The interreg V “Vlaanderen-Nederland” program, a program for transregional collaboration with financial support from the European Regional Development Fund. More info : www.grensregio.eu. A special thanks to thank Jana De Bodt, Dries Duchi, and Stijn Verweire for the help in the lab. Also thank you to thank Bart De Gusseme, Charlotte De Rudder, and Cristina García-Timermans for improving the manuscript.

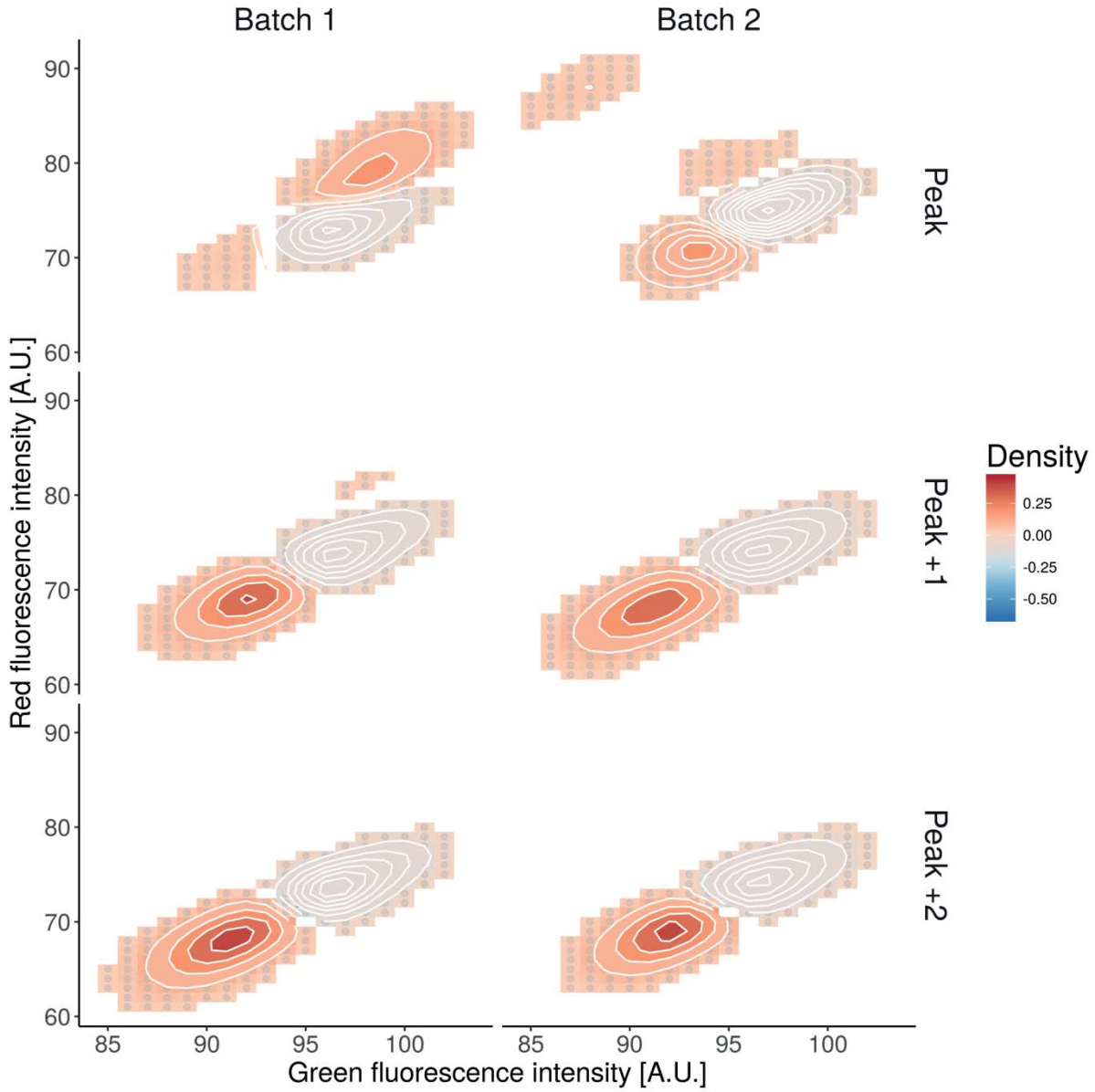
7 Appendix – Supplementary information for chapter 4



Appendix Figure 4 - 1: *Illustration of the gating strategy. The gates to enumerate the intact cells (**A**) and the permeabilized cells (**B**) are shown for an SGPI sample. The universal gates for SGPI (**C**) and SG (**D**) samples were used to reduce the background from the samples prior fingerprinting. A heat-killed sample was used to determine the position of the permeabilized cells (**E**) and a filtered sample of fresh medium was used to position the background (**F**).*



***Appendix Figure 4 - 2**: Hill number diversity index D_2 to describe phenotypic diversity based on the flow cytometry data after SG staining are expressed in arbitrary units in function of the time since the start of fermentation for batch 1 (**left**) and batch 2 (**right**) (**B**). Respiration is illustrated by the concentration of CO_2 expressed as percentage of air volume (**A**). The diversity indices are estimated averages based on bootstrapping (n=100) and the standard error (n=3) of the triplicates was added. Prior analysis, background fluorescence was removed by gating.*



***Appendix Figure 4 - 3**: Contrast of the density plots of the peak samples with the density plots of the samples with low phenotypic diversity for both batch 1 (**left**) and batch 2 (**right**). First, only the peak sample is considered (**above**) but then the next sample (**peak +1**) and the sample after that is contrasted (**peak +2**). The red color indicates a region with higher density in the samples with high phenotypic diversity D_2 in contrast to the samples with low phenotypic diversity D_2. Inversely, the regions in blue indicate regions of lower density in the samples with high phenotypic diversity D_2 in contrast to the samples with low phenotypic diversity D_2. Samples were stained with SG and background was removed prior analysis. If the difference between the two communities is lower than 0.02, no contrast value is shown on the graphs, which causes the appearance of different clusters.*

CHAPTER

5

SINGLE-CELL RAMAN SPECTROSCOPY FOR GENOTYPIC AND PHENOTYPIC DIFFERENTIATION

CHAPTER

5

SINGLE-CELL RAMAN SPECTROSCOPY FOR GENOTYPIC AND PHENOTYPIC DIFFERENTIATION OF BACTERIA POPULATIONS

1 Abstract

The increasing focus on bacterial individuality is leading to the development of new single-cell technologies. Research in the last decade often showed the importance of phenotypic heterogeneity for microbial cultures which adds an extra dimension in the microbial community analysis. In order to better understand the relevance of phenotypic heterogeneity in different types of microbial communities, a quick and easy way of determining microbial genotypes and phenotypes is required. Single-cell Raman spectroscopy is a potential solution for this as it makes a chemical fingerprint of every cell. In this chapter, we compare several workflows with Raman spectroscopy to discriminate among microbial genotypes and phenotypes. We demonstrate the sensitivity to microbial physiology of the method by comparing different phenotypes. Finally, we applied a novel method to discriminate and quantify different microbial taxa from a single measurement of an *in vitro* and *in silico* synthetic community.

Chapter redrafted after:

Buysschaert, B., Khalenkow, D., García-Timermans, C., Skirtach, A., Boon, N. Single cell Raman spectroscopy for genotypic and phenotypic differentiation. *In preparation*

2 Introduction

The capacity of microbial communities to adapt themselves to almost every known environment has led to the discovery of many metabolic pathways that could be used as the potential solution to 21st century problems. As a consequence, bacteria are used in a wide variety of industrial technologies such as the pharmaceutical industry, the food industry or in environmental technology. Because those bacterial populations and communities are fundamental to the success of those processes, a good understanding of the community is the key to a good management of the system. With this in mind, Verstraete *et al.* (2007) introduced the concept of microbial resource management (MRM) where the approach to better understand the microbial community is condensed to three questions: “Who is there?”, “What do they do?” and “Who is doing what with whom?”. The technical evolutions of the last half century made it possible to easily identify on the molecular level which bacteria are present. Initially, labor intensive clone libraries were used but later, next generation sequencing methods made it faster, easier and cheaper than ever before to identify the community members. Despite the technical advances some concessions had to be made regarding the depth at which communities are characterized. While clone libraries often used whole 16S rRNA gene sequencing, next generation sequencing is often based on partial 16S rRNA gene sequencing reducing the identification depth from the species or strain level to the family or genus level (Mizrahi-Man *et al.*, 2013). While knowing who is present can elucidate many questions, it does not explain everything as bacteria can behave differently depending on their environment. To understand that behavior and the potential interactions, a microbial community can be analyzed in different ways such as proteomics or metabolomics. Typically, the correlation between a certain functionality and community structure is made to attribute roles to a group of organisms (Kuypers and Jorgensen, 2007).

A common caveat of all previously mentioned methods is that they are bulk methods. The potential danger of a bulk analysis is that it may mask subpopulations by calculating population wide averages on a large number of cells (Davis and Isberg, 2016). As a consequence of this awareness, the focus of current research shifts increasingly towards cellular individuality (Ackermann and Schreiber, 2015) and a variety of single-cell tools and technologies are under development, going from single-cell sequencing (Gawad *et al.*, 2016) to single-cell transcriptomics (Wu *et al.*, 2014). This focus on individuality already highlighted the importance of phenotypic heterogeneity for bioprocess technology regarding both the productivity and robustness of starter cultures, probiotics, and bioprocesses (Delvigne and Goffin, 2014). Cells in starter cultures and probiotics can be both live and dead and a fraction of the cells are in the so-called viable but nonculturable (VBNC) state which can impact the communities functionality or productivity (Davey and Hexley, 2011). Next to viability, other

phenotypes have been shown to be important for bioprocess productivity (Sundstrom *et al.*, 2004, Hewitt *et al.*, 2007, Tracy *et al.*, 2008, Alonso *et al.*, 2012). These findings are not only relevant for bioprocess technology but also in the wider field of microbiology as also toxin production for example is related to phenotypic heterogeneity (Ceuppens *et al.*, 2013). This indirectly broadens the field of microbial ecology as even isogenic populations can be considered as a community of phenotypes. In this respect, phenotypic heterogeneity is, as deeper level of community organization, worthwhile investigating for community characterization as it is inherent to a community and its functionality.

To evaluate the importance of phenotypic heterogeneity for microbial community functionality, a method able to detect phenotypes or phenotypic changes on the single-cell level is necessary. Different approaches are possible, but the best-known technique is microscopy. A first approach is to monitor heterogeneity in function of cell motility or growth (Balaban *et al.*, 2013, Spudich and Koshland, 1976). Alternatively fluorescent probes or gene expression reporters can be used (Ackermann, 2015, Smits *et al.*, 2006). The availability and the possibility to simultaneously assess multiple traits of single cells makes it a good method but it lacks the high throughput modern research often requires. The introduction of flow cytometry has resolved this issue but both methods are unable to identify species or to differentiate phenotypes if they are not related to the morphology of the cell. Although fluorescent labels can be used to highlight specific physiological features, their use is not without limitations. In a first instance because labels can bind non-selectively to other components than the targets, skewing the results of the analysis or making the sample preparation more complex. In a second instance, labeling narrows the observations as only labeled features are measured. The latter is not *per se* a problem but requires knowledge on what has to be measured prior analysis. Raman spectroscopy is a non-invasive and non-destructive optical method and makes profiles of the chemical bonds present in the cell as the inelastic scatter of the incident Raman laser is detected. This results in a unique chemical signature for each cell and, in contrast to the previously mentioned methods, without labeling. Therefore, the Raman spectrum can say something about the cell phenotype and, being label free, has the potential to explain the detected differences without prior knowledge of cause. Other advantages are that the method is non-destructive and requires little or no sample preparation. Scientific research extensively reported the use of Raman spectroscopy for species discrimination (Hutsebaut and Moens, 2005, Jarvis and Goodacre, 2008, Stockel *et al.*, 2016) and also to determine species interactions by using SIP or deuterium (Berry *et al.*, 2015, Cui *et al.*, 2017).

In this chapter we compare different workflows and demonstrate how single-cell Raman spectroscopy can be used to discriminate among different microbial genotypes and

phenotypes. For this, two strains of *E. coli* were compared. To increase genotypic and phenotypic heterogeneity, one strain was transformed with a TOL plasmid tagged with a green fluorescent protein (GFP) marker and the other strain was grown in two different media to induce different phenotypes. To take the effect of sample handling and preservation into account, an old sample was compared to a freshly fixed and older fixed sample. Several data processing methods were compared, which showed that all cell types could be differentiated with multiple methods albeit with varying accuracy. To conclude, we present a novel algorithm to simultaneously quantify and identify bacteria taxa in a sample based on one single Raman spectrum. We illustrate the power of this method by applying our method both on *in sillico* and *in vitro* microbial communities.

3 Materials and methods

3.1 Bacteria

Four different *E. coli* cultures were prepared for the experiments. The first two cultures contained different *E. coli* strains; *i.e. E. coli* LMG 8063 and *E. coli* DH5α DSMZ 6897. Both were cultivated in Luria Bertani (LB) broth (Oxoid) and incubated at 37°C for 24 hours to reach stationary phase. To induce phenotypic heterogeneity an *E. coli* DH5α harboring a TOL plasmid with a mini-Tn5-$P_{A1-04/03}$::*gfp*mut3-cassette (Christensen et al., 1998; Haagensen et al., 2002) according to Boon *et al.* (2006) was used. To select the transformed cells, a kanamycin (Km) resistance gene was inserted and 50 mg/L of Km was added to the Luria Bertani broth prior incubation at 37°C for 24 hours. *E. coli* LMG 8063 was also cultivated in a minimal medium and incubated at 37°C for 24 hours. The minimal medium contained glucose as only carbon source and was made according to De Mey *et al.* (2007). To assess the effect of sample handling and preservation, two *E. coli* DH5α were fixed with formaldehyde as described by Read and Whiteley (2015). One sample was measured immediately after resuspension (fixed sample), the other sample was measured after a few hours of sample preparation (old fixed sample). In the same way, an unfixed sample of *E. coli* DH5α was measured a few hours after sample preparation (old sample). To quantify the relative abundance in a synthetic community, a *Delftia acidovrans* DSMZ 14801, *Lactobacillus casei* subsp. casei LMG 6904, *Citrobacter werkmanii* DSMZ 17579, *Pseudomonas fluorescens* LMG 1794^T, *Cupriavidus necator* LMG 1190, and *Micrococcus luteus* LMG 3293 were cultivated in Luria Bertani (LB) broth (Oxoid) and incubated at 37°C for 24 hours to reach stationary phase. *In vitro* mixtures were made after cell enumeration with flow cytometry as described previously (Buysschaert *et al.*, 2016). The mixing ratios were 0, 25%, 50%, 75%, and 100% of one species with the other.

3.2 Sample preparation

1 mL of cell suspension was centrifuged for five minutes at 5000 rcf at room temperature. After centrifugation the pellet was washed three times and resuspended in 0.22 µm-filtered water (Sartorius) to remove all medium components. A droplet of 10 µl of cellular suspension containing approximately 10^8 cells/mL was spotted on a CaF_2 glass slide and air-dried for a few minutes.

3.3 Raman spectroscopy

The Raman spectra of each single bacterial cell were measured with WITec Alpha300R+ confocal Raman microscope using 785 nm excitation diode laser (Toptica) and an UHTS 300 spectrometer with a -60°C cooled CCD camera (iDus 401 BR-DD, ANDOR). The 100x/0.9 NA (Nikon) objective was used. Laser power was measured before the objective and was set to 200 mW. The integration time was 45 seconds for each single bacterium spectrum acquisition.

3.4 Data analysis

The obtained spectra were imported as SPC files in R (R Core Team, 2015) for preprocessing and analysis. First, the most relevant region between 600 – 1800 cm^{-1} was selected using the *Hyperspec* package v0.98.20161118 (Beleites and Sergo, 2016). Next, the baseline was estimated using the SNIP algorithm with ten iterations and corrected by subtraction (**Figure 5 - 1**). The data was also normalized with the area under the curve (AUC) algorithm. Both functions are implemented in the *MALDIquant* package v1.16.2 (Gibb and Strimmer, 2012). For classification two unsupervised and three supervised methods were compared. For the first unsupervised method the similarity between all spectra was measured using the quantitative Jaccard distance measure (Ružička index) as implemented in the function *vegdist* from the *vegan* package v2.4.3 (Oksanen *et al.*, 2016). The resulting similarities were clustered based on agglomerative clustering with Ward's minimum variance method (ward.D2 from the *hclust* package) as linkage from the *stats* package (R Core Team, 2015) and visualized in a dendrogram with the iToL software (Letunic and Bork, 2016). The second unsupervised method is a visualization by means of t-SNE (van der Maaten and Hinton, 2008) using the *tsne* package v0.1.3. Subsequently the samples were classified by k-means hierarchical clustering from the package *stats* v3.4.1. Perplexity of the t-SNE algorithm was optimized by comparing the results of the classification and a final perplexity of 30 was selected. Partial least squares (PLS), partial least squares with subsequent linear discriminant analysis (PLS-LDA) and random forests were compared as examples of

supervised methods. All three were implemented in R using the packages *pls*, *MASS* and *randomForest* respectively (Mevik and Wehrens, 2016, Venables and Ripley, 2002, Liaw and Wiener, 2002). For all supervised methods the data was randomly divided in a training dataset and a test dataset. The training dataset was three times larger than the test dataset and the same for all implemented methods. Based on the random forest classification and based on a custom made peak selection function, the most important features contributing to the difference between the samples were selected and identified by comparing the results with existing tables from literature (**Table 5 - 1**). To estimate the relative abundance of species in a synthetic community, a linear unmixing model was assumed and the proportion of each species, or endmember, was estimated based on the spectrum of an axenic culture. For hyperpectral unmixing, the package *hsdar* v0.5.1 was used (Lehnert *et al.*, 2015). Spectra of both *in silico* mixtures of the two endmembers and *in vitro* mixtures were unmixed with the algorithm. The spectra of the endmembers and the mixtures were calculated as the average of 50 single-cell spectra

***Table 5 - 1**: Overview of wavenumbers and their assignments according to literature.*

Wavenumber [cm^{-1}]	Assignment	Reference
615	Monosubstituted benzenes	Jarvis *et al.* (2004)
620	Phenylalanine (skeletal)	Maquelin *et al.* (2002)
640	Tyrosine (skeletal)	Maquelin *et al.* (2002)
665	Guanine	Maquelin *et al.* (2002)
720	Adenine	Maquelin *et al.* (2002)
730	A ring stretching	Uzunbajakava *et al.* (2003)
752	T ring stretching	Uzunbajakava *et al.* (2003)
778 - 782	Nucleic acids (phosphoric acid esters)	Schuster *et al.* (2000b)
778 - 782	Nucleic acids (cytidine, uracil)	Schrader (1997)
778 - 785, 792	Cytosine, uracil (ring, stretching)	Maquelin *et al.* (2002)
798	Ring breathing	Jarvis *et al.* (2004)
810 - 820	Nucleic acids (C-O-P-O-C in RNA backbone)	Schuster *et al.* (2000b)
813	A-type helices in RNA	Uzunbajakava *et al.* (2003)
828	Aliphatic	Jarvis *et al.* (2004)
829	Exposed tyrosine	Maquelin *et al.* (2002)
830, 850	Tyrosine (in protein)	Schrader (1997)
835	DNA	Deng *et al.* (1999)
852	Buried tyrosine	Maquelin *et al.* (2002)
858	C-C stretching, C-O-C 1,4-glycosidic link	Maquelin *et al.* (2002)
897	C-O-C stretching	Maquelin *et al.* (2002)
1004	Phenylalanine	Schuster *et al.* (2000b)
1004	Phenylalanine	Maquelin *et al.* (2002)
1030 - 1130	Carbohydrates, mainly -C-C- (skeletal) and C-O, C-O-H deformation	Schuster *et al.* (2000b)
1032	Phenylalanine, C-N stretching	Uzunbajakava *et al.* (2003)
1054	Nucleic acids, C-O stretching, protein, C-N stretching	Uzunbajakava *et al.* (2003)

1061	C-N and C-C stretching	Maquelin *et al.* (2002)
1085	C-O stretching	Maquelin *et al.* (2002)
1098	Phosphate, C-C skeletal and C-O-C stretching from glycosidic link	Maquelin *et al.* (2002)
1100	Glass background	Schuster *et al.* (2000b)
1102	> PO_4^{3-} stretching (symmetric)	Maquelin *et al.* (2002)
1130	=C–C= (unsaturated fatty acids in lipids)	Schrader (1997)
1150	*n*-alkanes	Jarvis *et al.* (2004)
1175	Tyrosine, phenylalanine	Uzunbajakava *et al.* (2003)
1209	Tyrosine, phenylalanine, protein, amide III	Uzunbajakava *et al.* (2003)
1214, 1240, 1254	Thymine, cytosine, adenine, ring v	Uzunbajakava *et al.* (2003)
1220 - 1290	Amide III, lipids	Schuster *et al.* (2000b)
1249	N–H, C–N, amid III random coil	Schuster *et al.* (2000b)
1254	Adenine, amide III	Uzunbajakava *et al.* (2003)
1267	Lipids	van Manen *et al.* (2005)
1268	Amide III - α helix	Jarvis *et al.* (2004)
1295	CH_2 deformation	Maquelin *et al.* (2002)
1304	Adenine, amide III	Uzunbajakava *et al.* (2003)
1320	Amide III	Schuster *et al.* (2000b)
1320	Amide III, C-H deformation	Schuster *et al.* (2000b)
1336 - 1339	Adenine	Uzunbajakava *et al.* (2003)
1344	Adenine, guanine	Jarvis *et al.* (2004)
1375	Thymine, adenine, guanine	Uzunbajakava *et al.* (2003)
1401	α-amino acids	Jarvis *et al.* (2004)
1421 - 1427	Adenine, guanine	Uzunbajakava *et al.* (2003)
1431 - 1481	Protein marker band 1451	Uzunbajakava *et al.* (2003)
1441	Lipids	van Manen *et al.* (2005)
1450	C-H_2 deformation	Schrader (1997)
1468	C-H deformation	Maquelin *et al.* (2002)
1482 - 1487	Nucleic acids (C-O-P-O-C in RNA backbone)	Schuster *et al.* (2000b)
1505, 1518, 1532, 1578	Adenine, cytosine, guanine	Uzunbajakava *et al.* (2003)
1510	Adenine	Uzunbajakava *et al.* (2003)
1521	Pigment	Jarvis *et al.* (2004)
1572	Amide II	Jarvis *et al.* (2004)
1573	C=C	Schuster *et al.* (2000b)
1573	C=C, N-H deformation and C-N stretching (amide II)	Schuster *et al.* (2000b)
1575	Guanine, adenine (ring stretching)	Maquelin *et al.* (2002)
1606	Phenylalanine	Maquelin *et al.* (2002)
1614	Tyrosine	Maquelin *et al.* (2002)
1627	Olefinic	Jarvis *et al.* (2004)
1650 - 1680	Amide I	Maquelin *et al.* (2002)
1658	Unsaturated lipids	van Manen *et al.* (2005)
1660 - 1670	Amide I	Schuster *et al.* (2000b)
1740	> C=C ester stretching	Maquelin *et al.* (2002)

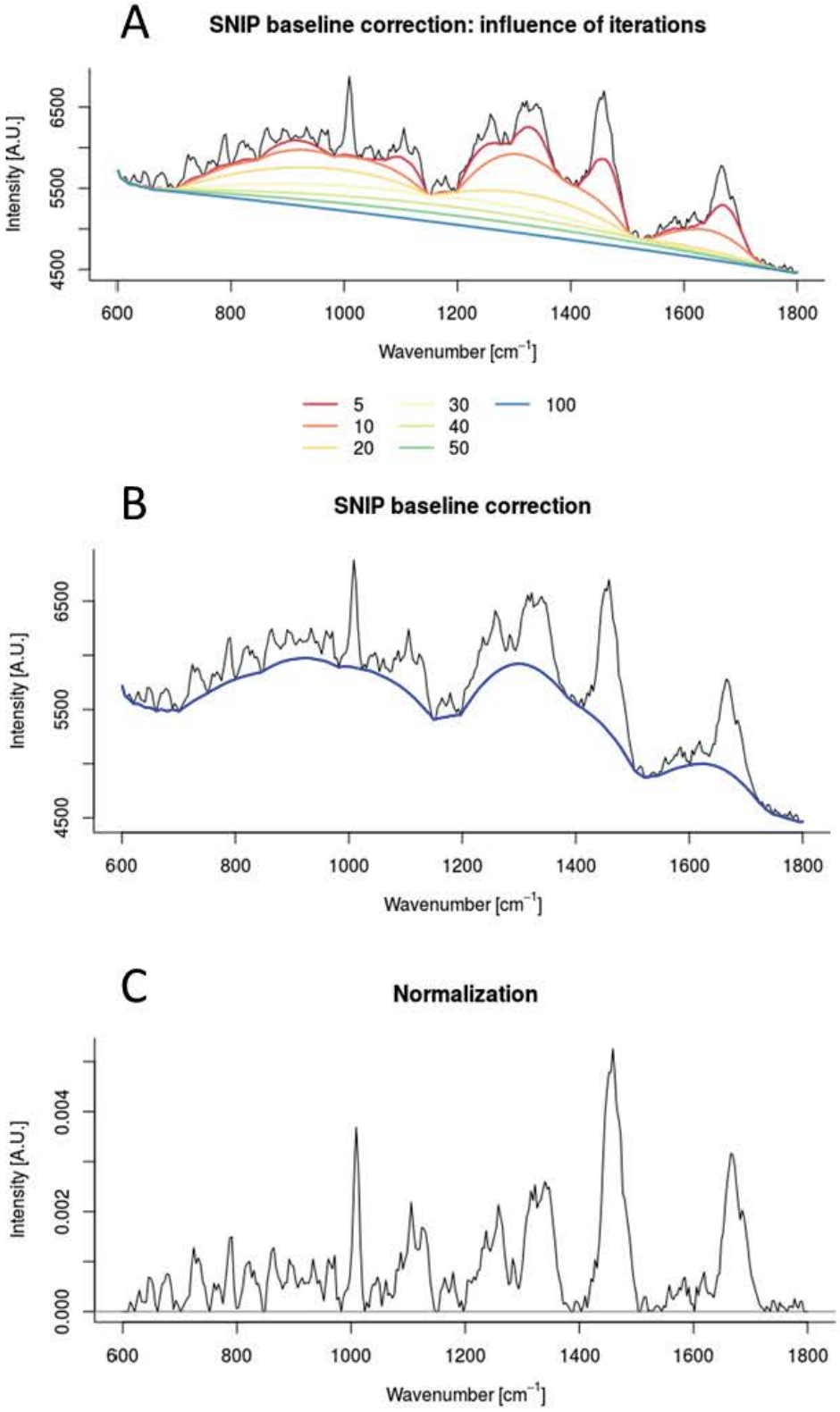


***Figure 5 - 1**: Baseline correction and normalization of the spectra. Based on the SNIP algorithm, several baselines were constructed depending on the number of iterations (**A**). ten iterations were selected as the curve was neither over-fitted nor under-fitted (**B**). Subsequently, the data was normalized using the area under the curve normalization (**C**).*

4 Results and discussion

To determine how single-cell Raman spectroscopy can be used to determine phenotypic differences, phenotypic differentiation was induced in different ways in two *E. coli* strains. Thus, a wild type *E. coli* DH5α was compared to its *gfp*-labeled counterpart and another *E. coli* strain was cultured both in a complex and a minimal medium. Furthermore, an old sample, a freshly fixed sample, and an old fixed sample of the wild type *E. coli* DH5α were compared to assess the effects of sample handling and preservation. As all these samples are either different due to their genotypes and/or external factors, they will be referred to as different phenotypes within this chapter.

4.1 Supervised and unsupervised methods

Both supervised and unsupervised methods can be used to analyze high dimensional Raman spectroscopy data. While it is generally known that supervised methods perform better, the unsupervised methods have the advantage that no knowledge of the clusters in the data is required prior analysis. A first unsupervised method is the t-SNE visualization algorithm (**Figure 5 - 2**).The results show that the data clearly clusters into five groups. The groups are all based on the type of organism and the culturing method. No difference can visually be found between the two types of fixed sample. The difference between an old *E. coli* DH5α sample and the *E. coli* LMG 6803 cultured in minimal medium is small as well. t-SNE proves to be a useful tool to visualize the large amount of information in two dimensions. The caveat of the method is that artificial clusters can be formed if the parameters are not optimized. Due to the nature of t-SNE, only a visual interpretation can be drawn from the data and no density based algorithms, such as clustering, can be used.

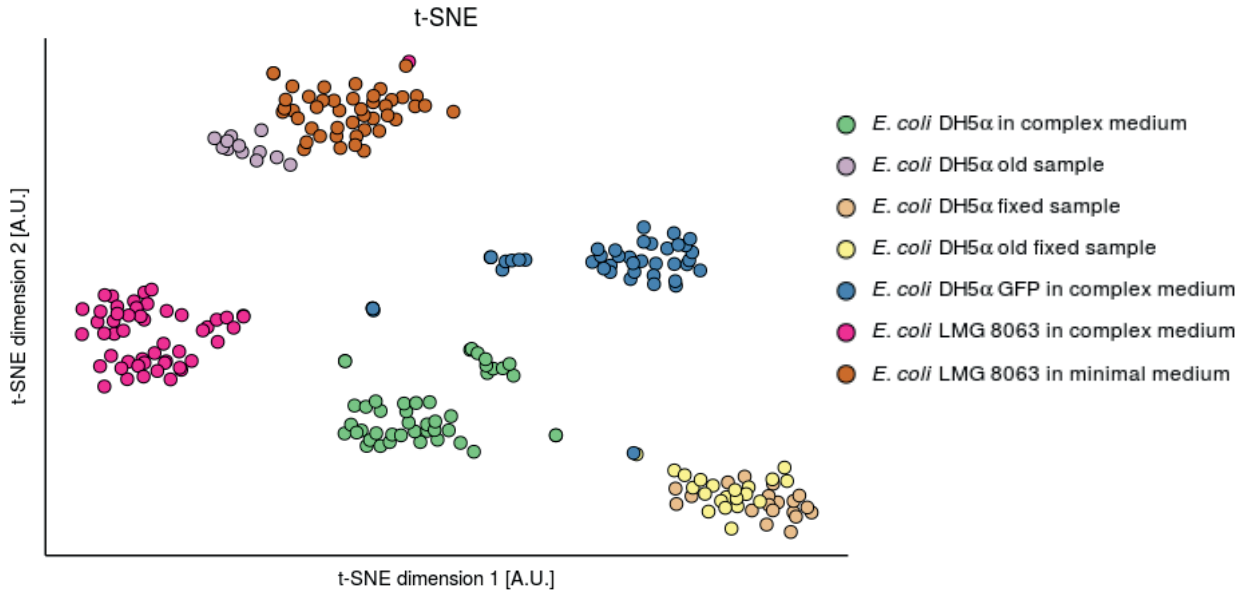


***Figure 5 - 2**: t-SNE visualization of the single-cell spectra of the different* E. coli *phenotypes and strains. Some* E. coli *DH5α samples were prepared differently to assess the impact of sample pre-treatment. All spectra were preprocessed in the same way.*

In order to find an automated and objective way to quantify clusters in the data, different and more conventional methods can be used. The most common examples are hierarchical clustering and PCA. In this chapter, we combined principal component analysis with k-means clustering to determine the clusters. For clustering, the ten first principal components were used as they retained more than 70% of the variability of the dataset. An increase in the number of principal components did not provide different results after clustering (data not shown). The optimal number of clusters was determined with the silhouette plots and seven clusters were found in the data. The results show that five clusters corresponded with the expected phenotypes and two clusters contained a small number of misclassified spectra (**Figure 5 - 3**). Similarly to t-SNE, no difference can be noticed between two types of fixed samples and both the old *E. coli* DH5α sample and the *E. coli* LMG 6803 cultured in minimal medium cluster together. All spectra of the old *E. coli* DH5α sample were therefore considered to be misclassified. The resulting accuracy is 85% if the two types of fixed samples are considered as one cluster.

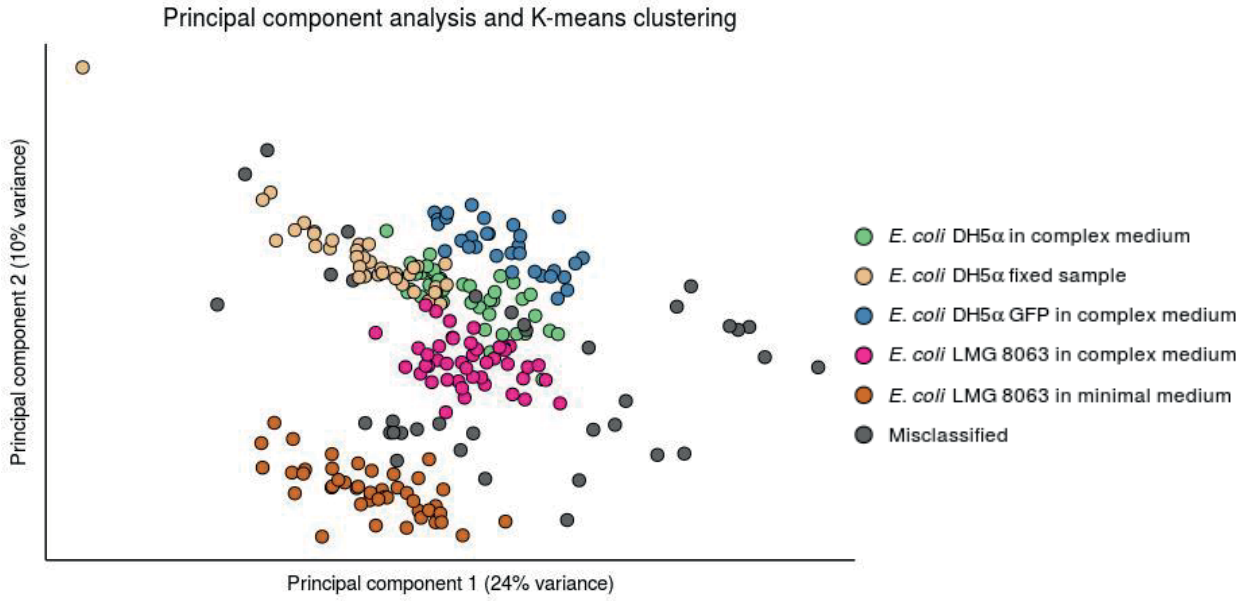


***Figure 5 - 3**: Two dimensional visualization of the principal component analysis on all single-cell spectra. Clustering was performed with k-means clustering, based on ten principal components. Colors were assigned to the clusters corresponding to the known phenotypes. Misclassified samples were considered as another cluster. The accuracy of the method is 85% for this dataset.*

A last unsupervised method is based on the density based Jaccard distance measure (Ružička index) between the samples and the subsequent hierarchical clustering with Ward's algorithm. The results demonstrate that the phenotypes can be discriminated from one another with the exception of the two types of fixation (**Figure 5 - 4**). Again, the old *E. coli* DH5α cells clusters together with the *E. coli* LMG 8603 cells grown in minimal medium and were considered to be misclassified. 19 misclassifications were found in total out of 236 spectra. The accuracy is thus 92% and noticeably better than PCA combined with k-means clustering. Results were similar with other distance measures such as the Bray distance (19 misclassifications; 92% accuracy) or the Euclidean distance and the spectral contrast angle (Wan *et al.*, 2002) (21 misclassifications; 91% accuracy). However, results altered considerably when other clustering algorithms were found. Ward's method showed to result in the highest accuracy (data not shown).

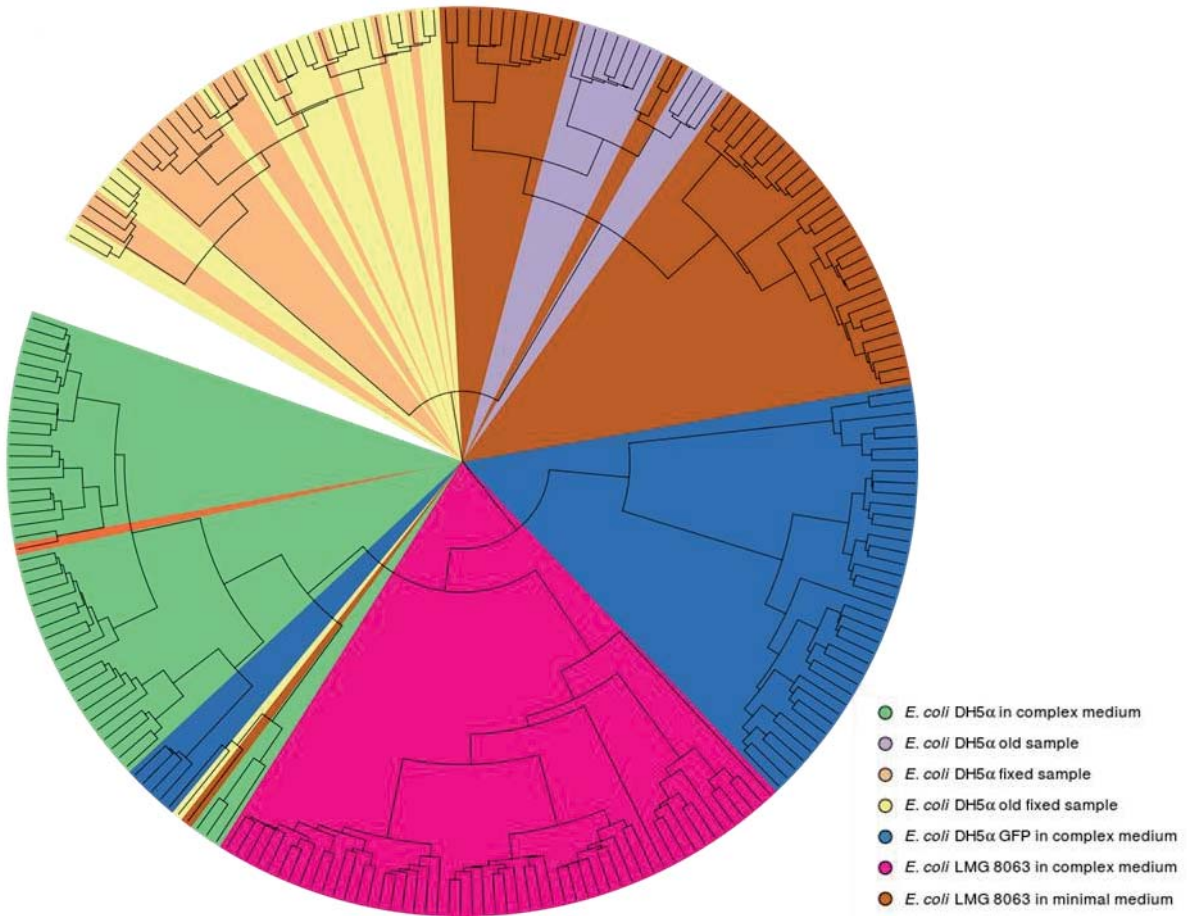


***Figure 5 - 4**: Dendrogram of the Raman spectra of the different strains and phenotypes. Similarity was calculated with the density based Jaccard distance measure and clustering was performed with Ward's hierarchical clustering algorithm. Sample preprocessing was the same for all samples. The accuracy of the method is 92%.*

Aside from unsupervised methods, also supervised methods were assessed to increase the performance to discriminate among phenotypes and the range of applications of phenotyping. The methods selected were partial least squares (PLS), partial least squares and subsequent linear discriminant analysis (PCA-LDA), and random forests. Although the intention of PCA and PLS is to reduce the dimensionality of the data, PLS is designed for datasets with large number of correlated variables (*e.g.* data points in a spectrum) and uses the labeled data to emphasize the directions necessary for discrimination. The PLS classification of the test set in function of the two first components or latent variables shows no clear visual distinction between the clusters although 58% of the spectra were classified correctly (**Figure 5 - 5**). The old sample of *E. coli* DH5α was only present twice in the test set and was twice misclassified. The accuracy is lower than the unsupervised methods which is partially related to the smaller size of the test set.

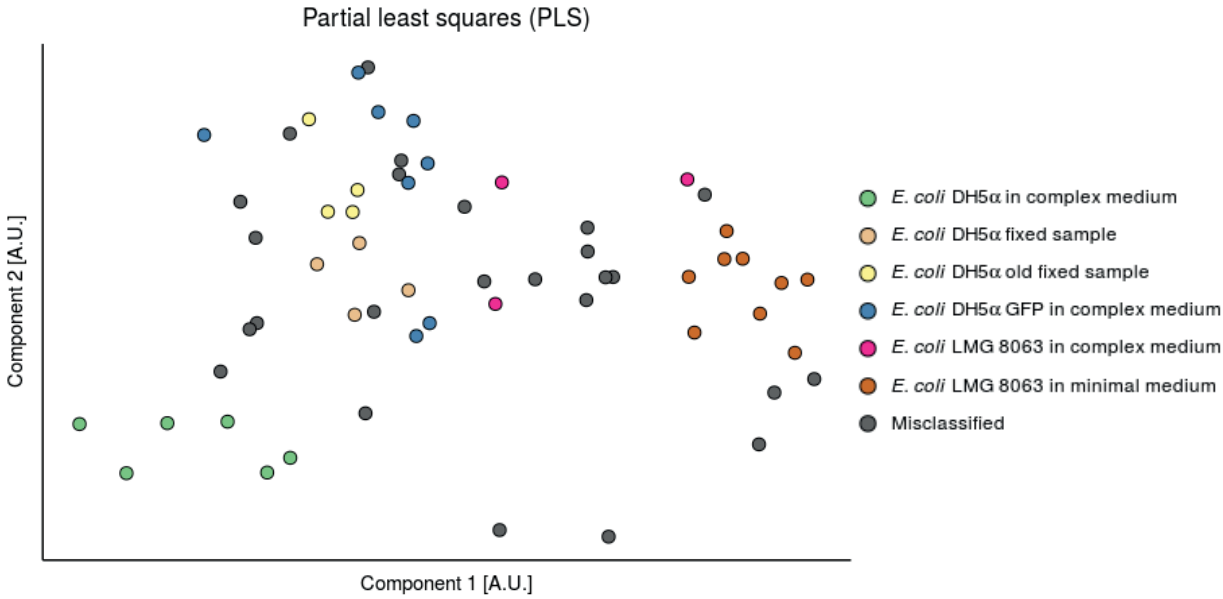


Figure 5 - 5: *Two dimensional visualization of the partial least square classification (PLS) on all single cell spectra of the test set. The model was built with a training set comprising two third of the data set and validated with the leave-one-out principle. Colors were assigned to correctly classified spectra and grey to all misclassified spectra. The accuracy of the method is 58%.*

PLS can also be used as a preprocessing step before LDA to improve accuracy. For LDA, six PLS components were selected as they resulted in the smallest root mean square error of prediction (RMSEP). Subsequent discriminant analysis was performed on the same test set after transformation by the PLS model mentioned above. The clustering of the different phenotypes and strains in function of the first two discriminants improved as compared to PLS (**Figure 5 - 6**). The accuracy increased to 86% for the test set and was therefore comparable but slightly better than PCA and k-means analysis. But, in contrast to the unsupervised methods, a distinction can be made between the two different types of fixation. The last supervised method tested here, random forests, was trained and tested with the same training set and test set. The model was constructed by training 1000 decision trees for classification. The method proved to be the most accurate with a classification accuracy of 97%. Because of this, the most important features or wavenumbers contributing to the classification were extracted to understand the biological differences causing the discrimination. Mainly the amide III signal representative for proteins (Raman bands 1269, 1272, 1280, 1308, 1319 and 1322 cm^{-1}), tyrosine and phenylalanine (Raman bands 637, 1009, 10031, 1179 cm^{-1}), and thymine, adenine, and guanine, representative for the nucleic acid content (Raman bands 1308, 1372, 1375 cm^{-1}) were found. This suggests that the phenotypic differences are related primarily to protein content and concentration and, to a lesser extent, nucleic acid concentration. Although not confirmed by an independent method, it is very likely that proteins are responsible for the difference between the phenotypes.

Especially with the *gfp*-labeled bacteria, where the fluorescent protein expression could be monitored visually. Furthermore, the difference between *E. coli* cultivated in a rich or in a minimal medium will probably be translated in the cellular protein profiles as well because of the different pathways required to grow on both media.

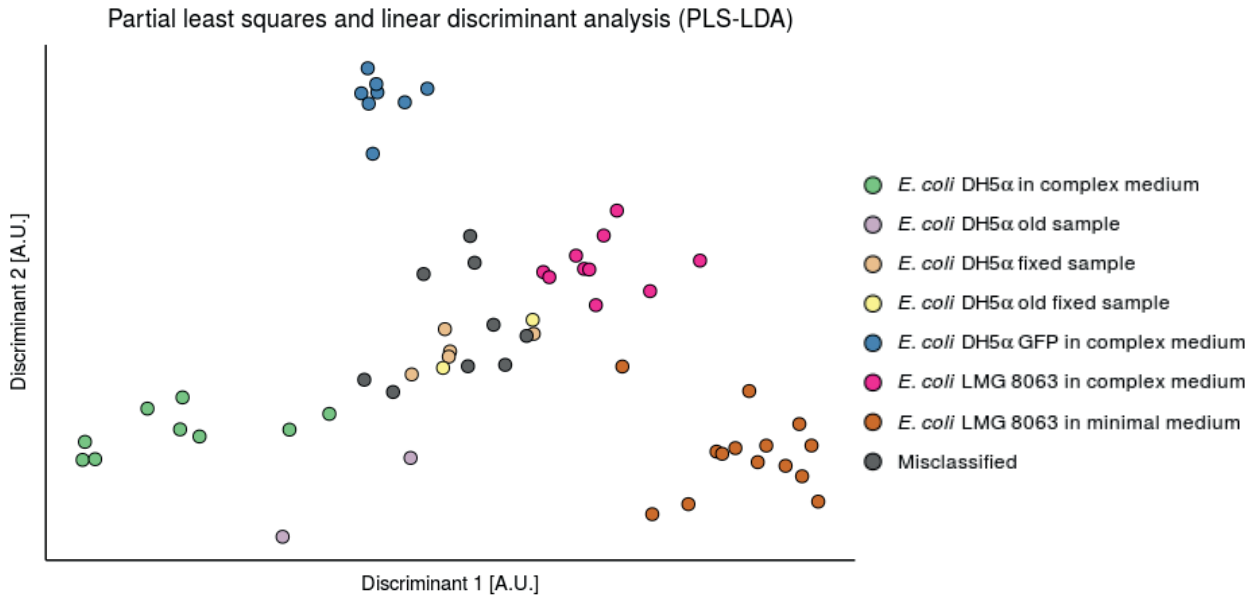


***Figure 5 - 6**: Two dimensional visualization of the partial least square classification (PLS) on all single-cell spectra and subsequent linear discriminant analysis (LDA) of the test set. The PLS classifier was built with a training set comprising two third of the data set and validated with the leave-one-out principle. Six components were selected for based on the minimal RSMEP for subsequent LDA modeling. Colors were assigned to correctly classified spectra and grey to all misclassified spectra. The accuracy of the method is 86%.*

The results presented in this chapter show that discrimination between phenotypes and genotypes until the strain level is possible with either supervised learning algorithms or unsupervised clustering. Similar attempts to discriminate among strains or to detect phenotypes have been reported before (Jarvis and Goodacre, 2008, Stephen *et al.*, 2012, Hutsebaut and Moens, 2005, Maquelin *et al.*, 2000, Rosch *et al.*, 2005, Stockel *et al.*, 2016). Different ways of processing the data are proposed and, similar to our findings, also unsupervised clustering algorithms were found to work well to discriminate among microbial strains (Harz *et al.*, 2005). The clear distinction between strains with Raman spectroscopy can be explained by the important differences that can occur between strains from the same genus (Yoon *et al.*, 2012). As Raman spectroscopy is a holistic method, differences in nucleic acid or protein profiles can be detected easily. Also supervised machine learning classification tools, such as support vector machines, were used to improve the data. The accuracy of the random forest classification is comparable but better than the accuracies of 94.1% (Harz *et al.*, 2005) and 89.2% (Rosch *et al.*, 2005) found with support vector

machines. It is important to note that comparison is difficult as the data sets differ. Next to this, the approach differs as our purpose is to characterize phenotypic heterogeneity while Rosch *et al.* (2005) and Harz *et al.* (2005) tested how robust their taxonomic classification model was against phenotypic heterogeneity. Inter- and intracellular heterogeneity of a pathogenic strain have been recorded and quantified by Hermelink *et al.* (2009). Similarly, Schuster *et al.* (2000a) and Huang *et al.* (2004) showed that bacterial phenotypes change in function of their growth stage. Both Hermelink *et al.* (2009) and Schuster *et al.* (2000a) defined the variability on a select number of peaks with a known biological meaning but did not exploit the holistic nature of Raman spectroscopy. Read *et al.* (2013) also used a limited number of peaks prior hierarchical clustering to determine the host origin of isolates and used automatic relevance determination (ARD) algorithm for this purpose. Other groups also used supervised learning such as PCA-LDA to discriminate among extra- and intracellular phenotypes of a pathogenic bacteria with a classification accuracy of 85% (Grosse *et al.*, 2015), to discriminate phenotypic reactions as consequence of antibiotic treatments with a classification accuracy of 83.6% (Athamneh *et al.*, 2014) or to determine bacterial metabolic history (Huang *et al.*, 2007).

