JYU DISSERTATIONS 418

Lara Dutra

Expanding the Toolbox for Molecular Microbial Studies

From Communities to Single Cells

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS
AND SCIENCE

JYU DISSERTATIONS 418

Lara Dutra

Expanding the Toolbox for Molecular Microbial Studies

From Communities to Single Cells

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi syyskuun 2. päivänä 2021 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
on September 2, 2021 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2021

Editors
Varpu Marjomäki
Department of Biological and Environmental Sciences, University of Jyväskylä
Ville Korkiakangas
Open Science Centre, University of Jyväskylä

ISBN 978-951-39-8799-2 (PDF)
URN:ISBN:978-951-39-8799-2
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-8799-2

ABSTRACT

Dutra, Lara
Expanding the toolbox for molecular microbial studies: from communities to single cells
Jyväskylä: University of Jyväskylä, 2021, 50 p.
(JYU Dissertations
ISSN 2489-9003; 418)
ISBN 978-951-39-8799-2 (PDF)
Diss.

Recently, the molecular revolution and nanotechnologies have provided tools for expanding the knowledge of microorganisms. This thesis had its starting point in utilizing existing methods to identify specific microbes. Polymerase chain reaction (PCR)-based assays were employed for the detection of zoonotic virus (*Orthopoxvirus*) in capybara's (*Hydrochoerus hydrochaeris*) faecal samples in an attempt to examine the spread of viruses and a possible route for viral outbreaks. Here, the existing method was sufficient to answer the question and find that viral genetic material is indeed present in these stool samples. The established PCR and sequencing-based methods were also applied to study bacteria responsible for the removal of highly chlorinated phenols from contaminated sites. Despite detecting bacterial species from samples *in situ* or from microcosm experiments, it became evident that the used method did not clarify which microbes specifically were responsible for bioremediation. The abundance of pentachlorophenol hydroxylase gene (*pcpB*), the major gene involved in the bioremediation process, was detected with quantitative PCR and the overall bacterial diversity with sequencing using the universal marker – the 16S rRNA gene. Yet, the real players of the process remained unknown. This led to the quest to specify the microorganisms that are carrying a specific gene. As a result, a novel method based on droplet microfluidics technology was developed to investigate microbes at single-cell resolution. Within these droplets, PCR is used to fuse 16S rRNA and a second gene of interest into concatemers. Through sequencing, the second gene can be associated with specific bacterial species. To make the method reliable, several individual steps in the procedure required both engineering and careful method validation. These included the elimination of errors stemming from the existence of DNA outside of the cells, the efficient breaking of the emulsion and hence the collection of DNA, the blocking of unwanted PCR amplification that was confounding the sequencing results, and ensuring that the sample was studied at single-cell resolution. The established methods may be applied to a variety of research where it is necessary to identify the microorganism carriers of specific genetically encoded functions, and therefore highlight the potential actors in the environmental processes.

Keywords: Methods in microbiology; NGS; PCR; single-cell technology; methods in microbiology.

Lara Dutra, University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

TIIVISTELMÄ

Dutra, Lara
Molekyylimikrobiologian tutkimusmenetelmien laajentaminen: yhteisöistä yksittäisiin soluihin
Jyväskylä: Jyväskylän yliopisto, 2021, 50 p.
(JYU Dissertations
ISSN 2489-9003; 418)
ISBN 978-951-39-8799-2 (PDF)
Diss.

Viime aikoina molekyylivallankumous ja nanoteknologia ovat tarjonneet työkaluja mikro-organismeja koskevan tiedon laajentamiseen. Myös tämän tutkielman lähtökohta oli hyödyntää olemassa olevia menetelmiä tiettyjen mikrobien tunnistamiseen. Polymeraasiketjureaktioon (PCR) perustuvia määrityksiä käytettiin zoonoottisen viruksen (*Orthopoxvirus)* havaitsemiseen vesisian (*Hydrochoerus hydrochaeris*) ulostenäytteistä, jonka avulla pyrittiin löytämään virustautiepidemioiden aiheuttajien mahdollisia leviämisreittejä. Olemassa olevat menetelmät soveltuivat virusten geneettisen materiaalin löytämiseen ulostenäytteistä. Vakiintuneita PCR- ja sekvensointimenetelmiä sovellettiin myös tutkittaessa biohajotukseen kykeneviä bakteereita kloorifenoleillä saastuneessa pohjavedessä. Huolimatta bakteerilajien havaitsemisesta saastuneissa aineksissa ja mikrokosmoskokeissa, kävi selväksi, etteivät laajasti käytettävissä olevat menetelmät selvittäneet todellisia biopuhdistuksesta vastanneiden mikrobien identiteettiä. Funktionaalisen geenin sisältävää DNA:ta havaittiin kvantitatiivisella PCR-menetelmällä ja bakteerien kokonaisdiversiteetti määritettiin sekvensoimalla yleinen merkkigeeni – 16S rRNA geeni. Silti prosessin todelliset toimijat eli toiminnan kannalta kriittisiä geenejä kantavat mikro-organismit säilyivät tuntemattomina. Tutkimuksessa siirryttiin kehittämään uutta pisaramikrofluidistiikka-tekniikkaan perustuvaa menetelmää, joka mahdollistaa mikrobien geeniyhdistelmien tutkimuksen yksittäisen solun tasolla. Öljyn erottamissa pisaroissa PCR-menetelmää käytetään sulauttamaan ja monistamaan 16S rRNA sekä funktionaalisesta geenistä syntyviä konkatemeereja. Sekvensoinnin avulla funktionaalinen geeni voidaan liittää tunnettuihin bakteeritaksoneihin. Menetelmän luotettavuuden varmistamiseksi useat yksittäiset vaiheet vaativat suunnittelua ja huolellista hienosäätöä. Näihin sisältyivät solujen ulkopuolisesta DNA:sta aiheutuvien virhelähteiden poistaminen, emulsion tehokas rikkominen ja DNA:n eristys, sekvensointituloksia vääristävän ei-toivotun PCR-monistuksen estäminen, sekä solutasoisen erottelukyvyn varmistaminen. Tätä optimoitua menetelmää voidaan soveltaa tutkimuksiin, joissa halutaan tunnistaa mikro-organismit, joiden merkitys on suuri tietyissä geneettisesti koodatuissa toiminnoissa.

Avainsanat: Mikrobiologian menetelmät; NGS; PCR; yksisoluteknologia.

Lara Dutra, Jyväskylän yliopisto, Bio- ja ympäristötieteiden laitos PL 35, 40014 Jyväskylän yliopisto

Author's address Lara Dutra
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
laambros@student.jyu.fi

Supervisors Professor Marja Tiirola
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Senior lecturer Matti Jalasvuori
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Postdoctoral Researcher Reetta Penttinen
Department of Biology
FI-20014 University of Turku
Finland

Reviewers Associate professor Tarja Pitkänen
Department of Food Hygiene and Environmental Health
P.O. Box 66 (Agnes Sjöbergin katu 2)
FI-00014 University of Helsinki
Finland

Senior lecturer Manu Tamminen
Department of Biology
FI-20014 University of Turku
Finland

Opponent Professor Stefan Bertilsson
Department of Aquatic Sciences and Assessment
P.O. Box 7070,
SE-750 07 Uppsala
Sweden

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following original papers, which will be referred to in the text by their Roman numerals I–IV.

I have participated in a substantial manner in the development of the research that resulted in the papers listed below. In paper I, the original idea was a collaboration between Gabriel Almeida, Giliane Trindade and me. The experimental design and result analysis were done by GA, with whom I share the first authorship, and me. GA, Jonatas Abrahao and I conducted the field work. Graziele Oliveira and I have done the laboratory work. I prepared the first draft of the manuscript. All authors have contributed to the final version of the manuscript. In paper II, I was involved in all the steps of the microcosm experiment: sample collection (with Anu Mikkonen), microcosm set-up (with AM, Pauliina Salmi, Marja Tiirola), data collection along the incubation period (with AM, PS) and molecular biology work (DNA extraction, qPCR and NGS library preparation were all done by me). AM prepared the final analysis, combination of results and the manuscript. The original idea in paper III came from MT and me. Experimental design, laboratory work and analysis were done by me. Ole Franz helped in laboratory work, and Veli-Mikko Puupponen helped with interpretation of part of the data. I wrote the manuscript to which all authors provided comments. The original idea for paper IV came from MT. The study design was done by me in collaboration with MT, Matti Jalasvuori and Reetta Penttinen. The laboratory work was done by RP and me, with help from OF, PS and Kimi Nurminen. RP and I analysed the results. The manuscript was written in collaboration of MJ, RP and me.

I Lara Ambrosio Leal Dutra, Gabriel Magno de Freitas Almeida, Graziele Pereira Oliveira, Jonatas Santos Abrahao, Erna Geessien Kroon & Giliane de Souza Trindade 2017. Molecular evidence of *Orthopoxvirus* DNA in capybara (*Hydrochoerus hydrochaeris*) stool samples. *Archives of Virology* 162: 439–448.

II Anu Mikkonen, Kati Yläranta, Marja Tiirola, Lara Ambrosio Leal Dutra, Pauliina Salmi, Martin Romantschuk, Shelley Copley, Jukka Ikäheimo & Aki Sinkkonen 2018. Successful aerobic bioremediation of groundwater contaminated with higher chlorinated phenols by indigenous degrader bacteria. *Water Research* 138: 118–128.

III Lara Dutra, Ole Franz, Veli-Mikko Puupponen & Marja Tiirola 2020. DNA recovery from Droplet Digital™ PCR emulsions using liquid nitrogen. *BioTechniques* 69: 450–454.

IV Lara Dutra, Matti Jalasvuori, Ole Franz, Paulina Salmi, Kimi Nurminen, Marja Tiirola & Reetta Penttinen 2021. Single-cell resolution genetic association analysis of heterogeneous bacterial communities by utilizing droplet digital PCR. Manuscript.

ABBREVIATIONS

ARG	Antibiotic resistance gene
bla	Betalactamase gene
bp	Base pairs
CFU	Colony forming unit
CP	Chlorophenol
ddNTP	Dideoxy-nucleotide triphosphate
ddPCR	Droplet digital PCR
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
dPCR	Digital polymerase chain reaction
dsDNA	Double stranded DNA
exDNA	External DNA
HCP	Highly chlorinated phenol
HGT	Horizontal gene transfer
kHz	Kilohertz
LN2	Liquid nitrogen
MAG	Metagenomics assembled genome
MDA	Multiple displacement amplification
mRNA	Messenger ribonucleic acid
NGS	Next generation sequencing
nm	Nanometer
OPV	*Orthopoxvirus*
OTU	Operational taxonomic unit
pcpB	Pentachlorophenol hydroxylase gene
PCR	Polymerase chain reaction
PFO	1H, 1H, 2H, 2H-Perfluoro-1-Octanol
pfu	Plaque forming unit
PMA	Propidium monoazide
qPCR	Quantitative polymerase chain reaction
RFU	Relative fluorescence unit
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SAG	Single amplified genome
SBS	Sequencing by synthesis
SMRT	Single molecule real-time
SMS	Single molecule sequencing
v/v	Volume per volume
VACV	*Vaccinia virus*
VACV-BR	*Vaccinia virus* Brazilian strain
VACV-WR	*Vaccinia virus* Western Reserve strain
w/v	Weight per volume
µl	Microliter
µM	Micro molar

From first, I made and mixed, and to think not I thought. I didn't have the deadlines. I lived pulling difficult from difficult, fish alive on griddle: who lives asp'rously, no fantasies. But, now, fete of fate to me comes, and sans little disquietudes, I'm from creaky net. And myself invented in this like, of to speculate ideas.

Guimarães Rosa in Grande Sertão: Veredas
(translation F. W. Martinez, 2013)

1 INTRODUCTION

1.1 Microbes, the endless enigma

Microorganisms are major players in the processes that shape life on earth, being found in all kinds of environments. There are 2.5 x 10^{30} bacterial and archaeal cells – the prokaryote microbes – on Earth (Kallmeyer *et al.* 2012), composing 15% of the planet's biomass (Bar-On *et al.* 2018). Meanwhile, viral abundance in the world is estimated to surpass the magnitude of 10^{30} (Suttle 2007). These microorganisms carry out processes that are fundamental to supporting the life of other organisms. They can be found in soil, oceans and fresh water as well as in sediments, where their metabolism orchestrates pathways for nutrient cycling (Whitman *et al.* 1998; Rojo *et al.* 2017). They can also be found in close association with animals and plants, being part of either the healthy microbiota or a cause of disease (Rojo *et al.* 2017). The spread of microbes on the planet is so vast that they can be found even in the most inhospitable environments, such as extremely acidic caves (Johnson 1998) or around submarine volcanoes (Danovaro *et al.* 2017). Yet today, most of the diversity of microorganisms is unknown. They can be considered to make up the so-called biological dark matter, in reference to the unknown ('dark') matter of the universe (Thompson *et al.* 2017).

Discovery, identification, and the elucidation of microorganisms' role in the environment has faced many challenges that have called for the development of new tools and ideas (Fig. 1). The discovery of microbes was prompted by the intellectual and technological advances of the scientific revolution. The rudimentary optical microscopes of the 17th century were the first tools to reveal the – at the time – unseen world of microbes. Two centuries later, the introduction of cultivation techniques contributed to a period of remarkable advances in the study of microbial organisms. Rigorous cultivation, together with the establishment of better tools for observation, built the context for key discoveries of the nature of microbes and their roles in human diseases and other processes. Louis Pasteur used such techniques to demonstrate that microbial growth in sterile solutions is due to contamination and not spontaneous generation,

therefore refuting the view that life arises in matter simply by being exposed to the air and supporting the germ theory of disease. Soon after, Robert Koch used growth experimentation to show that specific infectious diseases were caused by specific microorganisms. This period marks the beginning of the microbiology, the science dedicated to the study of microscopic organisms.