Regardless of all promising results, Raman spectroscopy so-far failed to be incorporated as a standard microbial technique. The most important issue is that the method still lacks the high throughput modern research requires today. In comparison to molecular analysis Raman spectroscopy is quick but with an acquisition time of more or less 60 seconds per cell, the method can hardly be considered fast. Several solutions can be implemented to decrease the acquisition time and subsequently increase the throughput. The most well documented attempts have been made by using SERS which was successfully applied for species discrimination (Jarvis and Goodacre, 2008). The time gain can, in combination with microfluidic devices, also lead to automation (Walter *et al.*, 2011). For more subtle differences such as phenotypic heterogeneity, SERS could prove unreliable as only the spectra of the molecules in the vicinity of the SERS particles are amplified (Hering *et al.*, 2008) and as only the signal of the molecules on the surface of the bacteria are amplified (Pahlow *et al.*, 2012). More recent Raman techniques such as coherent anti-Stokes Raman spectroscopy (CARS) and stimulated Raman spectroscopy reduce the acquisition time and are reported to provide robust signals, 100 times stronger than conventional Raman spectroscopy (Petrov *et al.*, 2007, Opilik *et al.*, 2013, Hansen *et al.*, 2015). Another difficulty is the plethora of methods available for data analysis. Though many researchers use similar methods, comparison is often difficult because of different instrumentation and purposes. To address this issue, we tried to compare multiple methods in this chapter and found that most commonly used methods are useful for the purpose of phenotypic fingerprinting though with

varying accuracy. Hutsebaut *et al.* (2005) and Rodriguez *et al.* (2011) propose some solutions to overcome issues related to instrumental drift or to standardize data processing in order to make spectra comparable, regardless of the instrumentation.

4.2 Estimation of the relative abundance of species in a synthetic community

The capability of Raman spectroscopy to discriminate microbial taxa can be used to monitor the population dynamics. For this purpose we demonstrate with *in silico* and *in vitro* mixtures of bacteria pairs how hyperspectral unmixing algorithms can be used. Three pairs of bacteria species were combined for this purpose and the results show that hyperspectral unmixing predicts almost perfectly the correct abundances of both species when mixed *in silico* (**Figure 5 - 7**). For the *in vitro* mixes, the results are less accurate yet show to differ depending on the microbial pair analyzed. The different performance of the algorithm between the *in silico* and *in vitro* mixes can be explained by the assumption of the unmixing algorithm that the spectrum resulting of a mixture, is a linear combination of both spectra of the species present. When mixing two bacteria *in vitro*, physiological adaptation might occur and the spectra of the axenic culture will not correspond with the spectra of the microbial community members. This phenomenon will be discussed and studied in more detail in **Chapter 6**. Also, similarly to flow cytometry, physiological differences dependent on the growth stage could interfere with the accuracy of the predictions. In this set-up, stationary phase cultures were used to reduce the biological variability. A last important factor influencing the results is the reproducibility of the Raman spectroscopy measurements. The technique's sensitivity to phenotypes is an advantage in comparison to other methods yet at the same time, technical variability can make the reproducibility or the detection of small differences difficult. In the results shown above we demonstrate that leaving an old sample of the same bacteria changes significantly and is, regardless of the method used, wrongfully classified as a separate subpopulation. Additional research towards standardization could help to reduce or estimate the still unknown impact of these technical biases.

Thus, to improve the results, a non-linear unmixing algorithm could be implemented. But, to our knowledge, no mathematical models are available describing how Raman spectra combine to a single summary spectrum. This is in part also due to the complex physical and biological interactions influencing the outcome. Alternatively, machine learning algorithms can also be used. In contrast to hyperspectral unmixing, this approach does not rely on a mathematical assumption yet deduces a model based on the spectra. A drawback of machine learning is that, for each set-up, a model should be trained and validated. Rubbens *et al.* (2017) published comparable research based on flow cytometry and used both LDA

and random forests to discriminate different microbial species of both *in silico* and *in vitro* mixtures of two bacteria populations. For more than half of the analyzed samples they identified single cells with >90% accuracy. Moreover, similarly to our findings, they report different performances depending on the bacteria pair. Our comparable findings show that, next to flow cytometry, also Raman spectroscopy can be used to identify and quantify microbial species in a microbial community. A fundamental difference between the techniques is the sensitivity. While flow cytometry only acquires around 24 parameters per cell, Raman spectroscopy acquires hundreds of features per cell. This difference might not be reflected by the performance of unmixing algorithms on a mixture of two bacteria populations but certainly will prove to be an advantage when more species are mixed and when different physiological states have to be taken into account.

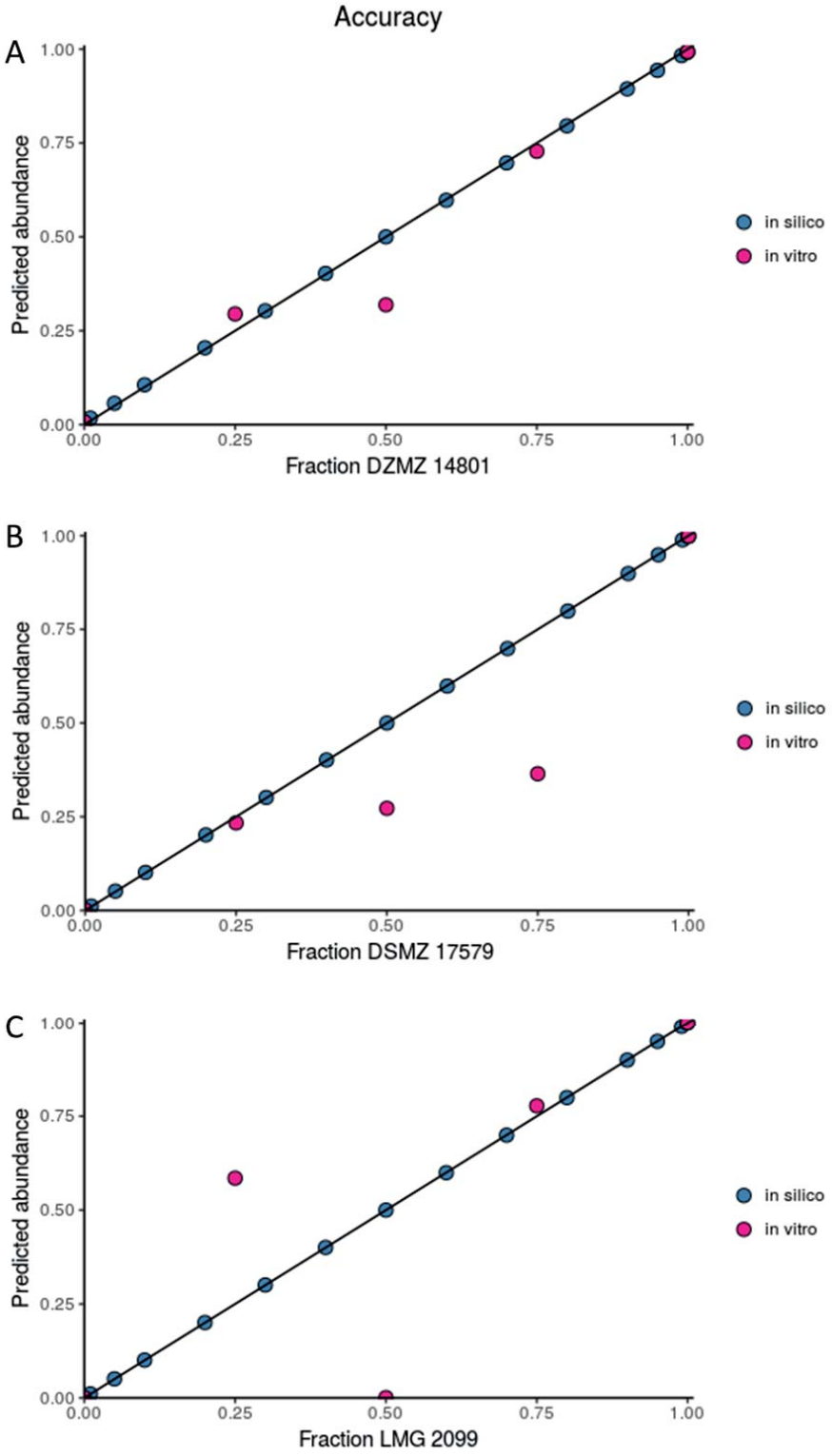


Figure 5 - 7*: The predicted abundance in function of the* in silico *(blue) or* in vitro *(red) fraction for three pairs of bacteria.* Delftia acidovrans *DSMZ 14801 and* Lactobacillus casei *subsp. casei LMG 6904 (**A**),* Citrobacter werkmanii *DSMZ 17579 and* Pseudomonas fluorescens *LMG 1794^{T} (**B**), and* Cupriavidus necator *LMG 1190 and* Micrococcus luteus *LMG 3293 (**C**).*

5 Conclusions and perspectives

In this chapter we demonstrate that single-cell Raman spectroscopy can used to differentiate among microbial phenotypes. Furthermore, we compared several common and new data analysis methods and show that not all are equally successful to resolve the different phenotypes. Based on our results, we conclude that hierarchical clustering is the best unsupervised method although PCA also showed satisfying results. When supervised methods are used, random forests are clearly superior to the other methods tested in this paper. Our findings are in line with the findings of other researchers and confirm that Raman spectroscopy is the best method developed so far for microbial phenotyping. However, to meet the demands of modern research, higher throughput is necessary. Recent advances in other types of Raman spectroscopy are promising and potentially the solution to this issue. Further research towards the implementation of these novel techniques should be encouraged. Finally, we also demonstrate that Raman spectroscopy of microbial communities can be used to identify and quantify the present bacteria populations. In this chapter we proposed hyperspectral unmixing and showed that relatively good results could already be achieved. The main drawback of this approach is the underlying assumption of linearity which does not take biological and technical variations into account. Better results could be achieved using machine learning algorithms yet further research is necessary to prove this.

6 Acknowledgements

This work was supported by the project grant SB-131370 of the IWT Flanders and the Geconcerteerde Onderzoeksactie (GOA) of Ghent University (BOF09/GOA/005). The authors would like to express their gratitude to Frederiek-Maarten Kerckhof and Jana de Bodt for providing the cultures and help in the lab. The authors also thank Ruben Props, Peter Rubbens and Charlotte De Rudder for their valuable input and feedback.

CHAPTER

6

PHENOTYPIC PLASTICITY IN COCULTURES

CHAPTER

6

PHENOTYPIC PLASTICITY OF COCULTURES

1 Abstract

With the advent of single-cell technologies, the scientific focus in microbiology is increasingly on the individual cells. The multiple reports about the impact of phenotypic heterogeneity on microbial cultures emphasize the importance of this deeper level of microbial organization. It is now well-established that several mechanisms are responsible for the occurrence of phenotypic diversity and that this diversity is useful for bacteria to survive and to cope with environmental changes. Most observations and research of phenotypic heterogeneity is based on observations of axenic cultures while in nature, bacteria do not live as axenic populations but rather as communities. Furthermore, detecting phenotypic diversity is challenging and for practical reasons, a fast and single-cell technique should be used. Flow cytometry and Raman spectroscopy are ideal for this purpose and could help to understand how phenotypic heterogeneity affects cell populations in a community. In this chapter, we demonstrate that both flow cytometry and Raman spectroscopy are suitable to detect phenotypic changes and that both methods are complementary. For this we developed microcosms which allowed individual cell populations to interact while remaining physically separated. We found that both species adapted to the presence of the other and that the level of phenotypic plasticity[1] was different for both species.

Chapter redrafted after:

Heyse, J., **Buysschaert, B.,** Props, R., Rubbens P., Skirtach A., Boon, N. Phenotypic plasticity in cocultures. *In preparation*

[1] Phenotypic plasticity refers to the changes in an organism's phenotype due to its adaptation to a unique environment.

2 Introduction

Phenotypic diversity has developed into a field of interest in microbiological research and a range of questions have emerged, such as whether this diversity is negligible in comparison to taxonomic diversity or diversity caused by environmental gradients, and how the properties of microbial communities are shaped by the fact that microbes act as individuals (Ackermann, 2013, Ackermann and Schreiber, 2015).

Several mechanisms have been reported to induce cell-to-cell heterogeneity. A first mechanisms is related to the genetic differences but, as phenotypic differentiation occurs at rates higher than any known mutational mechanism and as it is robust against the suppression of mutational mechanisms, it cannot be considered as the most prominent cause (Ackermann, 2015). Another and more important mechanism is the stochasticity or noise in gene expression (Ansel *et al.*, 2008, Fraser and Kaern, 2009, Avery, 2006). The level of stochasticity is an evolved trait that reflects the potential costs and benefits related to the gene expression. For example, proteins that signal responses to environmental perturbations are noisier compared to proteins for synthesis (Newman *et al.*, 2006). Heterogeneity in cell properties causes some individuals to exhibit features that will allow them to persist during fluctuating or adverse conditions, such as during exposure to antibiotics (Balaban *et al.*, 2004). This way heterogeneity might serve as a survival strategy. Another advantage is the potential to divide labor between phenotypes which then interact (Ackermann *et al.*, 2008). For example, Veening *et al.* (2008) observed that only a part of a *Bacillus subtilis* population excreted an exoprotease while the entire population could benefit from the exoprotease activity. Even in 'simple' microcosms containing only a single carbon source, clonal populations can differ in gene-expression and metabolic activity (Nikolic *et al.*, 2013). Environmental factors, such as nutrient status, influence phenotypic heterogeneity in clonal populations as well (Schreiber *et al.*, 2016). These studies have revealed that isogenic bacterial populations are not homogeneous populations, but rather communities consisting of different phenotypic subgroups, which can differ from each other both in a quantitative (*i.e.* continuous variation in phenotypic traits) and qualitative (*i.e.* distinct phenotypic states) way.

Phenotypic diversity is a population property which manifests itself at the level of individuals. To assess this fine scale diversity, tools that can reliably measure characteristics of single cells without disturbing their biochemical state are necessary. Flow cytometry is a laser-based technology that analyses individual cells by sending them through a beam of light by a fluid stream. Two types of optical properties can be detected in the flow cytometer, that is scattered light and fluorescence. In general, the scattered light provides information about the basic characteristics of the cells (*e.g.* size, intracellular properties) and the fluorescent

light provides additional information of the cell features that were stained. Another technique is single-cell Raman spectroscopy which assesses the chemical composition of a sample by evaluating the Raman scatter of the molecules (*e.g.* proteins, nucleic acids, fatty acids, *etc.*). This results in a very complex spectrum that can be interpreted as a chemical fingerprint of the cell (van de Vossenberg *et al.*, 2013), which can be used for phenotypic characterization of bacteria (Read *et al.*, 2013).

In nature, bacteria are not encountered as axenic cultures, but they are part of a community where many microorganisms coexist. Phenotypic heterogeneity has been shown to play a role in the functionality and productivity of industrial cultures (Delvigne and Goffin, 2014, Muller *et al.*, 2010). But, to our knowledge, little research has been performed on the role and occurrence of phenotypic diversity in natural microbial communities (Ackermann, 2013). In this chapter, we compared two bacteria strains under axenic conditions, as cocultures where both species were grown together while being physically separated by a membrane, and as a truly mixed culture.

3 Materials and methods

3.1 Isolates

An *Enterobacter* sp. and a *Pseudomonas* sp. were selected from a set of drinking water isolates which were provided by Pidpa (Provinciale en Intercommunale Drinkwater-maatschappij der Provincie Antwerpen, Belgium). The selection was based on two criteria: both bacteria had distinctly different cytometric fingerprints (Rubbens *et al.*, 2017) and could reach the stationary phase within 24 hours. The isolates were identified with Sanger sequencing.

3.2 Experimental set-up

Before the start of the experiment, both bacteria were plated on nutrient agar plates (Oxoid). From each plate, a single colony was picked and transferred to liquid minimal medium (M9 with 200 mg/L glucose as carbon source). After two days of incubation at 28°C, cell densities in the liquid cultures were determined by flow cytometry and the cultures were diluted to the desired starting cell densities in fresh medium. The required dilution was high enough to neglect differences in volume of fresh medium, and thus resources for growth, that were needed to prepare the cultures. The starting cell densities were set to have an initial cell density of 10^6 cells/mL in each microcosm.

Microcosms were prepared in transwell plates (Corning® Costar® 6-well cell culture plates, Corning Incorporated) where apical and basal compartments were created using cell culture inserts (ThinCert™ Cell Culture Inserts with pore diameter 0.4 µm, Greiner Bio-One). The membranes of the culture inserts were replaced by membranes with smaller pore sizes to avoid migration of bacteria between the two phases (Cyclopore® polycarbonate and polyester membranes with 0.2 µm pore size, Whatman). Four microcosms were created (**Figure 6 - 1**). Both isolates were grown in axenic conditions as a non-interacting reference. A coculture with a membrane between both taxa was prepared to study the community members separately while they could interact through the membrane. Lastly, a mixed culture without physical separation, representing 'full interaction', was made. Each microcosm was prepared in triplicate and randomized over the plates to account for plate effects. The plates were incubated at 28°C and gently shaken (25 rpm) to aid diffusion of the metabolites between the compartments. The communities were monitored over a period of 72 hours. Every 24 hours samples were analyzed by flow cytometry. After 72 hours, samples were also fixed and subsequently analyzed with Raman spectroscopy.

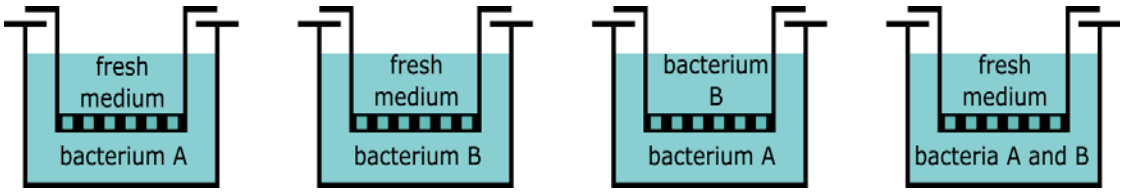


Figure 6 - 1: *Illustration of the experimental set-up. Bacteria in apical and basal phase can interact while they are physically separated by a membrane with a 0.22 µm pore size at the bottom of the cell culture inserts. Four microcosms were created: two axenic cultures, a coculture and a mixed culture. For each synthetic community biological triplicates were prepare.*

3.3 Raman spectroscopy

The fixation protocol for Raman spectroscopy was adapted from a previously described protocol (Read and Whiteley, 2015). 1 mL of cell-suspension was centrifuged for five minutes at room temperature and 5000 g. The supernatant was discarded and the cell pellet was resuspended in cold, 0.22 µm-filtered PBS (4°C). The cell-suspension was again centrifuged for five minutes at room temperature and 5000 g. The supernatant was discarded and the cell pellet was resuspended in the fixative, 0.22 µm-filtered 4% (v/v) paraformaldehyde in PBS (pH 7.2). The sample was allowed to fix for one hour at room temperature, in the dark. The fixative was removed by centrifuging for five minutes at room temperature and 5000 g and resuspending the pellet in cold, 0.22 µm-filtered PBS (4°C), twice. The fixed sample was stored at 4°C. Prior to analysis, the fixed sample was centrifuged for five minutes at room

temperature and 5000 g and the pellet was resuspended in 0.22 µm-filtered milli-Q (4°C). 10 µL of cell suspension was spotted onto a CaF_2 slide (Crystran Ltd.) and allowed to dry. The dried sample was analyzed immediately using a WITec Alpha300 R+ confocal Raman microscope with a 100x/0.9NA objective (Nikon), a 785 nm excitation diode laser (Toptica) and a UHTS 300 spectrometer with a -60°C cooled iDus 401 BR-DD CCD camera (Andor Technology Ltd.). Laser power before the objective was measured daily and was about 150 mW. Spectra were acquired in the range of 110-3375 cm^{-1} with 300 grooves/mm diffraction grating. For each single cell spectrum, the Raman signal was acquired over 40 seconds. All Raman samples were analyzed within one week after sampling, with minimal time between them to limit possible differences caused by differences in duration of the storage. For each population between 51 and 55 single-cell spectra were measured from a single biological replicate. To equalize the sample size, 51 spectra of each sample were selected for further analysis. Spectra with the lowest intensity were assumed to be of lesser quality, therefore the spectra with the lowest average intensity were discarded.

3.4 Flow cytometry

For flow cytometric analysis, the samples were diluted and stained with SYBR® Green I (SG, 100x concentrate in 0.22 µm-filtered DMSO, Invitrogen). Staining was performed as described previously (Prest *et al.*, 2013), with incubation for 20 min at 37°C in the dark Samples were analyzed immediately after incubation on a FACSVerse™ flow cytometer (BD Biosciences) with nine fluorescence detectors (527/32nm, 783/56 nm, 488/15 nm, 586/42 nm, 700/54 nm, 660/10 nm, 783/56 nm, 528/45 nm and 488/45 nm), a scatter detector and a blue laser (20 mW, 488 nm), a red laser (40 mW, 640 nm), and a violet laser (40 mW, 405 nm).

3.5 Data analysis

3.5.1 Phenotypic diversity analysis

The data was imported in R v3.3.1 (R Core Team, 2015) using the *flowCore* package v1.40.3 (B. Ellis). A quality control of the datasets was performed using the *flowClean* package v1.12.0 (Fletez-Brant *et al.*, 2016). After quality control, the background of the fingerprints was removed by manually creating a gate on the primary fluorescent channels. The *PhenoFlow* package v1.1 (Props *et al.*, 2016) was used to assess the phenotypic community structure of the bacterial populations. In short, for each bivariate parameter combination (e.g. scatter and fluorescence parameters) a binning grid is applied, this binning grid discretizes

the parameter space with each bin representing an operational phenotype. For each bin a kernel density estimation is applied. All density estimations are summed to the total density estimation of the community. The density values for each of the bins are then concatenated into a 1D-vector, which is called the 'phenotypic fingerprint'. From this fingerprint, the alpha diversity (*i.e.* within sample diversity) is calculated by means of the Hill diversity numbers and beta diversity (*i.e.* between sample diversity) is evaluated using the Bray-Curtis dissimilarity. Prior to diversity estimation, all populations were subsampled to 20 000 cells in order to have similar uncertainty levels.

3.5.2 *In silico* communities

After gating, the data was exported from R under Flow Cytometric Standard (FCS) format. The files were converted to comma-separated values (CSV) files to be further analyzed in Python, using the *InSilicoFlow* pipeline (Rubbens et al., 2017). In short, a fingerprint of the axenic cultures that make up the synthetic community is made. Next, the data of the axenic cultures is aggregated to a so-called '*in silico* community'. This *in silico* community consists of labeled data, which allows the use of supervised machine learning techniques. A classifier is trained to learn the difference between the fingerprints of the community-members. The label to be predicted is the taxon and the predictors are the scatter and fluorescence parameters. Once this classifier has been trained on the dataset, it can be used to predict the relative abundances of the taxa in a mixture. For training of the random forest, the biological replicates were pooled together and 10 000 cells of both taxon A and taxon B were randomly sampled. The data was partitioned into a balanced (*i.e.* the cell numbers for taxon A and taxon B are equal in these datasets) training and test set of 70% and 30% respectively. For this the cytometric fingerprints of the cocultures at the corresponding time points were used.

3.5.3 Raman data analysis

The data was analyzed in R v3.3.1. Spectral preprocessing was adapted from the study of Berry *et al.* (2015), using the package *MALDIquant* v1.16 (Gibb and Strimmer, 2012). In short, baseline correction was performed using the statistics-sensitive non-linear iterative peak-clipping (SNIP) algorithm. Next, the biologically relevant part of the spectrum (600-1800 cm^{-1}) was selected, over which the spectra hold 333 data points. The spectra were normalized by surface normalization and the necessity for peak alignment was evaluated.

4 Results and discussion

To assess the influence of cell-cell interactions on the phenotypic diversity of bacteria in microbial communities, microcosms with an *Enterobacter* sp. (bacterium A) and a *Pseudomonas* sp. (bacterium B) were prepared. One microcosm was prepared as coculture where the taxa were physically separated with a membrane but which allowed cell-cell communication. Another microcosm contained a mixed culture without physical separation between both taxa. The taxa were also cultured in axenic conditions as reference. To assess the changes in phenotypic diversity, both flow cytometry and Raman spectroscopy were used.

4.1 Flow cytometry

A comparison of the cytometric fingerprints based on the Bray-Curtis similarity measure shows that both taxa were clearly distinguishable when cultured in axenic conditions while the mixed culture was situated between both (**Figure 6 - 2**). The populations showed a significant shift in their cytometric fingerprints through time (p = 0.001, r^2 = 0.158). In addition, there was a significant difference in the fingerprints of taxon A as axenic culture compared to taxon A as coculture (p = 0.001, r^2 = 0.412). For taxon B the differences in the fingerprints between the axenic culture and the coculture were not significant (p = 0.089, r^2 = 0.168). The mixed culture shifted from a community with a cytometric fingerprint that was more similar to taxon A after 24 hours, towards a community with a cytometric fingerprint that was more similar to taxon B after 48 and 72 hours of incubation. Using the method described by Rubbens *et al.* (2017), an estimation of the relative abundance of taxon A and taxon B in the sample could be made with a model trained with the cytometric profiles of the cocultures at the corresponding time points. The abundance of taxon A was estimated to be 63% after 24 hours but decreased to 33% and 29% after 48 and 72 hours respectively. When the relative abundance of both taxa of the mixed culture were estimated with a model based on the cytometric fingerprints of the pure cultures at the corresponding time points or with a model based on the cytometric fingerprints of the pure cultures after 24 hours of incubation, the results showed to be different though all estimated a decrease over time of taxon A in the mixed culture suggesting that taxon B outcompeted taxon A in the mixed culture. Although this result showed to be consistent regardless of how the model was trained, care should be taken when considering the estimated relative abundance of both taxa as contact-dependent interactions could not be taken into account. In the cocultures no important decrease in cell density was observed over time for both taxa.

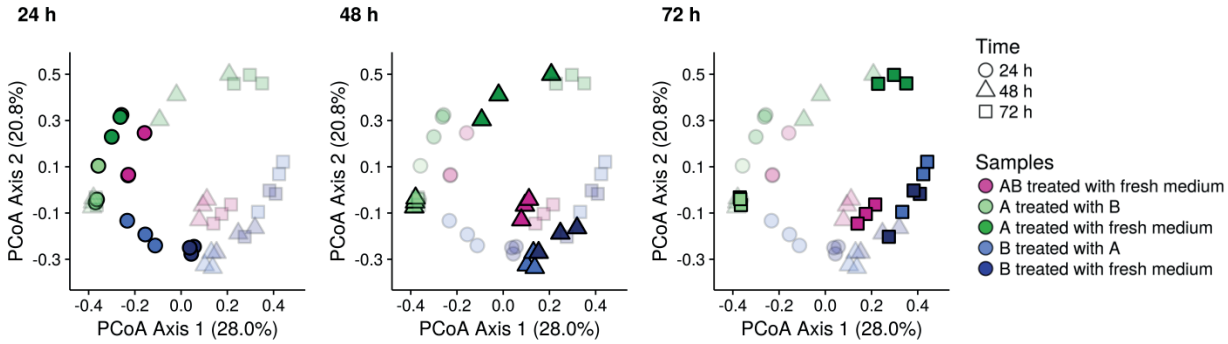


Figure 6 - 2: *PCoA ordination of the Bray-Curtis dissimilarities between the phenotypic fingerprints for all individual bacterial taxa in communities of axenic cultures, cocultures and mixed cultures. There were biological triplicates for each community. 'AB treated with fresh medium' refers to the mixed community without physical separation between the taxa. 'A treated with B' refers to the cytometric fingerprint of taxon A in coculture with taxon B and 'A treated with fresh medium' refers to the fingerprints of taxon A as axenic culture. Similar nomenclature was used for taxon B.*

The changes of the of the Hill number diversity indices in function of time per taxon reveal that the cytometric diversity in axenic cultures was larger than the cytometric diversity per taxon in coculture (**Figure 6 - 3**). Moreover, the differences between the diversity indices of the axenic cultures and the cocultures became larger over time. This result suggests that also the phenotypic diversity per taxon was lower in cocultures than in axenic cultures. A contrast analysis of the cytometric fingerprints between the cocultures and the axenic cultures in function of time shows that the cell population of taxon A in coculture shifted towards a smaller and more fluorescent population as compared to the axenic culture (**Figure 6 - 4**). For taxon B the difference was limited, with a small enrichment of a population with lower fluorescence intensity. Both the similarity and the Hill number diversity indices (D_0 and D_2) showed that the effect of microbial interaction on the phenotypic diversity was more pronounced for taxon A than for taxon B, indicating that both bacteria show phenotypic plasticity but that the taxa have different phenotypic responses to the interaction.

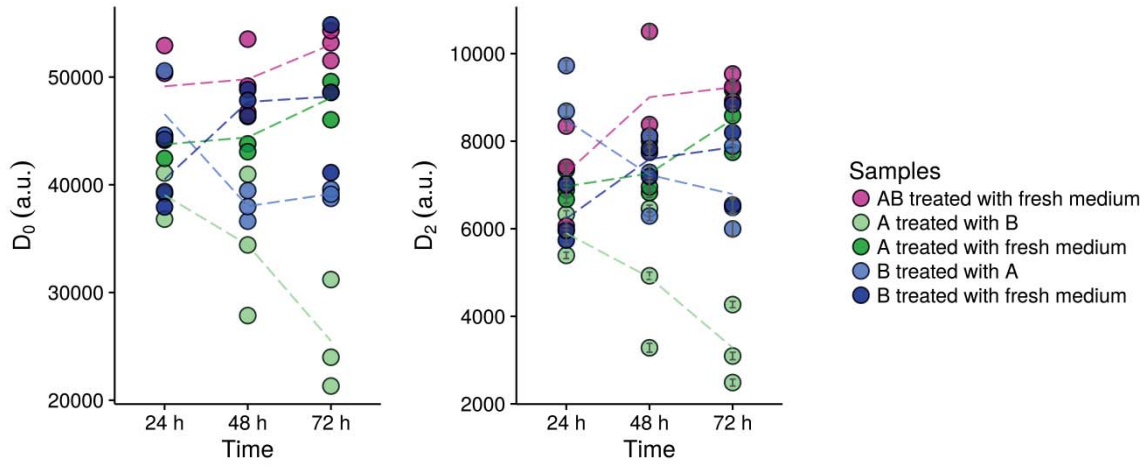


***Figure 6 - 3**: Hill diversity parameters D_0 (richness) and D_2 (diversity) for both individual bacterial taxa in communities of axenic cultures, cocultures and mixed cultures. Error intervals on the D_2 are generated by bootstrapping (999 bootstraps). There were biological replicates (n = 3) for each community. The dashed lines indicate the average trend of the replicates. Results for D_1 are not shown since D_1 is highly correlated with D_2 (r_p = 0.97).*

The contrast analysis of the cytometric fingerprints showed that the differences in scattering patterns were limited for both taxa, thus, there were no large changes in cell morphology (data not shown) but that the differences between the phenotypes was related to a higher fluorescence intensity of taxon A, and lower fluorescence intensity of taxon B. Since SG staining is a nucleic acid dye, a higher (or lower) fluorescence signal is directly related to a changed nucleic acid profile (Johnson and Spence, 2010). Two possible mechanisms could cause an increased fluorescence intensity, and inversely, a decreased fluorescence intensity. On one hand, the DNA copy number could have increased, increasing the fluorescence signal and suggesting cell replication. Although no growth was observed for both bacteria, it is possible that, under environmental stress, the bacteria adapt their cell cycle behaviour and chromosome content while maintaining a constant growth rate (Lieder *et al.*, 2016). On the other hand, RNA yields weaker fluorescent signals upon binding with SYBR green I in comparison to dsDNA (Solopova *et al.*, 2014, Johnson and Spence, 2010) and an increased fluorescence intensity could be attributed to a lower RNA content, indicating a shift in their gene expression. It has been established before that the gene expression shifts when bacteria are cocultured but no clear evidence whether the overall RNA content per cell increases or decreases has been reported (Sadabad *et al.*, 2015, Gonzalez-Torres *et al.*, 2015). Metabolic specialization could explain how certain genes can be upregulated while inducing an overall decrease in RNA concentration per cell (Johnson *et al.*, 2012).

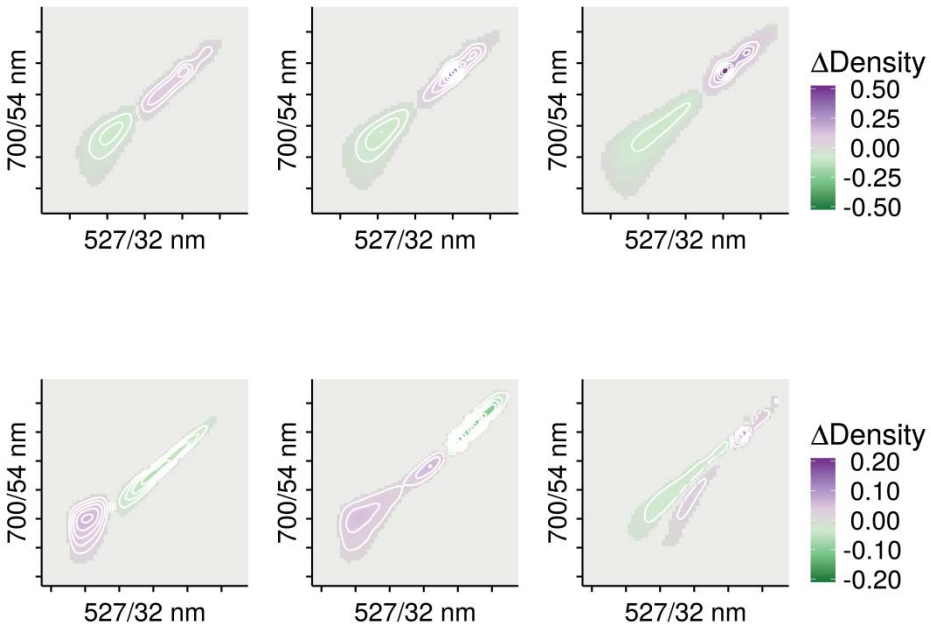


Figure 6 - 4: *Contrast analysis of the cytometric fingerprints between the axenic cultures and coculture members for each time point. Contrasts were calculated for taxon A (**A, upper row**) and taxon B (**B, lower row**) and were averaged over the biological triplicates. The color gradient indicates whether populations in the coculture increased (purple) or decreased (dark green) relative to their respective axenic culture at the specified time point. Pale green indicates no or very limited changes. If the difference between the two communities is lower than 0.01 no contrast value is shown on the graphs, which causes the appearance of different clusters.*

4.2 Raman spectroscopy

In parallel with flow cytometry, Raman spectroscopy was used to assess the phenotypic plasticity of both microbial taxa by measuring single-cell spectra of both taxon A and taxon B in the axenic cultures and in the coculture. A large peak in the range of 810 - 1030 cm^{-1} was present in the spectra of taxon A in the axenic culture, while this peak was not observed in any of the other populations. Intensity values showed large variability for this region. Since this might be the result of technical issues during fixation or storage of the sample, this region was excluded for further analysis (**Figure 6 - 5**).

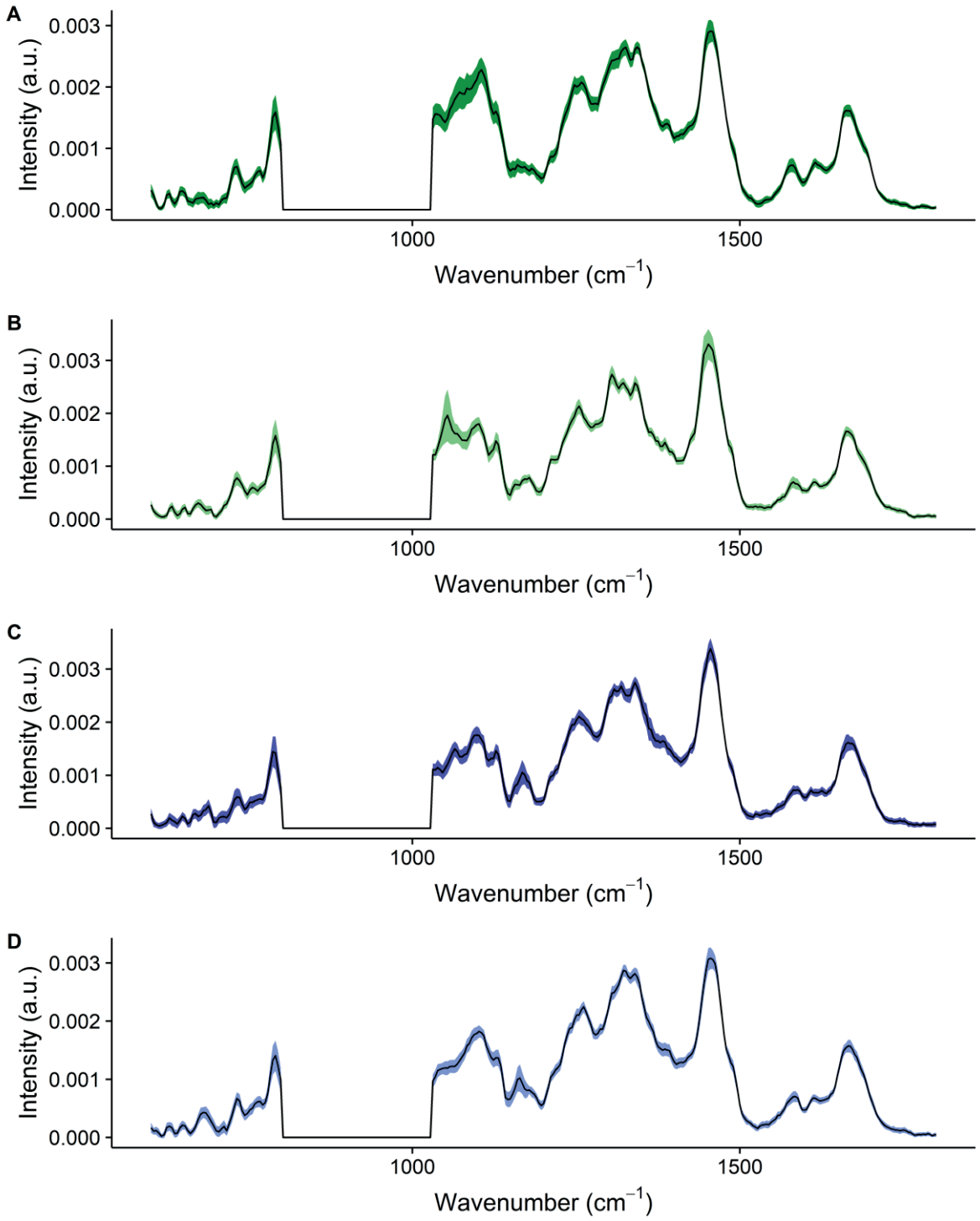


Figure 6 - 5: *Average Raman spectra of the single-cell measurements. The region of 810 - 1030 cm^{-1} was excluded for analysis, since technical issues during fixation or storage of the sample might have influenced this region. Taxon A in axenic culture (**A**), taxon A in coculture (**B**), taxon B in axenic culture (**C**) and taxon B in the coculture (**D**) are illustrated. Colored bands indicate the standard deviations. All average spectra are based on 51 single cell measurements.*

To gain insight in the difference between the phenotypes of each microcosm, spectra were visualized with PCA (**Figure 6 - 6**). The spectra of both taxa are separated well and the spectra of the same taxon with a different treatment are separated relatively well for taxon A, but not for taxon B. However, when performing PCA for each taxon separately, cells from each microcosm can be separated well (**Figure 6 - 6b and c**). Similar to the flow cytometric results, the differences were larger for taxon A than for taxon B.

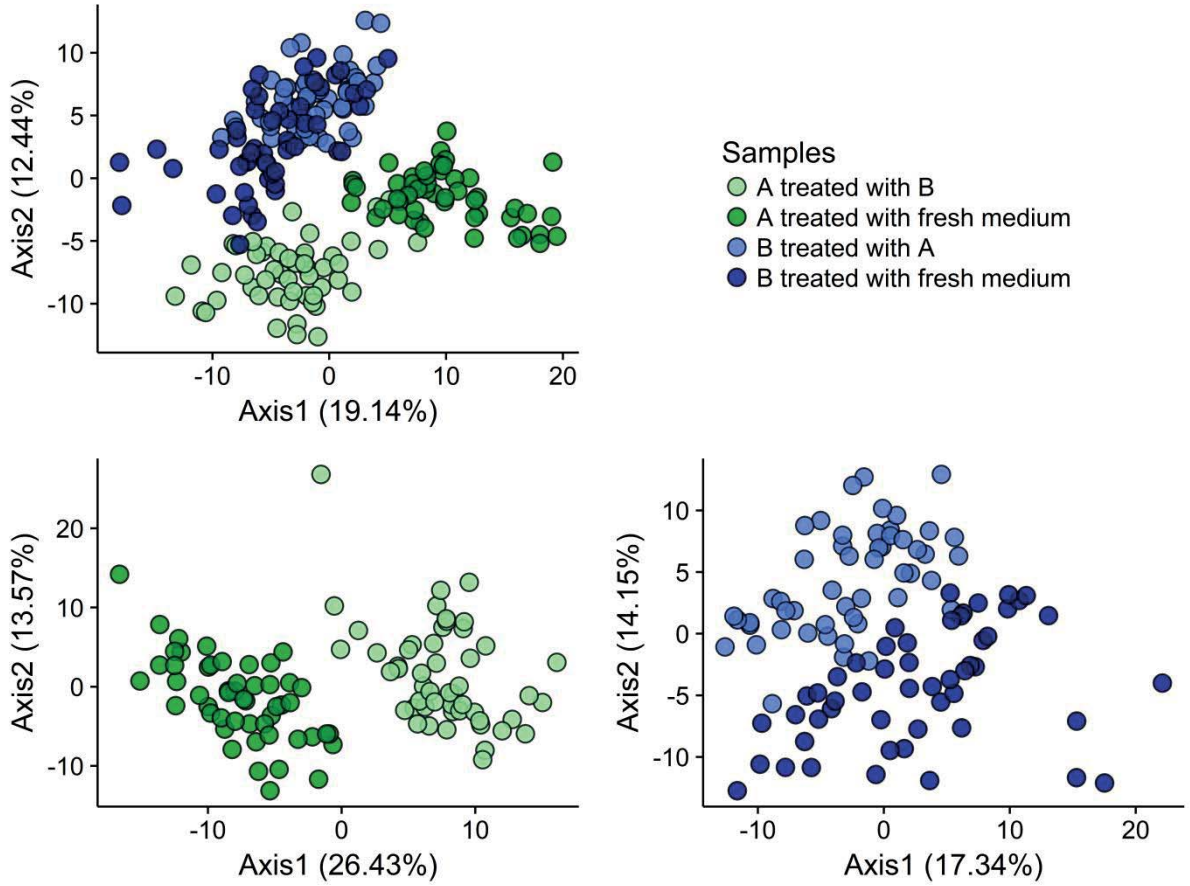


Figure 6 - 6: *PCA for all single cell Raman spectra (**A**), for taxon A (**B**) and taxon B (**C**). There are 51 single cell measurements for each population. Spectra were scaled and centred before performing PCA.*

A comparison of the peak profiles between the taxa as axenic culture and their coculture counterpart provides an insight in the biological causes of the observed differences. For taxon A, a higher level of nucleic acids was observed for the coculture in comparison with its axenic counterpart (wavenumbers 1053 cm^{-1}: nucleic acids; 1254 cm^{-1}: thymine, cytosine, adenine, and amide III; 1304 cm^{-1}: adenine, and amide III; 1486 cm^{-1}: nucleic acid backbone). The higher fluorescence intensity detected by flow cytometry combined with evidence of an overall higher nucleic acid content suggests a higher cell replication of taxon A in coculture. For taxon B, a lower concentration of nucleic acids, lipids, and proteins was observed in comparison with its axenic counterpart (wavenumbers 665 cm^{-1}: guanine; 779 cm^{-1}:nucleic acids; 1064 cm^{-1}: C-N and C-C stretching; 1297 cm^{-1}: CH_2 deformation; 1441

cm^{-1}: lipids ; 1675 cm^{-1}: amide I) (**Table 5 – 1, Chapter 5**). Especially wavenumbers 1064 cm^{-1} and 1441 cm^{-1} were important suggesting that especially the lipid concentration in of the cells was lower when taxon B was cocultured. These changes could be caused by a reduction of polyhydroxyalkanoic acids (PHAs) in the cells. PHAs are storage lipids and *Pseudomonas* sp. are known to synthesize them under several conditions.

The results of both flow cytometry and Raman spectroscopy confirm that both taxa exhibit a different level of phenotypic plasticity. For flow cytometry, the phenotypic traits contributing to the fingerprints are both the scatter signals (cell size and morphology) and fluorescence (nucleic acid content) but contrast analysis of those fingerprints revealed that only the nucleic acid content caused the detection of the phenotypic plasticity. Taking only these traits into account is an abstraction of the phenotypic diversity of the bacteria. Moreover, any phenotypic trait which is not related to the nucleic acid content or cell size and morphology could potentially remain undetected. Raman spectroscopy is a holistic method as it provides a single-cell fingerprint based on all molecules in the cell. This makes it possible to give a better biological interpretation of the observed differences. The comparison of the spectra showed that taxon A had a higher concentration of nucleic acids when cocultured and that taxon B had a lower concentration of nucleic acids, lipids and possibly proteins when in coculture. Results were similar to the results provided by flow cytometry despite the higher information content of the spectra. This suggests that both nucleic acid composition and the physiology are correlated and that the sensitivity of SG for small changes of nucleic acid suffices for the detection of small changes in nucleic acid profiles. Although Raman signals are weaker for each molecule, the multitude of molecules contributing to the fingerprints compensate for the signal weakness. Several methods could be implemented to improve the Raman signal strength and subsequently the sensitivity of the method (**Introduction**). Another downside of the method is that Raman spectroscopy still suffers from a low throughput. Phenotypic plasticity has been observed before and a well-known example is the adaptation of rhizobia bacteria which differentiate in the roots of leguminous plants (Fabaceae) to nitrogen fixing phenotypes (Koch *et al.*, 2010). In many cases, phenotypic plasticity is determined on one or a subset of specific genes or traits (Kummerli *et al.*, 2009, Corno and Jurgens, 2006). Alternatively, transcriptomics or proteomics can be used (Koch *et al.*, 2010, Barauna *et al.*, 2017). Read *et al.* (2013) used Raman spectra of *Campylobacter* cells could be classified according to their host of origin and is to our knowledge, the only publication reporting phenotypic plasticity with Raman spectroscopy.

5 Conclusions and perspectives

Flow cytometry and Raman spectroscopy have been proposed as fast and single-cell techniques for the phenotypic characterization of bacteria. In this chapter, we used both methods to determine the phenotypic diversity of two strains in a coculture. We showed that the bacteria exhibit phenotypic plasticity and that the adaption was different for the different species. Furthermore, we showed with flow cytometry that the phenotypic switch is related to a change in nucleic acid content though we could not precisely determine how the nucleic acid profile of the populations changed. Raman spectroscopy confirmed that the difference between the two populations of taxon A was due to an higher concentration in nucleic acids of the population grown in coculture which suggests a higher replication rate of taxon A in coculture. For taxon B, mainly a lower concentration in lipids was observed when in coculture. Finally, we conclude that Raman spectroscopy and flow cytometry are both useful for the phenotypic characterization of microbial cells.