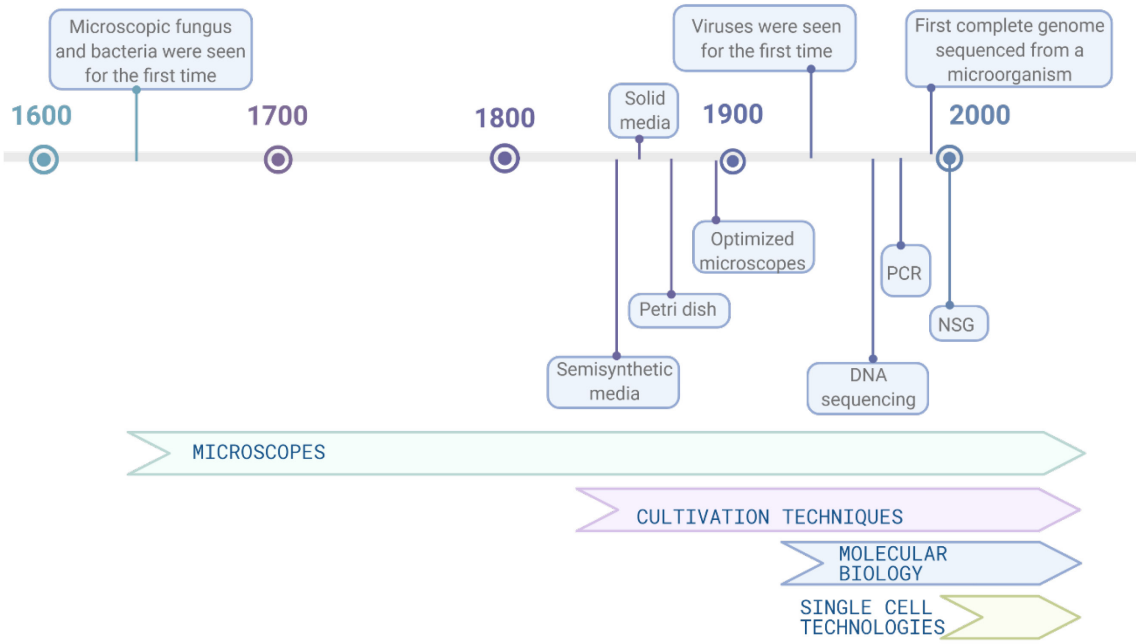


FIGURE 1. Timeline of development of methods in microbiology. Created with Biorender.

The second major advance for the study of microbes came with the development of molecular biology and methods for studying nucleic acids and proteins. The observation of similarities in nucleic acid sequences led to the definition of a universal taxonomic marker for prokaryotes, and therefore to the reorganization of taxonomic groups for all cellular life (Fox *et al.* 1977; Woese and Fox 1977; Albers *et al.* 2013). The use of molecular techniques such as cloning and polymerase chain reaction (PCR), and the identification of a molecular marker, ribosomal RNA (rRNA) genes, to sort out bacteria and archaea into phylogenetic groups extended the boundaries of microbiology. Eventually, studies on the molecular level revealed that the majority of microbes cannot yet be grown under laboratory conditions and in currently known media compositions. This naturally poses a limitation for laboratory techniques (Amann *et al.* 1995) and, consequently, to our ability to precisely identify microbes and their properties. Therefore, highlighting the need for alternative methods to study microbes and their properties.

While molecular methods and the use of rRNA genes as a marker have improved our ability to identify bacteria, the absence of a universal marker in viruses has imposed limitations to explore the viral diversity. However, in the past few decades, advances in sequencing methods have provided tools to finally explore phylogenetic relationships and the overall diversity of this group of microorganisms. Moreover, new technologies such as microfluidics and single

molecule sequencing (SMS) allow applications of molecular methods at single-cell and single molecule level, allowing us to zoom our research into the cellular level, hence unveiling the interactions between microbes and the role of each microorganism in their natural environments (Marcy *et al.* 2007; Gawad *et al.* 2016).

1.2 The rise of molecular methods for studying DNA

Molecular methods emerged in mid-20th century after a series of discoveries around DNA and its role in biological processes. Early in the 1950s, independent studies demonstrated that DNA carried the genetic code. Experiments with *Streptococcus pneumoniae* showed that DNA was the cellular fraction capable of inducing 'transformation' of mild strains into virulent ones (Avery *et al.* 1944). Meanwhile, the investigation of phage infection in *Escherichia coli* showed that DNA was the main molecule from the viral particle to be released into the bacterial cell and, therefore, to have some role in the production of progeny (Hershey and Chase 1952). Soon after, in 1953, the structure of DNA was determined (Watson and Crick 1953). The proposal of a double helix structure was made based on the x-ray crystallography study conducted by Rosalind Franklin (Zallen 2003). In 1958, DNA replication was shown to be a semiconservative process, *i.e.* during replication one strand of the parental double-stranded DNA (dsDNA) will be present in each newly generated copy (Meselson and Stahl 1958a, b). DNA polymerase was identified as the molecule responsible for incorporating nucleotides in the growing DNA chain (Lehman *et al.* 1958). By the end of the decade, the mechanism for DNA replication was characterized.

Early in the 1960s, the genetic code was cracked, and the roles of nucleic acids and proteins in biochemical process were identified. In 1961, Crick and collaborators introduced a study showing that sequences of three nucleotides in the DNA chain encode for one amino acid in a protein (Crick *et al.* 1961). Moreover, the study showed that the code was degenerated (i.e., more than one combination of three nucleotides encodes for the same amino acid) and non-overlapping (i.e., in a nucleotide chain, the sequence for the next code starts always after three bases) (Crick *et al.* 1961; Yanofsky 2007). In that same year, two groups identified messenger RNA (mRNA) to be the molecule responsible for carrying genetic information from DNA to the machinery that produces proteins (Cobb 2015). These advances in knowledge led to a concept that explained the flow of genetic information from DNA to enzymes – from the code to the biological process – the central dogma of biology. Soon after, it was proposed that the genetic code is universal, i.e., the same code exists in all organisms and therefore small variations between carriers of genetic material could provide the basis for taxonomic classification (Hinegardner and Engelberg 1963).

These advances in molecular biology and the understanding of nucleic acids and proteins allowed biology to be approached from a new perspective. For

instance, in microbiology, the analysis of the ribosomal RNA (rRNA) molecule sequence was the key for establishing the modern classification of prokaryotes. In the 1970s, a collection of studies was published suggesting that rRNA, as a well-conserved molecule among microbes, could be used as a marker to trace phylogenetic relationships between groups of microbes (Fox *et al.* 1977; Woese and Fox 1977; Albers *et al.* 2013). It was a milestone in the development of microbiology and provided insights into the evolutionary history of life by defining the three major kingdoms (Albers *et al.* 2013; Zhulin 2016). Analysis based on 16S and 18S rRNA sequences revealed that prokaryote microbes are divided in two major groups, bacteria and archaea, and that archaea are more closely related to eukaryotes than they are to bacteria.

1.2.1 Polymerase chain reaction: basics and improvements

Polymerase chain reaction (PCR) is a molecular technique that allows the *in vitro* generation of numerous copies of a specific DNA sequence. It relies on the enzymatic reaction carried out by DNA polymerase to replicate a DNA fragment from a template molecule (Saiki *et al.* 1985, 1988; Mullis and Faloona 1987). The reaction is prepared with the enzyme (and cofactors), deoxynucleotide triphosphates (dNTPs), oligonucleotides and the template molecule. Changes in temperature prompt the steps of the reaction: (i) dissociation of the dsDNA molecule into single strands at high temperature, (ii) oligonucleotides annealing to the specific region at specific annealing temperature, (iii) polymerization of the elongating strand at DNA polymerase optimal temperature (72 °C). After each cycle, the number of amplified DNA fragments doubles, and after, e.g., 40 cycles, millions of molecules have been generated. The technique was created in 1985 to solve the need for a sufficient amount of starting material for DNA analysis in samples that initially contained very few copies of the DNA target. At that time, DNA sequencing was not yet available as a routine procedure. Moreover, the detection of specific DNA sequences relied on methods that required a considerable amount of DNA. Therefore, researchers devised a method to generate copies of the specific fragment containing the specific variation that was being studied (Saiki *et al.* 1985).

New possibilities for applying the method were recognized very soon after its introduction. As early as in 1987, the PCR was used for diagnosing HIV (Kwok *et al.* 1987; Guatelli *et al.* 1989). Early in 1990s, the PCR was introduced to the microbial phylogenetic studies. The amplification and analysis of the DNA sequence encoding the rRNA molecule was suggested as an alternative to analysis of rRNA molecule sequence (Chen *et al.* 1989; Böttger 1989; Wilson *et al.* 1990; Weisburg *et al.* 1991). By using PCR, the DNA molecules encoding the target sequence could be amplified using cultivation-free methods and therefore generating enough material for analyses. Subsequently, the analysis of rRNA genes became a standard method to study diversity (Thompson *et al.* 2017; Hugenholtz *et al.* 2021).

In addition to using PCR for diagnostics and studying diversity, improvements to allow quantification further increased its importance in

microbiological studies. Guatelli *et al.* (1989) had earlier suggested that PCR could be employed for quantitative purposes. They argued that comparing the frequency of viral and host endogenous genes (those with constant copy numbers) could be useful for studying pathogenesis of HIV. Quantitative or real-time PCR (qPCR) appeared in the 1990s and provided a means for quantifying the amount of DNA in a given sample. In 1993, Higuchi *et al.* (1993) introduced a method to monitor DNA accumulation over the PCR cycles. The fluorescence of ethidium bromide was monitored by microcameras and used to reveal dsDNA. In each cycle, the number of copies of the target DNA would duplicate. And by the end of each cycle, the fluorescent signal indicating the presence of dsDNA would increase. The increasing amount of dsDNA over the cycles would reveal the accumulation of copies of DNA. Later, in 1996, the use of molecular probes and a more consistent description of the mechanism for quantification was published (Gibson *et al.* 1996; Heid *et al.* 1996). Thereafter, modifications have been made to improve the qPCR assay, and specialized instruments were built to automate the process.

Recently, a novel PCR-based technology emerged to provide a more precise quantitative analysis. The principle of the digital PCR (dPCR) was adapted from the need to detect single DNA molecules individually (not in a bulk as in PCR or qPCR). The technology has its roots back in 1992 when Sykes and collaborators elaborated a method using principles of limiting dilution and target distribution according to Poisson distribution to detect and quantify the molecules of target DNA in PCR (Sykes *et al.* 1992). But because it was more laborious and provided lower throughput than qPCR – the competing technology – it was forgotten. Later, in 1999, the idea of quantification by detecting single DNA molecules re-emerged in work by Vogelstein and Kinzler. By using multi-well PCR plates, they provided reactions with up to 384 partitions, and simplified the analysis by adopting DNA-binding dye to detect amplification (Vogelstein and Kinzler 1999). In the last two decades, automation has improved the partitioning and analysis of dPCR. The use of microfluidics technology enabled partitioning into thousands of partitions while shrinking the volume of reactions.

The droplet microfluidic system Droplet Digital PCR™ (ddPCR™) is a commercial droplet-based dPCR platform. It relies on droplet microfluidics to generate numerous small partitions from a bulk PCR reaction (Hindson *et al.* 2011; Pinheiro *et al.* 2012). In this system, each droplet works as an independent reaction chamber where DNA amplification occurs. In the partitioning process, the target molecules are randomly dispersed in the original volume and trapped into droplets in a stochastic manner. The probability of confining single target molecules inside droplets is higher in more diluted samples. Ideally, it is expected that just a small fraction of the droplets in a sample will carry the target molecule, and within this group the majority will have a single target. Thus, Poisson distribution is used to quantify the target. The detection of the target molecule is revealed by the fluorescence emitted by DNA binding dye or a probe. The fluorescence level of each droplet is measured using two fluorescence channels and expressed in relative fluorescence units (RFU). In droplets where DNA is amplified, the end point RFU is higher than in droplets where

amplification does not occur. Therefore, droplets can be assigned as positive or negative in terms of whether or not they contain the target. The quantification of the targets in the studied sample is based on the number of positive droplets among the total number of droplets.

The qPCR and dPCR methods have similarities, but also retain many differences. Both methods are used to quantify the target DNA and employ fluorescent DNA dyes and probes to detect accumulation of DNA along the cycles. However, in qPCR the fluorescence is measured along the exponential phase, while in ddPCR the fluorescence is measured at the endpoint. Moreover, in qPCR the quantification is based on the comparison of the fluorescence relative to a standard curve, where the number of copies of the target molecule is known. In dPCR there is no need for standard curve. That is possible because in dPCR the PCR mix volume is compartmentalized in tiny wells in plates or droplets by microfluidics, and each compartment is expected to be either empty or contain a single starting molecule. Therefore, the fluorescence intensity is expected to differ considerably between compartments where a template molecule enables the accumulation of copies in comparison to compartments with no target molecule present. That allows the definition of a threshold line for the fluorescence level to distinguish the target-containing compartments from the empty ones.