6 Acknowledgements and author contributions

This work was supported by the project grant SB-131370 of the IWT Flanders, by the IMPROVED project, subvented by The interreg V “Vlaanderen-Nederland” program, a program for transregional collaboration with financial support from the European Regional Development Fund. More info : www.grensregio.eu, by the Geconcerteerde Onderzoeksactie (GOA) of Ghent University (BOF09/GOA/005), and by the Ernest Dubois prize 2016 of the King Baudouin foundation. The authors want to thank Katrien De Maeyer (Pidpa) for providing the cultures, and Jana De Bodt and Tom Bellon for the help in the lab.

For this work, J. Heyse performed the experiments, analyzed the data and helped writing the manuscript. B. Buysschaert designed the experiment, helped for the data analysis and the writing the manuscript. R. Props, P. Rubbens, A. Skirtach and N. Boon helped to design the experiments and improved the manuscript.

CHAPTER

7

SUBSTRATA DEFINE BIOFILM ECOLOGY IN WATER DISTRIBUTION SYSTEMS

CHAPTER

7

SUBSTRATA DEFINE BIOFILM ECOLOGY IN WATER DISTRIBUTION SYSTEMS

1 Abstract

Microorganisms in drinking water distribution systems play an important role in the water quality. Biofilms harbor a large fraction of these bacteria and can cause problems to the water industry as they can be a potential source of bacterial contamination, affect taste and odor, and cause biocorrosion of pipe materials. Since sampling biofilms from a water distribution systems is difficult or impossible in practice, we investigated if the bulk water can be used to describe the biofilms. In this chapter, we attempted to establish the relationship between the flow cytometric fingerprints of the bulk water and the biofilm. We show that flow cytometry can be used to monitor biofilm growth and demonstrated that both microbial communities follow different dynamics. Furthermore, the influence of the source of water and the pipe materials on the cytometric fingerprints was illustrated in batch tests and, in the case of the piping materials, also with a lab-scale flow-through system. Results show that plastic materials supported more biofilm growth than metallic materials. Finally, we show that *Enterobacter amnigenus* type II, a common drinking water contaminant, was able to invade biofilms and grow in water.

Chapter redrafted after:

Buysschaert, B., Minne, M., Boon, N., De Gusseme, B. Substrata define biofilm ecology in water distribution systems . *In preparation*

2 Introduction

Microbial quality of drinking water is of major importance to water distribution utilities as bacteria can affect the taste, color, and odor of drinking water. More importantly, pathogenic bacteria can lead to disease outbreaks and threaten public health. Between 2000 and 2007 there were 354 outbreaks of waterborne diseases related to drinking water, resulting in over 47 617 episodes of illness reported by 14 European countries (Miettinen, 2009). However, the presence of bacteria in drinking is inevitable. Typically, a distinction is made between three spatial locations for bacteria: the bulk water phase, the biofilms formed on the pipe surfaces and the sediments formed as a result of particle deposition (Prest *et al.*, 2016). The bulk or planktonic water bacteria are the smallest fraction and concentrations between 10^3 and 10^6 bacteria/mL, measured with flow cytometry, have been observed without adverse effect on human health (Hammes *et al.*, 2008). Biofilms are putatively the largest fraction of drinking water bacteria (Flemming, 2002), but recent research showed that also the sediments contain an important fraction of the bacteria (Liu *et al.*, 2013). Biofilm cell density can vary significantly with cell numbers in the range of 10^4 to 10^8 cells/cm^2 (Prest *et al.*, 2016). The formation of a biofilm starts with the initial attachment of the planktonic microorganisms to the pipe surface. After attachment to the surface, the bacteria undergo further adaptation and the biofilm growth and maturation begins. During the maturation, mushroom-like structures are formed in the biofilms and planktonic organisms detach from the biofilm and seed the bulk water (Garrett *et al.*, 2008). At a certain point, a steady state is reached at which the biofilm itself does not grow but where the embedded bacteria actively grow, and disperse in the bulk water. Inversely, planktonic bacteria from the bulk water or pathogens can attach to preexisting biofilms where they integrate and survive for a prolonged period of time (Flemming and Wingender, 2010). In this way, biofilms can potentially act as a source of harmful organisms. To avoid the dispersion of unwanted organisms in the drinking water distribution systems (DWDS), DWDS are disinfected with biocides such as chlorine, chloramine or chlorine dioxide (White, 1988, Berry *et al.*, 2015). However, the spatial structure of a biofilm, the occurrence of persister cells, and encapsulation of the bacteria in an extracellular polymeric substance make biofilms very resistant to disinfection or other types of treatment (Emtiazi *et al.*, 2004, Tachikawa *et al.*, 2005, Gagnon *et al.*, 2005). Also, biocides can result in unwanted by-products (Bull, 1982) or deterioration of taste and odor (Bryan *et al.*, 1973).

The community composition of the biofilms varies depending on pipe materials, disinfection strategies, temperature, and the age of the biofilms. The most common materials used for DWDS are: galvanized steel, cast iron, copper, polyvinylchloride (PVC), unplasticized polyvinylchloride (UPVC), chlorinated polyvinylchloride (CPVC), polybutylene (PB) and

polyethylene (PE) (WHO, 2006). Metallic materials have been used to transport drinking water for quite a long time all over the world. Galvanized piping is often used around the world because of its low cost, but its popularity is declining due its susceptibility for damage in case warm water is used (Nielsen and Yding, 1983). Similarly, cast iron pipes often lead to water discoloration due to corrosion, especially in case of higher residence times. Copper pipes suffer less from corrosion than iron pipes (WHO, 2006). In contact with most drinking water supplies, copper is able to develop a protective layer of copper oxides and carbonates on the inside of the pipe. This layer limits the amount of copper that can dissolve in the water and ensures that the amount stays below the maximum of 2 mg/L, as mentioned in the Drinking Water Directive (EU, 1998). In general, metallic plumbing is mainly accepted for outdoor use, but, because of its high mass and inflexibility, it is not often applied for internal water plumbing. Also, a wide variety of non-metallic materials are applied in DWDS. CPVC and UPVC, both PVC derivatives, are more suitable for drinking water applications due to their smooth internal surface. They are also able to transport water for long distances without problems of pressure loss, pitting or scaling. Both plastics can become brittle when exposed to UV for a prolonged time which makes them more suitable for underground piping. Polyethylene (PE) is a commonly used type of material. It is very light and flexible and is the preferred material for long-distance transportation of drinking water (WHO, 2006). Besides pipe material, many other factors control the microbial growth in DWDS. A key aspect are the available nutrients in the water (e.g. carbon, nitrogen, and phosphorus) and microbial growth can thus be controlled by removing and limiting the available nutrients with adapted water treatment processes (Vanderkooij *et al.*, 1982). Though drinking water distribution systems are commonly underground, they are still subject to temperature fluctuations (Liu *et al.*, 2013). These fluctuations can also have a significant effect on the biofilm formation and microbial growth. Pinto *et al.* (2014) showed that microbial communities in drinking water change seasonally, which is related to the competitive advantage of specific species in defined temperature ranges (Prest *et al.*, 2016). Furthermore, also flow rate variations, pH, the age of the distribution system, and the source of water (e.g. ground water or surface water) play a role in the composition of microbial communities and in the structure of the biofilms.

Because bacteria are ever-present and because disinfection alone is insufficient, a good monitoring of the microbial quality is necessary. Monitoring of bacterial communities can be done in a number of ways but conventionally, and for legislation, plating methods are used. Though plating methods are useful, they are not without limitations (De Roy *et al.*, 2012, Wang *et al.*, 2010) and alternative methods are explored. Flow cytometry has shown to be a promising method as it a fast and cheap method, able to quantify bacteria. Moreover, with

additional 'fingerprinting' algorithms, extra information can be extracted from each measurement (Prest *et al.*, 2013, Props *et al.*, 2016, Buysschaert *et al.*, 2017). Props *et al.* (2016) and Koch *et al.* (2013c) both showed that when microbial communities changed, the subsequent fingerprint also changed. In this way, flow cytometry can be used to monitor the community dynamics of the bulk water in the DWDS. In practice biofilms are often impossible to sample for monitoring, yet their role in the drinking water quality is significant and it is hypothesized that they seed the microbial community of the bulk water by dispersion of bacteria from the biofilm. In this chapter, we attempted to establish the relationship between the flow cytometric fingerprints of the bulk water and the biofilm and demonstrate that both microbial communities follow different dynamics. Furthermore, the influence of the source of water and the pipe materials on the cytometric fingerprints was illustrated in batch tests and, in the case of the piping materials, also with a lab-scale flow-through system. The dynamic and protective nature of biofilms increases the risk for drinking water contaminants to colonize the biofilm and to subsequently jeopardize the water quality, regardless of the water treatment. We showed that *Enterobacter amnigenus* type II, a drinking water contaminant isolated from a DWDS, was able to colonize biofilms and to grow in clean water.

3 Materials and methods

3.1 Set-up

3.1.1 Batch test with different sources of tap water

A batch test was set up in glass vials. First, the vials were cleaned to remove the assimilable organic carbon (AOC) according to a previously described method (Charnock and Kjonno, 2000). Briefly, vials and screw caps were washed once with detergent and once without, rinsed three times with milli-Q water, soaked overnight in 0.2 M HCl and again rinsed three times with milli-Q water. The glass vials were covered with aluminum foil and heated to 550°C in a muffle oven for six hours to remove all trace organics. The screw caps were soaked in a 10% sodium persulphate solution at 60°C for at least one hour, rinsed three times with milli-Q water and finally air-dried. The AOC-free vials were filled with 15 mL of potable tap water from six different drinking water production plants. These production plants produce potable water of slightly different qualities from ground water in different regions (**Table 7 - 1**). To each type of water, coupons of PVC, CPVC, cast iron, and copper were added. Each coupon had a diameter of 12.7 mm and a surface of 1.26 cm². A series of blanks without coupon was included as well. The experiment was performed in quadruplicates for all materials. The vials were incubated at 21°C on a shaker at 90 rpm,

which resulted in an estimated water velocity of 0.15 m/s, for four weeks. After four weeks of incubation, the coupons were transferred in 50 mL milli-Q water and the biofilms were detached by sonication (Sonifier 250, Branson Ultrasonics) with a power of 200 W and an amplitude of 20 kHz for 80 seconds. Both the bulk water and the detached biofilms were then subsequently analyzed with flow cytometry.

Table 7 - 1: *Chemical and biological quality parameters from the water of the six different production plants.*

Region		A	B	C	D	E	F
Origin		ground water	ground water	ground water	ground water	ground water	ground water
E. coli	CFU/ 100 mL	<1	<1	<1	<1	<1	<1
Enterococci	CFU/ 100 mL	<1	<1	<1	<1	<1	<1
Total Coliformes	CFU/ 100 mL	<1	<1	<1	<1	<1	<1
HPC (22°C)	CFU/ 100 mL	<1	4	3	3	4	<1
Aeromonas	CFU/ 100 mL	absent	absent	2	absent	absent	absent
Conductivity	µS/cm	401	223	320	466	389	479
Temperature	°C	12.7	12.9	12.9	12.5	12.4	11.8
pH		7.7	8.2	7.7	7.8	8.1	7.5
Free chlorine	µg/L Cl2	<50	<50	<50	<50	<50	<50
Total chlorine	µg/L Cl2	93	<50	<50	<50	<50	90
Nitrate	mg/L	0.9	1.2	1.3	2	1.6	1.6
UV254	abs/m	2.6	2.1	6.2	4.4	5.1	2.7

3.1.2 Batch test with different materials

Similarly to the previous test, a batch test was set up in AOC-free vials and filled with 15 mL of tab water. In this experiment, only one source of tap water was used. Coupons of PVC, CPVC, two types of UPVC, PE, PE80, PE100, PE100-RC, cast iron, and copper were added to the vials. Vials without coupons served as blanks. The experiment was performed in quadruplicates for all materials with the exception of UPVC type 2, PE80, PE100 and PE100-RC, which were analyzed in duplicates. The vials were incubated at 21°C on a shaker (90 rpm) for four weeks. The water was refreshed twice a week and subsequently analyzed with flow cytometry. After four weeks of incubation, the coupons were transferred in 50 mL freshly filtered milli-Q water and the biofilms were detached by sonication. The detached biofilms were also analyzed with flow cytometry.

3.1.3 Lab-scale flow-through system

A lab-scale flow-through system was built with six different pipes: PVC, CPVC, UPVC, PE, iron, and copper. All pipes had an inner diameter between 11.15 and 12.45 mm and were cut to the same length of approximately 1.5 m. The set-up was connected with an automatic

water distributor to a potable water tap. Every day, the six pipes were successively flushed for 15 minutes. The experiment was carried out for ten weeks (68 days). After a start-up period of 12 days, every two weeks, the stagnant bulk water inside the pipes was collected and biofilm samples were taken from every material. Biofilm samples were taken by cutting four pieces of 1.5 cm per pipe and biofilms were detached as described above. Both biofilm and bulk samples were analyzed in technical quadruplicates with flow cytometry.

3.1.4 Microbial invasion

A batch test was set up with four materials: PVC, CPVC, cast iron and copper. Again, AOC-free vials were filled with a volume of 15 mL tap water. Of every material, four series of seven replicates were used: four replicates for flow cytometric analysis, one for fluorescence microscopy, a vial with autoclaved tap water and a blank vial without coupon. The experiment was performed in two parts. In the first part, a *gfp*-labeled *Enterobacter amnigenus* type II (see below), was added in vials containing clean and sterile coupons, while in the second part, the experiment was repeated with coupons that were already colonized by a four-week old biofilm and briefly rinsed with a 1.5 % (v/v) NaOCl solution before the experiment. All vials were incubated at 21°C on a shaker (90 rpm). After an initial incubation period of 12 days, the water of each vial was refreshed twice a week and analyzed on the flow cytometer. After 12 days and 19 days also the biofilms were analyzed with microscopy and flow cytometry (**Figure 7 - 1**). This was repeated for both parts of the experiment.

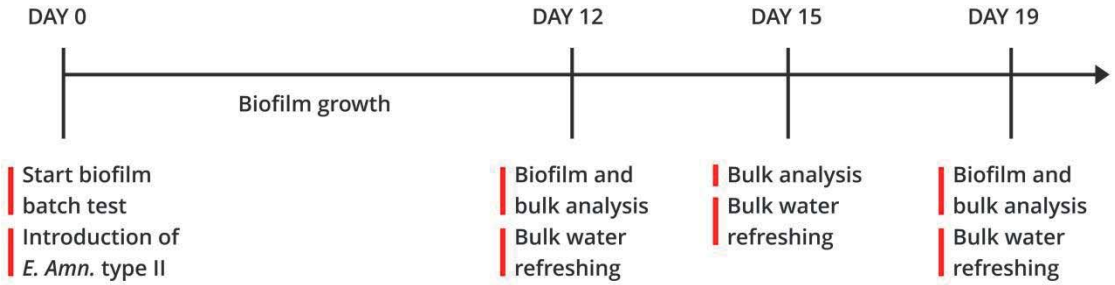


Figure 7 - 1: *Sampling frequency of the batch test for assessing* Enterobacter amnigenus *type II invasion of biofilms. This scheme was repeated for clean coupons and pre-colonized coupons.*

3.2 Transformation

An *Enterobacter amnigenus* type II was isolated from a DWDS and identified by the drinking water company Pidpa (Belgium). To differentiate the strain from native drinking water bacteria, a fluorescent marker was inserted in the genome by introducing a gene that

encodes for a green fluorescent protein (GFP) with biparental bacterial conjugation. The *Enterobacter amnigenus* acceptor strain was first made resistant to 100 mg/L rifampicin (Rf) in Luria-Bertani (LB, Oxoid) broth medium. An *Escherichia coli* S-17 λ pir strain containing a pUT-miniTn5-Km plasmid with a GFP marker and resistance genes for kanamycin (50 mg/L Km) and ampicillin (50 mg/L Amp) acted as donor (Delorenzo *et al.*, 1990). The donor was inoculated on a non-selective LB agar medium together with the resistant acceptor and incubated for 24 hours at 37°C. Subsequently, colonies were picked and transferred on LB agar medium with 50 mg/L Km (Sigma-Aldrich) and 100 mg/L Rf (Sigma-Aldrich). Successfully transformed bacteria were able to grow on both antibiotics. Fluorescence was assessed with UV illumination and with fluorescence microscopy.

Before the experiment, the strain was grown on a selective LB agar medium with 50 mg/L Km and 100 mg/L Rf to ensure purity. After 24 hours of incubation at 37°C, the bacterial strain was incubated twice for 24 hours in M9 minimal medium containing 400 mg/L glucose. Subsequently, the bacterial suspension was centrifuged for ten minutes at 1500 g. The supernatant was removed and the bacterial pellet was resuspended in an equal amount of autoclaved and filtered tap water. This was repeated three times to ensure completely wash-out of the growth media. After this washing step, 1 mL of the *Enterobacter* suspension, with a concentration of $5*10^8$ cells/mL, was added to each vial. The *Enterobacter* concentration was assessed with flow cytometry during the test by comparing an unstained sample with an SG-stained aliquot.

3.3 Flow cytometry

All samples were measured with a benchtop Accuri C6 cytometer (BD Biosciences). The stability of the instrument was controlled daily using 3 µm calibration beads (05-4018, Sysmex-Partec) and the instrument was calibrated according to the manufacturers standard. The blue laser (488 nm) was used for the excitation of the stains. The filters for the (fixed gain) photomultiplier detectors used during the measurements were 533 nm with a bandpass of 30 nm for the green fluorescence (FL-1) and 670 nm longpass filter for the red fluorescence (FL-3). The threshold was set on the 533/30 nm (FL-1) detector at the arbitrary unit of 500. Samples were stained with 10 µL/mL SYBR Green I staining solution (SG) to visualize the cells (Prest *et al.*, 2013).

3.4 Data analysis

Unless mentioned otherwise, all data was extracted from the proprietary Accuri C6 Csampler software version 1.0.264.21 in the flow cytometry standard (FCS 3.0) format and

subsequently imported into R version 3.4.0 (R Core Team, 2015) through the functionality offered by the *flowCore* package v1.42.2 (B. Ellis *et al.*). Data was first log transformed and subsequently normalized by dividing all values by the maximum fluorescence intensity signal. No compensation was applied. Gating to reduce the background was performed in R studio using the *flowCore* package on all samples. A 0.22 µm-filtered control was used to determine the position of the background. Additionally, a stained sample of the dilution buffer was used to assess the quality of the dilution buffer and of the stain. Next, a single-step discretization ('binning') and Gaussian bivariate density estimation was performed on the selected parameters (green and red fluorescence, FSC-H and SSC-H) using the *KernSmooth* package (Wand, 2015). A binning grid of 128 x 128 was fixed for each bivariate density estimation (Props *et al.*, 2016). All bivariate density estimations were concatenated to a one-dimensional feature vector, which we refer to as the fingerprint. Subsequently, the dissimilarity of the fingerprints was calculated using the quantitative Jaccard distance measure (Ružička index) as implemented in the function *vegdist* from the *vegan* package v2.4-3 (Oksanen *et al.*, 2016). Similarities were visualized with principal coordinate analysis (PCoA). Confidence ellipses were constructed at a confidence level of 95% and assumed a Gaussian distribution of the data. Cytometric diversity indices were calculated using the *phenoflow* package v1.1 (Props *et al.*, 2016). Statistical significance between the samples was calculated using an analysis of variance (ANOVA) and, in the event of a significant contribution, pairwise comparisons were computed. The statistical significance was computed using Tukey's post-hoc test and all differences resulting in p-values < 0.05 were considered significant.

4 Results and discussion

The impact of different sources of drinking water and piping materials on the flow cytometric fingerprints was assessed to evaluate how flow cytometry can be used as monitoring tool for DWDS. For this, batch tests with different types of drinking water and coupons of different commonly-used piping materials were set up. To mimic an actual distribution system, also a lab-scale flow-through system was built and used to monitor the changes of the biofilm and bulk water microbial community for different piping materials for a longer time-period. Each time both the cell density and the similarity between the cytometric fingerprints was evaluated. Finally, we studied the colonization potential of *Enterobacter amnigenus*, an indicator organism for drinking water quality, in the biofilms on different piping materials.

4.1 Source of water in relation to the type of piping material

After incubating PVC, CPVC, copper, and cast iron coupons with water from six different drinking water production plants and sources, the cell density of both the bulk and biofilm microbial community were compared (**Figure 7 – 2, a to f**). Statistical analysis revealed that for both phases, the type of material contributed significantly to the cell density, regardless of the water source. For the bulk phase, CPVC showed significantly higher concentrations than copper and cast iron, while for the biofilm, both CVPC and PVC showed significantly higher concentrations than copper and cast iron. The source of water did not impact cell density of the biofilms though region E showed significantly higher concentrations in the bulk phase than regions A and B. For some sources of water, also blank samples without coupons were incubated and analyzed. The bacterial concentration in the bulk phase was consistently and significantly higher for the blanks than for the samples incubated with pipe coupons though the sum of all bacteria per batch, regardless of their phase, was higher when coupons were added to the water. The cell density in the blank samples after four weeks showed a similar pattern as the initial cell concentration in the different waters: higher concentrations for water sources C and E compared to sources D and F (**Figure 7 - 2a**). These differences could be explained by the differences in UV_{254} which are indicative of the concentration aromatic carbon molecules. For each source of water, the differences in cell density in both phases suggest that bacterial growth is promoted by the presence of pipe materials and especially plastic materials such as UPVC and, to a lesser extent, PVC.

Plastic materials may be a source of organic compounds, such as plasticizers, stabilizers, softeners or coloring agents. These compounds can leach from the materials to the water and may be used by bacteria to support their growth in oligotrophic conditions and increase their capability to form biofilms (Rozej *et al.*, 2015). CPVC is produced from PVC resin after an additional chlorination step. Due this additional treatment step, more additives are introduced in the material, possibly responsible for a higher organic compound leaching and subsequent microbial growth. However, other studies have observed less growth on polymeric pipes compared to corrosion-prone materials, such as cast iron (Kerr *et al.*, 1999, Niquette *et al.*, 2000). Similarly to our findings, van der Kooij *et al.* (2014) report a higher biofilm growth potential of PVC in comparison to copper and cast iron in a batch experiment without water refreshment. The authors also illustrate the importance of the treatment of the material with a cleaning step which could also explain why we found CPVC to have a higher biofilm production potential than previously reported. In contrast to the smooth surface of the polymeric materials, the pitted surface of metallic materials, due to corrosion, can enhance biofilm growth by protecting bacteria from physical perturbations and from chemical disinfection (Liu *et al.*, 2016). Besides, the roughness of the iron surfaces can promote

microbial attachment and colonization due to the greater surface area. In contrast to the batch tests with smooth and new coupons performed in this study, the studies mentioned previously were also performed with a continuous flow which increases the abrasion force and creates a higher roughness of the iron materials. In addition, dissolved iron corrosion products, such as iron (oxy)hydroxides, can support the growth of specific biofilm-forming bacteria. The corrosion products can even retain nutrients, including carbon and phosphorus (Morton *et al.*, 2005). Although the iron coupons used for this experiment were corroded after a few days, no dissolved iron was detected in the bulk water according to atomic absorption spectroscopy (AAS). This may explain the 'lower' biofilm cell density on iron coupons as no dissolved iron was present in the water.

A comparison of the flow cytometric fingerprints with PCoA showed a varying distinction between the metallic and plastic coupons depending of the source of water (**Figure 7 - 3**). Especially cast iron coupons showed to be more different than the other materials which can be explained by the corrosion during the experiment. No clear distinction can be made between biofilm and bulk samples, which can be the result of an established equilibrium between both phases since the bulk water was not refreshed during the experiment. These results show that, apart from quantitative differences, that there are also qualitative differences between the cytometric fingerprints of the biofilms and the bulk water in function of the piping material and type of water. This suggests that the microbial community composition is different as flow cytometric fingerprints are known to differ according to the communities composition (Props *et al.*, 2016), thus explaining the different growth potential. Correspondingly, De Roy *et al.* (2012) showed that different cytometric fingerprints were found for different brands of bottled water. To confirm the putative differences in community composition, sequencing-based analysis should be performed and results compared to the flow cytometric data. A combination of both techniques also enables in the absolute quantification of the microbial taxa which improves the understanding of the microbial dynamics (Props *et al.*, 2017).

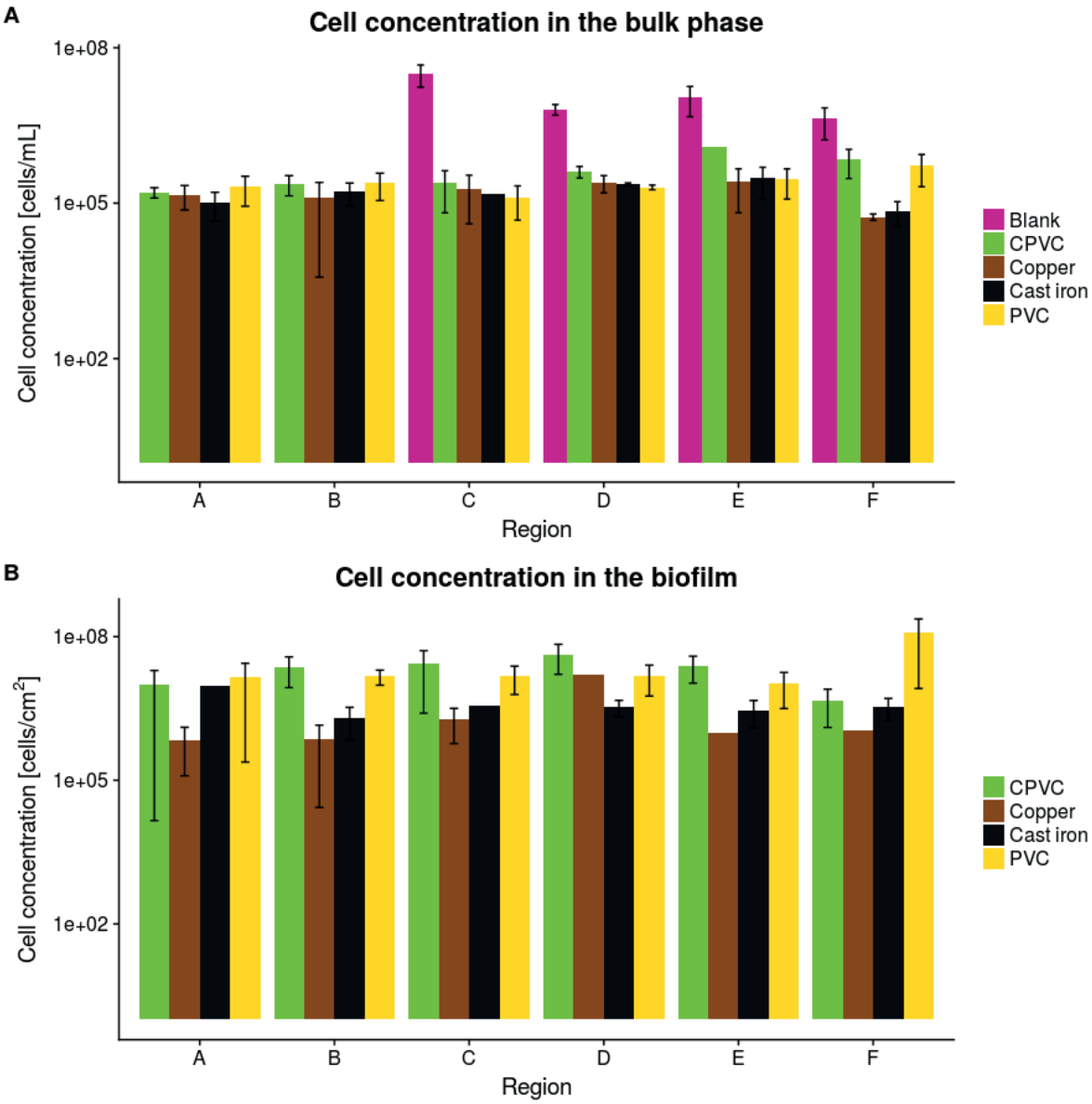


Figure 7 - 2: *Cell concentration in the bulk phase (**A**) and in the biofilms (**B**) after four weeks of incubation without refreshing the bulk water. The cell concentrations are visualized on a logarithmic scale with the corresponding error bars (n=4) for the six different water treatment plants and sources. Blank samples contained no coupons and for two regions blank samples are missing.*

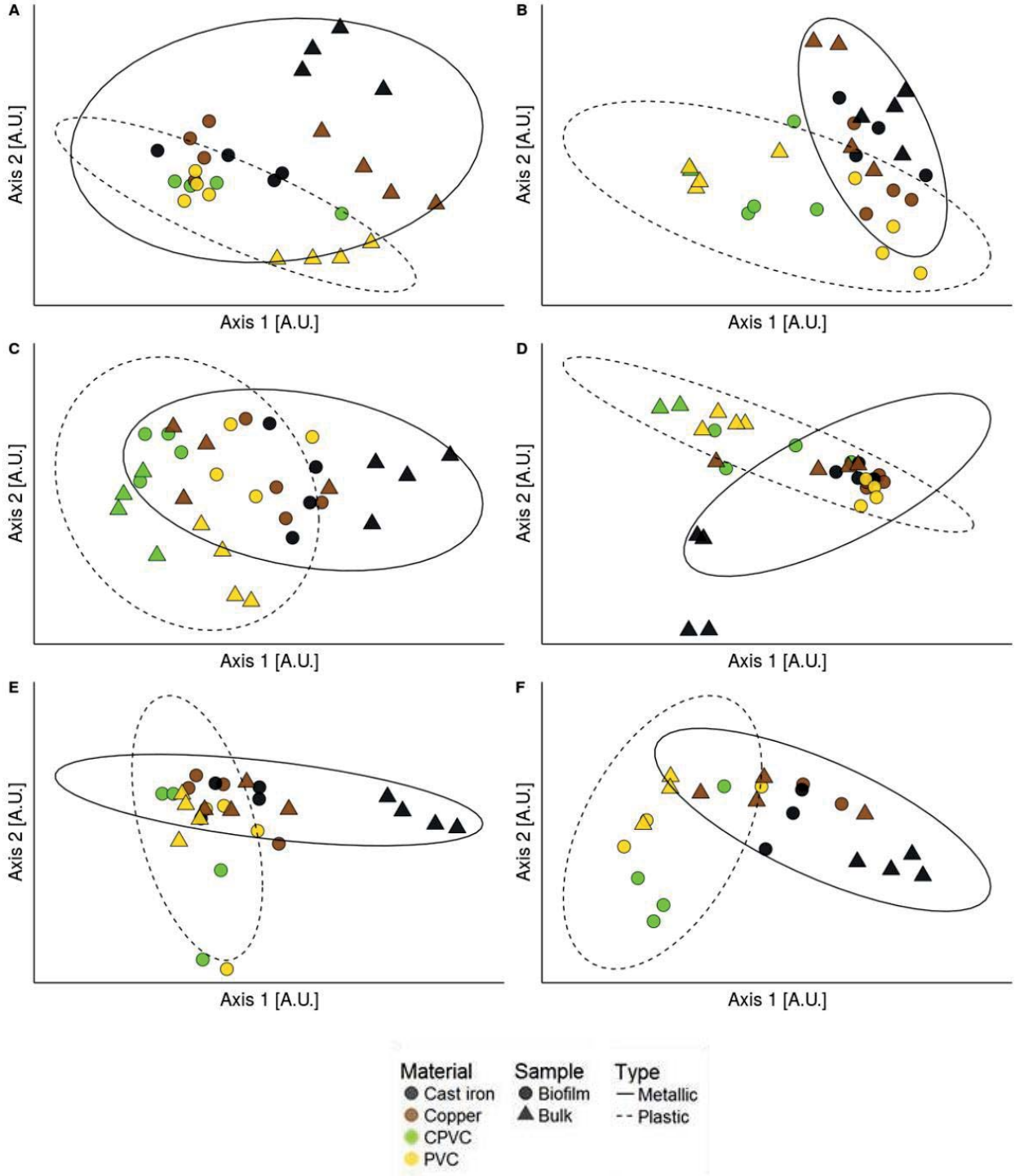


***Figure 7 - 3**: Principal coordinate analysis (PCoA) of all biofilm and bulk cytometric fingerprints for the six different sources of water (**A - F**) after four weeks of incubation. The bulk water was refreshed twice a week. Similarity was calculated with the density-based Jaccard distance measure (Ružička index). The colors represent the material of the coupons and the shapes distinguish between bulk water samples and biofilm samples. Confidence ellipses were constructed with a confidence level of 95% for the metallic and plastic materials.*

4.2 Effect of piping material

In a second experiment, the cell density in the bulk water was evaluated each time the water was refreshed. For all plastic materials, the changes in cell density in bulk water over time were comparable with the changes in cell density of the blank samples without coupon, suggesting little or no effect of the coupon on the bulk water cell density. The samples with a cast iron or copper coupon, the bulk cell densities were generally higher than the other samples, though not significantly (**Figure 7 – 4a**). The absence of a growth-promoting effect on the bulk water can be explained by the frequent replacement of the bulk water. Also the cell densities of the biofilms showed few significant differences after four weeks due to the high variability between replicates of PVC, PE80, and PE100. However, when comparing the average values, results are comparable to previous experiment where both PVC and CPVC yield more biofilm than copper or cast iron. The highest average density was found for the CPVC coupons ($4.12*10^7 \pm 1.07*10^7$ cells/cm²) (**Figure 7 – 4b**). Cell densities on the PVC, UPVC, PE80, PE100, and PE100-RC coupons varied around approximately $2*10^7$ cells/cm². Only the cell density on the CPVC coupons showed to be significantly higher than PE and UPVC type 1. Biofilm growth on UPVC was lower compared to PVC and CPVC. UPVC is unplasticized polyvinyl chloride and is totally free of plasticizers, thus reducing leaching of organic compounds, which may be responsible for the lower biofilm formation potential. Polyethylene (PE) showed noticeably lower biofilm cell density although Niquette *et al.* (2000) found a similar density of biomass growing on PVC and PE materials though after several months. No distinction between PE80, PE100, and PE100-RC could be made but, as the only difference between the materials is the polymerization process resulting in different flexibility, no big differences were expected. The repartition of the bacteria between the biofilm and the bulk did not provide a clear pattern distinguishing the different types of materials. In contrast to previous experiment, the bulk water was refreshed twice a week, which might explain why in this experiment less significant differences were observed altogether.

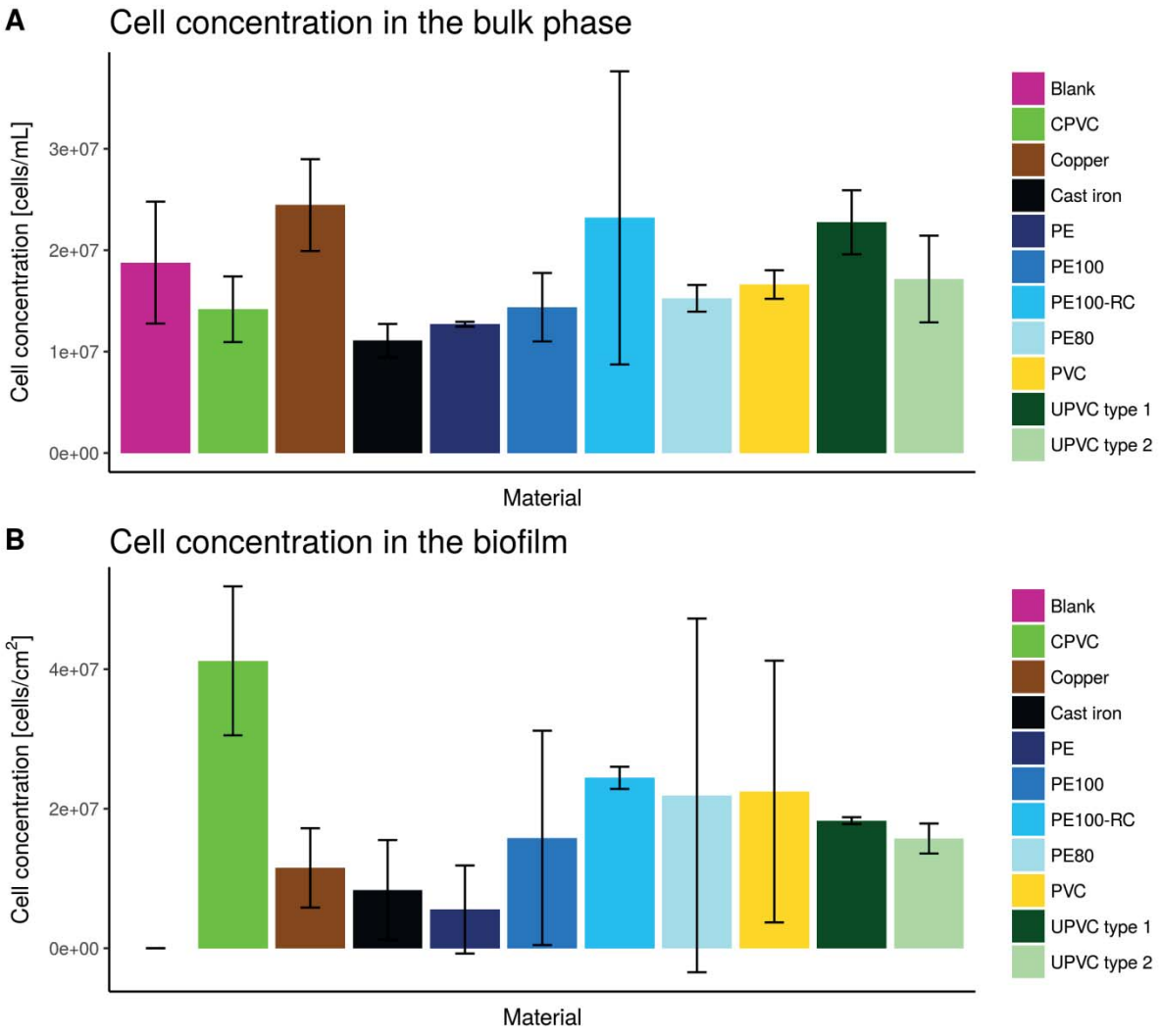


***Figure 7 - 4**: Cell concentration in the bulk phase (**A**) and in the biofilms (**B**) after four weeks of incubation. The cell concentrations are visualized on a linear scale with the corresponding error bars (n=4) for all coupon materials. Blank samples contained no coupons.*

When the flow cytometric diversity indices (D_2) of all samples from the bulk water are compared over time, a distinction can be made between the water incubated with plastic or metallic coupons. Plastic coupons show dynamics similar to the blank samples, with the exception of PVC for the first seven days and PE80 between 11 and 14 days (**Figure 7 - 5b,c and d**). Copper and especially cast iron show different dynamics after seven days (**Figure 7 - 5a**). The differences between cast iron and, to a lesser extent, copper with the other materials is also illustrated by the similarity between the fingerprint of all samples after 27 days (**Figure 7 - 6**). Furthermore, a good distinction can be made between the group of bulk samples and the group of biofilm samples, suggesting a different microbial community composition in the two phases. However, no relation can be established between the biofilm and the bulk water for each material. As the bulk water in combination with plastic materials seemed the least affected by the biofilm growth, based on the cytometric diversity, the plastic materials could be considered as the best materials to transport clean water. For cast iron, a

noticeably higher phenotypic diversity was observed, especially at the end of the experiment. As the cast iron coupons were completely corroded and little pieces of iron were released in the water, possibly also parts of the biofilm were released in the bulk water. Since biofilms are considered to have a higher taxonomic diversity due to a higher number of niches, and since a higher taxonomic diversity is partially linked with a higher phenotypic diversity, this can explain the higher diversity indices (Stewart and Franklin, 2008, Props *et al.*, 2016). On the other hand, for bulk waters with copper coupons, phenotypic diversity was noticeably lower compared to phenotypic diversity of the blanks after day 11. A possible explanation is that only specific phenotypes can grow in bulk water in the presence of copper coupons, which are known to have antimicrobial properties (Beeton *et al.*, 2014). In general, the cell concentrations and phenotypic diversity indices of the bulk samples showed a high variability, indicating that the microbial both biofilm and bulk water did not stabilize. This suggests, similar to literature, that biofilm growth is a long-term process and that longer test periods are needed for stable conditions (Martiny *et al.*, 2003).

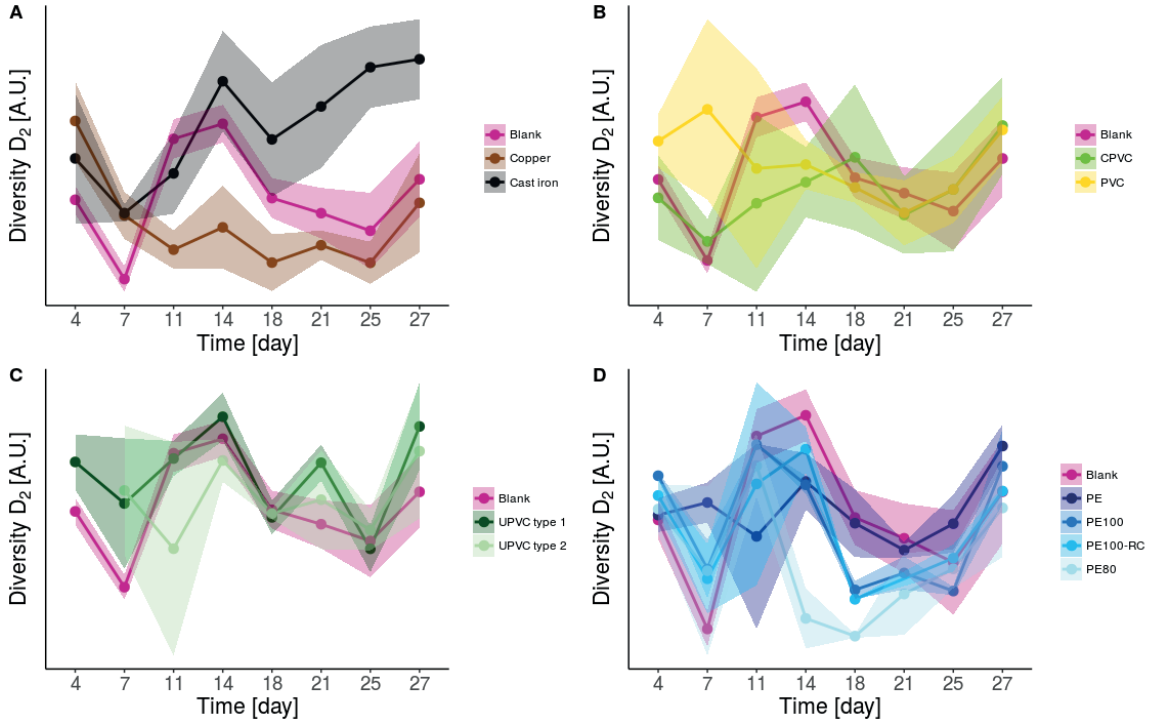


Figure 7 - 5: *Diversity indices D_2 of the bulk water for cast iron and copper (**A**), PVC and CPVC (**B**), UPVC type 1 and 2 (**C**), and PE, PE80, PE100 and PE100-RC (**D**). The blank samples were incubated without coupon and included in each plot to compare dynamics (**A-D**). The colored ribbons represent the standard deviation (n=4).*

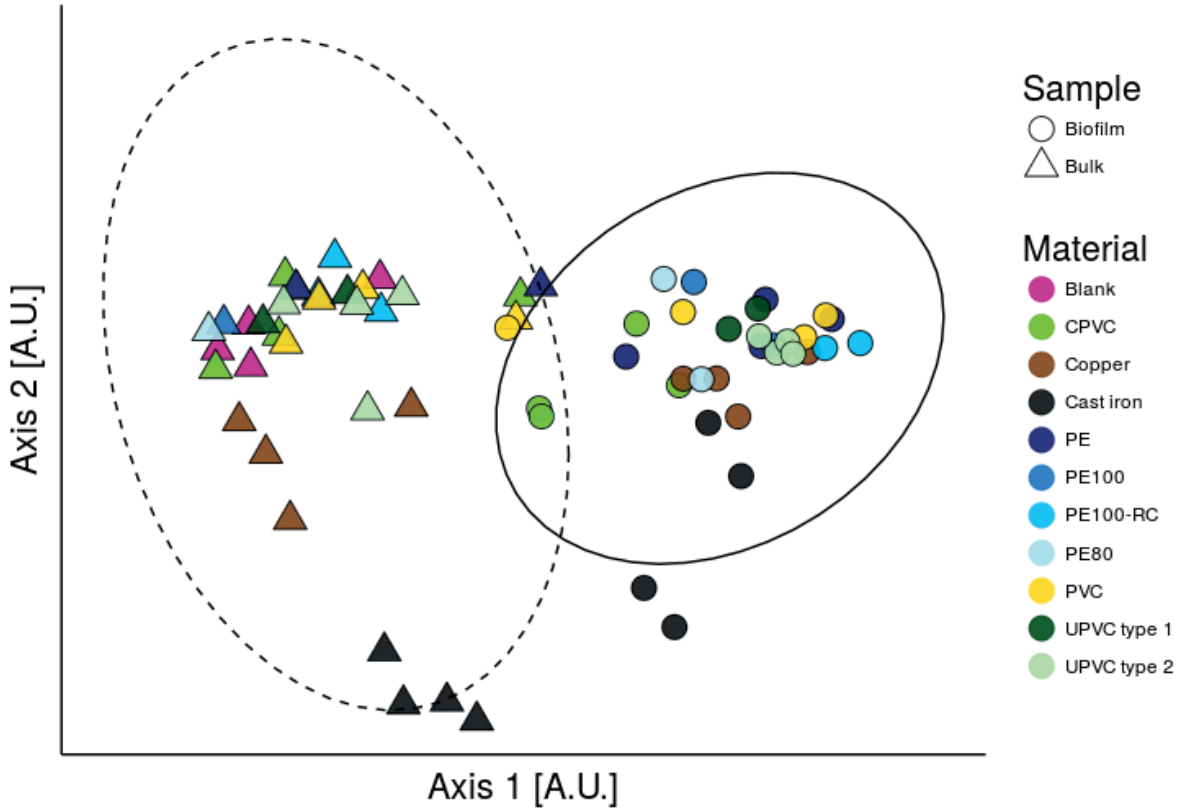


***Figure 7 - 6**: Principal coordinate analysis (PCoA) of both biofilm and bulk samples of all materials after 27 days. Similarity was calculated with the density-based Jaccard distance measure (Ružička index). The colors represent the material of the coupons and the shapes distinguish between bulk water samples and biofilm samples. Ellipses illustrate the distinction between bulk samples (dashed line) and biofilm samples (full line).*

4.3 Effect of the piping material in a flow-through system

Similarly to the batch test, the cell densities of the bulk water and the biofilms were measured with flow cytometry for every time point in a lab-scale flow-through system. The cell densities of the bulk water in combination with copper, cast iron, and PE pipes were in generally higher than the cell densities of the bulk water in combination with PVC, UPVC, and CPVC pipes. On the other hand, the biofilm density on the CPVC, UPVC, PE, and iron pipes was generally higher than for PVC. Biofilm growth on copper coupons was noticeably lower than other materials for all the time points, possibly due to the antimicrobial effect of copper. No clear increase in cell density was observed for the bulk water or biofilms between day 12 and 68, except for the biofilm on the cast iron pipes. Large standard deviations and no significant differences were observed in the cell density of either the bulk or the biofilms for any materials at the start of the experiment. In time, progressively more significant differences in the cell concentrations were observed. This suggests that more stable conditions can be created and that more relevant conclusions can be made when the biofilm is more mature.

The ratio between the cell density of the biofilm and the bulk fluctuated at all times around one for iron, PE, and PVC. CPVC showed to have proportionally more biofilm while copper showed, at all times, less biofilm with a ratio of 0.9 (**Figure 7 - 7a**). UPVC showed to have more biofilm in the beginning of the test and at the end, an equal ratio between both was measured.