1.2.2 DNA sequencing

The first generation of sequencing technologies for DNA was based on the analysis of DNA fragment size (Maxam and Gilbert 1977; Sanger *et al.* 1977; Heather and Chain 2016; Shendure *et al.* 2017). The approach was similar to what had been described for sequencing RNA in the previous years (Holley *et al.* 1965; Sanger *et al.* 1965). In 1968, the first DNA molecule was sequenced. A study on the genome of bacteriophage lambda demonstrated that sequencing can be successfully carried out from short DNA fragments with cohesive ends by using radiolabelled nucleotides (Wu and Kaiser 1968). The method employed the *in vitro* polymerization of DNA, and the reaction was carried out by DNA polymerase isolated from *Escherichia coli*. The cohesive ends of the phage DNA were used to prime the reaction. Growth of the DNA chain was marked by labelled nucleotides, and the reactions were carried out with one type of nucleotide at a time. Therefore, the incorporation of labelled nucleotides into the molecule was observed by analysing the resulting fragment lengths and hence the lengths allowed to deduce the sequence. In the following years, the procedure was facilitated by a series of optimizations. Synthetic oligonucleotides were introduced to initiate polymerization (Wu 1972; Sanger *et al.* 1973). The methods were simplified when polyacrylamide electrophoresis was adopted to analyse the fragments instead of electrophoresis, which was followed by chromatography (Sanger and Coulson 1975; Maxam and Gilbert 1977; Heather and Chain 2016).

In 1977, the introduction of two methods marked the beginning of the first generation of sequencing technologies. The Maxam-Gilbert sequencing method was introduced early in 1977 (Maxam and Gilbert 1977). The method was based

on the treatment of DNA molecules with chemical reagents in order to break the chain at specific bases. The fragmented lengths of molecules obtained after the treatment with chemicals of different 'cutting' profile were used to identify nucleotide position. Later that year, the dideoxy or chain-termination method was published. The method was based on the use of nucleotides with modified 3' ends, dideoxy-nucleotides (ddNTP's), to prevent the incorporation of the following nucleotide, therefore ending chain elongation (Sanger *et al.* 1977). The use of trace amounts of one type of dideoxy-nucleotides during *in vitro* polymerization would produce DNA fragments of variable lengths that always ended with a modified nucleotide. Therefore, chain-termination could reveal the position of this specific type of nucleotide in the chain. In this method, four reactions were needed to solve the sequence, one for each nucleotide.

The dideoxy method (also known as Sanger sequencing) became the method of choice in the following years after optimizations to the protocol (Heather and Chain 2016). Maxam and Gilbert's method was not taken much further, due to the technical demands (Heather and Chain 2016; Shendure *et al.* 2017). However, the detection of all four terminating nucleotides in a single reaction was enabled by replacing the radiolabelled ddNTPs with fluorescent ones. Moreover, the introduction of capillary electrophoresis improved the detection limit of the technique. The development of PCR and recombinant DNA technologies was also important for the success of the chain-termination method. These techniques facilitated the production of massive amounts of pure DNA with varying lengths and increased the power of sequencing into large-scale *de novo* sequencing. These techniques were employed to sequence large genomes despite the limitation of a 1 kb sequencing read length. Shotgun sequencing permitted researchers to reconstruct genomes by sequencing and assembling overlapping short fragments (Heather and Chain 2016). Ambitious genome projects attempted to optimize the throughput and automate the process (Shendure *et al.* 2017). By the end of 90s, the chain-termination method had been successfully used to sequence the genome of *Haemophilus influenza* (1995), *Saccharomyces cerevisiae* (1996) and *Caenorhabditis elegans* (1998). The speed of the process increased data yield that, in turn, propelled advances in computational analysis.

'Next-generation sequencing' (NGS) technology – or second generation of sequencing technology – was developed as an alternative to facilitate sequencing analysis (Shendure *et al.* 2017). NGS methods were based on sequencing-by-synthesis (SBS), i.e., the sequence was solved by identifying the nucleotides that were incorporated into the chain during polymerisation instead of analysing the lengths of DNA fragments. Another feature provided by the NGS methods was the possibility for multiplexing, i.e., the sequencing of distinct fragments could be done in the same reaction by introducing specific barcodes to the sequenced DNA strain. This resulted in higher sequencing capacity compared to the dideoxy method where just one sequence could be resolved per reaction. The availability and commercialization of different sequencing methods began in the early 2000s. Among the prominent NGS platforms were 454 pyrosequencing (discontinued in the last decade), Solexa/Illumina and Ion Torrent. In all of these

methods, copies of the target DNA are produced and attached to a solid surface to allow the spatial detection of nucleotide incorporation. In 454 pyrosequencing and Solexa/Illumina technologies, the incorporation is detected by emitted light or fluorescence, respectively, while in Ion Torrent the detection is observed by semiconductor sensors after the change of pH in the microenvironment. Several commercial platforms of NGS became available within a decade, and eventually sequencing became accessible for many types of studies and interests. NGS has been useful for metagenomics and amplicon-based sequencing for microbial communities and has revealed a diversity of microorganisms from various environments (Thompson *et al.* 2017; Straiton *et al.* 2019).

In the last two decades, a new generation of sequencing technologies has been in development. The third sequencing generation employs nanotechnology to resolve nucleotide sequences by single molecule sequencing (SMS), i.e., the sequencing is done from an individual molecule instead of analysing an assembly of clonal molecules. The first method was launched by Helicos LifeScience in the early 2000s and was based on microimaging a DNA molecule to detect fluorescence after incorporation of each base (Heather and Chain 2016). However, access to the method has been hindered by the costs. Recently, two novel methods have been launched commercially. PacBio is a single-molecule real-time sequencing (SMRT) technology that detects incorporation of each nucleotide in the growing DNA chain through fluorescence (Heather and Chain 2016). A single DNA molecule is confined in a very small compartment, and several compartments are analysed in parallel to improve the throughput. MinION has a different approach, so-called nanopore sequencing (Heather and Chain 2016). The detection of different nucleotides in a DNA chain is based on changes in the electric current while a DNA molecule flows through the nanopore of a bilipid layer. MinION devices have the advantage of portability and were, e.g., used to sequence the genome of the Ebola virus during the outbreak in 2015 (Quick *et al.* 2016).

1.3 Unravelling the world of microbes: from pure culture to microbial communities, and back to single cells

Historically, there is a need to identify microbes and understand what they do, *i.e.* their role in the environment. The commitment to elucidating the microbial world has motivated minds to develop and improve tools and methods by which microbes can be approached. These advances allow unlocking layers in the study of microorganisms and, therefore, expanding further the frontiers of microbiology. The toolbox for the microbial studies has expanded from visualization of microbial cells to the description of pure cultures, microbial communities, and recently, back to single cells.

Initial attempts to comprehend the nature of microorganisms, back in the 17th century, were ruled by observation of microbial cells with microscopes. Early

studies were limited to the collection of sizes and shapes of microorganisms together with the determination of the environments where these organisms could be found. It did not reveal much about microbes since microbial cells with similar shape could be found in different habitats, and there were no other methods available to differentiate microbes and their roles in these distinct places. Besides, the limitation of both instruments and human comprehension hindered the advances of microbial studies. However, the advances in optical theories, microscopic design, and the development of microscopy techniques, between the 17th and 20th century, supported the progress in the visualization of microbes (Lister 1830; Drews 2000, Sanderson 2019). Combinations of lenses and immersion techniques were developed using Abbe's principles in order to correct the distorted images and increase the resolution (Drews 2000). In parallel, Robert Koch and his colleagues introduced the staining techniques to observe microbiological specimens (Drews 2000; Summers 2009; Blevins and Bronze 2010). Techniques for affixing bacterial specimens to glass slides and the use of various dyes to enhance the contrast of bacterial cells were explored and permitted their identification. For example, methylene blue staining was developed to visualize tuberculosis bacilli in clinical specimens (Blevins and Bronze 2010). Gram-staining, for instance, was introduced by Hans Christian Gram in 1884. Gram-stained bacterial cells were stained using crystal violet to improve their contrast under the microscope. Later, safranin was adopted as a counter staining, and bacterial cells could then be classified based on their cell wall structure.