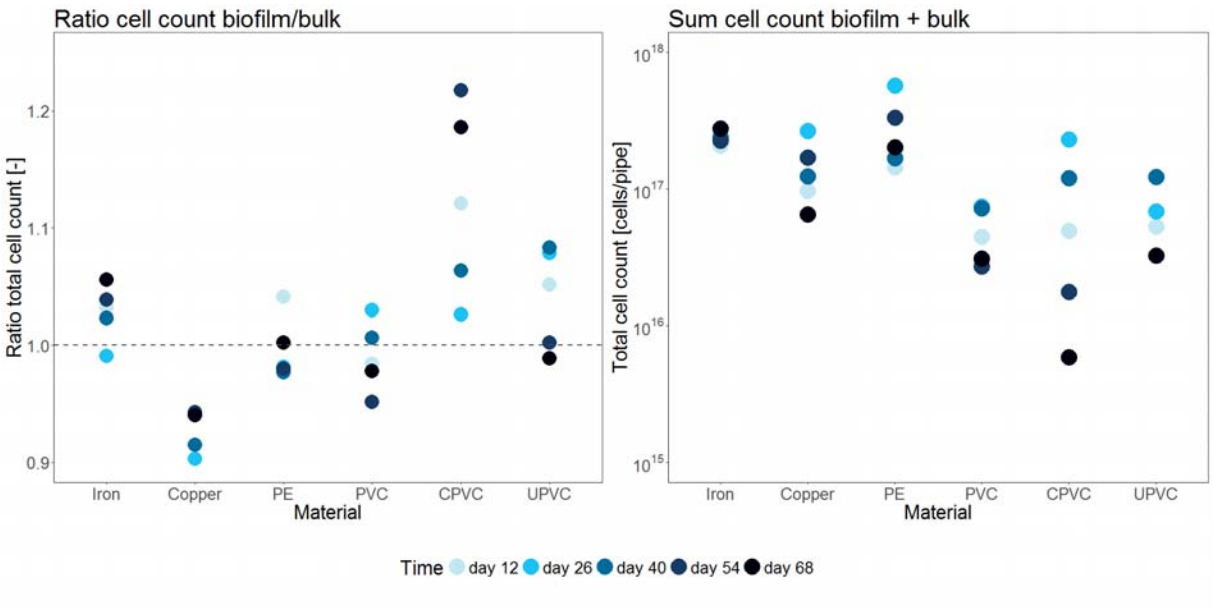


***Figure 7 - 7**: Ratio of the total cell density in the biofilm to bulk, calculated per pipe section for every material (**left**). Sum of the cell density in the biofilm and the bulk water per pipe section (**right**).*

Dynamics in community composition can be assessed with flow cytometric diversity index (D_2) and show, similarly to previous test, different dynamics for both copper and cast iron in comparison to the other materials (**Figure 7 - 8a,b**) (Props *et al.*, 2016). The bulk water with the PVC, CPVC, and UPVC coupons first showed similar dynamics relative to the inflowing water, though they were more diverging towards the end of the experiment (**Figure 7 - 8d,e and f**). In the same way, the bulk water incubated with the PE coupon initially showed a trend similar to the blank sample, yet cytometric diversity increased markedly towards the end of the experiment (**Figure 7 - 8c**). Biofilms on the coupons of all materials showed different dynamics for each coupon material, yet cytometric diversity was always higher than the diversity of the bulk water or the inflowing water which is in line with the expected higher taxonomic diversity of biofilms (Besemer, 2015). When the similarities between cytometric fingerprints are visualized with principal coordinate analysis, a clear distinction between biofilm and bulk fingerprints can be made at all times (**Figure 7 - 9**), but no relation between the diversity dynamics in bulk and biofilm can be established per material. In comparison to other materials, cast iron yielded very different fingerprints for both bulk and biofilm water in comparison to all other materials. The corrosion of the iron pipes caused for more

background fluorescence in the cytometric fingerprints and could not be removed completely. No clear difference between the plastic materials can be observed at all times based on the similarity of their cytometric fingerprints. The bacteria concentration in bulk was comparable for both plastic and metallic materials, although the materials influenced the biofilm formation during the experiment. When considering the dynamics of the cytometric fingerprints, the plastic materials showed to remain be more similar to the incoming water than the metallic materials in this flow-through experiment. This suggests that the microbial community composition is less altered by the plastic materials than by the metallic materials for the duration of this experiment. But, as this experiment was performed for a short period of time and because there are big quality differences between plastics or between manufacturers, results should be interpreted with caution. Furthermore, confirmation with sequencing methods should confirm this hypothesis.

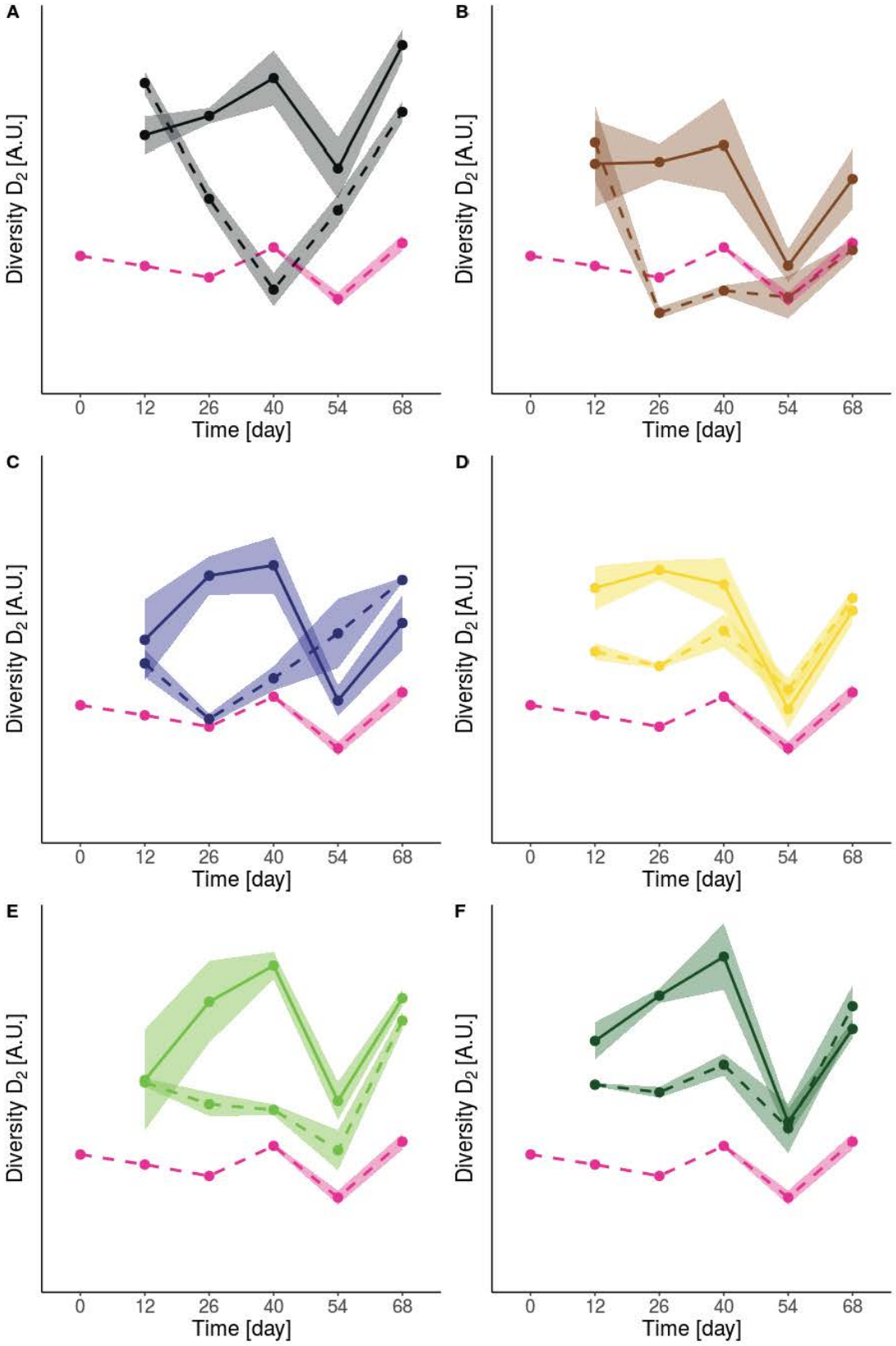


Figure 7 - 8: *Phenotypic diversity (D_2) dynamics in time for both biofilm and bulk water. Phenotypic diversity of the incoming water was included as blank (pink). The standard deviation is represented as colored ribbons (n=4). Results are shown separately for the biofilms (full line) and the bulk water (dashed line) for iron (**A**), copper (**B**), PE (**C**), PVC (**D**), CPVC (**E**), and UPVC (**F**).*

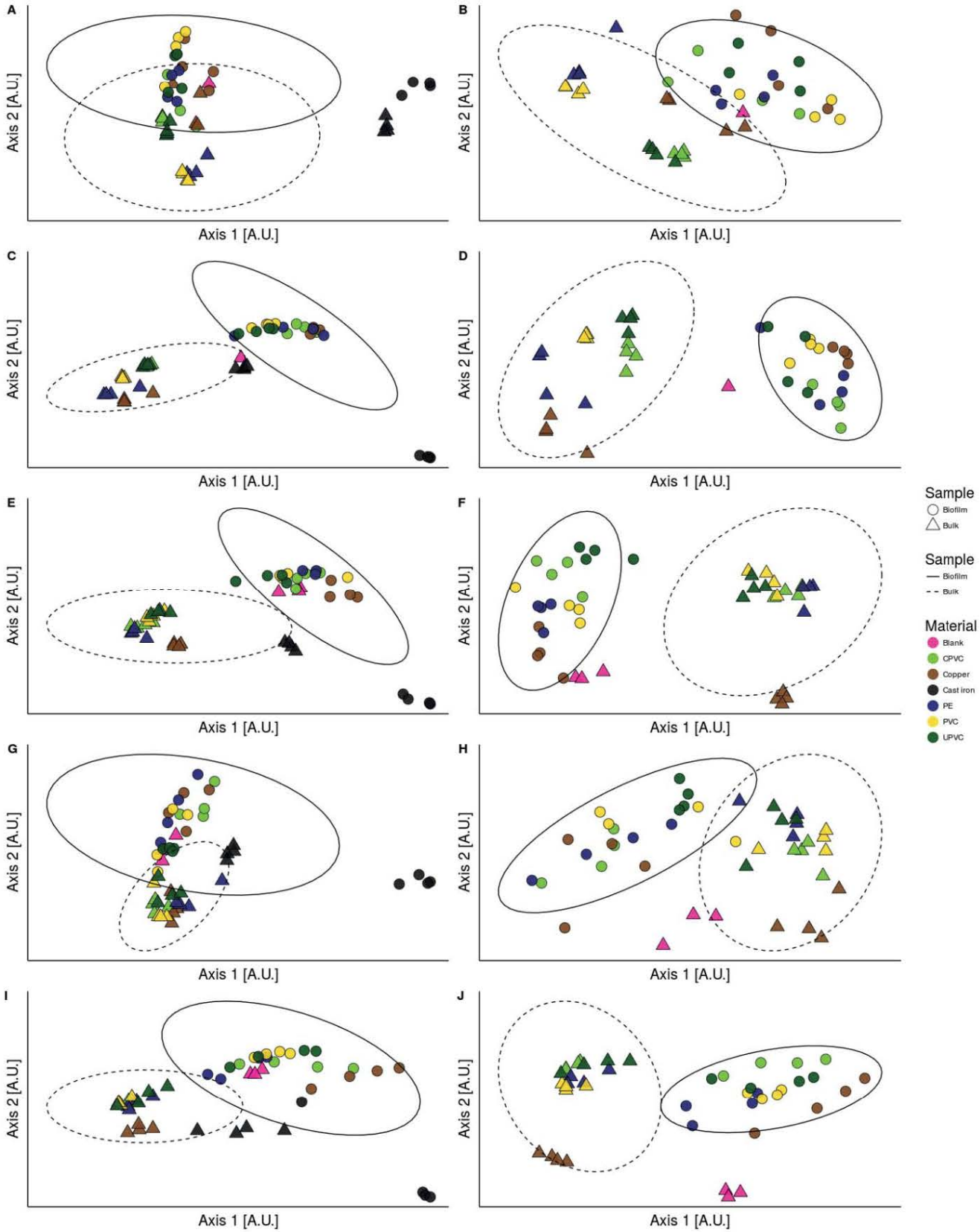


***Figure 7 – 9**: PCoA of the bulk and the biofilm samples for the six different pipe materials used of the lab-scale piping experiment (n=4). Similarity was calculated with the density-based Jaccard distance measure (Ružička index). Results are shown with iron (left) or without iron (right) for 12 (**A, B**), 26 (**C, D**), 40 (**E, F**), 54 (**G, H**), and 68 (**I, J**) days.*

4.4 Microbial invasion in biofilms

Bacteria from the bulk water or pathogens can attach to biofilms where they can integrate and survive for a prolonged period of time which can impact the water quality (Flemming and Wingender, 2010). An *Enterobacter amnigenus* type II has been isolated from a DWDS and caused recurrent water contamination. To assess the biofilm-formation capacity of *E. amnigenus*, a batch test was set up with clean coupons of four different materials: cast iron, copper, PVC, and CPVC. The bulk water and biofilms were both analyzed after 12 and 19 days of incubation. After the initial incubation period of 12 days, the bulk water was refreshed twice a week. A genomic *gfp* construct helped to distinguish *E. amnigenus* from the native microbial community.

Flow cytometric cell density measurements showed that biofilms grew for all coupons during the experiment, yet no statistically significant different biofilm concentration between the materials could be found (**Figure 7 - 10b**). In the bulk water, no clear microbial growth was noticed between 12 and 19 days for any sample, though copper supported the highest growth in the bulk water and PVC or CPVC the least (**Figure 7 - 10a**). *E. amnigenus* was present in all samples after 12 days of incubation in both the biofilm and the bulk water, suggesting the *Enterobacter* successfully colonized the biofilm. When the bulk water was refreshed twice a week, the concentration of *Enterobacter* in the bulk water decreased and, after 19 days, no *Enterobacter* could be detected with flow cytometry. Also, no autofluorescent cells were measured in the biofilm with flow cytometry. Microscopy showed that *Enterobacter* was still present in the biofilms though in low abundance (**Figure 7 - 11**). Copper and cast iron supported a visually higher concentration of *Enterobacter* in the biofilm. As flow cytometry cannot detect cell concentrations below approximately 10^3 cells/mL with our settings and instrumentation (Hammes *et al.*, 2012), we assume the cell concentration was below this level. The EU drinking water legislation states that no Enterococci should be detected with plating in a sample of 100 mL (EU, 1998) and, although the concentrations of *Enterobacter* in our samples were small and could not be detected with flow cytometry, they would lead to positive plate counts considering the high culturability of *Enterobacter*. In case autoclaved tap water was used to refresh the water, PVC and CPVC showed no wash-out of *Enterobacter* in both the bulk and the biofilm. Cast iron and copper showed no *Enterobacter* biofilm and the concentration in the bulk water decreased over time but remained detectable with flow cytometry suggesting *Enterobacter amnigenus* grew in the bulk water. This is in contrast to the vials with non-sterile tap water, where fluorescent *Enterobacter* was not always observed. An explanation is that *Enterobacter* can fully develop in the autoclaved tap water since no other living bacteria were present. The autoclaved tap water offers an ideal environment for the *Enterobacter* without competitors for nutrients. Moreover, the autoclaved

tap water contained dead bacterial biomass which *Enterobacter amnigenus* could use as substrate and nutrient source to grow in bulk water or develop biofilm structures on the different materials. This phenomenon has been reported before by Temmerman *et al.* (2006) with *Legionella pneumophila* and was referred to it as necrotrophic growth. This suggests that unsuccessful disinfection could enhance the growth of an introduced contaminant by making more nutrients available. Furthermore, the presence of other bacteria in the water could be important to suppress the growth of specific drinking water contaminants by competing for common resources. These hypotheses match with the established concept of biologically stable water, where biological growth in the drinking water distribution systems is mitigated by reducing and limiting the available nutrients, rather than by disinfection (Prest *et al.*, 2016).

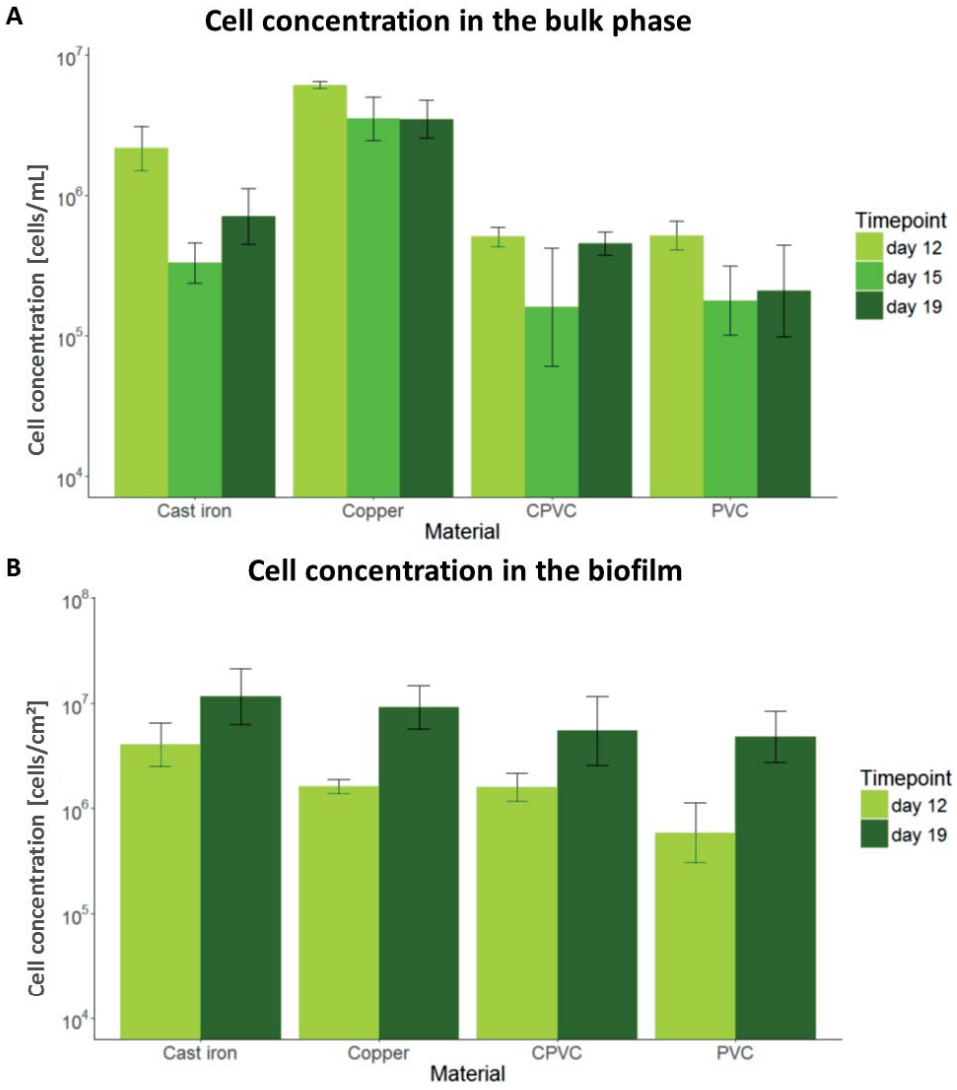


Figure 7 - 10: *Total cell concentration in the bulk phase (**A**) and in the biofilms (**B**) after 12, 15, and 19 days of incubation. the concentrations are visualized on a logarithmic scale with the corresponding error bars (n=4) for all materials.*

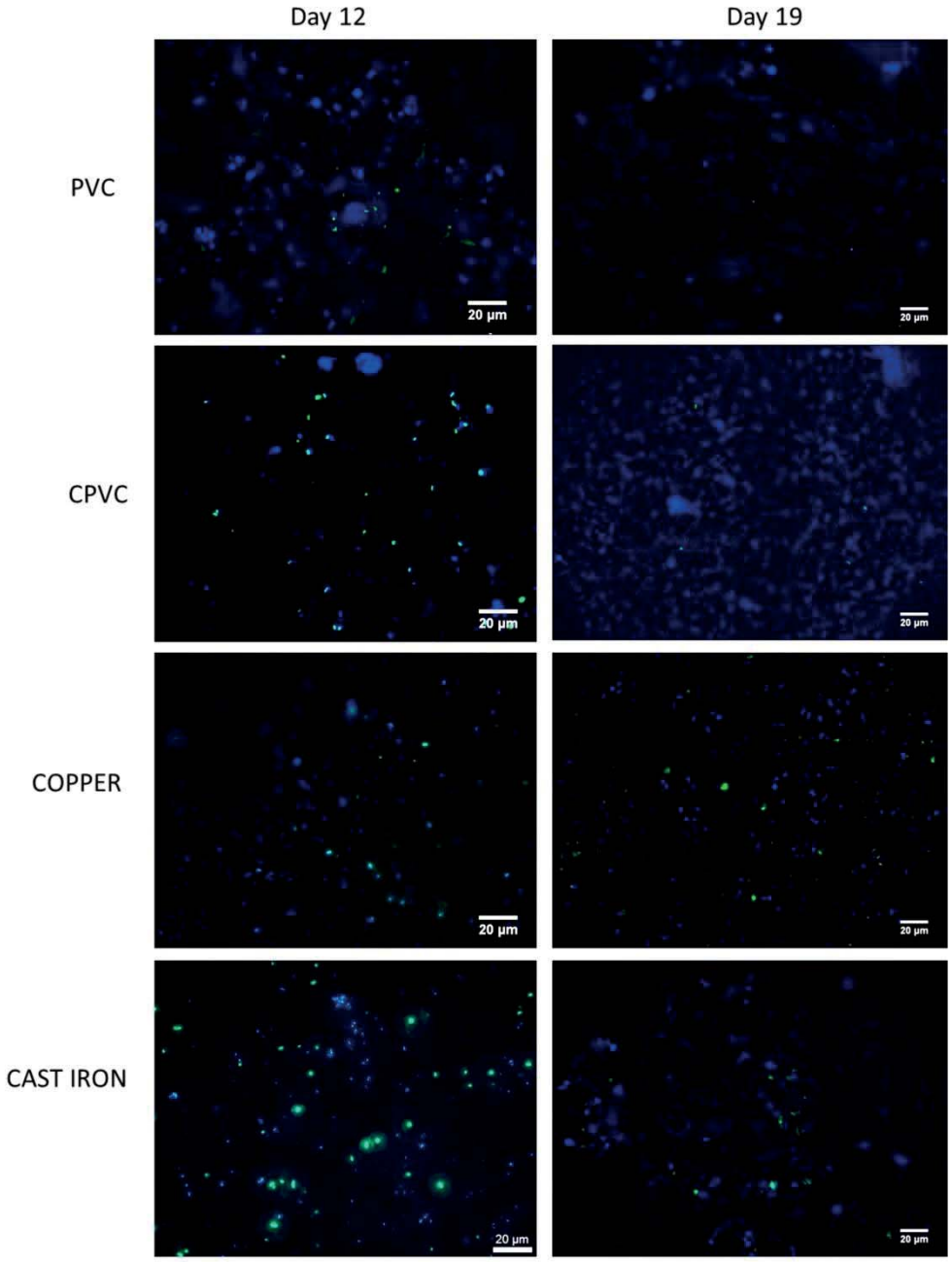


***Figure 7 - 11**: Overlay microscopic pictures of biofilms grown on the four different piping materials. Cells were stained with DAPI and appear blue. The green cells are fluorescent* Enterobacter amnigenus *type II cells. All coupons were cleaned prior this experiment. Water was refreshed twice between day 12 and day 19.*

In a second part, the biofilms were briefly disinfected before repeating the experiment. The *Enterobacter* concentration in the bulk water was higher for PVC, CPVC, and cast iron after 12 days of incubation, though also decreasing in time. Higher total cell densities were found for all biofilms after disinfection, which could be explained by the presence of dead cells on the pre-colonized biofilm. These dead cells could either be incorrectly counted as live cells or they could have been used as carbon source and substrate, resulting in additional growth. Microscopy revealed the presence of *Enterobacter* in all biofilms with a decreasing

concentration over time (**Figure 7 - 12**). The copper and cast iron biofilms showed a visually higher concentration in the same way as without pre-colonization. No *Enterobacter* cells were detected with flow cytometry in the samples where autoclaved tap water was added. This result is in contradiction with previous results. These results suggest that *E. amnigenus* can colonize biofilms in DWDS and that extensive disinfection can promote the invasion of pathogens instead of preventing it. To confirm this, further research is necessary.

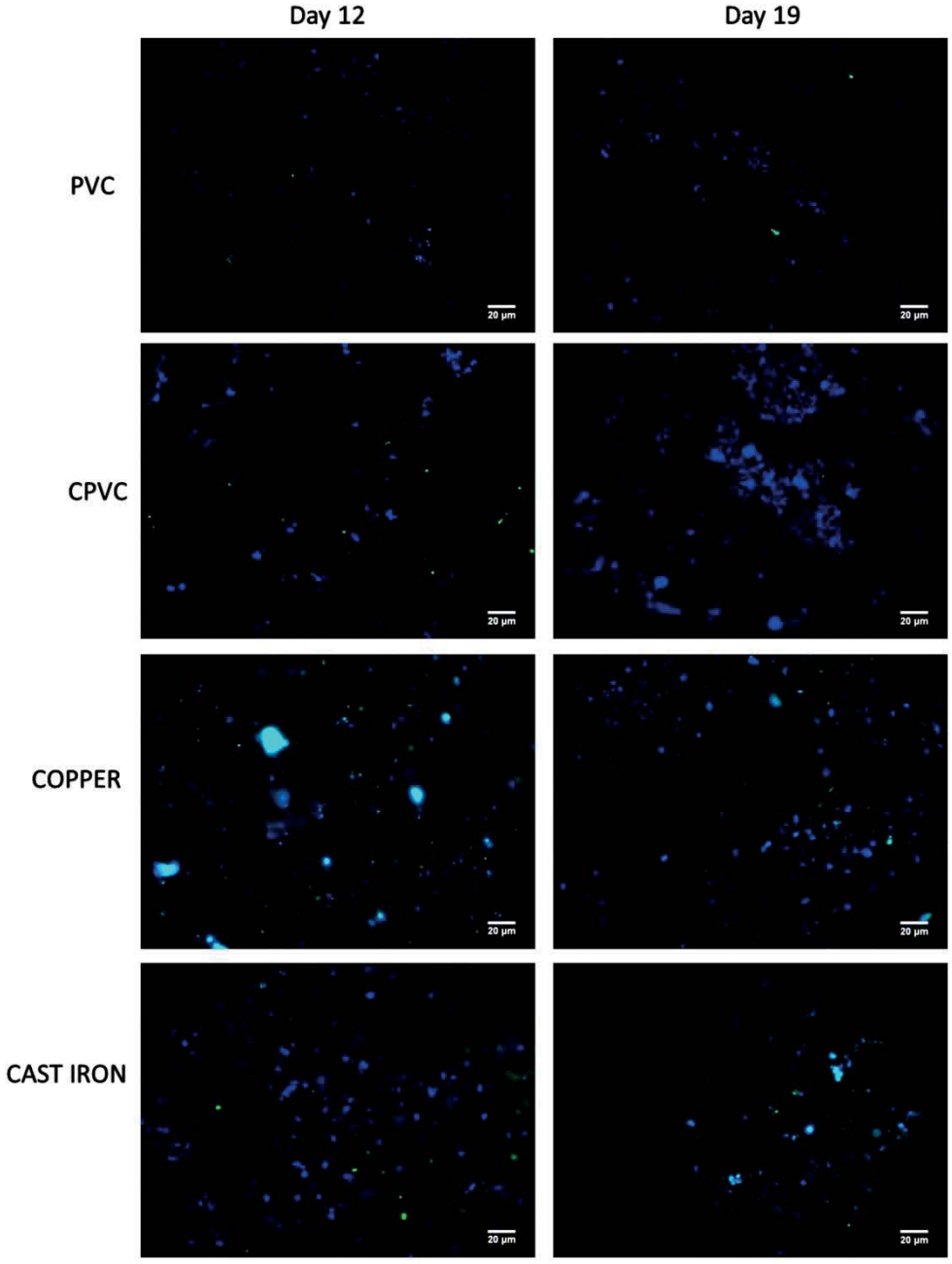


***Figure 7 - 12**: Overlay microscopic pictures of biofilms grown on the four different piping materials. Coupons were pre-colonized for 20 days and then briefly rinsed with 1.5 % NaOCl before this experiment. Cells were stained with DAPI and appear blue. The green cells are fluorescent* Enterobacter amnigenus *type II cells. Water was refreshed twice between day 12 and day 19.*

5 Conclusions and perspectives

In this chapter, the purpose was to determine if flow cytometric fingerprinting could be used to monitor the biofilm and bulk microbial community in drinking water distribution systems. Our findings showed that biofilm growth could be monitored with flow cytometry provided a proper sample preparation by means of sonication. Although cytometric background noise was important for biofilm samples, results were in line with comparable research in literature.

Several factors showed to affect the flow cytometric fingerprints of both the bulk and the biofilm samples. For example, drinking waters produced with different treatment methods or coming from different sources showed to have a different bacterial growth potential for both the biofilm and the bulk water during a period of four weeks. Flow cytometric fingerprinting and cell density measurements showed that, not only the water itself, but also the pipe material influenced the microbial communities in the biofilm and in the bulk water, both in terms of the community composition and cell concentration. We can conclude that, during the limited period of our experiments, plastic material induced more biofilm growth than the metallic materials, though less in the bulk water phase. The results from the flow-through experiment show that the difference in cell concentrations between the different materials changed over time. The phenotypic diversity of both the biofilms and the bulk waters also varied over time but was noticeably higher for the biofilms. This indicates that the microbial communities were continuously adapting to the changing conditions in the drinking water systems. Because of the short duration of our experiments, we cannot extrapolate our results to water distribution systems with a mature biofilm. For this, more research with older biofilms should be done.

Another objective of this study was to establish the relation between biofilm and bulk microbial communities with flow cytometry in order to determine if cytometric fingerprinting of the bulk water could be used to monitor the microbial community in the biofilms. However, our results showed no correlation between the dynamics of the bulk and biofilm fingerprints between four weeks to maximum 68 days. This suggests that the phylogenetic dynamics and composition of the biofilm and the bulk microbial communities are different. This is in line with previous research and could be validated by sequencing the samples of this experiment. The behavior of a typical drinking water contaminant *Enterobacter amnigenus* was evaluated during a batch test. The *Enterobacter* seemed to grow better in biofilms on metallic compared to plastic materials, but the difference was not statistically significant. The influence of disinfection on the *Enterobacter* invasion potential was unclear, and requires further research but suggests that the *Enterobacter* could grow better when more nutrients

are available and when no competition is present. This finding is in accordance with the concept of biologically stable water and illustrates the usefulness of biologically stable water.

6 Acknowledgements

This work was supported by the project grant SB-131370 of the IWT Flanders and by the IMPROVED project, subvented by The interreg V “Vlaanderen-Nederland” program, a program for transregional collaboration with financial support from the European Regional Development Fund. More info : www.grensregio.eu (in Dutch). The authors would like to thank Mike Taghon for his help with the construction of the flow-through experiment. We also would like to thanks Silvia Hidalgo for her help with the microscopy.

CHAPTER

8

ONLINE FLOW CYTOMETRIC MONITORING FOR MICROBIAL QUALITY ASSESSMENT IN A FULL-SCALE WATER TREATMENT PANT

CHAPTER

8

ONLINE FLOW CYTOMETRIC MONITORING FOR MICROBIAL QUALITY ASSESSMENT IN A FULL-SCALE WATER TREATMENT PLANT

1 Abstract

The ever-increasing need for high-quality drinking and process waters and the growing public awareness about possible contaminations drive the efforts to further development of automated control of water treatment plants. For example, membrane filtration processes and reverse osmosis in particular, are generally regarded as a safe barrier for inorganic, organic and microbial contamination. Yet, to ensure the final water quality and to increase the confidence of the end-user, intensive and preferably online monitoring should be further implemented as an early-warning tool to control membrane integrity and to prevent microbial regrowth in the distributing network, especially when non-traditional water resources such as wastewater are (re)used. In previous chapter, flow cytometry has been used for the monitoring of drinking water quality. In this chapter, we tested how a fresh water microbial community responded to the addition of different types of nutrients and demonstrate the applicability of flow cytometry and cytometric fingerprinting for a full-scale water treatment plant.

Chapter redrafted after:

Buysschaert, B., Vermijs, L., Baetens V., Naka, A., Boon, N., De Gusseme, B. Online flow cytometric monitoring for microbial quality assessment in a full-scale water treatment plant. *In preparation*

Buysschaert, B., Props, R., Vermijs, L., Baetens V., Naka, A., Boon, N., De Gusseme, B. Nutrient levels determine microbial community composition and dynamics. *In preparation*

2 Introduction

Bacteria in water are an important aspect of the water quality and may, when present in too high concentrations, lead to biofouling, microbiologically induced corrosion or even the spreading of pathogens. A close and online monitoring of the microbial quality of the water is thus a necessary tool to improve water quality and reduce downstream costs or to mitigate health hazards. One of the main challenges is the unwanted growth of biofilms on surfaces known as biological fouling (Chien *et al.*, 2012b) which occurs mainly in recirculating systems such as cooling towers. Especially open cooling water systems provide a favorable environment for microorganisms because they scrub microorganisms from the air and concentrate the nutrients present in remaining water by evaporation, resulting in faster microbial growth (Liu *et al.*, 2009). Biofilms can damage equipment through microbial induced corrosion (MIC), by clogging, and lead to an increased energy consumption due to decreased heat transfer (Meesters *et al.*, 2003, Bott, 1995). Biofouling and clogging leads to an increased pressure drop in ion exchangers and to increased resistance in membrane filters, which may also cause membrane breakthrough (Mcdonogh *et al.*, 1994). Furthermore, pathogenic bacteria can nestle in these biofilms and contaminate the water through the natural shedding cycle of biofilms (Wingender and Flemming, 2011). A well-known example is the spreading of the pathogen *Legionella pneumophila* in the form of aerosols (Dondero *et al.*, 1980, Keller *et al.*, 1996).

Membrane processes such as microfiltration (MF), ultrafiltration (UF), nanofiltration (NF) and reverse osmosis (RO) are commonly used in water treatment, including reuse applications (Jiang *et al.*, 2017). MF is designed to retain most bacteria and suspended solids in the range of 0.1 to 5 µm. The other methods have pore sizes ranging from maximum 20 nm (UF) to pore sizes smaller than 1 nm (NF and RO) and are therefore supposed to retain all bacteria. However, the passage of microorganisms through filtration membranes has been reported (Ghayeni *et al.*, 1999) and new research has demonstrated the existence of ultra-small bacteria in water (Brown *et al.*, 2015, Luef *et al.*, 2015). Also bacteria larger than the pore size are able to cross the membranes (Ghayeni *et al.*, 1999). Possible explanations are abnormalities in the membrane structure and oversized pores that were considerably larger than the manufacturers stated nominal pore size. Intensive use of membranes might lead to an enlargement of the pore sizes, and incompatible chemical cleaning (e.g. oxidative damage) can cause pore expansion as well (Goosen *et al.*, 2004, Hai *et al.*, 2014). In addition, bacteria themselves can change in size in respond to

changing environmental conditions (Chien *et al.*, 2012a) and may undergo size reduction in nutrient limited environments (Egli, 2010). The wall of most bacterial cells is not a rigid structure, but has a high flexibility and elasticity. This deformability, as well as the size of the bacteria, play an important role in their passage through filtration membranes (Wang *et al.*, 2008). A last possibility is related to the breakthrough of membranes due to mechanical or chemical stress, resulting in the passage of unfiltered water.

Biofilms are almost inevitable in piping systems and consist of complex and functionally organized microbial communities embedded in a matrix of extracellular polymeric substance (EPS) protecting the bacteria from environmental stresses (Flemming and Wingender, 2010). Typically, biocides such as 2,2-dibromo-3-nitrilopropionamide (DBNPA) or hypochlorite are added to the water to remove unwanted biofilms. Biocide dosage is not an ideal solution both because of the environmental impact of the biocides and because it enhances the corrosion of the equipment. Furthermore, a fraction of the bacteria in the biofilm are protected by the EPS layer and benefit from the nutrients from dead biofilm (Meesters *et al.*, 2003). In oligotrophic environments, nutrients availability is a critical cause of microbial growth and nutrient limitation is therefore an alternative strategy to mitigate issues caused by microbial growth (Flemming, 2002). Since regrowth rather than contamination is the main cause of microbiological issues in water treatment and distribution, the quality of the end product must be monitored properly. Different techniques exist that characterize the aquatic bacteria. While the most commonly used method in drinking water is the heterotrophic plate count method (HPC), this method would be unsuitable for industrial applications due to the long incubation times and its labor intensive nature (De Roy *et al.*, 2012). A more convenient technique is the is adenosine tri-phosphate (ATP) analysis which provides an estimation of the active and viable biomass (Hammes *et al.*, 2010). The speed, robustness, easiness, and low cost make it a very appealing technique (Hammes *et al.*, 2008, Velten *et al.*, 2007), yet ATP analysis is less frequently used than would be expected (Hammes *et al.*, 2010). A first issue is that ATP quantification is less precise for the low cell concentrations typically found water (Hammes *et al.*, 2010). Also, ATP measurements do not make the distinction between intra- and extracellular ATP. This can significantly alter the results as, in certain biological matrices, the extracellular ATP concentration can be several orders of magnitude higher than the intracellular ATP concentration (Hammes *et al.*, 2010, Sakakibara *et al.*, 1997). Finally, differences in species, cell sizes, and physiological states can alter the ATP concentration per cell making the conversion from ATP to biomass concentration difficult (Eydal and Pedersen, 2007).

Flow cytometry is a fast and robust method to determine bacteria concentration in liquids. With the use of appropriate dyes, it can determine viability and activity, similarly to ATP analysis. The single-cell resolution and the high throughput of the technique make it robust and by-passes the bias of converting the output to biomass concentration. In this chapter, we characterized the changes in the flow cytometric fingerprints in function of specific nutrients and for different microbial communities. Finally, we applied online flow cytometry to monitor the microbial communities and dynamics in a full-scale water treatment plant. The water was monitored at different stages in the treatment plant.

3 Materials and methods

3.1 Batch experiments

Set-up. Surface water from a river was collected for the experiments. For a first experiment, a dilution series of the water was prepared in 0.22 µm-filter sterilized bottled water (Evian). Immediately thereafter samples were taken for flow cytometry. Subsequently, nutrients were added as yeast extract (500 mg/L yeast extract or 195 mg/L C as final concentration, Oxoid) or as a mixture of acetate, NH_4Cl, and K_2HPO_4 with a controlled molar C:N:P ratio of 20:5:1 (200 mg/L C final concentration). The plates were then incubated at 28°C and after 24 hours, 500 µL was transferred from each sample and prepared for flow cytometric analysis. A blank with no added nutrients was included. All samples were prepared in triplicates.

For a second experiment, undiluted surface water was aliquoted in well-plates and subsequently nutrients or salt were added as NH_4Cl, K_2HPO_4, NaCl, glucose, and yeast extract. An overview of the nutrients and the salt with the final concentration is given in **Table 8 - 1**.

***Table 8 - 1**: The concentrations of all nutrients added to the second batch experiment were chosen sufficiently high to induce an effect.*

Stock solution	**NH_4^+**	**PO_4^-**	**NaCl**	**$C_6H_{12}O_6$**	**Yeast extract**
Concentrations [mg/L]	0.5	0.65	100	2	5
	1	1.3	200	20	50
	1.5	2	500	200	500
	2	2.6	1000		
		3.2			
		3.9			

TOC analysis. TOC analysis was performed using a TOC 5000 analyzer (Shimadzu). Basically, two standard series were made and measured on TC and TIC. Samples were diluted to a range of 1:50 mg/L TC and 1:2 mg/L TIC and analyzed on the TOC analyzer.

3.2 Online measurements

Online FCM measurements were performed at an industrial water treatment plant (**Figure 8 - 1**). At this production site, four different water qualities are produced from brackish surface water (Rodenhuizendok, Port of Gent, Belgium). During a first measuring period of six days, surface water was monitored after pre-filtering through 300 µm strainers. During a second period of seven days water after ultrafiltration (UF) was measured continuously. The UF racks are built each with 40 Microza hollow fiber UNA modules (Pall) with a pore size of 0.1 µm. UF rack 1 and 2 have one year old membrane modules, while UF rack three has membrane modules of more than three years. RO1D produces 90 m³/h single-pass RO permeate, using BW30XFR-400/34i brackish water membranes (Dow) in an two-stage configuration. The system is working at an 75% recovery rate, resulting in a <100 µS/cm demineralized water quality fit-for-use as cooling water. The cooling tower has an 1500 m³ open buffer, susceptible to contamination. After seven to eight times thickening, the cooling water (<800 µS/cm) is continuously discharged at about 7,5 m³/h from the tower and reused again as feed water for the UF in the water treatment plant (cooling water return). An OnCyt© staining robot coupled to an BD Accuri C6 Plus flow cytometer (BD Biosciences), as described by Besmer et al. (2014), was used for continuous FCM measurements. Every 16 minutes a fresh sample was taken and stained.

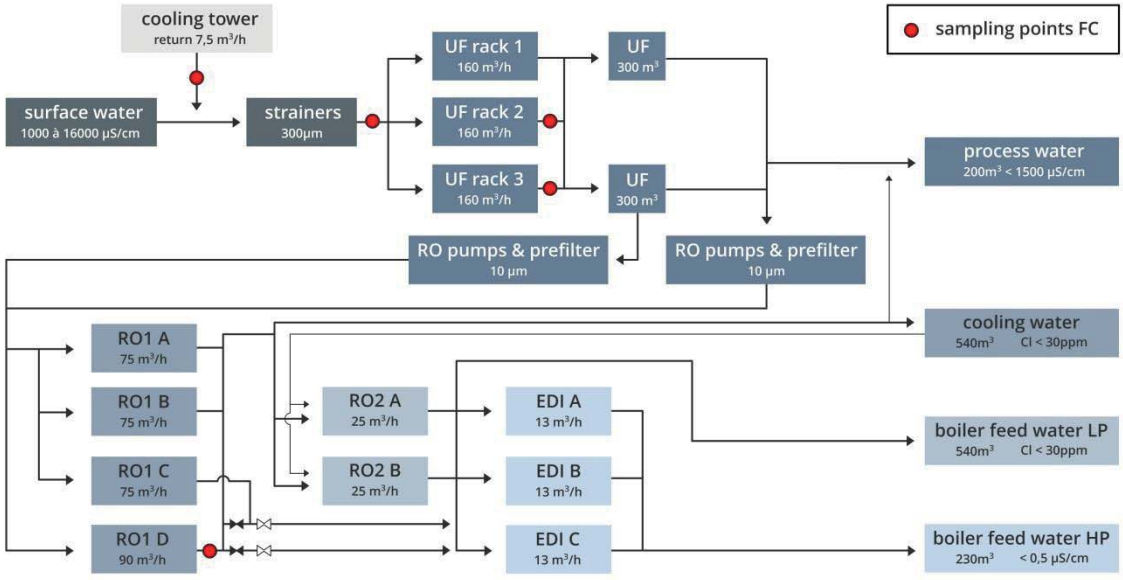


Figure 8 - 1: *Schematic overview of the water treatment plant with the flow cytometry sampling points.*

3.3 Flow cytometry

Staining protocol. An optimized staining protocol was used from Van Nevel *et al.* (2013). Bacteria were stained with 10 µL/mL of SYBR Green I (SG, Invitrogen, 100x diluted in DMSO from stock) for total cell counting. The samples were then incubated for 20 minutes at 37 °C inside the online robot to optimize the staining.

Data analysis. All data was extracted from the proprietary Accuri C6 CSampler software or FACSuite software in the flow cytometry standard (FCS 3.0) format and subsequently imported into R v3.4.0 (R Core Team, 2015) through the functionality offered by the *flowCore* package v1.42.2 (B. Ellis *et al.*). Data was first log transformed and then normalized by dividing all values by the maximum fluorescence intensity signal. No compensation was applied. Gating to reduce the background was performed in R studio using the *flowCore* package on both SG and SGPI stained samples. A 0.22 µm-filtered control was used to determine the position of the background and a heat-killed sample was used to determine the position of the permeabilized cell population (Berney *et al.*, 2007). Based on this, a universal gate was constructed to remove as much background as possible. The data quality was evaluated and improperly acquired data was cleaned using the *flowAI* package v1.4.3 (Monaco *et al.*, 2016) in order to remove anomalies in the data related to changes in flow rate, unstable signal or outliers in the lower limit of the dynamic range. Samples which failed the QC were removed from the dataset. This was done after background was removed with the universal gate. Next, a single-step discretization ('binning') and Gaussian bivariate density estimation was performed on the selected parameters (green and red fluorescence, FSC-H and SSC-H) using the *KernSmooth* package v2.23.15 (Wand, 2015). An equally spaced grid (binning grid) of 128 x 128 was fixed for each bivariate density estimation using the *flowFDA* package v1.0 available at https://github.com/lievenclement/flowFDA. All bivariate density estimations were concatenated to a one-dimensional feature vector which we refer to as the fingerprint. Subsequently, phenotypic alpha diversity was calculated according to the publication of Props *et al.* (2016) where Hill diversity indices are applied to describe the diversity of conceptual phenotypes within and between samples. The code is available at https://github.com/rprops/Phenoflow_package.

4 Results and discussion

Initially, lab experiments were performed to determine the ability of flow cytometry to detect fresh water community dynamics. For this, fresh water microbial communities were spiked with different nutrients and concentrations. Online flow cytometric measurements were then performed in a full-scale water treatment plant. Cell concentrations and Hill number diversity

indices were calculated for raw surface water after filtration with 300 µm strainers, after ultrafiltration, reverse osmosis, and finally also for recirculated wastewater from a cooling tower. The aim of this research was to demonstrate the usefulness of online flow cytometry to monitor and operate a full-scale water treatment plant.

4.1 Reactivity of the aquatic microbial community

Filtration resembles a dilution process in a way that less microorganisms will be present after dilution or filtration. Moreover, small tears in the filtration membranes can actually cause a dilution of the raw water. Therefore, surface water was incrementally diluted up to 1:5000 to approximate a filtration process and the different microbial communities were studied to understand how fast the flow cytometric fingerprints of the aquatic microbial communities reacts to the introduction of nutrients. Nutrients are the most important aspect inducing microbial regrowth and more specifically the presence of the elements C, N, and P are vital for microbial life. For this, a defined and complex medium were added to a freshwater community under the form of a C:N:P mixture with a ratio of 20:5:1 on one hand and yeast extract on the other hand. A comparison of the different diversity indices as described by Props *et al.* (2016) shows that D_0 is the only parameter that decreases with increasing dilution at the start of the experiment (**Figure 8 - 2**).

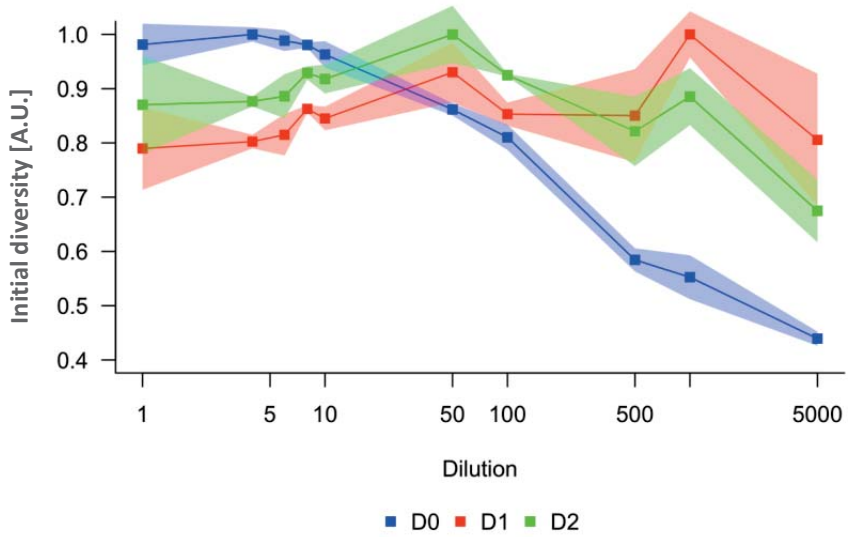


***Figure 8 - 2**: Alpha diversity parameters in function of initial dilution. The bands represent standard deviations, calculated from biological triplicates. Arbitrary units [A.U.] were rescaled to a [0-1] scale.*

As D_0 represents the number of positive bins in the flow cytometry bi-plots, it is logical that it is most sensitive to the decreased cell density. D_1 and D_2 on the other hand also take the density into account and give more weight to the most abundant phenotypes. The fact that

they are not influenced by the dilution illustrates that throughout the dilution series, the most abundant microbial phenotypes are still present. For the first 60 hours of incubation, samples were taken regularly to monitor the microbial growth and results showed that the effect of the nutrient addition was most visible after 20-30 hours and that it was the result of microbial growth. However, this incubation time can differ across microbial communities and we found that the aquatic microbiota from the same river reacted differently across seasons (data not shown).