Nevertheless, the cornerstone for the progress of microbiology was the development of methods to manipulate, isolate and grow microbes in controlled environments. The employment of cultivation methods enabled more meticulous investigation of microbial diversity as well as the physiological and ecological aspects of the lives of microorganisms. The improvements of microscopes and the introduction of staining procedures to aid the visualization of microbes happened in parallel, and were combined with cultivation methods. Phenotypical and physiological characteristics of microorganisms were catalogued and used to formulate the first classification system for bacteria (Drews 2000). The isolation and identification of microorganisms was the key to linking specific microbes to the onset of infectious diseases (Blevins and Bronze 2010). Between the 19th and 20th century, the cultivation of microbes in pure cultures became the strategy of choice for dissecting the individual characteristics and ecological roles of microbes in their respective environments. Many microorganisms were described, and niches of microbes revealed. Yet, the identification and characterization were limited by the fact – previously unknown – that the laboratory conditions available are not suitable for growing the majority of living microorganisms. Therefore, this has limited the view of the extent of the microbial world.

Over the past fifty years, the introduction of molecular methods and sequencing technologies have provided the means to analyse characteristics encoded in the genome, thereby adding a layer of information into the study of microbes. The use of molecular techniques allowed researchers to study microbes

that could not be cultivated in a laboratory. Moreover, these techniques have facilitated the detection of known microorganisms in distinct habitats and samples. For instance, the introduction of molecular detection of pathogens became the gold standard for diagnosing viral and bacterial infections. The molecular methods also prompted studies that focused on microbial communities rather than on one microorganism. In particular, NGS technologies have become a routine way to peek into microbial communities (Abreu and Taga 2016). These methods revealed that the great majority of microbes had not yet been identified and described. Analysis based on ribosomal RNA sequences (16S and 18S rRNA) recovered from environmental samples started in the late 1980s (Lane *et al.* 1985) and soon the establishment of PCR and DNA-sequencing methods simplified the analyses (Böttger 1989; Giovannoni *et al.* 1990; Wilson *et al.* 1990; Weisburg *et al.* 1991). The work was facilitated by the introduction of NGS, which allowed the massive parallel sequencing of different DNA fragments. Moreover, these methods enabled the comparison of microbial community structures in distinct habitats and the correlations of microbial groups with certain habitats and functions.

Notably, NGS tools have allowed for so-called metagenomics by sequencing of all DNA from environmental samples. The analysis of molecular data has enabled the investigation of microbial communities as a whole. Even complete microbial genomes have been reconstructed using such data. The reconstruction of genomes based on metagenomic data, or metagenomics assembled genomes (MAGs), provides us with insights on the diversity and functionalities of microbial dark matter. For example, Anantharaman and collaborators reconstructed over 2500 MAGs from aquifer sediments, revealing the presence of previously unknown lineages of bacteria and archaea (Anantharaman *et al.* 2016).

However, increasing knowledge of microbial communities has also revealed limitations on rRNA amplicon-based sequencing and metagenomics. Amplification kinetics in PCR targeting for 16S rRNA have been shown to impose a bias toward more abundant sequences, therefore missing or underrepresenting rare taxonomic groups in heterogeneous samples (Suzuki and Giovannoni 1996; Lee *et al.* 2012). Besides that, universal primers can be biased towards what is already known of microbial diversity. Therefore, dissimilar ribosomal sequences can be missed from the studies. Likewise, the functional analysis of communities describes the microbial lifestyles and metabolic pathways in a given environment but does not reveal the role of specific microbes. The reconstruction of assembled genomes from metagenomic data partially fill the gap, but fail to reveal structural variants of genomes, including rearrangements, gene insertions, duplication, horizontal gene transfer (HGT), mutations, and mobile genetic elements such as plasmids. Therefore, there has been a growing interest in new tools and strategies for the study of microbes at single cell resolution.

Single-cell genomic sequencing was introduced to microbiology sixteen years ago and has been considered a promising tool for filling in the missing pieces in the studies of microbial diversity (Kaster and Sobol 2020). The feasibility of sequencing genomic DNA from single bacterial cells was first shown using

laboratory strains by Raghunathan and collaborators. DNA fragments corresponding to 30% of the genome, including a fragment of 662 bp of 16S rRNA gene, were recovered from single cells after sequencing amplicons produced by multiple displacement amplification (MDA) using random primers and allowed the identification of genes (Raghunathan *et al.* 2005). Soon after, Marcy and collaborators sequenced the genomes from uncultivated microbes of the candidate phyla TM7 revealing a collection of over 1000 genes. These genes helped to illuminate the physiology of the members of the phyla (Marcy *et al.* 2007). Single-cell genome sequencing and single virus sequencing have been applied to identifying and characterizing groups of non-culturable microorganisms (Woyke *et al.* 2017, 2019; Martínez *et al.* 2020). And the analysis of single amplified genomes (SAGs) has proven useful to complement the metagenomics data and provide a more complete understanding of microbial communities (Abreu and Taga 2016; Doud *et al.* 2020; Kaster and Sobol 2020). Moreover, the single-cell studies have been useful in understanding the interactions of microorganisms within communities, particularly virus-host interaction. The gene composition observed from SAG and MAG was used as an indication of prophage infection (Labonté *et al.* 2015) and to follow the infection dynamics between viruses and an uncultured microbe (Roux *et al.* 2014).

Single-cell microbial analysis is still a novel field and has many technical challenges to be addressed (Kaster and Sobol 2020). In the past few years, novel methods and strategies have been developed to facilitate the research. Particularly, cell sorting has gained attention due to the need for detecting specific groups relevant for each study design (Lee *et al.* 2015; Woyke *et al.* 2017). Function-driven single-cell analysis emerged as a novel strategy to sort and facilitate the assessment of microorganisms related to specific processes (Woyke and Jarett 2015; Doud and Woyke 2017; Doud *et al.* 2020). Doud and collaborators devised a method to select certain microbes from a rare biosphere due to their affinity to cellulose and analysed their capacity to degrade the substrate (Doud *et al.* 2020). Another method focusing on functionalities of microbes is the epicPCR, developed by Spencer and collaborators. Instead of analysing the whole genome, epicPCR targets specific genes to detect and identify microbial hosts for specific encoded functionalities by trapping individual cells within a polyacrylamide compartment (Spencer *et al.* 2016). The method has been used, for example, to screen bacterial hosts for antibiotic resistance genes in wastewater treatment plants (Hultman *et al.* 2018) and to investigate virus-bacteria dynamics in a marine environment (Sakowski *et al.* 2021).

1.4 Answering biological questions

The process of understanding the microbial world has proven to be challenging and requires flexibility to choose and employ the right strategies and methods in order to unravel the puzzle at hand. In my thesis, the application of molecular methods and the expansion of toolbox for studying microbes is exemplified in

studies composing three distinct research topics. Below is a brief introduction to each topic.

1.4.1 Viral zoonosis epidemiology

Vaccinia virus (VACV) is a zoonotic virus that belongs to the genus *Orthopoxvirus* (OPV) and the *Poxviridae* family, a group of large double-stranded DNA viruses. VACV are naturally found in South America, where several outbreaks in human populations have been reported since the late 1990s (Trindade *et al.* 2007; Domingos *et al.* 2021). Wildlife is involved in the maintenance of the virus in nature, and outbreaks in livestock are sources of human infection. Small wild rodents have been shown to replicate and shed the virus through faeces, and are thus believed to be the source of outbreaks (Ferreira *et al.* 2008; Abrahão *et al.* 2009; D'Anunciação *et al.* 2012). Capybaras are large rodents and express habits that could be important in the viral transmission chain to humans. Previous studies have indicated the susceptibility of capybaras to VACV infection by recovering virus from experimentally infected animals (Moreira 1956) or by serological study of free ranging animals (Barbosa *et al.* 2014). However, VACV had not been detected in free-living animals before. Abrahão and collaborators have shown that the virus can be detected in PCR assay directly from clinical specimens (Abrahão *et al.* 2010). In this study, a qPCR-based assay is developed to detect VACV from capybara's stool samples.

1.4.2 Groundwater pollution and bioremediation by native microbial communities

Chlorophenols (CP) are a group of xenobiotic compounds that have been used as pesticides and wood preservatives in the past. These substances are toxic and have been reported as persistent contaminants of soils, lakes, and ground water worldwide. In Finland, chlorophenols have been detected in several aquifers even three decades after their use has been suspended, and no improvement of contamination levels has been observed during decades of monitoring. Since the late 90s, more effort has been put into removal of chlorophenols by using different means. Bacteria with potential to degrade CP have been detected in contaminated sites in Finland since late 90s (Männistö *et al.* 1999, 2001), and it was previously shown that CP degraders may be isolated from bacterial communities originating from polluted areas. The aerobic pathway for degrading highly chlorinated phenols (HCP) is initiated by the enzyme pentachlorophenol hydroxylase (PcpB, encoded by the *pcpB* gene), an inefficient reaction (Arora and Bae 2014). The first section of study II introduces the development and application of a qPCR-based assay to investigate the potential for aerobic degradation of highly chlorinated phenols (HCP) in polluted sites. The detection of the *pcpB* gene by qPCR is applied to spot the potential for the metabolic process from natural communities. In addition, the total bacterial community diversity, and the presence of Sphingomonads' (the only assumed HCP aerobic degrader group) is investigated through 16S rRNA gene survey.

However, a direct link to the presence of *pcpB* gene, potential HCP degrader groups and the reduction of CP is missing.

1.4.3 Application of molecular methods for single cell studies

The revelation of the complexity of microbial communities and interaction-networks has called for new methods to narrow research from the community to the cellular level and therefore to understand the role of each microbe in the equation. For instance, the study II investigates the aerobic dehalogenation of HCP and the used tools (PCR and NGS) provided a profile of the microbial community and the detection of putative degraders. Despite that, the methods were not enough to provide a complete picture of the microbes involved in the HCP degradation, nor to define the microbes triggering the process through the enzyme pentachlorophenol hydroxylase. Similar question – "Who is doing what?" can be applied to other study contexts and microorganisms residing in various habitats. For example, who are the microorganisms carrying of antibiotic resistance genes in wastewater treatment plants? Therefore, there has been a growing interest in the development of methods that can facilitate the analysis of the microorganism carriers of specific genetic functionalities within diverse microbial communities. The recent efforts have heavily focused on single cell-based studies. epicPCR is a well-engineered method that enabled the detection of "carrier microbes" without the need for generating extra data. Nonetheless, epicPCR data is limited to the identification of the "carrier microbes" and needs complementary methods to provide the profile of the entire studied community. Moreover, epicPCR would benefit by methods that produce monodisperse droplets such as droplets microfluidics.

2 AIMS OF THE STUDY

Despite the developments in microbiology that, for example, allow sequencing of single genomes, a complete understanding of microbial diversity and functions is not yet possible. As such, there is a constant demand for research resources that help build a better understanding of microbes via experimentation design and new methods. In my thesis I aimed to improve this understanding by:

i. employing PCR-based assay to detect and identify specific virus in relevant environmental samples (detection of *Orthopoxvirus* in capybara's stool samples);
ii. characterizing microbial communities and their potential in removing highly chlorinated phenols from groundwater;
iii. optimizing protocols for efficient DNA recovery from the emulsion PCR; and
iv. developing a new method for linking specific microbes and functionally important genes.

Although the contexts of these studies were different, together they highlight not only the development of tools for studying microbes, but how these tools can be combined and applied to answer important biological questions.

3 MATERIALS AND METHODS

The methods used in the articles I-IV that build this thesis are listed in the table below. Detailed descriptions of the methods can be found in the respective publications.

TABLE 1 Methods used in the publications.

Method	Publication
Field sampling	I, II
Experimental contamination of stool samples	I
DNA extraction	I, II, III, IV
qPCR	I, II
PCR	I, III, IV
ddPCR	III, IV
DNA electrophoresis analysis	I, II, III, IV
DNA sequencing and sequence analysis	I, II, IV
Bacterial culture	IV
Bacterial quantitation by plate count method	IV
Cell culture	I
Virus isolation	I

4 RESULTS AND DISCUSSION

4.1 Detecting microbes in complex samples using PCR-based assays

4.1.1 *Orthopoxvirus* in capybara's stool samples (I)

In this study, a non-invasive method to screen capybaras for VACV was developed using a qPCR assay to detect viral DNA from faecal samples. For optimization tests, capybaras' faecal samples were collected from an area with no previous report of circulating virus. Later, field samples collected from varied locations were used for the validation of the screening method. The sampling sites and sample characteristics are described in Table 1 and Figure 1 (I). Initial processing of the samples consisted of ten-fold dilution (*w/v*) in PCR-grade water followed by clarification by centrifugation. The following tests aimed to define optimal conditions for viral DNA detection in qPCR and the detection limit of the assay in such conditions.