When looking at the similarity between the cytometric fingerprints after 24 hours of incubation, the difference between the different sample types is clear (**Figure 8 - 3**). The further away the samples are from the control samples, the bigger the differences. This shows that yeast extract resulted in the biggest change of cytometric fingerprint which was also confirmed by ANOSIM (**Appendix Table 8 - 1**). Samples incubated with a defined nutrient source changed less than the samples incubated with a complex nutrient source. The controls changed after 24 hours of incubation.

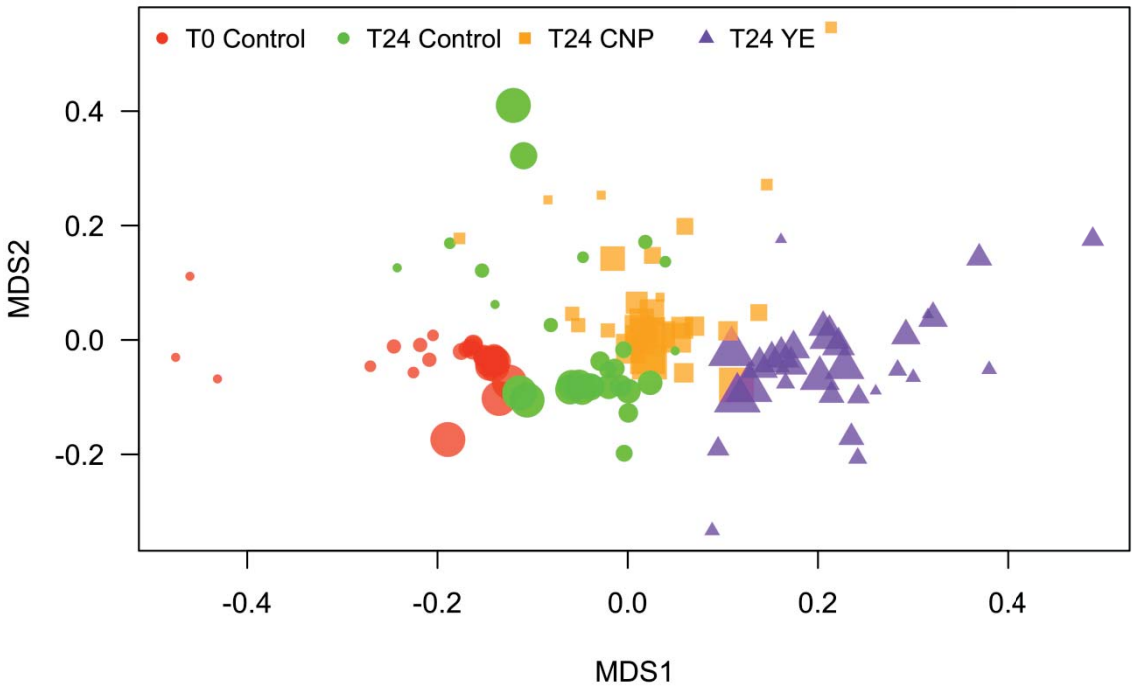


Figure 8 - 3: *Non-metric multidimensional scaling between samples on different time points and with different perturbations. "T24 YE" represents the yeast extract perturbation after 24h of incubation, "T24 CNP" the controlled perturbation with C:N:P ratio of 20:5:1 after 24h of incubation, "T24 Control" the control without perturbation after 24h of incubation and "T0 Control" the initial community before incubation. The smaller the shape size, the more diluted the initial sample.*

The changes in of the cytometric fingerprints can be explained by the microbial growth (**Figure 8 - 4**). The cell density of all control samples increased after 24 hours of incubation

due to growth on the nutrients already present in the water. However, when additional nutrients were added to the water, growth was six to ten times higher than without nutrient addition for the defined and complex nutrient source respectively. The discrepancy between both nutrient sources is likely due to the lack of certain elements or compounds (e.g. sulphur, magnesium) and the type of carbon-, nitrogen-, or phosphorus sources which do not promote optimal growth for all species present.

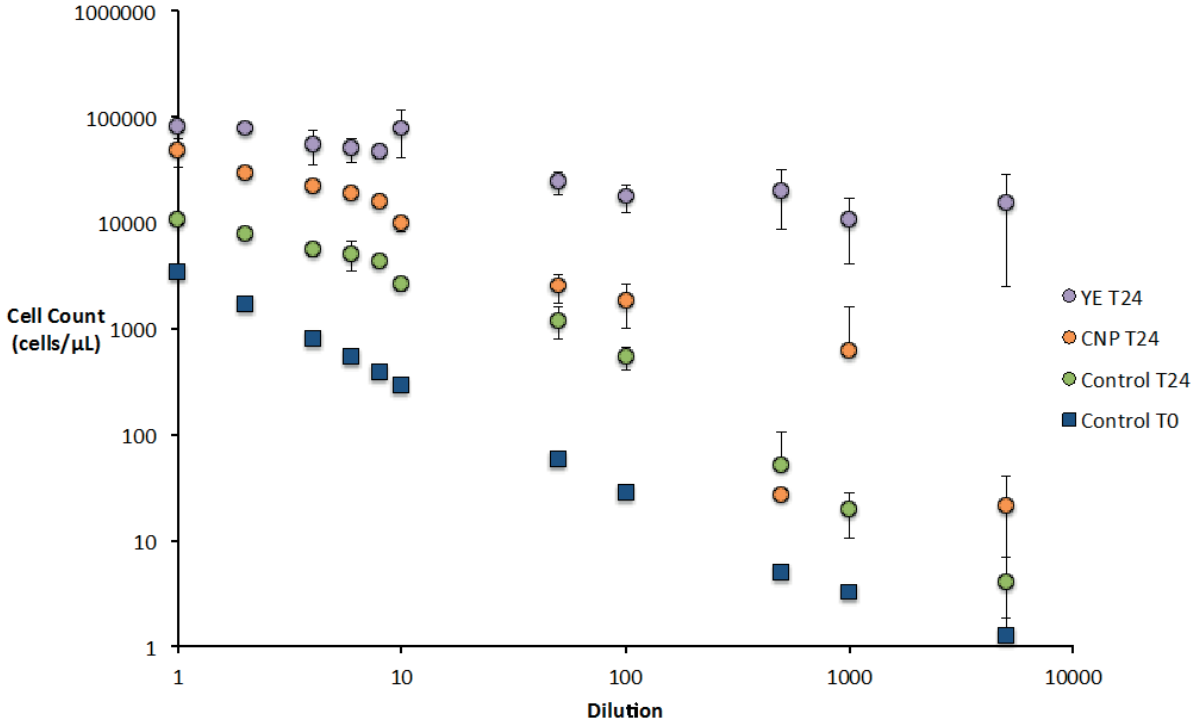


Figure 8 - 4: *Cell concentration in function of the dilution for different perturbations. "YE T24" represents the yeast extract perturbation after 24h of incubation, "CNP T24" the controlled perturbation with C:N:P ratio of 20:5:1 after 24h of incubation, "Control T24" the control without perturbation after 24h of incubation and "Control T0" the initial community before incubation.*

To look further into the effect of the dilution in relation to diversity, the cell concentration of all samples can be compared to the richness D_0 which was most sensitive to cell concentrations (**Figure 8 - 5**). The results show that, per condition, D_0 differs in function of the dilution. This is not always reflected in cell densities as can be seen for the samples incubated with yeast extract. After diluting the initial microbial community, different microbial communities were created by diluting out the least abundant species. Though the most diluted samples showed strong regrowth, this did not result in an important increase of phenotypes as no new species were introduced into the system. This illustrates that both the cell density and the diversity parameters are necessary to understand the dynamics of the microbial community.

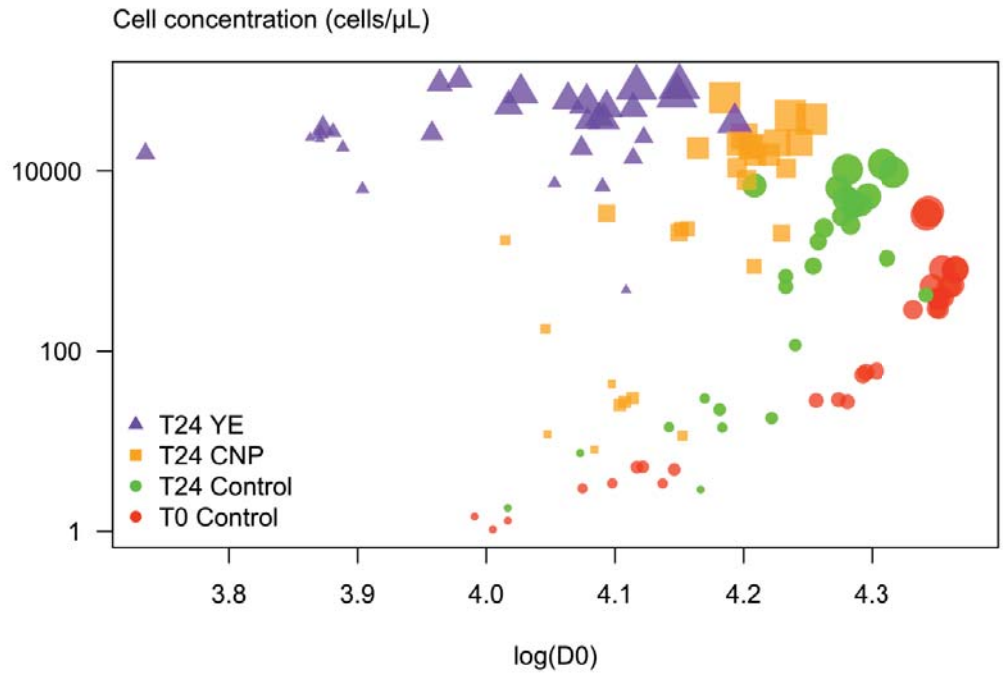


***Figure 8 - 5**: Cell concentration of all samples (in cells/µL) in function of the logarithm of the richness of each sample. "T24 YE" represents the yeast extract perturbation after 24h of incubation, "T24 CNP" the controlled perturbation with C:N:P ratio of 20:5:1 after 24h of incubation, "T24 Control" the control without perturbation after 24h of incubation and "T0 Control" the initial community before incubation. The smaller the picture size, the more diluted the initial sample.*

In a second experiment, the different compounds of the carbon source were tested individually and also different concentrations of the carbon sources were added to raw surface water to estimate how sensitive the microbial community was to a perturbation. The compounds added to the water were NH_4^+, PO_4^{3-}, glucose and yeast extract. Besides nutrients, also the influence of NaCl was assessed as it is a common indicator of RO break-through. The diversity indices D_0, D_1 and D_2, expressed relatively to the control (Δ diversity), for the samples where yeast extract or glucose was added, differed from the diversity of the control samples from 20-30 hours on (**Figure 8 - 6**). For all other nutrients, there was no visible change in diversity indices relative to the control samples (**Appendix Figure 8 - 1**). Also no important differences could be noticed between the two types of carbon sources though the highest concentration of yeast extract yielded a much stronger fluctuation in all indices than the other concentrations. In the case of glucose addition, the indices changed proportionally to the concentration. Both yeast extract and glucose contained approximately the same concentration of organic carbon. The lowest concentration of glucose or yeast extract increased the TOC concentration with 2 mg/L in water which already contained 13.09 ± 6.82 mg/L TOC. This shows that even small concentrations (± 15%) of TOC can lead to microbial growth in a period of 24 hours and that flow cytometry can be used to detect

microbial regrowth after minor organic contamination of a water sample which makes it applicable for water treatment monitoring as well.

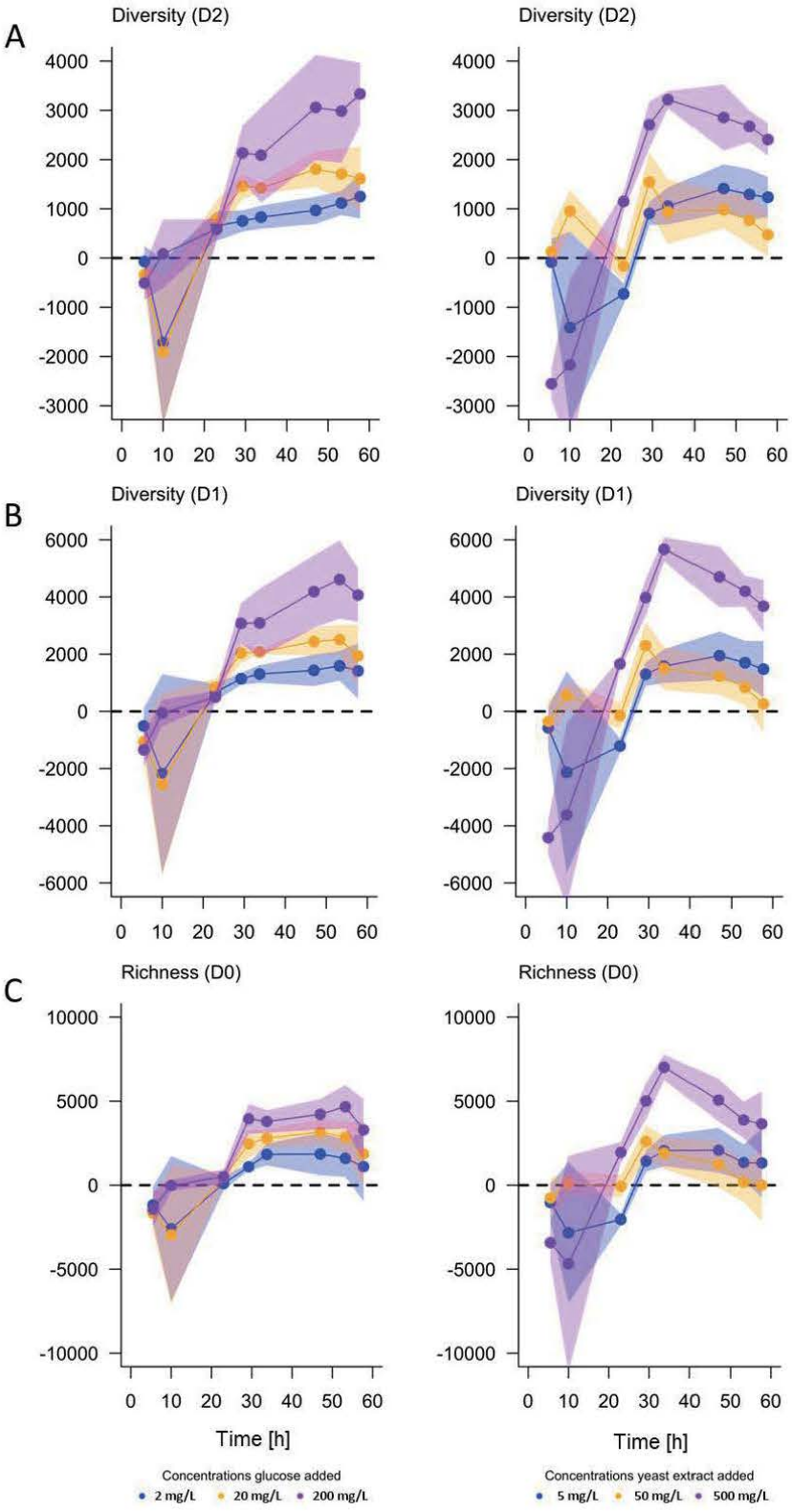
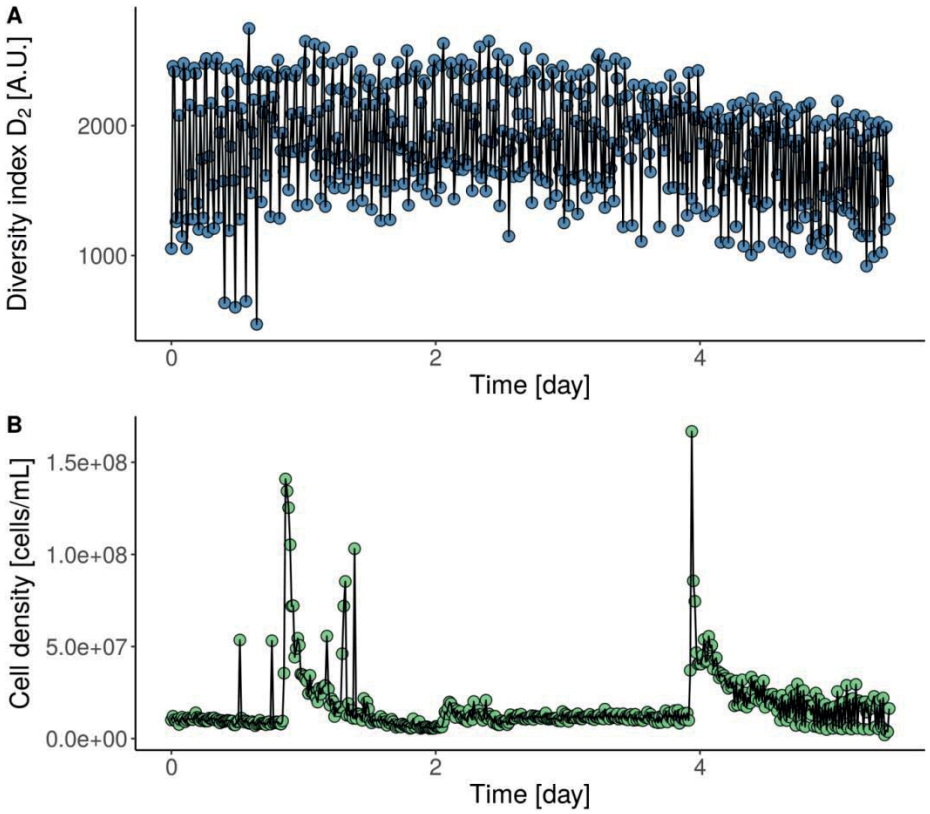


***Figure 8 - 6**: Diversity for both glucose (left) and yeast extract (right) in function of time for three different concentrations. The three different indices were calculated $\triangle D_0$ (**A**), $\triangle D_1$ (**B**) and $\triangle D_2$ (**C**). The colored bands represent the standard deviations (n=3). The dashed line represents the case in which no change relative to the control sample would be observed.*

4.2 Online monitoring in a full-scale water treatment plant

In a full-scale water treatment plant, the incoming surface water was monitored after passing through 300 µm strainers and prior to the ultrafiltration step (**Figure 8 - 1**). Cell concentrations fluctuated between $5.0*10^6$ and $1.7*10^8$ cells/mL (**Figure 8 - 7b**). After the first and the fourth day of measurement, two tailing peaks in cell density were observed. As the water is pumped from a dock, boats near the treatment plant's water inlet roiling the water could cause the sudden changes in cell density though this could not be confirmed by the turbidity measurements alone. However, the cytometric diversity showed not to be influenced by the increased cell concentration (**Figure 8 - 7a,c and d**). On the other hand, neither the peaks in the cell concentration measured by flow cytometry nor the peaks in feed water turbidity seem to result in an increased conductivity (**Figure 8 - 7c**). The conductivity measurements showed a large variability at the end of the experiments which is also reflected in the flow cytometry data. No clear explanation could be found for this observation. These results show that turbidity measurements and cytometric Hill diversity indices are not always correlated to cell density, and that conductivity and turbidity alone are not sufficient as parameters for monitoring the incoming water quality of a treatment plant. Other parameters influencing the microbial (re)growth, such as organic carbon (TOC and DOC), should be monitored as well.



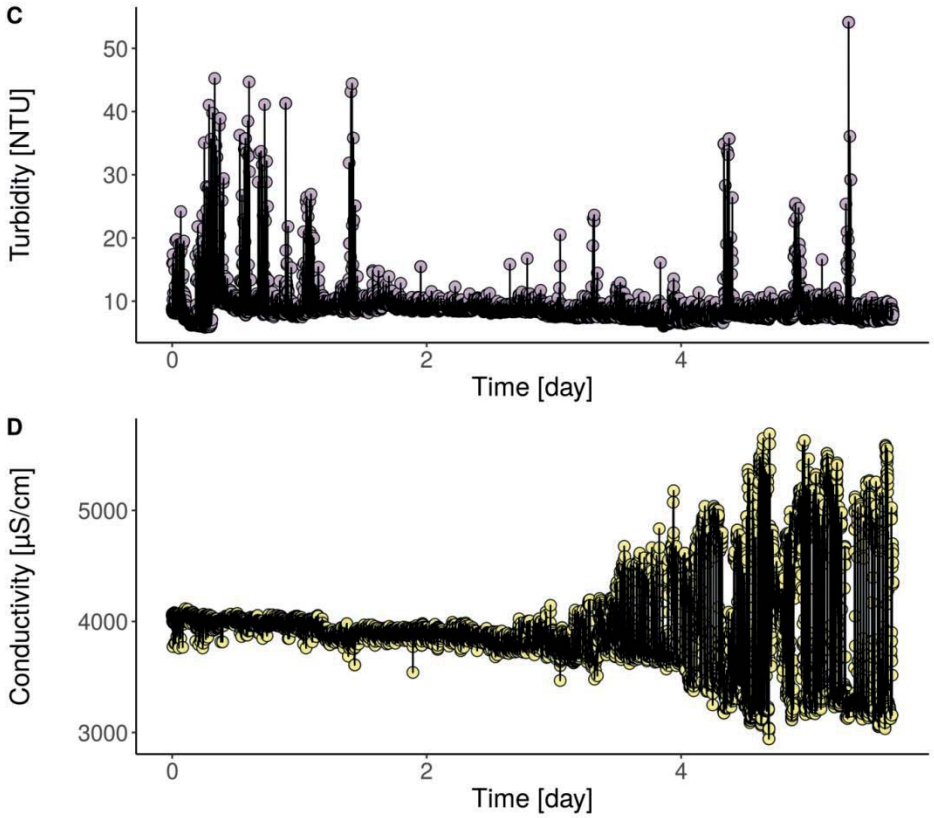


Figure 8 - 7: *Flow cytometric Hill number diversity D_2 (**A**) for surface water in function of time. Resulting cell concentration expressed in cells/mL (**B**), the turbidity (**C**) and conductivity (**D**). The surface water was strained with 300 µm sieves prior sampling. The sampling, staining and incubation was fully automated.*

In a second stage, the filtrate after UF was monitored for two different racks. The cell concentration after UF was approximately $3*10^4$ cells/mL for rack 2 and $2.5*10^4$ cells/mL for rack 3. Though the pore size of the racks was 100 nm, not all bacteria in the water could be removed. As mentioned above, there are several possible explanations for the occurrence of bacteria after filtration. Aside from an impaired membrane integrity, also the presence of small bacteria in the surface water could cause the higher concentrations of bacteria. Luef *et al.* (2015) showed that ultra-small bacteria or ultramicrobiota can be found in ground water, Ghai *et al.* (2013) showed the same for sea water, and Wang *et al.* (2007) demonstrated their presence in surface water. The range and dispersion of these small bacteria is still unclear but Ghai *et al.* (2013) estimated their relative abundance to be ~4% while Wang *et al.* (2007) estimated their relative abundance to be 0.2% of the aquatic bacteria. Assuming that all bacteria after UF are ultra-small bacteria, their relative abundance would be ~1% for this surface water entering the treatment plant, which is in accordance with results published previously. However, to proof this, confirmation with molecular techniques and microscopy is necessary. Furthermore, a periodic pattern in the cell concentration was observed on both racks (**Figure 8 - 8b,e**). For rack 2, a decrease in the concentrations was observed while for

rack 3, an increase in the cell concentration was observed. These peaks, for both racks, indicate a prolonged chemical cleaning of the membranes by recirculation of a cleaning solution. During the cleaning, the permeability increase (**Figure 8 - 8c,f**). The cleaning procedure typically lasts one hour and is done with a concentrated hypochlorite solution. Hypochlorite, an oxidizing agent, bleaches fluorochromes such as SG, making the bacteria undetectable with flow cytometry which explains the results for UF rack 2. Despite the cleaning, filtration units get dirtier over time and an accumulation of organic dirt, inorganic precipitates, and biofilms reduces the overall efficiency of the filtration unit and increases the need for maintenance as for rack 3. The latter was chemically cleaned every seven hours while UF rack 2 was only cleaned once daily. It is reflected in the lower permeability of UF rack 3 as well, an average 70 lmh/bar compared to 100 lmh/bar for UF rack 2. Because of the high degree of fouling on UF rack 3, it seems likely that, during the chemical cleaning, the accumulated organics react with the hypochlorite and that SG is subsequently not bleached. As dirt and bacteria detach from the filters, an increase in the cell concentration is visible in the recirculating cleaning solution. Aside from chemical cleaning, also backwashes are programmed regularly as can be seen by the frequent drops and surges in permeability (every 30 minutes for both racks). Yet, no effect of the backwashes was observed in microbiological parameters. Similarly to the surface water, the cytometric Hill number diversity did not reflect the fluctuations in the cell concentration and the permeability though after the last cleaning step of UF rack 1, a consistent decrease was noticed in the diversity suggesting a change in the community composition, which was not reflected in the cell number (**Figure 8 - 8a,d**). This change could however not be linked to any straightforward operational parameter.

Similarly, the cell concentration and the Hill number diversity indices were monitored over time for recirculated condensed cooling water and for the RO permeate. A concentration of approximately 10^7 cells/mL was measured for the recirculation water, which is likely caused by microbial regrowth after recirculation in the open cooling tower. In the RO permeate, a concentration of approximately 10^4 cells/mL was measured which is slightly lower, though comparable to the cell concentrations found after UF. In contrast to UF, it is unlikely that bacteria would pass through the membranes with pores smaller than 1 nm unless membranes are impaired. Dewettinck *et al.* (2001) and Kumar *et al.* (2007) reviewed the importance of integrity monitoring for RO membranes and reported that challenge tests resulted in a log removal value (LRV) for bacteria between 2.9 and 5.4. Considering that the bacteria concentration in the RO concentrate can mount up to 10^6 CFU/mL (Ridgway *et al.*, 1983), we estimate the bacteria concentration to be 10^8 cells/mL, assuming only 1% of the bacteria can be cultured (Ridgway *et al.*, 1983, Hammes *et al.*, 2008). The cell concentration

we measured after RO would result in a LRV of four which is in accordance to what was reported before (Kumar *et al.*, 2007). An alternative hypothesis for the presence of bacteria after RO filtration could be related to the regrowth of bacteria in the distribution system and the shedding of bacteria from biofilms formed on the permeate side of the filter units. Tang (2011) reported the presence of biofilms at the permeate side of RO membranes in dairy industry but, to our knowledge, no reports of biofilms on the permeate side of RO membranes have been published in the field of water treatment. To measure a concentration of 10^4 cells/mL in the bulk water, an estimated biofilm density of 10^7 cells/cm² is required considering that the RO module comprised 126 filter units, and assuming a biofilm growth rate of 0.03 day^{-1} (Boe-Hansen *et al.*, 2002). This biofilm density appears possible when comparing it to the biofilm density in drinking water distribution systems (Prest *et al.*, 2016). Finally, a possible contamination of the sampling line and sampling port cannot be excluded but a decrease in the cell concentration would then be expected in function of time. This was not noticed during the measurement period. A control test with the sample lines also gave no clear indication of contamination. For both RO and the recirculation water, no patterns were found in the cell concentrations. Furthermore, also no patterns could be discerned from the Hill number Diversity indices and no relationship could be established between the microbiological parameters and operational parameters (data not shown).

For all stages in the water treatment plant flow cytometry showed that bacteria were present in concentrations far above the theoretically expected level. Though membrane processes commonly used for water production and reuse applications cannot be expected to produce pure and sterile water as small particles and molecules can pass through the membranes in several ways as discussed above. Routine membrane integrity tests are thus necessary to ensure the proper functioning of the membranes. Direct integrity tests such as the pressure decay test, which measures the rate of decline of pressure across a membrane require the membrane unit to be offline. Therefore, indirect methods are chosen for continuous online monitoring during production. Turbidity is an example of such an indirect test but we show here that it is not sufficient to monitor the microbial quality of the water alone. Based on our findings, we argue that cell concentrations, measured with flow cytometry, give additional information about the water quality which other methods cannot provide. In this respect, flow cytometry could be used as an indirect membrane integrity test, and can serve as a tool to understand and control bacterial regrowth in the distribution network in order to ensure the final water quality for the end-user. Our results also showed that cytometric fingerprints, in the form of Hill number diversity indices, did not show important fluctuations. This is because of the short timespan at which the tests have been performed and because the water treatment operated properly during the whole experimental period. More variation should be

expected over longer periods of time (*i.e.* months) as research has shown that fresh water communities exhibit seasonal community composition fluctuations (Pinto *et al.*, 2014). In the lab experiments described before, we demonstrate with dilution that the diversity indices change when the microbial community changes which was also proven with by comparing the cytometric fingerprints with sequencing data of cooling water (Props *et al.*, 2016).

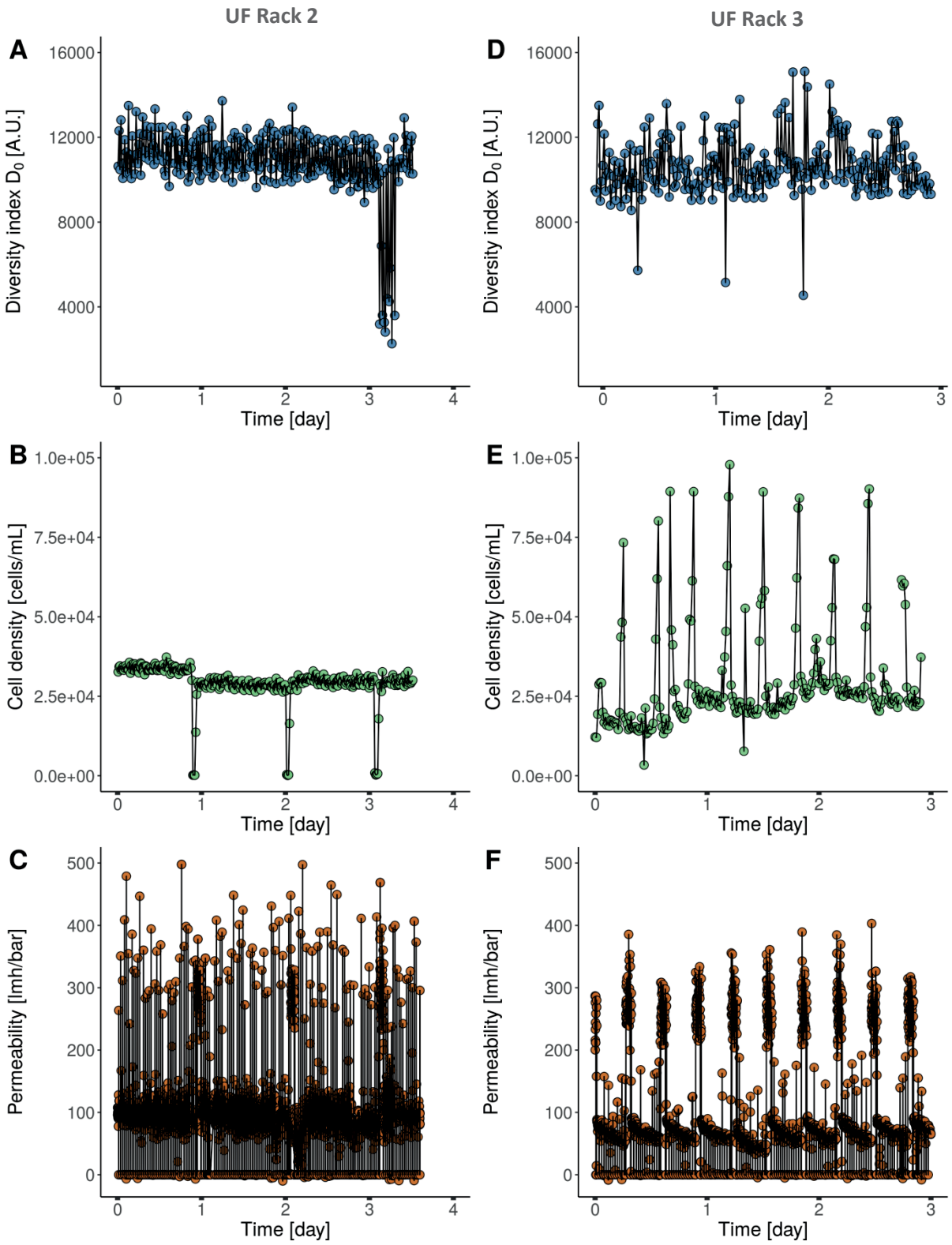


***Figure 8 - 8**: Flow cytometric Hill number diversity D_2 (**A, D**) of the UF filtrate in function of time. Resulting cell concentration expressed in cells/mL (**B, E**), and the permeability (**C, F**). The surface water was strained with 300 µm sieves prior sampling. Results are shown for UF rack 2 (**left**) and rack 3 (**right**). The sampling, staining and incubation were fully automated.*

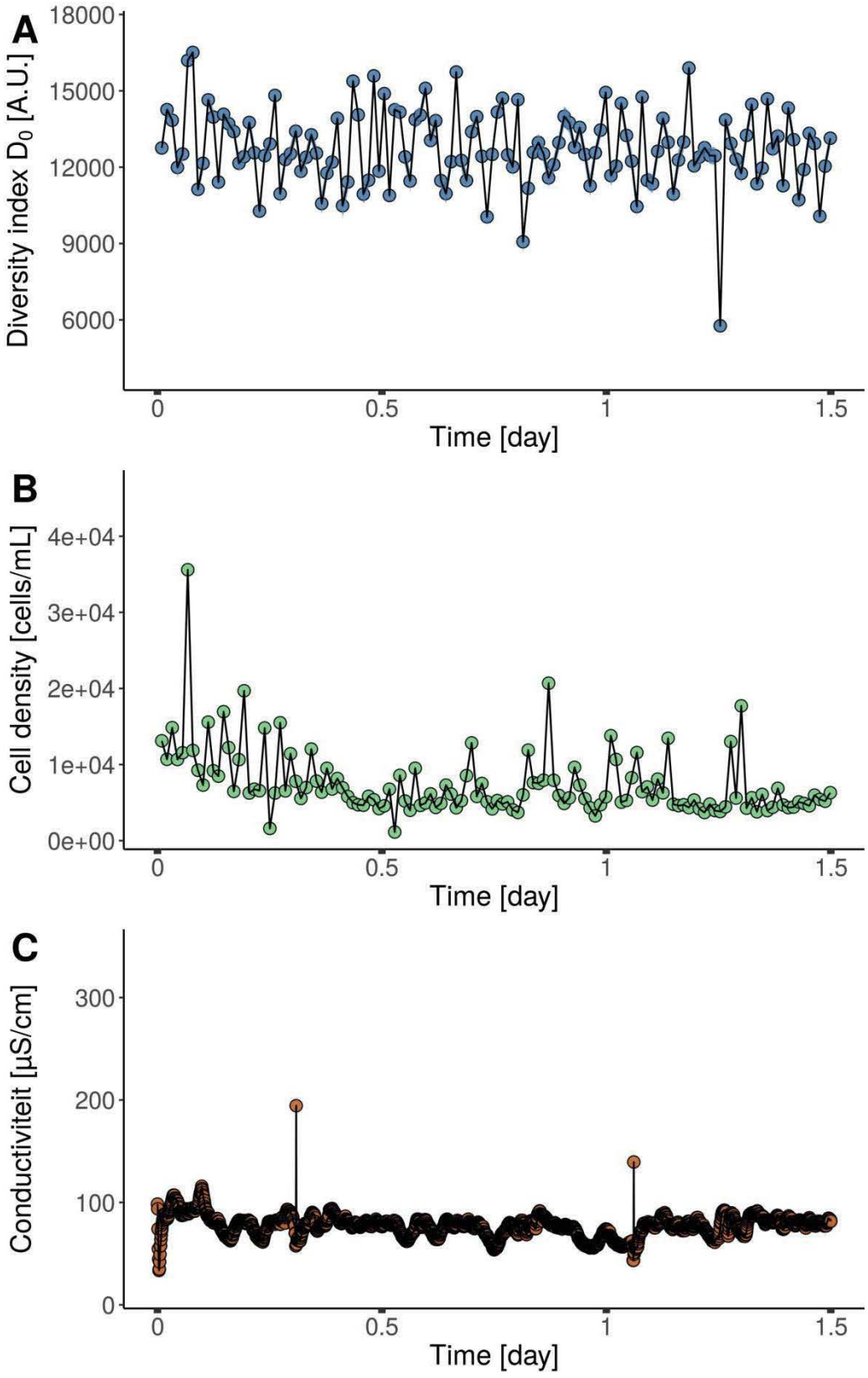


***Figure 8 - 9**: Flow cytometric Hill number diversity D_2 (**A**) of the RO permeate in function of time. Resulting cell concentration expressed in cells/mL (**B**), and the permeate conductivity (**C**).*

5 Conclusions and perspectives

In this chapter, we showed that flow cytometry is a good method to monitor fresh water communities online and automatically. We demonstrated in a full-scale water treatment plant that bacteria are present after every filtration step and that bacteria concentrations in the water are directly related to process operations. For example, we showed that prolonged and chemical membrane cleaning can lead to an increase of bacteria concentrations in the recirculation cleaning solution, while backwashing did not affect the cell concentration. As bacterial concentrations gave a better insight in the microbiological dynamics than e.g. turbidity measurements alone, we suggest that online flow cytometry could be used to indirectly monitor the membrane integrity and the changed microbial quality in the final water. With lab experiments we demonstrated that flow cytometry is able to detect both quantitative and qualitative changes in the fresh water communities. We also show that these changes are different depending on the type of microbial community and nutrient added.

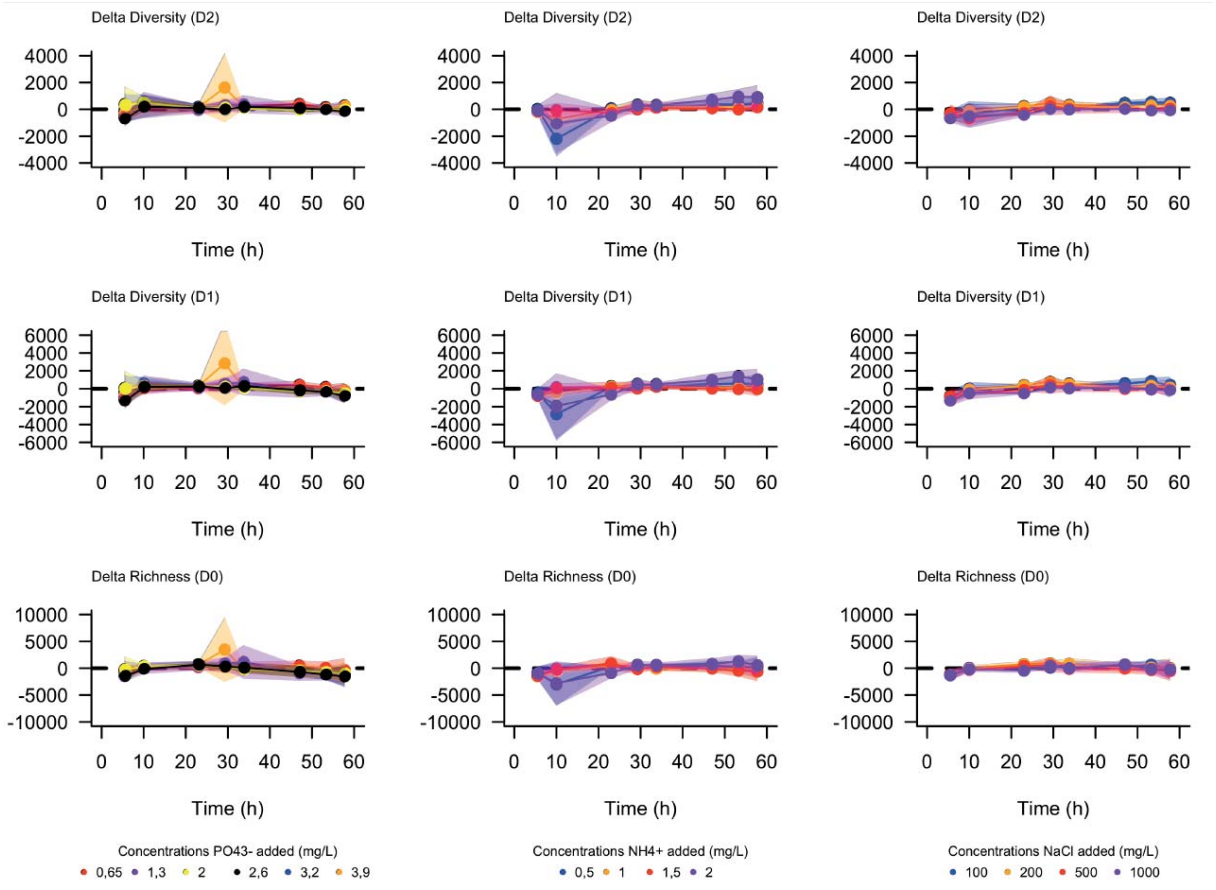
6 Acknowledgements

This work was supported by the project grant SB-131370 of the IWT Flanders and by the IMPROVED project, subvented by The interreg V “Vlaanderen-Nederland” program, a program for transregional collaboration with financial support from the European Regional Development Fund. More info : www.grensregio.eu (in Dutch). We would like to thank INDUSS for providing access to the water treatment plant.

7 Appendix– Supplementary information for chapter 8

Appendix Table 8 - 1: *Separation metric R between the control samples at the beginning of the experiment (T_0) and after 24h (T_{24}) between the control samples and samples perturbed with a controlled C:N:P ratio of 20:5:1 or a yeast extract perturbation.*

ANOSIM	T_0 Control	T_{24} Control
T_{24} Control	R = 0.3766 *P* < 0.001	NA
T_{24} C:N:P	R = 0.6005 *p* < 0.001	R = 0.2361 *p* < 0.001
T_{24} Yeast Extract	R = 0.8995 *p* < 0.001	R = 0.5902 *p* < 0.001



Appendix Figure 8 - 1*: Diversity (D_0, D_1, D_2) for three perturbations (**left:** PO_4^{3-}; **middle:** NH_4^+; **right:** NaCl) in function of time for different concentrations. The colored bands represent the standard deviation, calculated form biological triplicates. The striped horizontal line represents the case in which no change relative to the control sample would be observed.*

CHAPTER 9

GENERAL DISCUSSION

CHAPTER 9
GENERAL DISCUSSION

1 Positioning of the research

Microbial ecology is more relevant than ever before as, besides earth's biogeochemical, also many industrial applications are facilitated or affected by microorganisms. These applications can be based on complex microbial communities such as wastewater treatment, bioremediation, bioleaching or bioaugmentation (Bertrand *et al.*, 2015). Also more simple ecosystems such as pure culture, co-culture or multispecies fermentation play a crucial role in the biotechnological production of fine chemicals, food and pharmaceutical products. In the future, the importance of bioprocesses is likely to increase due to the technological, economical, and ecological benefits of microbiologically-driven techniques (Soetaert and Vandamme, 2010). On top of that, the sensitive and reactive nature of microbial communities makes them good indicators of how processes operate. For almost a century, water quality is monitored based on the absence or presence of certain indicator organisms.

To improve these bacteria-driven processes by making them more resistant or resilient, or to improve the monitoring of the communities to predict and avoid failures, the characterization of the microbial communities is vital. Verstraete *et al.* (2007) proposed the concept of microbial resource management (MRM) for this purpose. To put this concept into practice on an industrial scale, a method to quickly characterize microbial communities, regardless of their taxonomic or phenotypic complexity, is necessary. The advent of molecular tools has significantly improved our understanding of how microbial communities work but they remain complex, time-consuming, and expensive (Singer *et al.*, 2017). Flow cytometry is a fast and cheap method that can analyze 10 000 cells in a few seconds without requiring complex sample preparation and can be used as alternative to rapidly characterize microbial communities. In this doctoral research, we further developed the cytometric fingerprinting pipeline as described by De Roy *et al.* (2014a) and Van Nevel (2014), and tested the sensitivity at which flow cytometry could characterize microbial communities. Furthermore, we applied our fingerprinting technique for the monitoring of the water quality in a drinking water network (Van Nevel *et al.*, 2016a, Van Nevel *et al.*, 2016b) and in a full-scale industrial water production plant. On the other hand, we also applied our flow cytometry pipeline for the

monitoring of an *E. coli* batch fermentation. In parallel, a Raman spectroscopy pipeline was developed for community characterization because it is non-destructive and non-invasive method, capable of providing complementary information to flow cytometry.

2 Main research outcomes

Microbial ecosystem all constitute of one or more species. Because pure cultures can be considered as the most simplified form of a microbial community, flow cytometry and Raman spectroscopy for community characterization were first tested and developed based on pure cultures and cocultures. In a second part, the applicability of flow cytometric fingerprinting for the monitoring of aquatic microbial communities as tool for water quality assurance was evaluated with lab experiments and finally, also on a full-scale water treatment plant.

2.1 PART 1: Pure and (co)cultures

Because of their small size, bacteria cannot be detected by flow cytometers unless they are stained with a fluorochrome. Many dyes have been developed through the years, but few staining protocols are standardized and few truly multicolor protocols have been developed with the available dyes. In **Chapter 2,** several multicolor protocols were developed and compared. This was done with both a Gram positive and a Gram negative bacteria species to make the protocol universal. Not all multicolor dyes were suitable for high throughput and automated analysis as some dyes showed poor stability over time. The most stable dyes were SG and SGPI and were therefore used for all other experiments in this work.

In **Chapter 3**, the flow cytometric fingerprinting toolbox was improved with a similarity based tool to compare the fingerprints. This fingerprinting technique was applied on the cytometric data of 29 taxonomically related *Lactobacillus* strains and species to demonstrate that the strains could be discriminated with high accuracy. Both a mixture of beads and a mixture of two bacteria populations show that changes between 1% and 5% of the events are detected by the algorithm. By repeating the experiment three times, we showed that the method is also reproducible when a standardized growth protocol is used. To illustrate the impact of the microbial growth phase on the fingerprints, a batch culture was monitored at different time points and results showed that the fingerprinting algorithm could make a distinction between the growth phases.

In **Chapter 4**, a classic *E. coli* batch fermentation was monitored with flow cytometry to find out why the cytometric fingerprints change in function of their growth stage. The comparison between the dynamics in the fingerprints and the dynamics of reactor parameters, suggests

that the cytometric fingerprints change due to a phenotypic change caused by substrate depletion. Furthermore, the phenotypic switch detected by flow cytometry occurred before the change in respiration rate, which is considered to be the fastest method to detect the metabolic activity of a fermentation culture. This demonstrates that flow cytometric fingerprinting is also able to detect changes in phenotypes and that it could be applied for the monitoring and operational control of bioreactors.

Fluorochromes are necessary to make bacteria detectable for flow cytometers. This additional step increases the analysis time significantly and reduces the physiological information acquired to those features highlighted by the fluorochromes. In **Chapter 5**, Raman spectroscopy, a label-free alternative to flow cytometry, was tested for the detection of phenotypic diversity. A comparison between several data-analysis pipelines showed that for both explorative and predictive experiments, Raman spectroscopy could successfully be used to discriminate among phenotypes. Finally, Raman spectroscopy was also successfully used to estimate the relative abundance of a mixture of two microbial strains.

In order to determine what the impact is of phenotypic diversity in microbial communities, fast single-cell techniques capable of revealing this lowest level of diversity are required. In **Chapter 6**, results show that both flow cytometry and Raman spectroscopy are suitable methods to detect phenotypic changes despite their different working principle. For this, microcosms were developed which allowed individual cell populations to interact while remaining physically separated. Both methods suggest that each species adapted to the presence of the other and that the level of phenotypic plasticity was different for both species.

2.2 PART 2: Environmental microbiology

Water distribution systems are fundamental for drinking water or process water distribution. In both cases, the water quality is important and microbial regrowth in the distribution system should be limited as much as possible. The biofilms on the pipe walls are an important fraction of the bacteria in the water but are hard or impossible to monitor. In **Chapter 7**, the possibility to monitor the microbial community in the biofilms by fingerprinting the planktonic microbial community in the bulk water was investigated. For this, a series of batch tests and a lab-scale flow-through experiment were set up. Also, different piping materials and water from different water treatment plants were used. The microbial community in bulk water and the biofilm showed different dynamics, suggesting that both phases evolve in a different way and have a different community composition. Furthermore, the pipe materials and the type of water impacted the cytometric fingerprint of the communities, also suggesting a different

community composition. Finally, we tested how a drinking water indicator organism, *Enterobacter amnigenus,* can colonize biofilms and eventually lead to a persistent contamination of the drinking water. Again the type of pipe material played an important role but more importantly, when sterile water was added, the bacteria grew much more and could form biofilms more easily. This supports the concept of biostable water, where the purpose is to mitigate microbial growth by reducing the available carbon sources and to reduce disinfection as the native microbial community acts as a buffer against invasive species (Prest *et al.*, 2016).

In **Chapter 8**, the sensitivity of flow cytometric fingerprinting towards the addition of elements which could be used as nutrients by the microbial community in the water was tested and results showed that the fingerprint changed when growth was induced by a carbon source. Other nutrient sources showed no effect. A dilution experiment also demonstrated that the type of microbial community has an effect on the reactivity of the fingerprints towards the addition of a carbon source. Finally, an online flow cytometer was installed in a full-scale water treatment plant to demonstrate the practical application of flow cytometry and cytometric fingerprinting at the different stages of the water treatment. Results showed that after every filtration stage, regardless of the type of membrane, bacteria could be detected and that the cell concentration could be related to the age of the filters and the operation of the treatment plant. This suggests that flow cytometry could be used as an online tool to monitor the performance of the filtration procedures and eventually also as an early-warning system.

3 Phenotypic diversity as extra dimension of community structure

The term phenotype refers to the observable characteristics of an individual resulting from the interaction of its genotype with the environment and was coined for the first time by Johannsen (1911) based on his observations of plants. Since then, phenotypic heterogeneity has been observed in all taxonomic domains. Genetically identical bacteria populations are known to exhibit a range of phenotypic differences, even under controlled laboratory conditions (Ceuppens *et al.*, 2013, Elowitz *et al.*, 2002). This diversity can greatly benefit the microbial population as a whole, especially if this population is confronted with sudden environmental changes. Subpopulations, resistant to these changes, can ensure the survival of the entire population. This survival strategy is referred to as bet-hedging and is independent of environmental cues. Alternatively, phenotypic heterogeneity can be driven by environmental changes encountered by a subset of the microbial community which adapts accordingly (Davis and Isberg, 2016). The adaptability of the population is referred to as phenotypic plasticity. Based on the definition of phenotypes, also the genotypes can be

considered as phenotypes and the phenotypic diversity could be viewed as a deeper level of community organization. Phenotypic heterogeneity is a potentially important component of biological diversity; it arises at the level of individual microbial cells and provides groups of microorganisms with added functionality (Ackermann, 2015). For pure cultures, the impact of phenotypic diversity has been shown before (Delvigne and Goffin, 2014, Muller *et al.*, 2010), but little research is done on the impact of phenotypic diversity on microbial communities. Questions such as whether this diversity is negligible in comparison to taxonomic diversity, and how the properties of microbial communities are shaped by the fact that microbes act as individuals should be answered. Given the high taxonomic diversity in natural samples, it seems evident that phenotypic diversity is a negligible level of diversity. But, natural environments are inherently heterogeneous; they are characterized by microenvironments and fluctuating conditions both in space and time. This already results in a certain level of phenotypic heterogeneity as result of the phenotypic plasticity of the bacteria. Moreover, also phenotypic heterogeneity caused by bet-hedging is expected. This additional level of diversity also increases the fitness of a microbial community to survive fluctuating conditions. In contrast to genotypic diversity, phenotypic diversity endures as long as the genotype is present (Ackermann, 2015). Another advantage is the possibility of an 'unfair' division of labor. Labor division and microbial collaboration can sometimes result in two populations of which one does not benefit, which reduces the growth and reproduction of that population (Ackermann *et al.*, 2008, Mrak *et al.*, 2007, Nikel *et al.*, 2014). In case this population is another genotype, it cannot persist. But in case this population is a subpopulation of distinct phenotype, the subpopulation can be replenished continuously as long as the genotype persists.

The complex origin of phenotypic diversity makes it difficult to hypothesize about the relationship between genotypic and phenotypic diversity in a natural ecosystem (**Figure 9 - 1**). As phenotypic diversity is the consequence of the genotypic make-up of a cell, at least as many phenotypes are expected. As bet-hedging is an inherent trait of a population, its contribution to the overall phenotypic diversity in a community could be considered as constant. Another factor influencing the diversity is the number of available niches in the environment. These niches are inherent to the environment (e.g. microenvironments) and limit the diversity as two phenotypes cannot coexist if they occupy exactly the same niche based on the competitive exclusion rule. But, an overlapping niche can also result in two niches based on resource partitioning (Hardin, 1960) and the number of niches is therefore expected to increase with increased diversity though at a decreasing rate. The total number of phenotypes is then always equal to, or lower than the available niches, regardless of the ecosystem complexity. When considering the phenotypic diversity at the level of a single

population within a microbial community, the phenotypic diversity is expected to decrease with an increased genotypic diversity. In case of highly diverse ecosystems, the contribution of phenotypic diversity due to bet-hedging is supposed to be unaltered, while the contribution of the phenotypic plasticity to the population diversity would be low or negligible. For communities with low genotypic diversity, the individual phenotypic diversity might increase as microbial interactions could create new niches for which new phenotypes can adapt. In **Chapter 6**, we also showed that the individual phenotypic diversity of two bacteria taxa decreased when the bacteria were grown as cocultures.

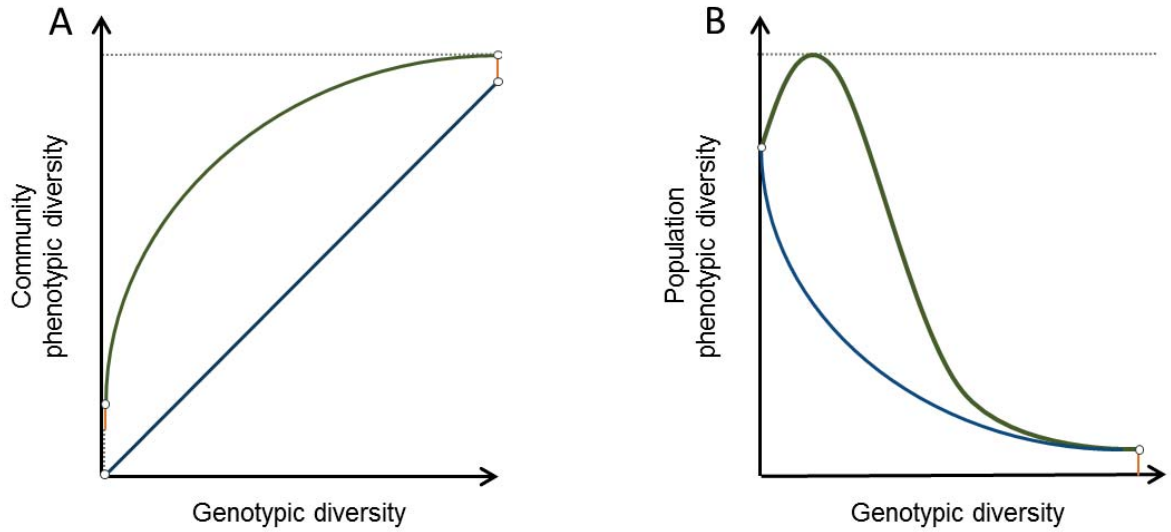


Figure 9 - 1: *The hypothesized relation between genotypic and phenotypic diversity on the community level (**A**) and on population level (**B**). **A**: The minimal phenotypic diversity is always equal to the genotypic diversity as the former is a consequence of the latter (blue line). The expected phenotypic diversity would be larger than the genotypic diversity due to both bet-hedging and phenotypic plasticity (green line). **B**: The phenotypic diversity of an individual population in a microbial community will decrease with increased genotypic diversity. This can either be a continuous decrease (blue line) or a show a local increase when the genotypic diversity results in the creation of additional niches due to microbial interactions (green line). In both cases, the minimal phenotypic diversity will be equal to the bet-hedging capacities of that population (orange line).*

Recent advances in single-cell technologies have allowed to assess the heterogeneity between microbial cells. Populations can, for example, be divided in subpopulations depending on their nucleic acid content, based on their Raman profiles or based on the expression level of certain target genes (**Chapter 6**). According to the method, different observable characteristics are accounted for and the phenotypic diversity is defined in a different way, which impedes direct comparison between methods and quantification of phenotypic diversity. On top of that, the underlying mechanisms regulating phenotypic heterogeneity will always result in a continuous spectrum of phenotypes which calls for a more practical definition of a phenotype. Similarly, microbial species are defined as bacteria

of which the 16S rRNA genes show a certain, and arbitrary level of similarity. This species definition is still debated but its value for research is inestimable (Doolittle and Papke, 2006). For this purpose, a holistic and single-cell technique is necessary such as Raman spectroscopy. However, the method shows too many sources of variability (**see section 5**) and more research should be done to resolve this issue. Alternatively, another technique could be proposed for this. Nevertheless, as shown in **Chapter 4,5, and 6**, both flow cytometry and Raman spectroscopy can be used to monitor changes in phenotypic heterogeneity and dynamics, which could help to answer the questions what the relation is between the genotypic and phenotypic diversity in communities, and which could also help to understand what the link is between community functionality and phenotypic diversity.

4 Flow cytometry

4.1 Financial aspects today and in the future

16S rRNA gene sequencing is considered as the golden standard technique for microbial community characterization because it allows to identify bacteria making up the microbial community and because it is a very sensitive and accurate technique. The technological advances have made sequencing equipment affordable and many research centers are able to perform sequencing, increasing the methods popularity. The possibility to outsource this task, could open the possibility of the wider application of sequencing for industry. Today, this is not the case because of the cost and the complexity of the method. In comparison, flow cytometry provides information of a different nature and at a lower resolution but it is cheaper and the implementation of an automated and online device is merely a technical challenge. To compare and estimate the applicability of both methods for industrial applications, the balance between price, information, time consumption, complexity, and accuracy should be made (**Table 9 - 1**).

***Table 9 - 1**: Comparison of next generation sequencing (Illumina) and flow cytometry in terms of price, time consumption, complexity, and accuracy.*

Technique	Price per sample	Time per sample	Sample preparation (96 samples)	Staff costs	Complexity	Accuracy
NGS (Illumina)	€ 65	> 35 hours	8 hours	€ 400	+++	+++
FCM	€ 0.11	< 25 min	1 hour	€ 50	+	++

Sequencing is, regardless of technological advances, still a complex technique. Though the actual analysis could be outsourced, a DNA extraction, amplification, and purification would be necessary. These steps are complex and time-consuming and require specialist skills to

perform. Furthermore, these preparation steps as well as the actual outsourcing would take some time, and it is not uncommon to wait several weeks after sampling to obtain results. We estimate that for 96 samples, eight working hours are required for the sample preparation. The overall process would require more time as also the incubation time, for the polymerase chain reaction (PCR) for example, should be taken into account. Considering an average labor cost of €50/h, the estimated cost is €400 for one batch. The cost of the actual sequencing can vary, depending on the number of samples but we estimate that €60 per sample is a realistic price. The combined cost of the DNA extraction and amplification is €5 per sample. Flow cytometry also requires some sample preparation and when done manually, one hour is sufficient for 96 samples. The cost of the consumables and reagents is about €0.11 per sample. For both methods the data analysis is difficult but, in both cases, an automated processing pipeline could be developed, facilitating the workflow. Based on the costs per sample, flow cytometry is by far the cheapest technique and it can already be automated (Besmer *et al.*, 2014, Brognaux *et al.*, 2013). Though sequencing is more accurate, the major constraint is the sample preparation, which is labor intensive and thus costly. Also, much time is lost when outsourcing the sequencing. For sequencing to be usable as monitoring technique, a simplified sample preparation in combination with an easy-to-operate device is necessary. The development of next-generation chip sequencing technology (Ion-torrent, minION) presents an opportunity for very cheap sequencers which could be used with minimal skills, although DNA still needs to be extracted, amplified, and purified to sequence the correct DNA fragments (Quail *et al.*, 2012, Ashton *et al.*, 2015, Rothberg *et al.*, 2011). Also flow cytometers could be developed further as early-warning systems; and even hand-held prototypes have been made though the resolution for microbiological applications is still insufficient (Im *et al.*, 2015, Zhu and Ozcan, 2015, Koydemir *et al.*, 2015).

4.2 What flow cytometry can and cannot reveal

Flow cytometry is an optical technique which relies, in this work, on nucleic acid dyes (*i.e.* SG and PI). Staining of bacteria is a complex interaction between the dye, the bacteria, and the matrix in which the bacteria are suspended. The information within the cytometric fingerprint is dependent on this interaction and provides more information about both the physiology and the taxonomy.

4.2.1 Taxonomy

The kingdom of bacteria is very diverse and bacteria do not only differ in their shape and size, but also in the size of their genome and their metabolic potential. In **Chapter 3** we showed that flow cytometric fingerprinting could discriminate among microbial species and strains. While flow cytometers cannot perceive the similarity between 16S rRNA gene sequences used for taxonomic classification, the fingerprints were different in their fluorescence intensity because the species selected for the experiment differed in their DNA content. Based on these results, we can state that flow cytometry can perceive taxonomic differences, provided the dye(s) used to stain the bacteria can visualize these differences. Rubbens *et al.* (2017) demonstrated that models, trained with *in silico* mixtures of bacteria pairs, could estimate the relative abundance of each species of an *in vitro* mixture. The precision differed for each pair and, similarly to our findings, these differences can be attributed to the differences in genome size. For example, a pair of *Pseudomonas fluorescens* and *Pseudomonas putida* cannot be resolved successfully as their genome size is very similar in size (approximately six million bp), while a pair of *Micrococcus luteus* and *Shewanella oneidensis* can be resolved due to the important difference in genome size (2.5 million bp and 5 million bp respectively).

For mixed microbial communities, where multiple species are present, the dynamics of the flow cytometric fingerprints can be used to approximate the dynamics in the community composition. Props *et al.* (2016) compared the diversity dynamics of sequencing data and flow cytometry fingerprints and showed a good relation between both (**Figure 9 - 2**). Similarly to the results presented in **Chapter 3** and by Rubbens *et al.* (2017), the relation between both can be explained by the fact that different species can have a different amount of nucleic acids, which is translated by subpopulations with different fluorescence intensities. A change in community composition will then result in changes of these subpopulations. However, flow cytometry and sequencing are fundamentally different and a perfect correlation cannot be expected.

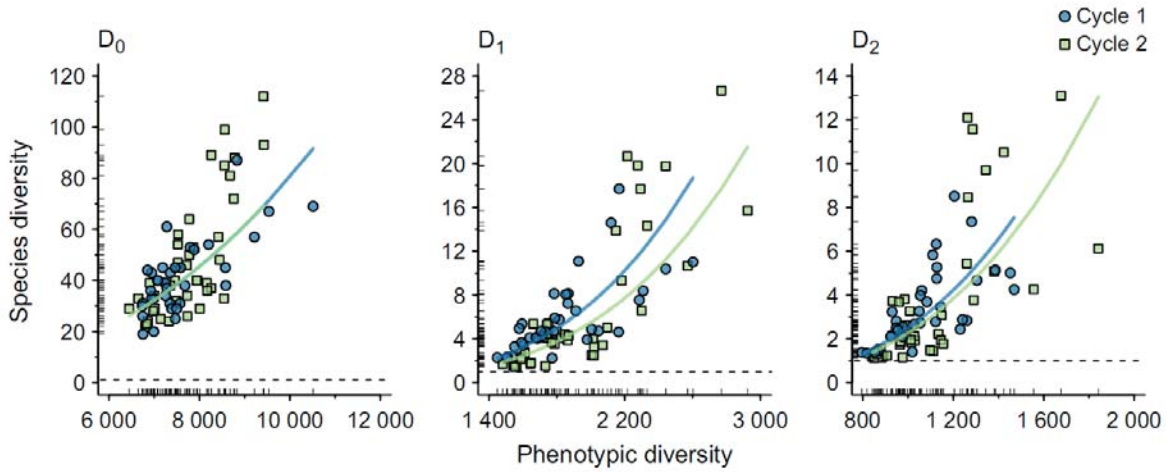


***Figure 9 - 2**: Relation between species diversity and phenotypic diversity for the three Hill number diversity indices (**D_0**, **D_1** and **D_2**). During the experiment, water of a cooling water tower was analyzed during two operation cycles. Species diversity was modeled by a generalized linear mixed model. The phenotypic diversity calculated based on the green and red fluorescence parameters and the scatter signals in the same way as all previous chapters. A significant correlation was found for all diversity indices (Props et al., 2016).*

On the same samples published by Props *et al.* (2016), we calculated the correlation between the diversity indices of sequencing, denaturing gradient gel electrophoresis (DGGE), and flow cytometry. Based on the second order Hill number diversity (D_2), no significant correlation between DGGE and sequencing could be established, while a good and significant correlation was established between flow cytometry and Illumina sequencing (Spearman's $\rho = 0.769$, p-value < 0.0001) (**Figure 9 - 3**). Based on these results we argue that flow cytometry is a better alternative than DGGE for community fingerprinting.

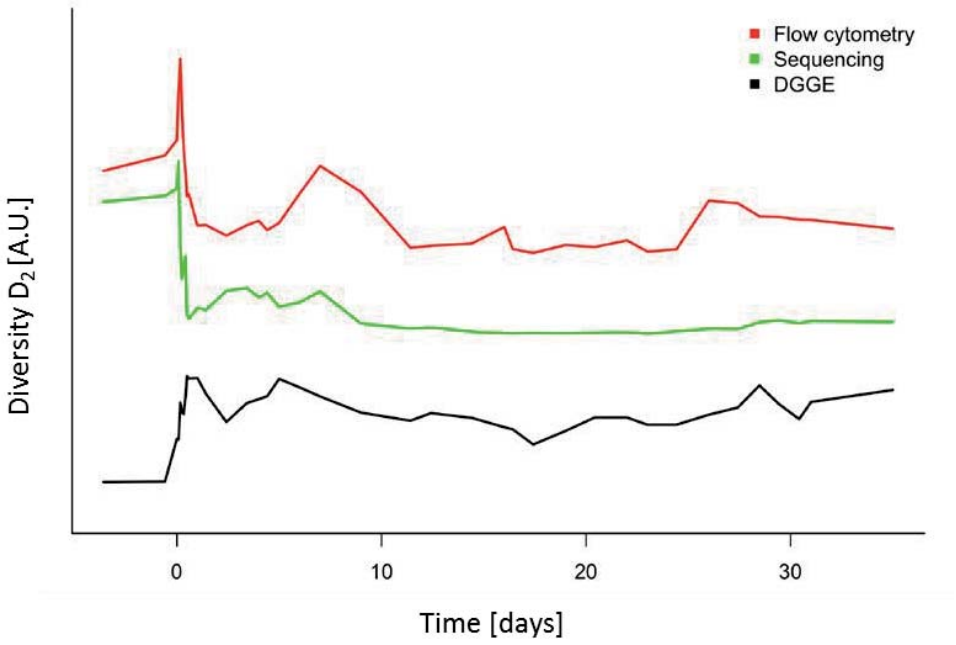


Figure 9 - 3: *The Hill number diversity index D_2 calculated for flow cytometry, sequencing, and DGGE on the samples of one cooling tower cycle as published by Props et al. (2016). Flow cytometry and sequencing showed a good Spearman's correlation (Spearman's ρ = 0.769,* p-*value < 0.0001) while sequencing and DGGE showed no significant correlation (Spearman's ρ = 0.182,* p-*value = 0.2725).*

This relationship between cytometric fingerprints and community composition was confirmed independently by Koch *et al.* (2013c), who compared the changes of the flow cytometric fingerprints in function of the community composition determined by tRFLP. Because Koch *et al.* (2013c) used a different instrument, fingerprinting approach, and nucleic acid dye, it can be stated that the correlation is robust. To improve the correlation between both methods, the dye characteristics and the sensitivity of the detectors are important. A promising combination of dyes would be acridine orange (AO) and 7-aminoactinomycin D (7-AAD) for example. Acridine orange is a nucleic acid dye with a strong affinity for AT-rich sequences and is green fluorescent when bound to DNA, while 7-AAD has a strong affinity for the GC-rich sequences and is red fluorescent when bound to DNA. As a result, the cell populations would be partitioned based on both the nucleic acid content and relative GC/AT ratio.

4.2.2 Physiology

The relationship between the flow cytometric fingerprints and the cell physiology is well-established and evident. A plethora of fluorescent dyes have been developed in the last decades targeting different aspects of the cell physiology. In **Chapter 2**, we attempted to optimize and develop multicolor protocols to increase the amount of physiological information of the fingerprints. The most important issue with dyes is their complex chemistry, resulting in

small changes of the cytometric fingerprints which are not important for the classical gate-based processing of the data, but which impact the fingerprinting approach considerably. As we demonstrated in **Chapter 3**, a change of 1.2% of the events is already sufficient for the fingerprints to be considered different. A better characterization and the development of more stable dyes could improve the resolution of the cytometric fingerprinting. Other important aspects of the dyes are the quantum yield, the sensitivity, and the spectrum of the fluorochromes. Fluorochromes with a high quantum yield produce brighter fluorescent signals, and are therefore better for the reduction of background fluorescence. Also the sensitivity of the fluorochrome is important because brightly fluorescent signals are not sufficient for the accurate detection of cellular features. SG for example, is a very bright and sensitive fluorochrome and both characteristics explain the success of SG for flow cytometric fingerprinting. Another important issue about multicolor protocols is the overlap in emission spectra, which severely restricts the number of possible combinations. Much research has been done on Qdots with very narrow emission spectra to make multicolor flow cytometry possible. But, Qdots have to be coupled to antibodies for targeted staining which makes staining protocols more costly and complex. Because staining is one of the most important sources of variability, and because staining and the subsequent incubation step are an important time loss for monitoring (**Chapter 4**), a label-free approach is favored. As most bacteria are not autofluorescent, flow cytometry should rely on the scattered light which we found to be insufficient for accurate fingerprinting in all the experiments. We also found that the scatter detectors were not the most important contributors to the fingerprints but that this depended on the instrument used (data not shown). Good quality scatter detectors could therefore make a difference.

4.3 Caveats

In order to perform fingerprinting in a reproducible way, several aspects should be taken into account which, when unaccounted for, may lead to incorrect results or decrease the quality of the results. A first important aspect which may influence the results, is the number of cells on which the fingerprints are based. While for some applications the relative concentration can be considered as a source of variability inherent to the set-up and meaningful for comparison, care should be taken when directly comparing the cytometric fingerprints. Props *et al.* (in preparation) demonstrated on lake water samples that the number of cells per sample can have an influence on the Hill number diversity indices and their standard deviation. A sample size of 10 000 cells is therefore recommended as a minimum, although for the higher order diversity indices (D_1 and D_2) 1000 cells could suffice (**Figure 9 - 4**). The practical implications of this effect are important for set-ups where samples with low cell

concentrations are measured such as clean water in a water treatment plant (**Chapter 8**). Experiments in which we measured the regrowth potential of milli-Q water illustrate this effect (**Figure 9 - 5**).

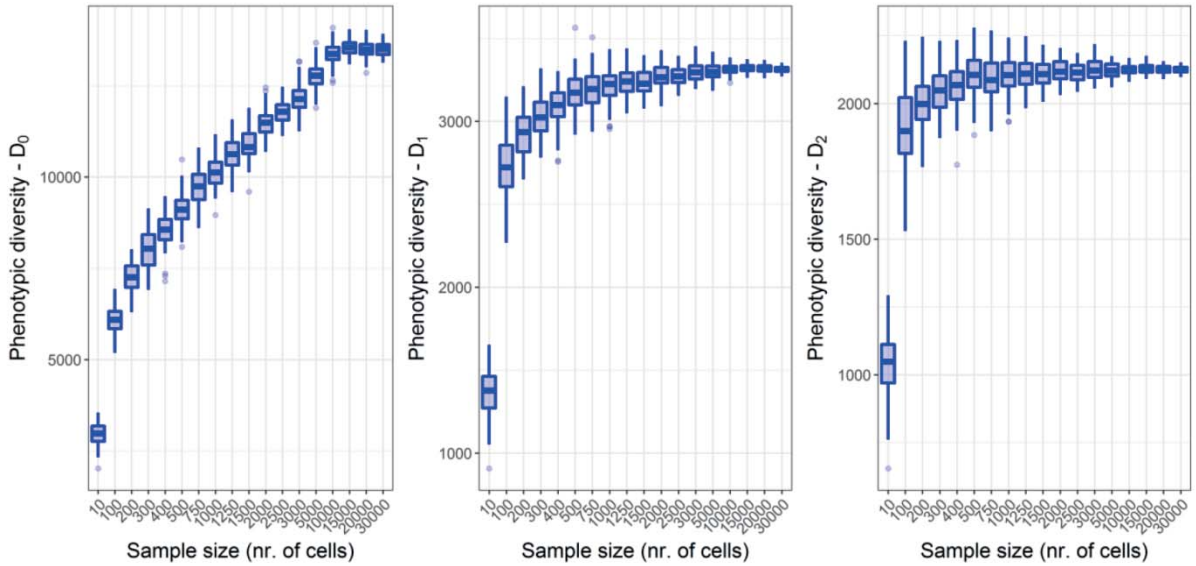


Figure 9 - 4: *Influence of the number of cells used for flow cytometric fingerprinting for robust calculation of all Hill number diversity indices. Cells were randomly selected from a same sample for each specific sample size and the Hill number diversity indices were calculated for 100 bootstrap samples. The calculation of the fingerprints was identical to the pipeline used in previous chapters (Props et al., in preparation).*

A second important aspect influencing the fingerprints, is the background. In **Chapter 3**, we demonstrated that the background reduces the accuracy of the classification of the 29 *Lactobacillus* species. The background is a very variable part of the cytometry data but cannot be ruled out completely when working with a standardized protocol. In **Chapter 7**, we reported that the background was important for most biofilm samples and that not all background could be removed with a universal gate. Based on our findings in **Chapter 3 and 7**, we argue that background reduces the accuracy of the fingerprinting because of its variability. At the same time, we argue that the cytometric fingerprinting algorithm were not affected much by the sometimes very important fraction of background. Because of this, a closer look into the causes of background can help explain the nature of background and its consequences for fingerprinting.

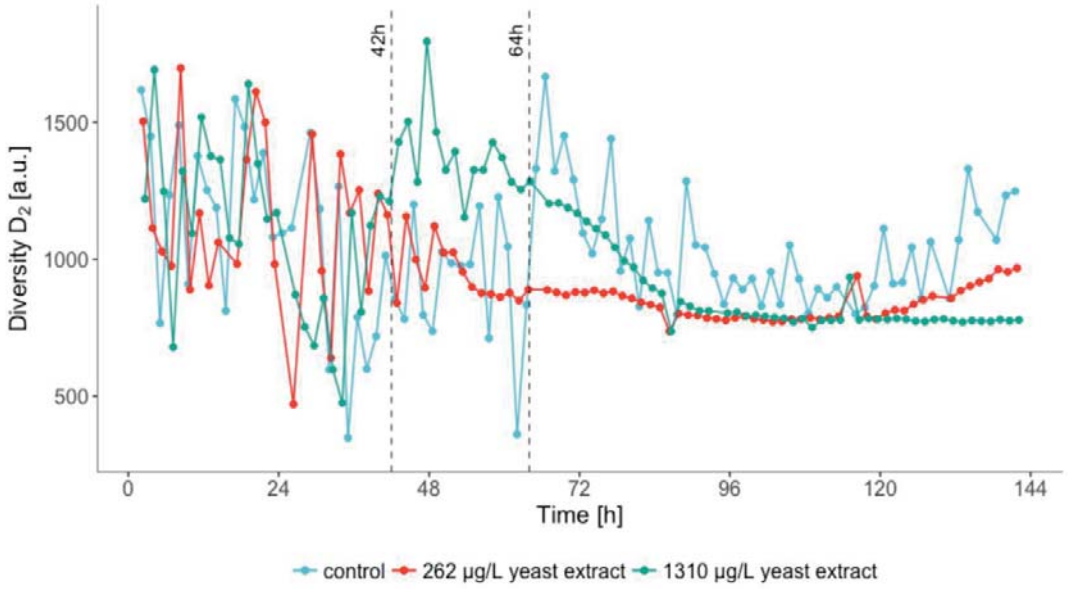


Figure 9 - 5: *Hill number diversity indices calculated for milli-Q water in function of time where 262 µg/L and 1310 µg/L of yeast extract was added. After 42 hours, growth was measured, based on the cell concentration, in both waters where yeast extract was added. After 64 hours, also some growth was measured in the water in which nothing was added. At first the cell numbers were too low to produce robust Hill number diversity indices. After approximately 64 hours, the Hill number diversity indices became less variable as the fingerprints were calculated on a sufficient number of cells.*

The background is a group name for all events originating from the instrument (instrument noise), from artifacts, which are added by accident to the sample, and from the sample itself. The instrument noise is generally constant and originates from the electronics but it can increase due to improper maintenance. Artifact background can be caused by many things and Van Nevel (2014) provides an overview and good illustrations of different causes of artifact background. Much of this background can be avoided by filtering all liquids used during the analysis and by using clean and disposable plastic lab consumables. Other sources of artifacts are experiment specific and will not be discussed further. The sample background is more complex to characterize as many things are thought to contribute. A complete characterization of the background is impossible but we found that in combination with the nucleic acid dyes SG and PI, mainly organic matter causes background noise. Van Nevel (2014) showed that free DNA, humic acids, sodium thiosulphate ($Na_2S_2O_3$), and complex media (e.g. Luria Bertani broth, yeast extract, casitone, and tryprone) contribute to a higher background in combination with SG and SGPI. Presumably because of the free DNA which also binds to the nucleic acid dyes, though with lower quantum yields. Similarly, we tested commercially available corrosion inhibitors and biocides for use in demineralized process water and also found an important contribution to the cytometric background.

Additionally, we also performed tests with $CaCO_3$, $CaSO_4$, $FeSO_4$, Na_2CO_3, NaCl, Na_2SO_4, NH_4Cl, and $(NH_4)_2SO_4$ with concentrations up to four times the maximum allowed concentration in drinking water and found no meaningful effect on the fingerprints. To conclude, background is undesirable as it reduces the performance of the fingerprinting. But, the majority of the background can be removed with a universal gate as we have done throughout this doctoral research. Nevertheless, preliminary experiments showed that the background was different for different compounds which suggests that there is more to background than expected. Preliminary results also showed that the background is very variable which makes the concept of background characterization difficult.

5 Raman spectroscopy

Whereas flow cytometry relies on staining protocols, Raman spectroscopy is a non-invasive, non-destructive, and label-free technique. On top of that, Raman spectroscopy can make single-cell fingerprints based on all molecules in the cell, which makes it more comprehensive, sensitive, and accurate than flow cytometry. A comparison between flow cytometry and Raman spectroscopy revealed that both methods were sensitive enough to detect the phenotypic plasticity of two bacteria taxa (**Chapter 6**). The most important difference between the methods is related to the number of cells measured for both. While with flow cytometry 10 000 cells can be measured, the throughput of Raman spectroscopy is much lower and approximately 50 cells per sample were measured (**Chapter 5 and 6**). There are two reasons for this lower throughput; the longer acquisition time and the lack of automation. The acquisition time was 45 seconds per cell in our experiments and is the consequence of the weak Raman signals. Shorter acquisition times are possible but the quality of the spectrum decreases and subtle differences between cells, proper to phenotypic differences, are not measured. Research has shown that an acquisition time of 10 seconds with spontaneous Raman scatter is sufficient for species characterization (Almarashi *et al.*, 2012). The possibility to increase the signal strength with SERS or CARS, could improve the methods throughput. Besides the acquisition time, also the method automation hampers the high throughput. Song *et al.* (2017) recently published a method for automated single-cell detection and sorting of bacteria. Yet, they also reported an acquisition time of five seconds per cell. Alternatives are laser tweezers Raman microspectroscopy (LTRS), which uses optical trapping of individual bacteria in an aqueous solution to focus the cells in the laser beam (Ma *et al.*, 2013, Xie *et al.*, 2005). Alternatively, chip-based methods can also be used (Walter *et al.*, 2011). Watson *et al.* (2008) published an interesting application of a chip-based system where both flow cytometry and Raman spectroscopy were combined. But, as microbial flow cytometry relies on fluorescent probes, and as Raman spectroscopy is

sensitive to fluorescent signals, a combination of CARS and flow cytometry would be most suitable for bacteria. Automation solutions to increase the throughput go hand in hand with signal enhancement. Attempts to increase the throughput have been made but the throughput remains low (*i.e.* approximately hundred cells) in comparison to methods such as flow cytometry. The sensitivity of Raman spectroscopy is the method's strength but also increases the necessity of standardized protocols to ensure reproducibility. Hutsebaut *et al.* (2005) provides a good overview on how to properly calibrate instruments but also the impact of sample preparation, fixation and storage should be researched (**Chapter 5**) (Read and Whiteley, 2015).

6 Applications

The characterization of microbial communities is useful in many applications and for different types of industry to gain insight in the microbial dynamics for monitoring. Single-cell optical techniques are fast methods and could therefore be implemented as early-warning methods, complementary to other established or more accurate methods. Depending on the complexity of the community and the goal of the characterization, different approaches can be proposed. Based on our findings, we argue that flow cytometry can, on the short term, be used as community characterization method for industrial applications. Raman spectroscopy still faces some technical difficulties to make it faster and high throughput, which makes it unsuitable for industrial applications at the moment. For research, both methods could already be used for community characterization.

6.1 Low complexity microbial communities

For research and for most types of fermentation in industrial biotechnology, microbial communities containing one or a limited number of species are used. Axenic isolates are the simplest type of microbial community and because the taxonomic diversity is known and invariable, monitoring such communities is putatively easier. But, research has shown that pure cultures are not as predictable and invariable as they seem and that that an isogenic bacteria population could be regarded as community of subpopulations (Elowitz *et al.*, 2002). Furthermore, the presence of inefficient subpopulations can significantly impact the process performance (Muller *et al.*, 2010). In **Chapter 4,** we illustrated the possibility to use flow cytometric fingerprinting for the characterization of and *E. coli* population during a batch fermentation. Flow cytometry was not only a fast, but also a very sensitive tool to detect the altered process conditions. Next to pure cultures, also cocultures (two or more species) are implemented for fermentation in industrial biotechnology because they offer the opportunity

to use cheap substrates, increase yields and product quality. Furthermore, the metabolic potential of a combination of organisms opens the possibility to develop new processes and products (Bader *et al.*, 2010). Combining two or more species is complex and microbial interactions can have both positive and negative effects on the overall process performance. In **Chapter 6,** we showed that bacteria change their physiology when cultured together and the combination of both Raman spectroscopy and flow cytometry revealed how both bacterial populations changed their physiology. Complementary to our cytometry pipeline, the method described by Rubbens *et al.* (2017) allows to estimate the relative abundance of the bacteria present. Single-cell Raman spectroscopy is capable of detecting both phenotypes of bacteria and genotypes (Hutsebaut and Moens, 2005) and could be used to determine the relative abundance of different species and their phenotypes in a community. Absolute quantification is more difficult due to the small sample size as at least 400 cells should be measured to be accurate (Bolter *et al.*, 2002). Relative species abundances could also be measured on bulk samples using unmixing algorithms with improved non-linear models. Alternatively, machine learning algorithms can be used to detect the relative amount of each bacteria in a group spectrum (**Chapter 5**).

The most important challenge for the application of flow cytometry for reactor monitoring is the nature of the substrate and the microbial concentration. Substrates used during fermentation are often very complex and concentrated, which could interfere with the staining of the cells. For example, the presence of salts is known to reduce the binding efficiency of SG to DNA (Zipper *et al.*, 2004) (**Chapter 2**). Furthermore, the presence of too much free DNA could also increase the background which would also reduce the fingerprinting accuracy (**section 4.3**). the high microbial concentrations are also a potential issue as concentrations above 10^6 cells/mL result in the inaccurate detection and counting of the cells by the flow cytometer. Furthermore, the chances of clogging the capillaries increase with highly concentrated samples. Generally, a dilution step is sufficient to reduce both the interfering chemical compounds and the microbial concentration. To summarize, flow cytometry allows the fast and online detection of physiological changes in the microbial communities and also allows the absolute quantification of the microorganisms.

6.2 High complexity microbial communities

Bacteria are present in water in the form of complex communities containing many species. Even in clean water, such as drinking water, bacteria are present (Luhrig *et al.*, 2015, Hammes *et al.*, 2008). Different qualities are required for the many industrial, agricultural, and domestic applications for which water is used. Microbiology is an important factor of the

water quality to reduce health hazards but also to avoid technical failures in industrial settings, such as microbially induced corrosion or biofouling. Monitoring of the microbial communities is typically done by using the heterotrophic plate count method (HPC) as it is also part of the EU and national drinking water quality regulations (EU, 1998). Additionally also the presence of indicator bacteria which, when present in water, indicate the poor water quality. For example, the presence of *E. coli* in water indicates that the water can be hazardous for human consumption. The presence of indicator organisms does not provide full certainty about the water quality as only some *E. coli* species are pathogenic and as not all *E. coli* species are indicators of human or animal contamination (Levy *et al.*, 2012). Furthermore, legislation does not provide an upper limit of the number of colony forming units (CFU) detected because this is highly variably and cannot be generalized. But it does stipulate that 'no abnormal changes' should be observed (Van Nevel, 2014). Considering that maximum 0.1% of the bacteria present in water can be cultured, the accuracy of the method and its ability to be used as a quality indicator is questionable (Hammes *et al.*, 2008). HPC is generally not used for monitoring of process water due to its slow speed and the labor requirements, and the microbial quality monitoring is thus often disregarded. Biocides or other disinfectants are generally added to mitigate microbiological issues.

Flow cytometry is a fast, cheap, and automated method to detecting and count accurately all bacteria in the water. Guidelines and a standardized protocol were also introduced in the Swiss water legislation (SLMB, 2012). Also outside Switzerland, flow cytometry has been applied for the monitoring of drinking water quality, which illustrates the utility of and high demand for the method. The method is not limited to drinking water but can be used for any type of water where microbial quality is a relevant and useful parameter to monitor. Different types of information about the microbial communities in water can be obtained with flow cytometry depending on the operation of the instrument and the data processing. Total cell concentrations are the most straightforward information obtained with flow cytometry. Because the bacteria concentration is not directly related to health risks or operational risks, no upper limit can be set to ensure good water quality. Instead, risk assessment should be customized for each type of water, its treatment process, and its purpose, similarly to the CFU counts in drinking water (Sartory, 2004). Good water quality can be assured in case no abnormal changes in the cell concentration are detected after a reliable baseline has been established. For example, in **Chapter 7** we showed that drinking water bacteria could grow when coupons of common pipe materials were added and that growth was dependent on the source of the water and the type of coupon. In **Chapter 8**, we saw that, during the short measurement period, the microbial load of the water entering the water treatment plant increased twice due to boats roiling the water in the docks. Furthermore, we could link the

bacteria concentration to the operation of the different filtration units in the treatment plant and suggested that the cell concentration could be used as an online tool to monitor the filtration efficiency and could serve as early-warning method in case of process failures such as membrane disruption. Also, in a drinking water distribution system we showed that cell concentrations could help to detect dead ends or points of contamination (Van Nevel *et al.*, 2016a, Van Nevel *et al.*, 2016b). Flow cytometry can also reveal more about the microbial physiology besides simple enumeration. The response of bacteria to disinfection can be evaluated by using dyes such as SGPI (Van Nevel *et al.*, 2016a). Finally, more information can be extracted about the aquatic microbial community by using fingerprinting approaches. Multiple methods are available but all intent to extract more information out of the complex cytometric patterns. With our fingerprinting pipeline we could demonstrate that, even when cell concentrations were comparable, the community composition was not always so similar. This can indicate that pipes containing dirty water should be flushed longer (Van Nevel *et al.*, 2016b) or that the water originates from a different source or came in contact with different materials (**Chapter 7**). Similarly to the bacteria concentration, a stable water quality could be guaranteed in case no abnormal deviations from the established baseline are observed. The combination of both cel concentration (*i.e.* regrowth or contamination) and fingerprinting (*i.e.* community dynamics) can be a useful approach for determining biological stability of water.

7 Conclusions

The increased necessity to monitor and characterize microbial communities for both research and industry calls for fast, cheap and uncomplicated methods for community characterization. Single-cell optical techniques such as flow cytometry and Raman spectroscopy are therefore promising tools. During this doctoral research, we further developed flow cytometric fingerprinting and added an explorative similarity based approach to the flow cytometric fingerprinting toolbox. As such, we showed that flow cytometry allows discriminating among different bacteria species and strains and that phenotypic diversity could be assessed with flow cytometric fingerprinting. We also showed that Raman spectroscopy is suitable for the detection of both taxonomic and phenotypic diversity but that Raman spectroscopy can also provide a biological explanation for the observed differences. For practical applications, flow cytometry is so-far the most convenient method and we successfully demonstrated the added value of this method as an early-warning system for the monitoring of microbial water quality in drinking water distribution systems and in a full-scale water treatment plant.

SUMMARY - SAMENVATTING

SUMMARY

Bacteria are ubiquitous on earth and typically form complex microbial communities of coexisting genotypes and phenotypes. These communities are important for our modern society as many different industrial applications are facilitated by bacteria. These applications range from the production of fine chemicals, pharmaceuticals, and food to wastewater treatment, bioremediation, bioaugmentation, and bioleaching. In the future, the importance of bioprocesses is likely to increase due to the technological, economical, and ecological benefits of microbiologically-driven techniques. Microbial communities are the result of the complex interactions among bacteria and between bacteria and their environment. They are therefore constantly in flux and react sensitively on changed conditions. As a consequence, microbial communities are also good indicators of how processes operate. For example, water quality is monitored based on the absence or presence of certain indicator organisms. To improve bioprocesses or to monitor microbial communities for quality purposes, techniques for community characterization are necessary. In this doctoral research we explored how flow cytometry could be used for this purpose. Existing fingerprinting methods were improved and the sensitivity of this technique was tested for the taxonomic and phenotypic characterization of microbial communities. Since flow cytometry relies on staining and since staining can decrease the resolution and speed of the method, Raman spectroscopy, a label-free alternative, was also investigated for community characterization.

Bacteria are very small and fluorescent dyes staining a specific feature of microbial cells are necessary to detect them. The type of dye dramatically impacts the cytometric fingerprints as only the stained features are assessed. In **Chapter 2** different multicolor staining protocols were compared to increase the amount of physiological information per cell and consequently to improve the fingerprints. We found that not only the number of dyes are important but also the stability of the dyes. Results showed that SG and SGPI are the most stable dyes and therefore favored for cytometric fingerprinting.

In **Chapter 3**, the cytometric fingerprinting toolbox was improved with a similarity based approach. To test the sensitivity of this unsupervised approach, 29 *Lactobacillus* species and strains were compared. Results showed that 27 out of 29 species and strains could be discriminated with SG or SGPI staining. The sensitivity of the method was found to be between 1% and 5% of change in the fingerprints, depending on the overlap between the bacteria fingerprints. The method is also reproducible but a standardized growth protocol is necessary. To illustrate the possible impact of the different growth stages on the fingerprints,

a comparison of the fingerprints of a batch culture were compared. The results showed that the fingerprints did change sufficiently for the fingerprinting algorithm.

To further investigate the impact of the microbial growth stage on the cytometric fingerprints, an *E. coli* batch fermentation was monitored over time (**Chapter 4**). Results showed that the fingerprints changed due to a phenotypic switch caused by substrate depletion. This switch occurred before conventional methods such as exhaust-gas analysis could detect the substrate depletion. As a consequence, we conclude that flow cytometric fingerprinting can be used as a monitoring tool for bioprocess operations. A contrast analysis shows that the sensitivity of the method is related to the sensitivity of the nucleic acid dyes to detect small changes in the nucleic acid composition.

Because labeling bacteria with fluorescent dyes reduces the amount of physiological information comprising the fingerprints, an alternative label-free method such as Raman spectroscopy was evaluated (**Chapter 5**). Several data processing pipelines were compared and results showed that phenotypic differences could be detected by both supervised and unsupervised methods with high accuracy. Moreover, we developed and tested a new algorithm which quantifies bacteria species in a community based on the average spectrum of the community. Quantification with this algorithm is possible but different results were found for different bacteria mixes.

To compare the sensitivity and the complementarity of flow cytometry and Raman spectroscopy, an experiment with microcosms was set up (**Chapter 6**). When two bacteria species were cocultured, they both changed their phenotypes though in different ways. This could be established with both flow cytometry and Raman spectroscopy but the combination of the two methods confirmed the observations. This also illustrates that the two methods are complementary.

Microbiology is an important aspect of water quality. To reduce health risks, bacteria in drinking water are mitigated by disinfection but it is impossible to remove all bacteria. Therefore a good monitoring of the microbial quality, especially in drinking water distribution systems is necessary. In **Chapter 7**, we compared the planktonic microbial communities with the microbial communities in the biofilm which are putatively the most important fraction of bacteria in DWDS. Results showed that both the type of materials in contact with the water and the origin and treatment of the water influenced the cytometric fingerprints and hence the microbial community. No relationship could be established between the bulk and biofilm fingerprints. Furthermore, we tested if an *Enterobacter amnigenus*, a drinking water contaminant, could colonize biofilms and subsequently contaminate drinking water. Results showed that the *Enterobacter* could do this, regardless of the type of material or of a pre-

existing biofilm. When *Enterobacter* was incubated with sterilized water, the bacteria grew better, probably because of necrotrophic growth or because of a lack of competition.

In **Chapter 8**, we tested the sensitivity of cytometric community fingerprinting by adding elements which could be used by the aquatic microbial community to grow. We showed that the fingerprints are mainly sensitive to (re)growth of the microbial community. Finally, we installed an online flow cytometer in a full-scale water treatment plant to illustrate the potential of the method as early-warning system and as water quality monitoring tool. Water after different filtration steps was monitored and we showed that after every filtration unit (UF, RO), bacteria could be detected. The cell density was related to the type of filter, the age, and fluctuated in function of the operation of the treatment plant.

To conclude, in this doctoral research we showed that flow cytometric fingerprinting can be used for community characterization and that it correlates with genotypic and phenotypic changes in the microbial community. The speed and automation of the method makes it an ideal candidate for industrial monitoring of bioreactors or water distribution systems. Raman spectroscopy is an alternative approach which is potentially more powerful as it is can easily discriminate among genotypes and phenotypes. Yet, Raman spectroscopy cannot be automated and more research is necessary to adapt it for industrial applications. However, the possibility to both characterize the species composition and the phenotypic composition of a community in less than a few hours makes it a promising tool for research.

SAMENVATTING

Bacteriën zijn wijdverspreid op aarde en vormen complexe gemeenschappen waar verschillende genotypen en fenotypen samenleven. Die gemeenschappen zijn belangrijk voor onze moderne samenleving omdat veel verschillende industriële toepassingen door bacteriën worden bewerkstelligd. Die toepassingen gaan van de productie van fijnchemicaliën, farmaceutische producten, en voeding tot het opzuiveren van afvalwater, bioremediatie, bioaugmentatie, en bioleaching. In de toekomst zal het belang van dergelijke technologieën toenemen door de ecologische, economische en technologische voordelen die hen kenmerken. Microbiële gemeenschappen zijn het product van de complexe interacties van bacteriën onderling en van de bacteriën met hun omgeving. Ze zijn dus dynamisch en reageren snel op externe veranderingen waardoor ze ook goede indicatoren zijn voor tot dusver onopgemerkte veranderingen. Een voorbeeld hiervan is de kwaliteitsborging van water die berust op de aan- of afwezigheid van bepaalde microbiële indicatororganismen. Om biotechnologische processen te kunnen verbeteren of om microbiële gemeenschappen te kunnen monitoren voor kwaliteitsborging, is een techniek die de microbiële gemeenschappen kan karakteriseren noodzakelijk. Tijdens dit doctoraatsonderzoek heeft men onderzocht hoe flow cytometrie hiervoor kan worden gebruikt. Bestaande fingerprinting methoden werden verbeterd en de gevoeligheid van de methode voor het karakteriseren van microbiële gemeenschappen werd bepaald. Omdat flow cytometrie berust op het kleuren van de cellen en omdat die kleuring de resolutie en de snelheid van de methode verminderen, heeft men een kleurenvrij alternatief, nl. Raman spectroscopie, onderzocht voor het karakteriseren van microbiële gemeenschappen.

Bacteriën zijn erg klein en fluorescente kleurstoffen zijn nodig om ze detecteerbaar te maken voor flow cytometrie. Het type kleurstof is cruciaal voor flow cytometrie en de fingerprinting algoritmen omdat enkel die gekleurde eigenschappen worden belicht. In **Hoofdstuk 2** heeft men verschillende 'multicolor' protocollen met elkaar vergeleken om de fysiologische informatie van de fingerprints te verhogen. Een van de belangrijkste bevindingen is dat 'multicolor' kleurstoffen vaak niet voldoende stabiel zijn voor fingerprinting en dat SG en SGPI de meest geschikte kleurstoffen zijn. In de rest van dit werk werd dan ook steeds voor deze kleurencombinaties gekozen.

In **Hoofdstuk 3** heeft men de mogelijkheden van flow cytometrische fingerprinting uitgebreid door de similariteit te berekenen tussen de fingerprints. Om de gevoeligheid van de methode te testen werden 29 verwante *Lactobacillus* species en stammen vergeleken. De resultaten gaven aan dat 27 van de 29 species en stammen van elkaar konden worden onderscheiden.

Met een combinatie van verschillende micropartikels en met een combinatie van verschillende soorten bacteriën werd de gevoeligheid van de methode geschat tussen de 1% à 5% veranderingen in de fingerprints. Mits een gestandaardiseerd groeiprotocol is de methode tevens reproduceerbaar. De gevoeligheid van de methode voor de fysiologische veranderingen tijdens de groei werd geïllustreerd door een cultuur op te volgen in de tijd.

Om te kunnen verklaren welke fysiologische veranderingen aan de basis liggen van een veranderde fingerprint in functie van de groeifasen werd in **Hoofdstuk 4** een *E. coli* batch fermentatie opgevolgd met flow cytometrie. De resultaten toonden aan dat de fingerprints veranderden op het moment dat het substraat op was. Die fenotypische switch werd eerder waargenomen dan de veranderingen gemeten door conventionele methoden zoals gas analyse. Verder heeft men met een contrastanalyse kunnen aantonen dat de gevoeligheid van de nucleïnezuurkleurstoffen voor kleine veranderingen in de samenstelling van de nucleïnezuren verantwoordelijk is voor de gevoeligheid van methode.

Het gebruik van kleurstoffen voor het detecteerbaar maken van bacteriën vermindert de fysiologische informatie in de fingerprints en Raman spectroscopie werd verder onderzocht als kleuringsonafhankelijk alternatief (**Hoofdstuk 5**). Verschillende dataverwerkingsalgoritmen werden vergeleken en resultaten hebben aangetoond dat zowel met een supervised als met een unsupervised algoritme de genotypische en fenotypische verschillen tussen populaties accuraat konden worden waargenomen. Daarenboven heeft men een algoritme voorgesteld om op basis van het gemiddelde spectrum van een gemeenschap de species binnen de gemeenschap te kwantificeren. De resultaten waren succesvol maar verschillend afhankelijk van de bacteriën in de gemeenschap.

Om de gevoeligheid en complementariteit van flow cytometrie en Raman spectroscopie te bepalen, werd een experiment met vereenvoudigd ecosysteem opgesteld (**Hoofdstuk 6**). Bacteriën die samen werden opgegroeid pasten hun fenotypen aan maar elk op een andere manier. Dit werd waargenomen door zowel flow cytometrie als Raman spectroscopie en de combinatie van beiden methoden bevestigde de waarnemingen hetgeen de complementariteit van beide illustreert.

Bacteriën zijn een belangrijk onderdeel van waterkwaliteit. Om gezondheidsrisico's te verminderen worden desinfectantia gebruik in het drinkwater distributienetwerk. Omdat de bacteriën nooit helemaal kunnen worden verwijderd, is een goede monitoring van de microbiële gemeenschap in het distributie netwerk noodzakelijk. In **Hoofdstuk 7** heeft men de fingerprints van de planktonische microbiële gemeenschap vergeleken met de fingerprints van de microbiële gemeenschap in de biofilms die verondersteld is de belangrijkste fractie bacteriën te bevatten in het drinkwaternetwerk. Resultaten toonden aan dat zowel het type

leidingmateriaal als de bron en productie methode van het water een rol spelen in de vormgeving van de microbiële gemeenschap in het water en de biofilms. Nochtans kon er geen verband worden vastgesteld tussen de fingerprints van de biofilms en het bulk water. Verder heeft men ook aangetoond dat *Enterobacter amnigenus*, een typische drinkwatercontaminant, biofilms kon koloniseren ongeacht het type leidingmateriaal. Indien geautoclaveerd water gebruikt werd, groeiden de *Enterobacter* beter. Wellicht door necrotrofische groei of door een gebrek aan concurrerende bacteriën.

In **Hoofdstuk 8** heeft men de gevoeligheid van flow cytometrische fingerprinting getest door elementen toe te voegen aan het water die door de aquatische microbiële gemeenschap als substraat kon worden gebruikt. Dit heeft aangetoond dat de methode voornamelijk gevoelig is voor (her)groei van de microbiële gemeenschap en dat die voornamelijk door koolstof tot stand wordt gebracht. Finaal heeft men met een online flow cytometer een industrieel waterproductiecentrum gemonitord om te illustreren hoe flow cytometrie als vroege indicator voor kwaliteitsborging kan worden gebruikt. Het water werd gestest na verschillende stappen in het proces en men heeft aangetoond dat ook na filtratie (UF of RO) bacteriën aanwezig zijn in het water. De bacterie concentratie was afhankelijk van het type filter, de leeftijd en de manier waarom het water productie centrum bestuurd werd.

In dit doctoraatsonderzoek kom men aantonen dat flow cytometrische fingerprinting een geschikte methode is voor het karakteriseren van microbiële gemeenschappen en dat het correleert met zowel genotypische als fenotypische veranderingen in de microbiële gemeenschap. De snelheid en automatisatie van de methode maken het reeds geschikt voor industriële toepassingen zoals voor het monitoren van bioreactoren of het bewaken van microbiële kwaliteit in een waterdistributienetwerk. Raman spectroscopie is een alternatieve methode die mogelijks krachtiger is omdat het moeiteloos genotypen en fenotypen kan onderscheiden. Echter, Raman spectroscopie kan nog niet worden geautomatiseerd en meer onderzoek is noodzakelijk eer het gebruikt kan worden voor industriële toepassingen. Maar de mogelijkheid om tegelijk de taxonomische en fenotypische samenstelling van een microbiële gemeenschap te bepalen in slechts enkele uren maakt deze methode uiterst geschikt voor onderzoek.

References

REFERENCES

Acar, M., Mettetal, J. T. & Van Oudenaarden, A. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics,* 40**,** 471-475.

Ackermann, J. U., Muller, S., Losche, A., Bley, T. & Babel, W. 1995. *Methylobacterium Rhodesianum* Cells Tend to Double the DNA Content under Growth Limitations and Accumulate Phb. *Journal of Biotechnology,* 39**,** 9-20.

Ackermann, M. 2013. Microbial individuality in the natural environment. *Isme Journal,* 7**,** 465-467.

Ackermann, M. 2015. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology,* 13**,** 497-508.

Ackermann, M. & Schreiber, F. 2015. A growing focus on bacterial individuality. *Environmental Microbiology,* 17**,** 2193-2195.

Ackermann, M., Stecher, B., Freed, N. E., Songhet, P., Hardt, W. D. & Doebeli, M. 2008. Self-destructive cooperation mediated by phenotypic noise. *Nature,* 454**,** 987-990.

Almarashi, J. F. M., Kapel, N., Wilkinson, T. S. & Telle, H. H. 2012. Raman Spectroscopy of Bacterial Species and Strains Cultivated under Reproducible Conditions. *Spectroscopy-an International Journal,* 27**,** 361-365.

Alonso, J. L., Mascellaro, S., Moreno, Y., Ferrús, M. A. & Hernández, J. 2002. Double-staining method for differentiation of morphological changes and membrane integrity of *Campylobacter coli* cells. *Applied and environmental microbiology,* 68**,** 5151-5154.

Alonso, S., Rendueles, M. & Diaz, M. 2012. Physiological heterogeneity of *Pseudomonas taetrolens* during lactobionic acid production. *Applied Microbiology and Biotechnology,* 96**,** 1465-1477.

Ambriz-Avina, V., Contreras-Garduno, J. A. & Pedraza-Reyes, M. 2014. Applications of Flow Cytometry to Characterize Bacterial Physiological Responses. *Biomed Research International*.

Amor, K. B., Breeuwer, P., Verbaarschot, P., Rombouts, F. M., Akkermans, A. D. L., De Vos, W. M. & Abee, T. 2002. Multiparametric Flow Cytometry and Cell Sorting for the Assessment of Viable, Injured, and Dead *Bifidobacterium* Cells during Bile Salt Stress. *Applied and Environmental Microbiology,* 68**,** 5209-5216.

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., Francois, J. & Yvert, G. 2008. Cell-to-cell Stochastic variation in gene expression is a complex genetic trait. *Plos Genetics,* 4.

Arku, B., Fanning, S. & Jordan, K. 2011. Flow cytometry to assess biochemical pathways in heat-stressed *Cronobacter* spp. (formerly *Enterobacter sakazakii*). *Journal of Applied Microbiology,* 111**,** 616-624.

Arora, R., Petrov, G. I. & Yakovlev, V. V. 2008. Analytical capabilities of coherent anti-Stokes Raman scattering microspectroscopy. *Journal of Modern Optics,* 55**,** 3237-3254.

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J. & Weightman, A. J. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology,* 71**,** 7724-7736.

Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J. & O'grady, J. 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature Biotechnology,* 33**,** 296-+.

Athamneh, A. I. M., Alajlouni, R. A., Wallace, R. S., Seleem, M. N. & Senger, R. S. 2014. Phenotypic Profiling of Antibiotic Response Signatures in *Escherichia coli* Using Raman Spectroscopy. *Antimicrobial Agents and Chemotherapy,* 58**,** 1302-1314.

Avery, S. V. 2006. Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology,* 4**,** 577-587.

B. Ellis, P. Haaland, F. Hahne, N. Le Meur, Gopalakrishnan, N. & Spidlen, J. flowCore: flowCore: Basic structures for flow cytometry data. . Bioconductor: R package version 1.34.3.

B. Ellis, P. H., F. Hahne, N. Le Meur, N. Gopalakrishnan, J. Spidlen and M. Jiang flowCore: flowCore: Basic structures for flow cytometry data. *r package version 1.34.10,* Bioconductor.

Baatout, S., De Boever, P. & Mergeay, M. 2005. Temperature-induced changes in bacterial physiology as determined by flow cytometry. *Annals of Microbiology,* 55**,** 73-80.

Bader, J., Mast-Gerlach, E., Popovic, M. K., Bajpai, R. & Stahl, U. 2010. Relevance of microbial coculture fermentations in biotechnology. *J Appl Microbiol,* 109**,** 371-87.

Balaban, N. Q., Gerdes, K., Lewis, K. & Mckinney, J. D. 2013. A problem of persistence: still more questions than answers? *Nature Reviews Microbiology,* 11**,** 587-591.

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. 2004. Bacterial persistence as a phenotypic switch. *Science,* 305**,** 1622-1625.

Barauna, R. A., Freitas, D. Y., Pinheiro, J. C., Folador, A. R. C. & Silva, A. 2017. A Proteomic Perspective on the Bacterial Adaptation to Cold: Integrating OMICs Data of the Psychrotrophic Bacterium *Exiguobacterium antarcticum* B7. *Proteomes,* 5.

Bashashati, A. & Brinkman, R. R. 2009. A survey of flow cytometry data analysis methods. *Adv Bioinformatics***,** 584603.

Beeton, M. L., Aldrich-Wright, J. R. & Bolhuis, A. 2014. The antimicrobial and antibiofilm activities of copper(II) complexes. *Journal of Inorganic Biochemistry,* 140**,** 167-172.

Beleites, C. & Sergo, V. 2016. hyperSpec: a package to handle hyperspectral data sets in R. R package version 0.98-20161118 ed.

Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. & Lilley, A. K. 2005. The contribution of species richness and composition to bacterial services. *Nature,* 436**,** 1157-1160.

Benincasa, M., Pacor, S., Gennaro, R. & Scocchi, M. 2009. Rapid and Reliable Detection of Antimicrobial Peptide Penetration into Gram-Negative Bacteria Based on Fluorescence Quenching. *Antimicrobial Agents and Chemotherapy,* 53**,** 3501-3504.

Berney, M., Hammes, F., Bosshard, F., Weilenmann, H. U. & Egli, T. 2007. Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. *Appl Environ Microbiol,* 73**,** 3283-90.

Berney, M., Vital, M., Huelshoff, I., Weilenmann, H. U., Egli, T. & Hammes, F. 2009. Rapid, cultivation-independent assessment of microbialviability in drinking water (vol 42, pg 4010, 2008). *Water Research,* 43**,** 2567-2567.

Berry, D., Mader, E., Lee, T. K., Woebken, D., Wang, Y., Zhu, D., Palatinszky, M., Schintimeister, A., Schmid, M. C., Hanson, B. T., Shterzer, N., Mizrahi, I., Rauch, I., Decker, T., Bocklitz, T., Popp, J., Gibson, C. M., Fowler, P. W., Huang, W. E. & Wagner, M. 2015. Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** E194-E203.

Bertrand, J.-C. E., Caumette, P. E., Lebaron, P. E., Matheron, R. E., Normand, P. E. & Sime-Ngando, T. E. 2015. *Environmental Microbiology: Fundamentals and Applications Microbial Ecology*.

Besemer, K. 2015. Biodiversity, community structure and function of biofilms in stream ecosystems. *Research in Microbiology,* 166**,** 774-781.

Besmer, M. D., Weissbrodt, D. G., Kratochvil, B. E., Sigrist, J. A., Weyland, M. S. & Hammes, F. 2014. The feasibility of automated online flow cytometry for in-situ monitoring of microbial dynamics in aquatic ecosystems. *Front Microbiol,* 5**,** 265.

Blainey, P. C. 2013. The future is now: single-cell genomics of bacteria and archaea. *Fems Microbiology Reviews,* 37**,** 407-427.

Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. 2003. Noise in eukaryotic gene expression. *Nature,* 422**,** 633-637.

Boe-Hansen, R., Albrechtsen, H. J., Arvin, E. & Jorgensen, C. 2002. Bulk water phase and biofilm growth in drinking water at low nutrient conditions. *Water Research,* 36**,** 4477-4486.

Bolter, M., Bloem, J., Meiners, K. & Moller, R. 2002. Enumeration and biovolume determination of microbial cells - a methodological review and recommendations for applications in ecological research. *Biology and Fertility of Soils,* 36**,** 249-259.

Bombach, P., Hubschmann, T., Fetzer, I., Kleinsteuber, S., Geyer, R., Harms, H. & Muller, S. 2011. Resolution of natural microbial community dynamics by community fingerprinting, flow cytometry, and trend interpretation analysis. *Adv Biochem Eng Biotechnol,* 124**,** 151-81.

Boon, N., Depuydt, S. & Verstraete, W. 2006. Evolutionary algorithms and flow cytometry to examine the parameters influencing transconjugant formation. *Fems Microbiology Ecology,* 55**,** 17-27.

Bott, T. R. 1995. *Fouling in heat exchangers,* Amsterdam, The Netherlands, Elservier Science B.V.

Bouix, M. & Ghorbal, S. 2015. Rapid assessment of *Oenococcus oeni* activity by measuring intracellular pH and membrane potential by flow cytometry, and its application to the more effective control of malolactic fermentation. *International Journal of Food Microbiology,* 193**,** 139-146.

Boye, E. & Lobner-Olesen, A. 1991. Bacterial growth control studied by flow cytometry. *Res Microbiol,* 142**,** 131-5.

Breeuwer, P., Drocourt, J. L., Rombouts, F. M. & Abee, T. 1996. A novel method for continuous determination of the intracellular pH in bacteria with the internally conjugated fluorescent probe 5 (and 6-)-carboxyfluorescein succinimidyl ester. *Applied and Environmental Microbiology,* 62**,** 178-183.

Brehm-Stecher, B. F. & Johnson, E. A. 2004. Single-cell microbiology: Tools, technologies, and applications. *Microbiology and Molecular Biology Reviews,* 68**,** 538-+.

Brognaux, A., Francis, F., Twizere, J. C., Thonart, P. & Delvigne, F. 2014. Scale-down effect on the extracellular proteome of *Escherichia coli*: correlation with membrane permeability and modulation according to substrate heterogeneities. *Bioprocess Biosyst Eng,* 37**,** 1469-85.

Brognaux, A., Han, S. S., Sorensen, S. J., Lebeau, F., Thonart, P. & Delvigne, F. 2013. A low-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors. *Microbial Cell Factories,* 12.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., Wilkins, M. J., Wrighton, K. C., Williams, K. H. & Banfield, J. F. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature,* 523**,** 208-U173.

Bryan, P. E., Kuzminsk.Ln, Sawyer, F. M. & Feng, T. H. 1973. Taste Thresholds of Halogens in Water. *Journal American Water Works Association,* 65**,** 363-368.

Budich, C., Neugebauer, U., Popp, J. & Deckert, V. 2008. Cell wall investigations utilizing tip-enhanced Raman scattering. *Journal of Microscopy-Oxford,* 229**,** 533-539.

Bull, R. J. 1982. Health-Effects of Drinking-Water Disinfectants and Disinfectant by-Products. *Environmental Science & Technology,* 16**,** A554-A559.

Bunthof, C. J. & Abee, T. 2002. Development of a Flow Cytometric Method To Analyze Subpopulations of Bacteria in Probiotic Products and Dairy Starters. *Applied and Environmental Microbiology,* 68**,** 2934-2942.

Bunthof, C. J., Van Den Braak, S., Breeuwer, P., Rombouts, F. M. & Abee, T. 1999. Rapid fluorescence assessment of the viability of stressed *Lactococcus lactis*. *Applied and Environmental Microbiology,* 65**,** 3681-3689.

Butler, H. J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood, N. J., Gardner, B., Martin-Hirsch, P. L., Walsh, M. J., Mcainsh, M. R., Stone, N. & Martin, F. L. 2016. Using Raman spectroscopy to characterize biological materials. *Nature Protocols,* 11**,** 664-687.

Buysschaert, B., Byloos, B., Van Houdt, R., Leys, N. & Boon, N. 2016. Reevaluating multicolor flow cytometry to assess microbial viability. *Applied Microbiology and Biotechnology*.

Buysschaert, B., Kerckhof, F. M., Vandamme, P., De Baets, B. & Boon, N. 2017. Flow cytometric fingerprinting for microbial strain discrimination and physiological characterization. *Cytometry A*.

Ceuppens, S., Timmery, S., Mahillon, J., Uyttendaele, M. & Boon, N. 2013. Small *Bacillus cereus* ATCC 14579 subpopulations are responsible for cytotoxin K production. *Journal of Applied Microbiology,* 114**,** 899-906.

Charnock, C. & Kjonno, O. 2000. Assimilable organic carbon and biodegradable dissolved organic carbon in Norwegian raw and drinking waters. *Water Research,* 34**,** 2629-2642.

Chen, S., Cao, Y., Ferguson, L. R., Shu, Q. & Garg, S. 2012. Flow cytometric assessment of the protectants for enhanced in vitro survival of probiotic lactic acid bacteria through simulated human gastro-intestinal stresses. *Appl Microbiol Biotechnol,* 95**,** 345-56.

Chen, Y.-C., Chen, L.-A., Chen, S.-J., Chang, M.-C. & Chen, T.-L. 2004. A modified osmotic shock for periplasmic release of a recombinant creatinase from *Escherichia coli*. *Biochemical Engineering Journal,* 19**,** 211-215.

Cheng, J. X., Jia, Y. K., Zheng, G. F. & Xie, X. S. 2002. Laser-scanning coherent anti-stokes Raman scattering microscopy and applications to cell biology. *Biophysical Journal,* 83**,** 502-509.

Chien, A. C., Hill, N. S. & Levin, P. A. 2012a. Cell Size Control in Bacteria. *Current Biology,* 22**,** R340-R349.

Chien, S. H., Hsieh, M. K., Li, H., Monnell, J., Dzombak, D. & Vidic, R. 2012b. Pilot-scale cooling tower to evaluate corrosion, scaling, and biofouling control strategies for cooling system makeup water. *Review of Scientific Instruments,* 83.

Cicerone, M. 2016. Molecular imaging with CARS micro-spectroscopy. *Current Opinion in Chemical Biology,* 33**,** 179-185.

Comas, J. & Vives-Rego, J. 1997. Assessment of the effects of Gramicidin, Formaldehyde, and surfactants on *Escherichia coli* by flow cytometry using nucleic acid and membrane potential dyes. *Cytometry,* 29**,** 58-64.

Corno, G. & Jurgens, K. 2006. Direct and indirect effects of protist predation on population size structure of a bacterial strain with high phenotypic plasticity. *Applied and Environmental Microbiology,* 72**,** 78-86.

Cronin, U. P. & Wilkinson, M. G. 2008a. *Bacillus cereus* endospores exhibit a heterogeneous response to heat treatment and low-temperature storage. *Food Microbiol,* 25**,** 235-43.

Cronin, U. P. & Wilkinson, M. G. 2008b. Monitoring growth phase-related changes in phosphatidylcholine-specific phospholipase C production, adhesion properties and physiology of *Bacillus cereus* vegetative cells. *J Ind Microbiol Biotechnol,* 35**,** 1695-703.

Cui, L., Yang, K., Zhou, G. W., Huang, W. E. & Zhu, Y. G. 2017. Surface-Enhanced Raman Spectroscopy Combined with Stable Isotope Probing to Monitor Nitrogen Assimilation at Both Bulk and Single-Cell Level. *Analytical Chemistry,* 89**,** 5794-5801.

Da Silva, T. L., Reis, A., Kent, C. A., Kosseva, M., Roseiro, J. C. & Hewitt, C. J. 2005. Stress-induced physiological responses to starvation periods as well as glucose and lactose pulses in *Bacillus licheniformis* CCMI 1034 continuous aerobic fermentation processes as measured by multi-parameter flow cytometry. *Biochemical Engineering Journal,* 24**,** 31-41.

Da Silva, T. L., Roseiro, J. C. & Reis, A. 2012. Applications and perspectives of multi-parameter flow cytometry to microbial biofuels production processes. *Trends Biotechnol,* 30**,** 225-32.

Dalwai, F., Spratt, D. & Pratten, J. 2006. Modeling shifts in microbial populations associated with health or disease. *Applied and environmental microbiology,* 72**,** 3678-3684.

Davey, H. M. & Hexley, P. 2011. Red but not dead? Membranes of stressed *Saccharomyces cerevisiae* are permeable to propidium iodide. *Environmental Microbiology,* 13**,** 163-171.

Davey, H. M. & Kell, D. B. 1996. Flow cytometry and cell sorting of heterogeneous microbial populations: The importance of single-cell analyses. *Microbiological Reviews,* 60**,** 641-&.

Davis, K. M. & Isberg, R. R. 2016. Defining heterogeneity within bacterial populations via single-cell approaches. *Bioessays,* 38**,** 782-790.

De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B. & Vandamme, P. 2011. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Systematic and Applied Microbiology,* 34**,** 20-29.

De Mey, M., Lequeux, G. J., Beauprez, J. J., Maertens, J., Van Horen, E., Soetaert, W. K., Vanrolleghem, P. A. & Vandamme, E. J. 2007. Comparison of different strategies to reduce acetate formation in *Escherichia coli*. *Biotechnology Progress,* 23**,** 1053-1063.

De Roy, K., Boon, N. & Thas, O. 2014a. *Microbial resource management : introducing new tools and ecological theories.* Thesis submitted in fulfillment of the requirements for the degree of doctor (PhD) in Applied Biological Sciences, UGent,.

De Roy, K., Clement, L., Thas, O., Wang, Y. & Boon, N. 2012. Flow cytometry for fast microbial community fingerprinting. *Water Res,* 46**,** 907-19.

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.

De Roy, K., Marzorati, M., Van Den Abbeele, P., Van De Wiele, T. & Boon, N. 2014b. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environmental Microbiology,* 16**,** 1472-1481.

Delorenzo, V., Herrero, M., Jakubzik, U. & Timmis, K. N. 1990. Mini-Tn5 Transposon Derivatives for Insertion Mutagenesis, Promoter Probing, and Chromosomal Insertion of Cloned DNA in Gram-Negative Eubacteria. *Journal of Bacteriology,* 172**,** 6568-6572.

Delvigne, F. & Goffin, P. 2014. Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. *Biotechnology Journal,* 9**,** 61-72.

Deng, H., Bloomfield, V. A., Benevides, J. M. & Thomas, G. J. 1999. Dependence of the raman signature of genomic B-DNA on nucleotide base sequence. *Biopolymers,* 50**,** 656-666.

Dewettinck, T., Van Houtte, E., Geenens, D., Van Hege, K. & Verstraete, W. 2001. HACCP (Hazard Analysis and Critical Control Points) to guarantee safe water reuse and drinking water production - a case study. *Water Science and Technology,* 43**,** 31-38.

Diaper, J. P., Tither, K. & Edwards, C. 1992. Rapid Assessment of Bacterial Viability by Flow-Cytometry. *Applied Microbiology and Biotechnology,* 38**,** 268-272.

Dieckmann, R., Graeber, I., Kaesler, I., Szewzyk, U. & Von Dohren, H. 2005. Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by Intact-Cell-MALDI-TOF mass spectrometry (ICM-MS). *Applied Microbiology and Biotechnology,* 67**,** 539-548.

Dittrich, W. & Gühde, W. 1973. *Flow-through chamber for photometers to measure and count particles in a dispersion medium*.

Dondero, T. J., Rendtorff, R. C., Mallison, G. F., Weeks, R. M., Levy, J. S., Wong, E. W. & Schaffner, W. 1980. An Outbreak of Legionnaires' Disease Associated with a Contaminated Air-Conditioning Cooling Tower. *New England Journal of Medicine,* 302**,** 365-370.

Doolittle, W. F. & Papke, R. T. 2006. Genomics and the bacterial species problem. *Genome Biology,* 7.

Egli, T. 2010. How to live at very low substrate concentration. *Water Research,* 44**,** 4826-4837.

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. 2002. Stochastic gene expression in a single cell. *Science,* 297**,** 1183-1186.

Emtiazi, F., Schwartz, T., Marten, S. M., Krolla-Sidenstein, P. & Obst, U. 2004. Investigation of natural biofilms formed during the production of drinking water from surface water embankment filtration. *Water Research,* 38**,** 1197-1206.

Eu 1998. COUNCIL DIRECTIVE 98/83/EC, on the quality of water intended for human consumption. *In:* UNION, E. (ed.).

Eydal, H. S. C. & Pedersen, K. 2007. Use of an ATP assay to determine viable microbial biomass in Fennoscandian Shield groundwater from depths of 3-1000 m. *Journal of Microbiological Methods,* 70**,** 363-373.

Fitzgerald, D. J., Stratford, M., Gasson, M. J., Ueckert, J., Bos, A. & Narbad, A. 2004. Mode of antimicrobial action of vanillin against *Escherichia coli*, *Lactobacillus plantarum* and *Listeria innocua*. *Journal of Applied Microbiology,* 97**,** 104-113.

Fleischmann, M., Hendra, P. J. & Mcquillan, A. J. 1974. Raman-Spectra of Pyridine Adsorbed at a Silver Electrode. *Chemical Physics Letters,* 26**,** 163-166.

Flemming, H. C. 2002. Biofouling in water systems - cases, causes and countermeasures. *Applied Microbiology and Biotechnology,* 59**,** 629-640.

Flemming, H. C. & Wingender, J. 2010. The biofilm matrix. *Nature Reviews Microbiology,* 8**,** 623-633.

Fletez-Brant, K., Spidlen, J., Brinkman, R. R., Roederer, M. & Chattopadhyay, P. K. 2016. flowClean: Automated Identification and Removal of Fluorescence Anomalies in Flow Cytometry Data. *Cytometry Part A,* 89a**,** 461-471.

Flores, E. & Herrero, A. 2010. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology,* 8**,** 39-50.

Forster, S., Snape, J. R., Lappin-Scott, H. M. & Porter, J. 2002. Simultaneous Fluorescent Gram Staining and Activity Assessment of Activated Sludge Bacteria. *Applied and Environmental Microbiology,* 68**,** 4772-4779.

Fraser, D. & Kaern, M. 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molecular Microbiology,* 71**,** 1333-1340.

Fuller, M. E., Streger, S. H., Rothmel, R. K., Mailloux, B. J., Hall, J. A., Onstott, T. C., Fredrickson, J. K., Balkwill, D. L. & Deflaun, M. F. 2000. Development of a vital fluorescent staining method for monitoring bacterial transport in subsurface environments. *Appl Environ Microbiol,* 66**,** 4486-96.

Gagnon, G. A., Rand, J. L., O'leary, K. C., Rygel, A. C., Chauret, C. & Andrews, R. C. 2005. Disinfectant efficacy of chlorite and chlorine dioxide in drinking water biofilms. *Water Research,* 39**,** 1809-1817.

Garrett, T. R., Bhakoo, M. & Zhang, Z. B. 2008. Bacterial adhesion and biofilms on surfaces. *Progress in Natural Science-Materials International,* 18**,** 1049-1056.

Gawad, C., Koh, W. & Quake, S. R. 2016. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics,* 17**,** 175-188.

Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A. & Rodriguez-Valera, F. 2013. Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci Rep,* 3**,** 2471.

Ghayeni, S. B. S., Beatson, P. J., Fane, A. J. & Schneider, R. P. 1999. Bacterial passage through microfiltration membranes in wastewater applications. *Journal of Membrane Science,* 153**,** 71-82.

Gibb, S. & Strimmer, K. 2012. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* 28**,** 2270-2271.

Gomes, A., Fernandes, E. & Lima, J. L. 2005. Fluorescence probes used for detection of reactive oxygen species. *J Biochem Biophys Methods,* 65**,** 45-80.

Gonzalez-Torres, P., Pryszcz, L. P., Santos, F., Martinez-Garcia, M., Gabaldon, T. & Anton, J. 2015. Interactions between Closely Related Bacterial Strains Are Revealed by Deep Transcriptome Sequencing. *Applied and Environmental Microbiology,* 81**,** 8445-8456.

Goosen, M. F. A., Sablani, S. S., Ai-Hinai, H., Ai-Obeidani, S., Al-Belushi, R. & Jackson, D. 2004. Fouling of reverse osmosis and ultrafiltration membranes: A critical review. *Separation Science and Technology,* 39**,** 2261-2297.

Gorenflo, V., Steinbuchel, A., Marose, S., Rieseberg, M. & Scheper, T. 1999. Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining. *Applied Microbiology and Biotechnology,* 51**,** 765-772.

Greenspan, P. & Fowler, S. D. 1985. Spectrofluorometric studies of the lipid probe, nile red. *J Lipid Res,* 26**,** 781-9.

Grosse, C., Bergner, N., Dellith, J., Heller, R., Bauer, M., Mellmann, A., Popp, J. & Neugebauer, U. 2015. Label-free imaging and spectroscopic analysis of intracellular bacterial infections. *Anal Chem,* 87**,** 2137-42.

Grun, B. & Leisch, F. 2008. FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. *Journal of Statistical Software,* 28**,** 1-35.

Guckert, F. T. J. & O'konski, C. T. 1947. A photoelectronic counter for colloidal particle. *Journal of the American Chemical Society,* 69**,** 2422-2431.

Hai, F. I., Riley, T., Shawkat, S., Magram, S. F. & Yamamoto, K. 2014. Removal of Pathogens by Membrane Bioreactors: A Review of the Mechanisms, Influencing Factors and Reduction in Chemical Disinfectant Dosing. *Water,* 6**,** 3603-3630.

Hammes, F., Berney, M. & Egli, T. 2011. Cultivation-independent assessment of bacterial viability. *Adv Biochem Eng Biotechnol,* 124**,** 123-50.

Hammes, F., Berney, M., Wang, Y., Vital, M., Koster, O. & Egli, T. 2008. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Res,* 42**,** 269-77.

Hammes, F., Broger, T., Weilenmann, H. U., Vital, M., Helbing, J., Bosshart, U., Huber, P., Odermatt, R. P. & Sonnleitner, B. 2012. Development and laboratory-scale testing of a fully automated online flow cytometer for drinking water analysis. *Cytometry A,* 81**,** 508-16.

Hammes, F. & Egli, T. 2010. Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. *Anal Bioanal Chem,* 397**,** 1083-95.

Hammes, F., Goldschmidt, F., Vital, M., Wang, Y. & Egli, T. 2010. Measurement and interpretation of microbial adenosine tri-phosphate (ATP) in aquatic environments. *Water Res,* 44**,** 3915-23.

Hansen, G., Johansen, C. L., Honore, A. H., Jensen, H. M., Jespersen, L. & Arneborg, N. 2015. Fluorescent labelling negatively affects the physiology of *Lactococcus lactis*. *International Dairy Journal,* 49**,** 130-138.

Hardin, G. 1960. The competitive exclusion principle. *Science,* 131**,** 1292-7.

Harrell, F. E. J. 2017. Hmisc: Harrell Miscellaneous. R package version 4.0-3 ed.

Harrison, D. E. F. 1978. Mixed Cultures in Industrial Fermentation Processes. *Advances in Applied Microbiology,* 24**,** 129-164.

Harz, M., Rosch, P., Peschke, K. D., Ronneberger, O., Burkhardt, H. & Popp, J. 2005. Micro-Raman spectroscopic identification of bacterial cells of the genus *Staphylococcus* and dependence on their cultivation conditions. *Analyst,* 130**,** 1543-1550.

Hennig, C. 2007. Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis,* 52**,** 258-271.

Hering, K., Cialla, D., Ackermann, K., Dorfer, T., Moller, R., Schneidewind, H., Mattheis, R., Fritzsche, W., Rosch, P. & Popp, J. 2008. SERS: a versatile tool in chemical and biochemical diagnostics. *Analytical and Bioanalytical Chemistry,* 390**,** 113-124.

Hermelink, A., Brauer, A., Lasch, P. & Naumann, D. 2009. Phenotypic heterogeneity within microbial populations at the single-cell level investigated by confocal Raman microspectroscopy. *Analyst,* 134**,** 1149-1153.

Herrera, G., Martinez, A., Blanco, M. & O'connor, J. E. 2002. Assessment of *Escherichia coli* B with enhanced permeability to fluorochromes for flow cytometric assays of bacterial cell function. *Cytometry,* 49**,** 62-9.

Hewitt, C. J., Nebe-Von Caron, G., Axelsson, B., Mcfarlane, C. M. & Nienow, A. W. 2000. Studies related to the scale-up of high-cell-density *E. coli* fed-batch fermentations using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnol Bioeng,* 70**,** 381-90.

Hewitt, C. J., Nebe-Von Caron, G., Nienow, A. W. & Mcfarlane, C. M. 1999. Use of multi-staining flow cytometry to characterise the physiological state of *Escherichia coli* W3110 in high cell density fed-batch cultures. *Biotechnology and bioengineering,* 63**,** 705-711.

Hewitt, C. J., Onyeaka, H., Lewis, G., Taylor, I. W. & Nienow, A. W. 2007. A comparison of high cell density fed-batch fermentations involving both induced and non-induced recombinant *Escherichia coli* under well-mixed small-scale and simulated poorly mixed large-scale conditions. *Biotechnology and Bioengineering,* 96**,** 495-505.

Hill, M. O. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology,* 54**,** 427-432.

Hoefel, D., Grooby, W. L., Monis, P. T., Andrews, S. & Saint, C. P. 2003. A comparative study of carboxyfluorescein diacetate and carboxyfluorescein diacetate succinimidyl ester as indicators of bacterial activity. *Journal of Microbiological Methods,* 52**,** 379-388.

Hong, W. L., Liao, C. S., Zhao, H. S., Younis, W., Zhang, Y. X., Seleem, M. N. & Cheng, J. X. 2016. In situ Detection of a Single Bacterium in Complex Environment by Hyperspectral CARS Imaging. *Chemistryselect,* 1**,** 513-517.

Hornbaek, T., Dynesen, J. & Jakobsen, M. 2002. Use of fluorescence ratio imaging microscopy and flow cytometry for estimation of cell vitality for *Bacillus licheniformis*. *Fems Microbiology Letters,* 215**,** 261-265.

Horvath, G., Petras, M., Szentesi, G., Fabian, A., Park, J. W., Vereb, G. & Szollosi, J. 2005. Selecting the right fluorophores and flow cytometer for fluorescence resonance energy transfer measurements. *Cytometry Part A,* 65a**,** 148-157.

Huang, W. E., Bailey, M. J., Thompson, I. P., Whiteley, A. S. & Spiers, A. J. 2007. Single-cell Raman spectral profiles of *Pseudomonas fluorescens* SBW25 reflects in vitro and in planta metabolic history. *Microbial Ecology,* 53**,** 414-425.

Huang, W. E., Griffiths, R. I., Thompson, I. P., Bailey, M. J. & Whiteley, A. S. 2004. Raman microscopic analysis of single microbial cells. *Analytical Chemistry,* 76**,** 4452-4458.

Huh, D. & Paulsson, J. 2011. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics,* 43**,** 95-U32.

Hutsebaut, D. & Moens, L. P. 2005. *Evaluation of Raman spectroscopy as identification tool for bacteria.* 2005.

Hutsebaut, D., Vandenabeele, P. & Moens, L. 2005. Evaluation of an accurate calibration and spectral standardization procedure for Raman spectroscopy. *Analyst,* 130**,** 1204-1214.

Hyka, P., Zullig, T., Ruth, C., Looser, V., Meier, C., Klein, J., Melzoch, K., Meyer, H. P., Glieder, A. & Kovar, K. 2010. Combined use of fluorescent dyes and flow cytometry to quantify the physiological state of *Pichia pastoris* during the production of heterologous proteins in high-cell-density fed-batch cultures. *Appl Environ Microbiol,* 76**,** 4486-96.

Hytonen, J., Haataja, S. & Finne, J. 2006. Use of flow cytometry for the adhesion analysis of *Streptococcus pyogenes* mutant strains to epithelial cells: investigation of the possible role of surface pullulanase and cysteine protease, and the transcriptional regulator Rgg. *Bmc Microbiology,* 6.

Im, H., Castro, C. M., Shao, H. L., Liong, M., Song, J., Pathania, D., Fexon, L., Min, C., Avila-Wallace, M., Zurkiya, O., Rho, J., Magaoay, B., Tambouret, R. H., Pivovarov, M., Weissleder, R. & Lee, H. 2015. Digital diffraction analysis enables low-cost molecular diagnostics on a smartphone. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 5613-5618.

Janda, J. M. & Abbott, S. L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology,* 45**,** 2761-2764.

Jarvis, R. M., Brooker, A. & Goodacre, R. 2004. Surface-enhanced Raman spectroscopy for bacterial discrimination utilizing a scanning electron microscope with a Raman spectroscopy interface. *Analytical Chemistry,* 76**,** 5198-5202.

Jarvis, R. M. & Goodacre, R. 2008. Characterisation and identification of bacteria using SERS. *Chemical Society Reviews,* 37**,** 931-936.

Jepras, R. I., Carter, J., Pearson, S. C., Paul, F. E. & Wilkinson, M. J. 1995. Development of a robust flow cytometric assay for determining numbers of viable bacteria. *Appl Environ Microbiol,* 61**,** 2696-701.

Jessup, C. M., Kassen, R., Forde, S. E., Kerr, B., Buckling, A., Rainey, P. B. & Bohannan, B. J. M. 2004. Big questions, small worlds: microbial model systems in ecology. *Trends in Ecology & Evolution,* 19**,** 189-197.

Jiang, S. X., Li, Y. N. & Ladewig, B. P. 2017. A review of reverse osmosis membrane fouling and control strategies. *Science of the Total Environment,* 595**,** 567-583.

Jiao, S., Zhang, Z. Q., Yang, F., Lin, Y. B., Chen, W. M. & Wei, G. H. 2017. Temporal dynamics of microbial communities in microcosms in response to pollutants. *Molecular Ecology,* 26**,** 923-936.

Johannsen, W. 1911. The Genotype Conception of Heredity. *The American Naturalist,* 45**,** 129-159.

Johnson, D. R., Goldschmidt, F., Lilja, E. E. & Ackermann, M. 2012. Metabolic specialization and the assembly of microbial communities. *Isme Journal,* 6**,** 1985-1991.

Johnson, I. & Spence, M. 2010. *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*, Life Technologies Corporation.

Keller, D. W., Hajjeh, R., Demaria, A., Fields, B. S., Pruckler, J. M., Benson, R. S., Kludt, P. E., Lett, S. M., Mermel, L. A., Giorgio, C. & Breiman, R. F. 1996. Community outbreak of legionnaires' disease: An investigation confirming the potential for cooling towers to transmit Legionella species. *Clinical Infectious Diseases,* 22**,** 257-261.

Keren, I., Shah, D., Spoering, A., Kaldalu, N. & Lewis, K. 2004. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. *Journal of Bacteriology,* 186**,** 8172-8180.

Kerr, C. J., Osborn, K. S., Roboson, G. D. & Handley, P. S. 1999. The relationship between pipe material and biofilm formation in a laboratory model system. *Journal of Applied Microbiology,* 85**,** 29s-38s.

Klitgord, N. & Segre, D. 2010. Environments that Induce Synthetic Microbial Ecosystems. *Plos Computational Biology,* 6.

Knabben, I., Regestein, L., Schauf, J., Steinbusch, S. & Buchs, J. 2011. Linear Correlation between Online Capacitance and Offline Biomass Measurement up to High Cell Densities in *Escherichia coli* Fermentations in a Pilot-Scale Pressurized Bioreactor. *Journal of Microbiology and Biotechnology,* 21**,** 204-211.

Koch, C. 2013. *Monitoring microbial community dynamics with flow cytometry.* Doctor rerum naturalium, Leipzig.

Koch, C., Fetzer, I., Harms, H. & Muller, S. 2013a. CHIC-an automated approach for the detection of dynamic variations in complex microbial communities. *Cytometry A,* 83**,** 561-7.

Koch, C., Fetzer, I., Schmidt, T., Harms, H. & Muller, S. 2013b. Monitoring functions in managed microbial systems by cytometric bar coding. *Environ Sci Technol,* 47**,** 1753-60.

Koch, C., Gunther, S., Desta, A. F., Hubschmann, T. & Muller, S. 2013c. Cytometric fingerprinting for analyzing microbial intracommunity structure variation and identifying subcommunity function. *Nature Protocols,* 8**,** 190-202.

Koch, C., Harnisch, F., Schroder, U. & Muller, S. 2014. Cytometric fingerprints: evaluation of new tools for analyzing microbial community dynamics. *Frontiers in Microbiology,* 5.

Koch, M., Delmotte, N., Rehrauer, H., Vorholt, J. A., Pessi, G. & Hennecke, H. 2010. Rhizobial Adaptation to Hosts, a New Facet in the Legume Root-Nodule Symbiosis. *Molecular Plant-Microbe Interactions,* 23**,** 784-790.

Konopka, A., Carrero-Colon, M. & Nakatsu, C. H. 2007. Community dynamics and heterogeneities in mixed bacterial communities subjected to nutrient periodicities. *Environmental Microbiology,* 9**,** 1584-1590.

Konopka, A., Lindemann, S. & Fredrickson, J. 2015. Dynamics in microbial communities: unraveling mechanisms to identify principles. *Isme Journal,* 9**,** 1488-1495.
Kotte, O., Volkmer, B., Radzikowski, J. L. & Heinemann, M. 2014. Phenotypic bistability in *Escherichia coli*'s central carbon metabolism. *Molecular Systems Biology,* 10.
Koydemir, H. C., Gorocs, Z., Tseng, D., Cortazar, B., Feng, S., Chan, R. Y. L., Burbano, J., Mcleod, E. & Ozcan, A. 2015. Rapid imaging, detection and quantification of *Giardia lamblia* cysts using mobile-phone based fluorescent microscopy and machine learning. *Lab on a Chip,* 15**,** 1284-1293.
Krafft, C., Dietzek, B. & Popp, J. 2009. Raman and CARS microspectroscopy of cells and tissues. *Analyst,* 134**,** 1046-1057.
Kumar, M., Adham, S. & Decarolis, J. 2007. Reverse osmosis integrity monitoring. *Desalination,* 214**,** 138-149.
Kummerli, R., Jiricny, N., Clarke, L. S., West, S. A. & Griffin, A. S. 2009. Phenotypic plasticity of a cooperative behaviour in bacteria. *Journal of Evolutionary Biology,* 22**,** 589-598.
Kuypers, M. M. M. & Jorgensen, B. B. 2007. The future of single-cell environmental microbiology. *Environmental Microbiology,* 9**,** 6-7.
Lahtinen, S. J., Ouwehand, A. C., Reinikainen, J. P., Korpela, J. M., Sandholm, J. & Salminen, S. J. 2006. Intrinsic properties of so-called dormant probiotic bacteria, determined by flow cytometric viability assays. *Applied and Environmental Microbiology,* 72**,** 5132-5134.
Larkin, P. 2011. *Infrared and raman spectroscopy : principles and spectral interpretation,* Amsterdam ; Boston, Elsevier.
Lawrence, J., Neu, T. & Swerhone, G. 1998. Application of multiple parameter imaging for the quantification of algal, bacterial and exopolymer components of microbial biofilms. *Journal of Microbiological Methods,* 32**,** 253-261.
Lebaron, P., Parthuisot, N. & Catala, P. 1998. Comparison of blue nucleic acid dyes for flow cytometric enumeration of bacteria in aquatic systems. *Appl Environ Microbiol,* 64**,** 1725-30.
Lebaron, P., Servais, P., Agogue, H., Courties, C. & Joux, F. 2001. Does the high nucleic acid content of individual bacterial cells allow us to discriminate between active cells and inactive cells in aquatic systems? *Appl Environ Microbiol,* 67**,** 1775-82.
Lee, Y. K., Ho, P. S., Low, C. S., Arvilommi, H. & Salminen, S. 2004. Permanent colonization by *Lactobacillus casei* is hindered by the low rate of cell division in mouse gut. *Applied and Environmental Microbiology,* 70**,** 670-674.
Legendre, P. & Legendre, L. 2012. Numerical ecology, 3rd English edition. Amsterdam: Elsevier.
Lehnert, L., Meyer, H. & Bendix, J. 2015. Hyperspectral Data Analysis in R - The new hsdar package.
Letunic, I. & Bork, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*.
Levy, K., Nelson, K. L., Hubbard, A. & Eisenberg, J. N. S. 2012. Rethinking Indicators of Microbial Drinking Water Quality for Health Studies in Tropical Developing Countries: Case Study in Northern Coastal Ecuador. *American Journal of Tropical Medicine and Hygiene,* 86**,** 499-507.
Lewis, C. L., Craig, C. C. & Senecal, A. G. 2014. Mass and Density Measurements of Live and Dead Gram-Negative and Gram-Positive Bacterial Populations. *Applied and Environmental Microbiology,* 80**,** 3622-3631.
Leys, N., Baatout, S., Rosier, C., Dams, A., S'heeren, C., Wattiez, R. & Mergeay, M. 2009. The response of *Cupriavidus metallidurans* CH34 to spaceflight in the international space station. *Antonie Van Leeuwenhoek,* 96**,** 227-245.
Liaw, A. & Wiener, M. 2002. Classification and Regression by randomForest. *R News,* 2(3)**,** 18-22.
Lidstrom, M. E. & Konopka, M. C. 2010. The role of physiological heterogeneity in microbial population behavior. *Nature Chemical Biology,* 6**,** 705-712.

Lieder, S., Jahn, M., Koepff, J., Muller, S. & Takors, R. 2016. Environmental stress speeds up DNA replication in Pseudomonas putida in chemostat cultivations. *Biotechnology Journal,* 11**,** 155-163.

Lindstrom, E. S. 2000. Bacterioplankton community composition in five lakes differing in trophic status and humic content. *Microbial Ecology,* 40**,** 104-113.

Linhova, M., Branska, B., Patakova, P., Lipovsky, J., Fribert, P., Rychtera, M. & Melzoch, K. 2012. Rapid flow cytometric method for viability determination of solventogenic clostridia. *Folia Microbiologica,* 57**,** 307-311.

Little, A. E. F., Robinson, C. J., Peterson, S. B., Raffa, K. E. & Handelsman, J. 2008. Rules of Engagement: Interspecies Interactions that Regulate Microbial Communities. *Annual Review of Microbiology,* 62**,** 375-401.

Liu, G., Verberk, J. Q. J. C. & Van Dijk, J. C. 2013. Bacteriology of drinking water distribution systems: an integral and multidimensional review. *Applied Microbiology and Biotechnology,* 97**,** 9265-9276.

Liu, S., Gunawan, C., Barraud, N., Rice, S. A., Harry, E. J. & Amal, R. 2016. Understanding, Monitoring, and Controlling Biofilm Growth in Drinking Water Distribution Systems. *Environmental Science & Technology,* 50**,** 8954-8976.

Liu, Y., Zhang, W., Sileika, T., Warta, R., Cianciotto, N. P. & Packman, A. 2009. Role of bacterial adhesion in the microbial ecology of biofilms in cooling tower systems. *Biofouling,* 25**,** 241-253.

Lopez-Amoros, R., Castel, S., Comas-Riu, J. & Vives-Rego, J. 1997. Assessment of *E. coli* and *Salmonella* viability and starvation by confocal laser microscopy and flow cytometry using rhodamine 123, DiBAC4(3), propidium iodide, and CTC. *Cytometry,* 29**,** 298-305.

Lopez-Amoros, R., Comas, J. & Vives-Rego, J. 1995. Flow cytometric assessment of *Escherichia coli* and *Salmonella typhimurium* starvation-survival in seawater using Rhodamine 123, propidium iodide and oxonol. *Appl Environ Microbiol,* 61**,** 2521-2526.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H. Y., Birarda, G., Thomas, B. C., Singh, A., Williams, K. H., Siegerist, C. E., Tringe, S. G., Downing, K. H., Comolli, L. R. & Banfield, J. F. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun,* 6**,** 6372.

Luhrig, K., Canback, B., Paul, C. J., Johansson, T., Persson, K. M. & Radstrom, P. 2015. Bacterial Community Analysis of Drinking Water Biofilms in Southern Sweden. *Microbes and Environments,* 30**,** 99-107.

Ma, H. F., Zhang, Y. & Ye, A. P. 2013. Single-cell discrimination based on optical tweezers Raman spectroscopy. *Chinese Science Bulletin,* 58**,** 2594-2600.

Maecker, H. T., Frey, T., Nomura, L. E. & Trotter, J. 2004. Selecting fluorochrome conjugates for maximum sensitivity. *Cytometry A,* 62**,** 169-73.

Mah, T.-F., Pitts, B., Pellock, B., Walker, G. C., Stewart, P. S. & O'toole, G. A. 2003. A genetic basis for *Pseudomonas aeruginosa* biofilm antibiotic resistance. *Nature,* 426**,** 306-310.

Mallon, C. A., Van Elsas, J. D. & Salles, J. F. 2015. Microbial Invasions: The Process, Patterns, and Mechanisms. *Trends in Microbiology,* 23**,** 719-729.

Maquelin, K., Choo-Smith, L. P., Van Vreeswijk, T., Endtz, H. P., Smith, B., Bennett, R., Bruining, H. A. & Puppels, G. J. 2000. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. *Anal Chem,* 72**,** 12-9.

Maquelin, K., Kirschner, C., Choo-Smith, L. P., Van Den Braak, N., Endtz, H. P., Naumann, D. & Puppels, G. J. 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. *Journal of Microbiological Methods,* 51**,** 255-271.

Marie, D., Vaulot, D. & Partensky, F. 1996. Application of the novel nucleic acid dyes YOYO-1, YO-PRO-1, and PicoGreen for flow cytometric analysis of marine prokaryotes. *Appl Environ Microbiol,* 62**,** 1649-55.

Martiny, A. C., Jorgensen, T. M., Albrechtsen, H. J., Arvin, E. & Molin, S. 2003. Long-term succession of structure and diversity of a biofilm formed in a model drinking water distribution system. *Applied and Environmental Microbiology,* 69**,** 6899-6907.

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.

Matos, C. T. & Lopes Da Silva, T. 2013. Using multi-parameter flow cytometry as a novel approach for physiological characterization of bacteria in microbial fuel cells. *Process Biochemistry,* 48**,** 49-57.

Mcdonogh, R., Schaule, G. & Flemming, H. C. 1994. The Permeability of Biofouling Layers on Membranes. *Journal of Membrane Science,* 87**,** 199-217.

Meesters, K. P., Van Groenestijn, J. W. & Gerritse, J. 2003. Biofouling reduction in recirculating cooling systems through biofiltration of process water. *Water Res,* 37**,** 525-32.

Mevik, B.-H. & Wehrens, R. 2016. Introduction to the pls Package. 2.6-0 ed.

Miettinen, I. 2009. Outbreaks of waterborne diseases. *In:* ENHIS, E. E. A. H. I. S. (ed.). Kuopio Finland: WHO Europe.

Mizrahi-Man, O., Davenport, E. R. & Gilad, Y. 2013. Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *Plos One,* 8.

Moldovan, A. 1934. Photo-electric technique for the counting of microscopical cells. *Science,* 80**,** 188-189.

Monaco, G., Chen, H., Poidinger, M., Chen, J. M., De Magalhaes, J. P. & Larbi, A. 2016. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics,* 32**,** 2473-2480.

Morton, S. C., Zhang, Y. & Edwards, M. A. 2005. Implications of nutrient release from iron metal for microbial regrowth in water distribution systems. *Water Research,* 39**,** 2883-2892.

Mrak, P., Podlesek, Z., Van Putten, J. P. M. & Zgur-Bertok, D. 2007. Heterogeneity in expression of the Escherichia coli colicin K activity gene cka is controlled by the SOS system and stochastic factors. *Molecular Genetics and Genomics,* 277**,** 391-401.

Muller, S. 2007. Modes of cytometric bacterial DNA pattern: a tool for pursuing growth. *Cell Prolif,* 40**,** 621-39.

Muller, S., Harms, H. & Bley, T. 2010. Origin and analysis of microbial population heterogeneity in bioprocesses. *Current Opinion in Biotechnology,* 21**,** 100-113.

Muller, S., Losche, A., Bley, T. & Scheper, T. 1995. A Flow Cytometric Approach for Characterization and Differentiation of Bacteria during Microbial Processes. *Applied Microbiology and Biotechnology,* 43**,** 93-101.

Muller, S. & Nebe-Von-Caron, G. 2010. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. *FEMS Microbiol Rev,* 34**,** 554-87.

Munzel, T., Afanas'ev, I. B., Kleschyov, A. L. & Harrison, D. G. 2002. Detection of superoxide in vascular tissue. *Arteriosclerosis Thrombosis and Vascular Biology,* 22**,** 1761-1768.

Murrant, C. L. & Reid, M. B. 2001. Detection of reactive oxygen and reactive nitrogen species in skeletal muscle. *Microscopy Research and Technique,* 55**,** 236-248.

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., Horreard, F., Amann, R., Jorgensen, B. B. & Kuypers, M. M. M. 2008. A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 17861-17866.

Mysara, M., Njima, M., Leys, N., Raes, J. & Monsieurs, P. 2017. From reads to operational taxonomic units: an ensemble processing pipeline for MiSeq amplicon sequencing data. *Gigascience,* 6.

Nebe-Von-Caron, G., Stephens, P. J., Hewitt, C. J., Powell, J. R. & Badley, R. A. 2000. Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. *Journal of Microbiological Methods,* 42**,** 97-114.

Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., Derisi, J. L. & Weissman, J. S. 2006. Single-cell proteomic analysis of *Saccharomyces cerevisiae* reveals the architecture of biological noise. *Nature,* 441**,** 840-846.

Newton, R. J., Jones, S. E., Eiler, A., Mcmahon, K. D. & Bertilsson, S. 2011. A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews,* 75**,** 14-49.

Nielsen, K. & Yding, F. 1983. Influence of Pipe Quality on Corrosion of Galvanized Steel Pipes for Domestic Water-Supply. *Werkstoffe Und Korrosion-Materials and Corrosion,* 34**,** 547-556.

Nielsen, T. H., Sjøholm, O. R. & Sørensen, J. 2009. Multiple physiological states of a *Pseudomonas fluorescens* DR54 biocontrol inoculant monitored by a new flow cytometry protocol. *FEMS microbiology ecology,* 67**,** 479-490.

Niemirycz, E., Gozdek, J. & Koszka-Maron, D. 2006. Variability of organic carbon in water and sediments of the odra river and its tributaries. *Polish Journal of Environmental Studies,* 15**,** 557-563.

Nikel, P. I., Silva-Rocha, R., Benedetti, I. & De Lorenzo, V. 2014. The private life of environmental bacteria: pollutant biodegradation at the single cell level. *Environmental Microbiology,* 16**,** 628-642.

Nikolic, N., Barner, T. & Ackermann, M. 2013. Analysis of fluorescent reporters indicates heterogeneity in glucose uptake and utilization in clonal bacterial populations. *Bmc Microbiology,* 13.

Niquette, P., Servais, P. & Savoir, R. 2000. Impacts of pipe materials on densities of fixed bacterial biomass in a drinking water distribution system. *Water Research,* 34**,** 1952-1956.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H. & Wagner, H. 2016. vegan: Community Ecology Package. R package version 2.3-4 ed.

Okuno, M., Kano, H., Leproux, P., Couderc, V., Day, J. P. R., Bonn, M. & Hamaguchi, H. 2010. Quantitative CARS Molecular Fingerprinting of Single Living Cells with the Use of the Maximum Entropy Method. *Angewandte Chemie-International Edition,* 49**,** 6773-6777.

Opilik, L., Schmid, T. & Zenobi, R. 2013. Modern Raman Imaging: Vibrational Spectroscopy on the Micrometer and Nanometer Scales. *Annual Review of Analytical Chemistry, Vol 6,* 6**,** 379-398.

Orphan, V. J., House, C. H., Hinrichs, K. U., Mckeegan, K. D. & Delong, E. F. 2001. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science,* 293**,** 484-487.

Pahlow, S., Marz, A., Seise, B., Hartmann, K., Freitag, I., Kammer, E., Bohme, R., Deckert, V., Weber, K., Cialla, D. & Popp, J. 2012. Bioanalytical application of surface- and tip-enhanced Raman spectroscopy. *Engineering in Life Sciences,* 12**,** 131-143.

Papadimitriou, K., Pratsinis, H., Nebe-Von-Caron, G., Kletsas, D. & Tsakalidou, E. 2007. Acid tolerance of *Streptococcus macedonicus* as assessed by flow cytometry and single-cell sorting. *Appl Environ Microbiol,* 73**,** 465-76.

Pestov, D., Wang, X., Ariunbold, G. O., Murawski, R. K., Sautenkov, V. A., Dogariu, A., Sokolov, A. V. & Scully, M. O. 2008. Single-shot detection of bacterial endospores via coherent Raman spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 422-427.

Petrov, G. I., Arora, R., Yakovlev, V. V., Wang, X., Sokolov, A. V. & Scully, M. O. 2007. Comparison of coherent and spontaneous Raman microspectroscopies for noninvasive detection of single bacterial endospores. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 7776-7779.

Petry, R., Schmitt, M. & Popp, J. 2003. Raman spectroscopy--a prospective tool in the life sciences. *Chemphyschem,* 4**,** 14-30.

Pinto, A. J., Schroeder, J., Lunn, M., Sloan, W. & Raskin, L. 2014. Spatial-Temporal Survey and Occupancy-Abundance Modeling To Predict Bacterial Community Dynamics in the Drinking Water Microbiome. *Mbio,* 5.

Pohlscheidt, M., Charaniya, S., Bork, C., Jenzsch, M., Noetzel, T. L., Luebbert, A. & Flickinger, M. C. 2009. Bioprocess and Fermentation Monitoring. *Encyclopedia of Industrial Biotechnology.* John Wiley & Sons, Inc.

Porter, J., Deere, D., Hardman, M., Edwards, C. & Pickup, R. 1997. Go with the flow - use of flow cytometry in environmental microbiology. *Fems Microbiology Ecology,* 24**,** 93-101.

Porter, J., Diaper, J., Edwards, C. & Pickup, R. 1995. Direct measurements of natural planktonic bacterial community viability by flow cytometry. *Appl Environ Microbiol,* 61**,** 2783-6.

Pot, B., Felis, G. E., Bruyne, K. D., Tsakalidou, E., Papadimitriou, K., Leisner, J. & Vandamme, P. 2014. The genus *Lactobacillus*. *Lactic Acid Bacteria.* John Wiley & Sons, Ltd.

Prest, E. I., El-Chakhtoura, J., Hammes, F., Saikaly, P. E., Van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. 2014. Combining flow cytometry and 16S rRNA gene pyrosequencing: A promising approach for drinking water monitoring and characterization. *Water Research,* 63**,** 179-189.

Prest, E. I., Hammes, F., Kotzsch, S., Van Loosdrecht, M. C. & Vrouwenvelder, J. S. 2013. Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res,* 47**,** 7131-42.

Prest, E. I., Hammes, F., Van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. 2016. Biological Stability of Drinking Water: Controlling Factors, Methods, and Challenges. *Frontiers in Microbiology,* 7.

Props, R., Kerckhof, F. M., Rubbens, P., De Vrieze, J., Sanabria, E. H., Waegeman, W., Monsieurs, P., Hammes, F. & Boon, N. 2017. Absolute quantification of microbial taxon abundances. *Isme Journal,* 11**,** 584-587.

Props, R., Monsieurs, P., Mysara, M., Clement, L. & Boon, N. 2016. Measuring the biodiversity of microbial communities by flow cytometry. *Methods in Ecology and Evolution***,** n/a-n/a.

Props, R., Schmidt, M., Heyse, J., Vanderploeg, H., Boon, N. & Denef, V. in preparation. Invasive dreissenid mussels induce shifts in bacterioplankton diversity through selective feeding on high nucleic acid bacteria. *in preparation*.

Pyle, B. H., Broadaway, S. C. & Mcfeters, G. A. 1995. A rapid, direct method for enumerating respiring enterohemorrhagic *Escherichia coli* O157:H7 in water. *Appl Environ Microbiol,* 61**,** 2614-9.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *Bmc Genomics,* 13.

R Core Team 2015. R: A language and environment for statistical computing. *R foundation for Statistical Computing.* Vienna, Austria.

Raman, C. V. & Krishnan, K. S. 1928. A new type of secondary radiation. *Nature,* 121**,** 501-502.

Rault, A., Bouix, M. & Beal, C. 2008. Dynamic analysis of *Lactobacillus delbrueckii subsp. bulgaricus* CFL1 physiological characteristics during fermentation. *Appl Microbiol Biotechnol,* 81**,** 559-70.

Rault, A., Bouix, M. & Beal, C. 2009. Fermentation pH influences the physiological-state dynamics of *Lactobacillus bulgaricus* CFL1 during pH-controlled culture. *Appl Environ Microbiol,* 75**,** 4374-81.

Read, D. S. & Whiteley, A. S. 2015. Chemical fixation methods for Raman spectroscopy-based analysis of bacteria. *Journal of Microbiological Methods,* 109**,** 79-83.

Read, D. S., Woodcock, D. J., Strachan, N. J. C., Forbes, K. J., Colles, F. M., Maiden, M. C. J., Clifton-Hadley, F., Ridley, A., Vidal, A., Rodgers, J., Whiteley, A. S. & Sheppard, S. K. 2013. Evidence for Phenotypic Plasticity among Multihost *Campylobacter jejuni* and *C. coli* Lineages, Obtained Using Ribosomal Multilocus Sequence Typing and Raman Spectroscopy. *Applied and Environmental Microbiology,* 79**,** 965-973.

Read, S., Marzorati, M., Guimaraes, B. C. M. & Boon, N. 2011. Microbial Resource Management revisited: successful parameters and new concepts. *Applied Microbiology and Biotechnology,* 90**,** 861-871.

Rezaeinejad, S. & Ivanov, V. 2011. Heterogeneity of *Escherichia coli* population by respiratory activity and membrane potential of cells during growth and long-term starvation. *Microbiol Res,* 166**,** 129-35.

Ridgway, H. F., Kelly, A., Justice, C. & Olson, B. H. 1983. Microbial Fouling of Reverse-Osmosis Membranes Used in Advanced Wastewater-Treatment Technology - Chemical, Bacteriological, and Ultrastructural Analyses. *Applied and Environmental Microbiology,* 45**,** 1066-1084.

Robertson, B. R. & Button, D. K. 1989. Characterizing Aquatic Bacteria According to Population, Cell-Size, and Apparent DNA Content by Flow-Cytometry. *Cytometry,* 10**,** 70-76.

Rodriguez, J. D., Westenberger, B. J., Buhse, L. F. & Kauffman, J. F. 2011. Standardization of Raman spectra for transfer of spectral libraries across different instruments. *Analyst,* 136**,** 4232-4240.

Roederer, M. 2002. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats (vol 45, pg 194, 2001). *Cytometry,* 48**,** 113-113.

Rogers, W. T. & Holyst, H. A. 2009. FlowFP: A Bioconductor Package for Fingerprinting Flow Cytometric Data. *Adv Bioinformatics***,** 193947.

Rosch, P., Harz, M., Schmitt, M., Peschke, K. D., Ronneberger, O., Burkhardt, H., Motzkus, H. W., Lankers, M., Hofer, S., Thiele, H. & Popp, J. 2005. Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: Application to clean-room-relevant biological contaminations. *Applied and Environmental Microbiology,* 71**,** 1626-1637.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y. T., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., Mckernan, K. J., Williams, A., Roth, G. T. & Bustillo, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature,* 475**,** 348-352.

Rozej, A., Cydzik-Kwiatkowska, A., Kowalska, B. & Kowalski, D. 2015. Structure and microbial diversity of biofilms on different pipe materials of a model drinking water distribution systems. *World Journal of Microbiology & Biotechnology,* 31**,** 37-47.

Rubbens, P., Props, R., Boon, N. & Waegeman, W. 2017. Flow Cytometric Single-Cell Identification of Populations in Synthetic Bacterial Communities. *Plos One,* 12.

Sadabad, M. S., Von Martels, J. Z. H., Khan, M. T., Blokzijl, T., Paglia, G., Dijkstra, G., Harmsen, H. J. M. & Faber, K. N. 2015. A simple coculture system shows mutualism between anaerobic faecalibacteria and epithelial Caco-2 cells. *Scientific Reports,* 5.

Sakakibara, T., Murakami, S., Hattori, N., Nakajima, M.-O. & Imai, K. 1997. Enzymatic Treatment to Eliminate the Extracellular ATP for Improving the Detectability of Bacterial Intracellular ATP. *Analytical Biochemistry,* 250**,** 157-161.

Salmond, G. P. C. & Whittenbury, R. 1985. Biology of Microorganisms, 4th Edition - Brock,Td, Smith,Dw, Madigan,Mt. *Nature,* 314**,** 49-49.

Sartory, D. P. 2004. Heterotrophic plate count monitoring of treated drinking water in the UK: a useful operational tool. *International Journal of Food Microbiology,* 92**,** 297-306.

Schmidt, T. M. 2006. The maturing of microbial ecology. *International Microbiology,* 9**,** 217-223.

Schrader, B. 1997. Die Möglichkeiten der Raman-Spektroskopie im Nah-Infrarot-Bereich, Teil II. *Chemie in unserer Zeit,* 31**,** 270-279.

Schreiber, F., Littmann, S., Lavik, G., Escrig, S., Meibom, A., Kuypers, M. M. M. & Ackermann, M. 2016. Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. *Nature Microbiology,* 1.

Schuster, K. C., Reese, I., Urlaub, E., Gapes, J. R. & Lendl, B. 2000a. Multidimensional information on the chemical composition of single bacterial cells by confocal Raman microspectroscopy. *Analytical Chemistry,* 72**,** 5529-5534.

Schuster, K. C., Urlaub, E. & Gapes, J. R. 2000b. Single-cell analysis of bacteria by Raman microscopy: spectral information on the chemical composition of cells and on the heterogeneity in a culture. *Journal of Microbiological Methods,* 42**,** 29-38.

Scully, M. O., Kattawar, G. W., Lucht, R. P., Opatrny, T., Pilloff, H., Rebane, A., Sokolov, A. V. & Zubairy, M. S. 2002. FAST CARS: Engineering a laser spectroscopic technique for rapid identification of bacterial spores. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 10994-11001.

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Burgmann, H., Huber, D. H., Langenheder, S., Lennon, J. T., Martiny, J. B. H., Matulich, K. L., Schmidt, T. M. & Handelsman, J. 2012. Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology,* 3.

Shapiro, H. 2003. *Practical Flow Cytometry,* United States of America, John Wiley & sons, Inc.

Shapiro, H. H. 2000. Microbial analysis at the single-cell level: tasks and techniques. *Journal of Microbiological Methods,* 42**,** 3-16.

Shi, L., Gunther, S., Hubschmann, T., Wick, L. Y., Harms, H. & Muller, S. 2007. Limits of propidium iodide as a cell viability indicator for environmental bacteria. *Cytometry A,* 71**,** 592-8.

Singer, E., Wagner, M. & Woyke, T. 2017. Capturing the genetic makeup of the active microbiome in situ. *Isme Journal,* 11**,** 1949-1963.

Slmb 2012. Determining the Total Cell Count and Ratios of High and Low Nucleic Acid Content Cells in Freshwater Using Flow Cytometry. Switserland: Federal Office of Public Health.

Smits, W. K., Kuipers, O. P. & Veening, J.-W. 2006. Phenotypic variation in bacteria: the role of feedback regulation. *Nat Rev Micro,* 4**,** 259-271.

Soetaert, W. & Vandamme, E. J. 2010. The Scope and Impact of Industrial Biotechnology. *Industrial Biotechnology.* Wiley-VCH Verlag GmbH & Co. KGaA.

Solopova, A., Van Gestel, J., Weissing, F. J., Bachmann, H., Teusink, B., Kok, J. & Kuipers, O. P. 2014. Bet-hedging during bacterial diauxic shift. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 7427-7432.

Song, Y., Kaster, A. K., Vollmers, J., Song, Y., Davison, P. A., Frentrup, M., Preston, G. M., Thompson, I. P., Murrell, J. C., Yin, H., Hunter, C. N. & Huang, W. E. 2017. Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea. *Microb Biotechnol,* 10**,** 125-137.

Spudich, J. L. & Koshland, D. E. 1976. Non-genetic individuality: chance in the single cell. *Nature,* 262**,** 467-471.

Starke, I. C., Vahjen, W., Pieper, R., Zentek, J., #Xfc & Rgen 2014. The Influence of DNA Extraction Procedure and Primer Set on the Bacterial Community Analysis by Pyrosequencing of Barcoded 16S rRNA Gene Amplicons. *Molecular Biology International,* 2014**,** 10.

Steen, H. B. & Boye, E. 1980. Bacterial growth studied by flow cytometry. *Cytometry,* 1**,** 32-6.

Stephen, K. E., Homrighausen, D., Depalma, G., Nakatsu, C. H. & Irudayaraj, J. 2012. Surface enhanced Raman spectroscopy (SERS) for the discrimination of *Arthrobacter* strains based on variations in cell surface composition. *Analyst,* 137**,** 4280-4286.

Stewart, P. S. & Franklin, M. J. 2008. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology,* 6**,** 199-210.

Stockel, S., Kirchhoff, J., Neugebauer, U., Rosch, P. & Popp, J. 2016. The application of Raman spectroscopy for the detection and identification of microorganisms. *Journal of Raman Spectroscopy,* 47**,** 89-109.

Strauber, H. & Muller, S. 2010. Viability States of Bacteria-Specific Mechanisms of Selected Probes. *Cytometry Part A,* 77a**,** 623-634.

Sundstrom, H., Wallberg, F., Ledung, E., Norrman, B., Hewitt, C. J. & Enfors, S. O. 2004. Segregation to non-dividing cells in recombinant *Escherichia coli* fed-batch fermentation processes. *Biotechnology Letters,* 26**,** 1533-1539.

Suzuki, R. & Shimodaira, H. 2015. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 2.0-0 ed.

Tachikawa, M., Tezuka, M., Morita, M., Isogai, K. & Okada, S. 2005. Evaluation of some halogen biocides using a microbial biofilm system. *Water Research,* 39**,** 4126-4132.

Tanaka, Y., Yamaguchi, N. & Nasu, M. 2000. Viability of *Escherichia coli* O157 : H7 in natural river water determined by the use of flow cytometry. *Journal of Applied Microbiology,* 88**,** 228-236.

Tang, X. 2011. *Controlling biofilm development on ultrafiltration and reverse osmosis membranes used in dairy plants.* doctor of philosophy, Massey University.

Tarpey, M. M., Wink, D. A. & Grisham, M. B. 2004. Methods for detection of reactive metabolites of oxygen and nitrogen in vitro and in vivo considerations. *Am J Physiol Regul Integr Comp Physiol,* 286**,** R431–R444.

Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J. T., Heijnen, J. J. & Van Gulik, W. M. 2010. Genome-Derived Minimal Metabolic Models for *Escherichia coli* MG1655 With Estimated In Vivo Respiratory ATP Stoichiometry. *Biotechnology and Bioengineering,* 107**,** 369-381.

Temmerman, R., Vervaeren, H., Noseda, B., Boon, N. & Verstraete, W. 2006. Necrotrophic growth of *Legionella pneumophila*. *Applied and Environmental Microbiology,* 72**,** 4323-4328.

Thijs, S., Op De Beeck, M., Beckers, B., Truyens, S., Stevens, V., Van Hamme, J. D., Weyens, N. & Vangronsveld, J. 2017. Comparative Evaluation of Four Bacteria-Specific Primer Pairs for 16S rRNA Gene Surveys. *Frontiers in Microbiology,* 8.

Tracy, B. P., Gaida, S. M. & Papoutsakis, E. T. 2008. Development and Application of Flow-Cytometric Techniques for Analyzing and Sorting Endospore-Forming Clostridia. *Applied and Environmental Microbiology,* 74**,** 7497-7506.

Tracy, B. P., Gaida, S. M. & Papoutsakis, E. T. 2010. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Curr Opin Biotechnol,* 21**,** 85-99.

Tyo, K. E., Zhou, H. & Stephanopoulos, G. N. 2006. High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp strain PCC6803. *Applied and Environmental Microbiology,* 72**,** 3412-3417.

Uzunbajakava, N., Lenferink, A., Kraan, Y., Willekens, B., Vrensen, G., Greve, J. & Otto, C. 2003. Nonresonant Raman imaging of protein distribution in single human cells. *Biopolymers,* 72**,** 1-9.

Van De Vossenberg, J., Tervahauta, H., Maquelin, K., Blokker-Koopmans, C. H. W., Uytewaal-Aarts, M., Van Der Kooij, D., Van Wezel, A. P. & Van Der Gaag, B. 2013. Identification of bacteria in drinking water with Raman spectroscopy. *Analytical Methods,* 5**,** 2679-2687.

Van Der Kooij, D., Van Der Wielen, P. W. J. J., Rosso, D., Shaw, A., Borchardt, D., Ibisch, R., Apgar, D., Witherspoon, J., Di Toro, D. M., Paquin, P. R., Mavinic, D., Koch, F., Guillot, E., Loret, J.-F., Hoffmann, E., Ødegaard, H., Hernandez-Sancho, F. & Molinos-Senante, M. 2014. *Microbial Growth in Drinking Water Supplies*, IWA publishing.

Van Der Maaten, L. & Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research,* 9**,** 2579-2605.

Van Manen, H. J., Kraan, Y. M., Roos, D. & Otto, C. 2005. Single-cell Raman and fluorescence microscopy reveal the association of lipid bodies with phagosomes in leukocytes. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 10159-10164.

Van Nevel, S. 2014. *CHARACTERISATION OF BACTERIA IN DRINKING WATER.* Ph.D., Gent University.

Van Nevel, S., Buysschaert, B., De Gusseme, B. & Boon, N. 2016a. Flow cytometric examination of bacterial growth in a local drinking water network. *Water and Environment Journal*.

Van Nevel, S., Buysschaert, B., De Roy, K., De Gusseme, B., Clement, L. & Boon, N. 2016b. Flow cytometry for immediate follow-up of drinking water networks after maintenance. *Water Res,* 111**,** 66-73.

Van Nevel, S., Koetzsch, S., Weilenmann, H. U., Boon, N. & Hammes, F. 2013. Routine bacterial analysis with automated flow cytometry. *J Microbiol Methods,* 94**,** 73-6.

Vandenabeele, P., Edwards, H. G. M. & Moens, L. 2007. A decade of Raman spectroscopy in art and archaeology. *Chemical Reviews,* 107**,** 675-686.

Vanderkooij, D., Visser, A. & Hijnen, W. a. M. 1982. Determining the Concentration of Easily Assimilable Organic-Carbon in Drinking-Water. *Journal American Water Works Association,* 74**,** 540-545.

Veening, J. W., Igoshin, O. A., Eijlander, R. T., Nijland, R., Hamoen, L. W. & Kuipers, O. P. 2008. Transient heterogeneity in extracellular protease production by Bacillus subtilis. *Molecular Systems Biology,* 4.

Velten, S., Hammes, F., Boller, M. & Egli, T. 2007. Rapid and direct estimation of active biomass on granular activated carbon through adenosine tri-phosphate (ATP) determination. *Water Research,* 41**,** 1973-1983.

Venables, W. N. & Ripley, B. D. 2002. *Modern Applied Statistics with S,* New York, Springer.

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.

Vidal-Mas, J., Resina-Pelfort, O., Haba, E., Comas, J., Manresa, A. & Vives-Rego, J. 2001. Rapid flow cytometry - Nile red assessment of PHA cellular content and heterogeneity in cultures of *Pseudomonas aeruginosa* 47T2 (NCIB 40044) grown in waste frying oil. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology,* 80**,** 57-63.

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.

Vriezen, J., De Bruijn, F. J. & Nüsslein, K. R. 2012. Desiccation induces viable but non-culturable cells in *Sinorhizobium meliloti* 1021. *AMB express,* 2.

Wagner, M. 2009. Single-Cell Ecophysiology of Microbes as Revealed by Raman Microspectroscopy or Secondary Ion Mass Spectrometry Imaging. *Annual Review of Microbiology,* 63**,** 411-429.

Walter, A., Marz, A., Schumacher, W., Rosch, P. & Popp, J. 2011. Towards a fast, high specific and reliable discrimination of bacteria on strain level by means of SERS in a microfluidic device. *Lab on a Chip,* 11**,** 1013-1021.

Wan, K. X., Vidavsky, I. & Gross, M. L. 2002. Comparing similar spectra: From similarity index to spectral contrast angle. *Journal of the American Society for Mass Spectrometry,* 13**,** 85-88.

Wand, M. 2015. KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995). http://CRAN.R-project.org/package=KernSmooth: R package version 2.23-15.

Wang, Y., Hammes, F., Boon, N. & Egli, T. 2007. Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 mu m pore size filters and shape-dependent enrichment of filterable bacterial communities. *Environmental Science & Technology,* 41**,** 7080-7086.

Wang, Y., Hammes, F., De Roy, K., Verstraete, W. & Boon, N. 2010. Past, present and future applications of flow cytometry in aquatic microbiology. *Trends Biotechnol,* 28**,** 416-24.

Wang, Y. Y., Hammes, F., Duggelin, M. & Egli, T. 2008. Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environmental Science & Technology,* 42**,** 6749-6754.

Wassenaar, T. M. & Lukjancenko, O. 2014. Comparative genomics of *Lactobacillus* and other LAB. *Lactic Acid Bacteria.* John Wiley & Sons, Ltd.

Watson, D. A., Brown, L. O., Gaskill, D. R., Naivar, M., Graves, S. W., Doorn, S. K. & Nolan, J. P. 2008. A flow cytometer for the measurement of Raman spectra. *Cytometry Part A,* 73a**,** 119-128.

White, R. K. 1988. Drinking-Water and Health - Disinfectants and Disinfectant by-Products, Vol 7 - Natl-Res-Council-Commiss-Life-Sci-B-Environm-Studies-Toxicol-Safe-Drinking-Water-Comm-Subcomm-Disinfectant-by-Product. *Risk Analysis,* 8**,** 159-159.

Who 2006. *Health aspects of plumbing,* Switzerland, World Health Organization.

Wick, L. M., Quadroni, M. & Egli, T. 2001. Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. *Environmental Microbiology,* 3**,** 588-599.

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., Shou, W. Y., Kettle, H., Flint, H. J., Haas, A. F., Laroche, B., Kreft, J. U., Rainey, P. B., Freilich, S., Schuster, S., Milferstedt, K., Van Der Meer, J. R., Grosskopf, T., Huisman, J., Free, A., Picioreanu, C., Quince, C., Klapper, I., Labarthe, S., Smets, B. F., Wang, H., Soyer, O. S. & Fellows, I. N. I. 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *Isme Journal,* 10**,** 2557-2568.

Wingender, J. & Flemming, H. C. 2011. Biofilms in drinking water and their role as reservoir for pathogens. *International Journal of Hygiene and Environmental Health,* 214**,** 417-423.

Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F. & Quake, S. R. 2014. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods,* 11**,** 41-6.

Xie, C., Mace, J., Dinno, M. A., Li, Y. Q., Tang, W., Newton, R. J. & Gemperline, P. J. 2005. Identification of Single Bacterial Cells in Aqueous Solution Using Confocal Laser Tweezers Raman Spectroscopy. *Analytical Chemistry,* 77**,** 4390-4397.

Yamaguchi, N. & Nasu, M. 1997. Flow cytometric analysis of bacterial respiratory and enzymatic activity in the natural aquatic environment. *Journal of Applied Microbiology,* 83**,** 43-52.

Yoon, S. H., Han, M. J., Jeong, H., Lee, C. H., Xia, X. X., Lee, D. H., Shim, J. H., Lee, S. Y., Oh, T. K. & Kim, J. F. 2012. Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biology,* 13.

Yue, S. H. & Cheng, J. X. 2016. Deciphering single cell metabolism by coherent Raman scattering microscopy. *Current Opinion in Chemical Biology,* 33**,** 46-57.

Zhu, H. & Ozcan, A. 2015. Opto-fluidics based microscopy and flow cytometry on a cell phone for blood analysis. *Methods Mol Biol,* 1256**,** 171-90.

Zipper, H., Brunner, H., Bernhagen, J. & Vitzthum, F. 2004. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Res,* 32**,** e103.

CURRICULUM VITAE

CURRICULUM VITAE

Personalia

Full name:	Benjamin François Charles Jacques Buysschaert
Date and place of birth	23rd July 1990, Gent, Belgium
Nationality:	Belgian
Email:	Benjamin.Buysschaert@ugent.be
	benjaminbuysschaert@gmail.com
Phone:	+32 494 57 12 46
Affiliation:	Center for microbial ecology and technology (CMET), Faculty of bioscience engineering, Ghent University
Workplace address:	Coupure Links 653, 9000 Gent

Education and working experience

2014 - 2018 **PhD Researcher**

Ghent University, Gent, Belgium
Center for Microbial Ecology and Technology (CMET)
PhD researcher with the IWT-project grant no. SB-131370

"*Single-cell optical fingerprinting for microbial community characterization*"

Promotors: Prof. dr. ir. Nico Boon
Prof. dr. ir. Bart De Gusseme

Apr. 2017 – Aug. 2017 **Visiting Scientist**

DC water and sewer authority, Washington DC, USA

Advisors: Dr. Haydée De Clippeleir and Dr. Sudhir Murthi

July 2013 – Sep. 2013 **Internship**

Institute of Plant Biotechnology (IPK), Gatersleben, Germany

Project: "*Impact of the PGPR* Raoultella terrigena *on plant growth and root architecture".*

Promotor: Prof. dr. Nicolaus Von Wirén

2008 - 2013 **Bachelor and Master of Bioscience Engineering, Cell & Gene Biotechnology**

Ghent University, Gent, Belgium

Master thesis: "*Development of the Skin Community Interaction (SCIN) Model to examine skin odour formation".*

Promotor: Prof. dr. ir. Nico Boon

Publications

Articles on ISI Web of Science (published, A1)

Buysschaert,B., Kerckhof, F.-M., Vandamme, P., De Baets, B. & Boon, N. (2017). Flow cytometric fingerprinting for strain discrimination. *Cytometry part A.*

Besmer, M.D., Sigrist, J.A., Props, R., **Buysschaert, B**., Mao, G.N., Boon, N. and Hammes, F. (2017) Laboratory-Scale Simulation and Real-Time Tracking of a Microbial Contamination Event and Subsequent Shock-Chlorination in Drinking Water. Frontiers in microbiology 8.

Cagnetta, C., D'Haese, A., Coma, M., Props, R., **Buysschaert, B.**, Verliefde, A.R.D. and Rabaey, K. (2017) Increased carboxylate production in high-rate activated A-sludge by forward osmosis thickening. Chem Eng J 312, 68-78.

Van Nevel S, **Buysschaert B**, De Roy K, De Gusseme B, Clement L, Boon N (2017). Flow cytometry for immediate follow-up of drinking water networks after maintenance *Water research*. 111: 66-73

Buysschaert B, Byloos B, Van Houdt R, Leys N, Boon N (2016). Reevaluating multicolor flow cytometry to assess microbial viability. *Applied microbiology and biotechnology*. 100:9037–9051

Van Nevel S, **Buysschaert B**, De Gusseme B, Boon N (2016). Flow cytometric examination of bacterial growth in a local drinking water network. *Water and Environment Journal*. 30: 167-176

Sieprath, T., Corne, T.D.J., Nooteboom, M., Grootaert, C., Rajkovic, A., **Buysschaert, B.**, Robijns, J., Broers, J.L.V., Ramaekers, F.C.S., Koopman, W.J.H., Willems, P.H.G.M. and De Vos, W.H. (2015) Sustained accumulation of prelamin A and depletion of lamin A/C both cause oxidative stress and mitochondrial dysfunction but induce different cell fates. Nucleus-Phila 6, 236-246.

Callewaert, C., **Buysschaert, B.**, Vossen, E., Fievez, V., Van de Wiele, T. and Boon, N. (2014) Artificial sweat composition to grow and sustain a mixed human axillary microbiome. Journal of Microbiological Methods 103, 6-8.

Articles intended for peer review (A1)

Buysschaert, B., Khalenkow, D., García-Timermans, C., Skirtach, A., Boon, N. Single cell Raman spectroscopy for genotypic and phenotypic differentiation. *Submitted,* Environmental mmicrobiology

Buysschaert, B., Props, R., De Mey, M., Boon, N. Using phenotypic diversity for microbial bioreactor monitoring. *In preparation*

Heyse, J., **Buysschaert, B.,** Props, R., Rubbens P., Skirtach A., Boon, N. Phenotypic plasticity in cocultures. *In preparation*

Buysschaert, B., Minne, M., Boon, N., De Gusseme, B. Substrata define biofilm ecology in water distribution systems . *In preparation*

Buysschaert, B., Vermijs, L., Baetens V., Naka, A., Boon, N., De Gusseme, B. Online flow cytometric monitoring for microbial quality assessment in a full-scale water treatment plant. *In preparation*

Buysschaert, B., Props, R., Vermijs, L., Baetens V., Naka, A., Boon, N., De Gusseme, B. Nutrient levels determine microbial community composition and dynamics. *In preparation*

Non peer-reviewed publications

Starckx, S., **Buysschaert, B.** "Wat leeft er in dat water?", Mens en Molecule, April 2017

Presentations at (inter)national conferences

The presenting author is underlined

García-Timermans, C., Buysschaert, B., Kerckhof, F.-M., Rubbens, P., Skirtach, A., & Boon, N. (2017). phenotypes with Raman spectroscopy. Presented at the FT-IR Workshop.

Buysschaert,B**.,** Kerckhof, F.-M., Vandamme, P., De Baets, B. & Boon, N. (2016). Flow cytometric fingerprinting for strain discrimination. Presented at 6th international conference on the analysis of microbial cells at the single cell level

Buysschaert B, Byloos B, Lerno, A.-S., Props, R., Van Houdt R, Leys N, Boon N (2015). Multicolor flow cytometry to assess microbial viability. Presented at 4th international How dead is dead conference (HDID-conference)

Callewaert, C., Buysschaert, B., Vossen, E., Van Gele, M., Van de Wiele, T., & Boon, N. (2013). The SCIN simulation: a novel in vitro technique to study microbial colonization in the axillary region. *Applied Biological Sciences, 18th National symposium, Abstracts*. Presented at the 18th National symposium on Applied Biological Sciences (NSABS 2013).

Callewaert, C., Buysschaert, B., Van Gele, M., Van de Wiele, T., & Boon, N. (2012). The SCIN simulation: a novel in vitro technique to study the odour generation of microbial colonization on the skin. *Belgian Society for Cell and Developmental Biology, Meeting abstracts*. Presented at the Belgian Society for Cell and Developmental Biology (BSCDB) Fall meeting 2012 : Epidermal cell biology.

Scientific awards

Ernest Dubois prize for innovative research about water (2016), King Baudouin Foundation, Belgium

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

L'homme se découvre quand il se mesure à l'obstacle

Antoine de Saint-Exupéry, Terre des hommes

Doing a PhD is more than four years of scientific research. It is a personal adventure and a challenge. Luckily, it wasn't a lonely adventure for me and I am grateful to many people for their guidance, support, and friendship.

First, I would like to thank the examination committee for critically reviewing and for improving the manuscript. Thank you also for the constructive feedback on how to make my thesis more accessible to a broader audience.

Verder wens ik zeker ook mijn beide promotoren te bedanken voor hun niet aflatende steun. Zonder jullie had ik dit nooit verwezenlijkt. Nico, nog voor ik effectief gestart was heb je me enorm geholpen om tijdens mijn buitenlandse stage het IWT projectvoorstel te schrijven. Ook tijdens de laatste vier jaar stond je deur steeds open voor mij voor allerhande grote en kleine problemen. Je positieve en opgewekte persoonlijkheid is enorm aanstekelijk waardoor ik telkens opnieuw weer met goesting er tegenaan ging. Bart, ook al ben je pas later officieel mijn promotor geworden, ook jij was er van in het begin bij met de experimenten in Brakel en Beersel samen met Sam. Jouw praktische input op mijn doctoraat is een enorme meerwaarde geweest. Ik apprecieer enorm jouw enthousiasme en dat je ondanks je drukke agenda zoveel tijd voor me vrij hebt gemaakt. Nico, Bart, het was werkelijk een plezier om met jullie beide samen te werken en ik kijk ernaar uit om in de toekomst verder met jullie te werken.

Naast de promotoren zijn er nog andere mensen die ik wil bedanken. Karen, Sam, dank u om mij te helpen bij de voorbereidingen van mijn IWT beurs, zonder jullie input zou ik die zeker nooit behaald hebben. Sam, de vorming die ik van jou heb gekregen in de eerste maanden van mijn doctoraat hebben mij een enorm vooruit geholpen. Zonder jouw hulp in het prille begin zou ik nooit zoveel hebben verwezenlijkt. Bo, we hebben samen uren doorgebracht in de kelder in het FCM labo om onze eerste paper te schrijven. Het was een lastig werkje maar ik hou er vooral goede herinneringen aan dankzij jou. FM, Ruben, ik heb ooit als grap gezegd dat ik ging vragen om jullie tot co-promotoren te benoemen maar dat illustreert hoeveel ik aan jullie te danken heb. Er is geen hoofdstuk of paper die ik heb geschreven waar jullie geen input hebben gegeven. Ik bewonder jullie kennis en ik dank jullie voor jullie onbaatzuchtig samenwerking. Cristina, Agathi, I am happy to have had the chance to tutor you as junior PhD students. It was a pleasure to do so and it did not take much time before I started to learn from you. I thank you also very much for helping me out and for the

good moments in the lab. Dmitry, thank you very much to introduce me to Raman spectroscopy and for helping me with my Raman experiments. Peter, jou wil ik bedanken voor je wiskundige input dat de kwaliteit van mijn werk enorm ten goede kwam.

Ik wil ook uitdrukkelijk het secretariaat en de ATP bedanken voor de buitengewone hulp. Jullie zijn samen met de proffen het hart van het labo en onontbeerlijk voor het slagen van ons werk. Christine, Regine, Sarah, dank u om geduldig mijn duizenden vragen te beantwoorden en voor de aangename babbels. Tim, Jana, Mike, Siska, Greet, Renée, Tom, ik dank jullie voor de toffe babbels in het labo en jullie dat ik steeds bij jullie terecht kon voor advies, praktische problemen, hulp in het labo, opstellingen bouwen, *etc*.

Om dit onderzoek uit te voeren heb ik ook de hulp gekregen van geweldige masterstudentes. Ik wil daarom ook Jasmine, Lotte, Marjolein, Valerie en An-Sofie bedanken. Jullie waren geweldige studentes en ik hou enkel goede herinneringen aan onze samenwerking. Jasmine, je bent nu zelf begonnen aan een doctoraat, ik weet dat je dat heel goed zal doen en ik kijk uit naar jouw verdediging.

Op het einde van mijn doctoraat heb ik de kans gekregen om naar de V.S. te gaan. Haydée, Siegfried, dankzij jullie heb ik een onvergetelijke en plezante ervaring in de Washington DC gehad. Ik had me geen betere afsluiter van mijn doctoraat kunnen voorstellen. Ook in België heb ik de hulp gehad van verschillende mensen uit de drinkwaterwereld. Hiervoor wil ik Katrien De Maeyer, Koen Huysman, Koen Joris, Kristin Van Hecke, en François Van De Velde bedanken.

Colleagues are an essential part of our work life and I would also like to thank the whole CMET team for their contribution to this work. Every time I was completely stuck at work, there was someone who could help me out with a solution or advice, and in case there was none, a good joke or a nice chat to lift my spirits. The good atmosphere at work has made it a real pleasure to come to work every day. In time I learned to know many of you personally and I consider the friendships that I gained after four years as the best souvenir from my PhD. Thank you!

I'd like also to thank all my friends who supported me and who gave me the opportunity to be leave work behind me for a moment. There are too many people and too many occasions for which I'm grateful that naming them here is an impossible task. Thank you to everyone who has been there for me and with whom I shared the good moments in life.

Pour terminer, j'aimerais remercier ma famille, grande et petite, qui a toujours été là pour moi. Je remercie en particulier Bonne-Mamy chez qui m'a aidé à étudier. Bonne-Mamy, merci d'être une grand-mère extraordinaire, c'est entre autre grâce à toi que j'ai pris goût aux sciences et aux études. Je tiens bien sûr aussi à remercier mes sœurs et mon beau-frère. Géraldine, Pauline, Pierre-Henri, je peux toujours compter sur vous et depuis que je suis petit vous avez veillé sur moi. J'admire votre intelligence et vous êtes un exemple pour moi. Que je suis arrivé si loin, est aussi partiellement grâce à vous. Papy et Mamy, comment puis-je exprimer ma gratitude pour tout ce que vous m'avez donné? Je pense que mon amour pour la nature et les sciences ont été nourri pendant mon enfance en lisant chaque soir avec vous les livres de 'Kinderen ontdekken'. Merci de m'avoir donné tout ce dont j'ai eu besoin et d'avoir fait de moi l'homme que je suis. Je suis triste que Mamy n'est plus là pour partager ce moment avec nous mais je me console avec les bons souvenirs et en me rappelant ce qu'elle m'a appris. Papy, je suis très reconnaissant de tout ce que tu fais encore pour moi. Sans ton soutien et ta sagesse je n'aurai pas pu arriver si loin! Merci !

Benjamin, Januari 2018