Minimum non-inhibitory dilution was tested by diluting the faecal samples from 20 up to 2500-fold (*w/v*) in PCR-grade water and adding each dilution with 10^4 plaque forming units per microliter (pfu/µl) of virus (I, Fig. 2a). Table 3 (I) shows the compilation of results showing that 200-fold dilution of faecal matter is enough to minimize inhibitory effect and detect viral DNA. An aliquot of each sample dilution previously added with virus was used for DNA extraction using the Phenol Chloroform Isoamyl alcohol (PCIA, 25:24:1) method (Sambrook 2012), and the yielded DNA used as a template in the qPCR reaction (I, Fig. 2b). A comparison of the amplifications showed that extraction of DNA works equally to diluted samples. Next, faecal samples were artificially introduced with viral loads of 1 to 10^4 pfu/µl to test the detection limit for the reaction using diluted samples. An aliquot was subjected to DNA extraction using PCIA and the results from both types of samples were compared (I, Fig. 2c and d). The results show that the virus can be detected from diluted samples in concentrations as low as 1

pfu/µl, while in samples where DNA was isolated the detection limit was 100 pfu/µl (I, Table 4).

Faecal samples are composed by substances that inhibit PCR reactions, and the faecal matrix composition is affected by diet (reviewed by Schrader *et al.* 2012). The dilution of samples, or the alternative DNA extraction, in this study had the purpose of diminishing the potential inhibitory effect caused by the faecal matrix. The dilution of the faecal matrix implies the dilution of the inhibitory substances, however the same holds true for the viral DNA present in the sample. On the other hand, the DNA extraction procedure can eliminate many of the inhibitory substances. However, DNA can be lost during the process and, if not executed correctly, inhibitory substances can be carried over from the extraction reagents or even from the faecal matrix. The use of diluted faecal samples for viral detection by PCR has been shown before (Ferreira *et al.* 2008; Abrahao *et al.* 2009, 2010). As for the DNA extraction, customized protocols and kits have been developed for the studied animal. In this study I compared the PCR detection of virus using both sample dilution and DNA extraction using the standard phenol chloroform protocol. The lower detection limit observed in DNA extracted samples could be explained by the inhibition of reaction due to carry over of inhibitory substances.

The above described protocol for sample preparation was validated with samples collected from urban and wild environments. In three samples, the screening resulted in amplification with qPCR. To rule out laboratory contamination, alternative OPV genes were targeted in qPCR and PCR as well (I, Table 2). Analysis of sequences obtained from fragments of two genes (C11R and A26L) demonstrated that the amplified DNA was highly similar to Brazilian VACV (VACV-BR) (I, Fig. 4a and b). The presence of a genetic marker in sequences obtained from a fragment of the gene A26L suggests clustering with VACV-BR group I, rather than VACV-BR group 2 or VACV-WR (I, Fig. 4c). However, it must be noted that several attempts to isolate the virus were unsuccessful.

OPV have been detected in various regions of Brazil and South America. Wild and domestic species have been implicated in the natural circulation of virus (Lima *et al.* 2019; Domingos *et al.* 2021). The susceptibility of capybaras to OPV infections has been shown earlier by laboratory infection (Moreira 1956). Recently, serum neutralizing assays from free living animals have indicated the presence of neutralizing antibodies specific for VACV in free living capybaras (Barbosa *et al.* 2014; Antunes *et al.* 2020). However, no molecular data had been obtained from capybaras' samples so far.

The PCR-based assay optimized here to detect OPV from capybaras' stool samples and showed the presence of DNA from *Orthopoxvirus* in wild capybaras. The detection of OPV DNA in samples collected from an urban park in Belo Horizonte, with no previous knowledge of virus circulation, has drawn attention for the silent circulation of OPV in such areas (Domingos *et al.* 2021). After the publication of article I, serological and molecular surveys of domestic and wild animals have confirmed the circulation of VACV in the same region (Costa *et al.* 2017, 2018; Miranda *et al.* 2017).

The qPCR assay to detect viral DNA from faecal samples has proven to be a convenient tool for screening OPV in wild capybaras population since no direct contact with the animal is needed. Several OPV targets have been detected through qPCR, and genetic markers resembling VACV-BR identified. The detection of DNA from free-ranging capybaras suggest that capybaras may play a role in the circulation of VACV between wild and domestic environments. Nonetheless, a better genetic characterization of these putative new viruses is necessary to assign them to VACV groups. Moreover, to evaluate the real role of capybaras in VACV transmission cycle and their potential to spread viruses is necessary to broaden the scope of study to an eco-epidemiological approach. Therefore, capturing the animals to collect other sample types might be needed in the future.

4.1.2 Examining microbial diversity from a successful bioremediation process (II)

The first section of paper II dealt with the development and application of a qPCR-based assay to investigate the aerobic degradation of highly chlorinated phenols (HCP) in polluted sites. The detection of pentachlorophenol hydroxylase gene (*pcpB*, encoding the first enzyme of the degradation pathway) by qPCR was extensively tested *in silico* and applied to spot the potential for the metabolic process from natural communities. In addition, the total bacterial community diversity, and the presence of Sphingomonads' (the only assumed HCP aerobic degrader group) was investigated through 16S rRNA gene survey.

The molecular detection of the gene *pcpB* in Pursiala site revealed the presence of potential CP degraders. Moreover, it suggested that the microorganisms carrying *pcpB* were more abundant the closer the sampling was done in respect to the most contaminated area (but not on the hotspot) (II, Table 2). The hotspot, however, presented 10-fold lower abundance compared to the next most contaminated well. The same trend was observed in Sphingomonad abundance, when group-specific 16S rRNA qPCR was used. Bacterial 16S rRNA gene counts were roughly similar in all five sampled spots. In Kärkölä, the investigation was carried out from a long term *in situ* experimental treatment. The treatment was based on water circulation to promote mechanical aeration (II, Fig. 1), and compared to the start level, results revealed a notable reduction in CP concentration (II, Fig. 3). The microbial analysis carried out after four and ten months of the beginning of the treatment revealed high relative abundance of *pcpB* genes and Sphingomonads – even higher than found in Pursiala (Table 3). These results suggested that aeration of groundwater could provide a stimulus to native degrader bacteria to increase CP mineralization rate by activation of the HCP pathway. However, without molecular data prior to the start of bioremediation process or RNA transcripts to indicate active metabolism, it was not possible to draw a direct connection to the observed changes. Therefore, to verify this assumption, an experiment was designed to test a natural bacterial community in a closed system.

A microcosm experiment was set up using groundwater and sediment collected from Pursiala, where high levels of CP and the *pcpB* gene were previously detected. Incubation bottles were filled with fully or partially aerated ground water and 1% (v/v) of sediment slurry sampled from the zone with the highest contamination level. The samples were kept at 15 ˚C in the dark for 21 days. Measurements taken during the experiment showed that oxygen was depleted in low aerated bottles after 10 days, while in fully aerated bottles the air saturation was at 64% by the end of the experiment. Inspections at the start and end of the experiment showed reduction of up to 96% of CPs, while the number of copies of the *pcpB* gene increased by 2400% and 360% in low and fully aerated samples, respectively (II, Fig. 2). Assessment of the microbial community by 16S rRNA gene sequencing revealed an increased proportion of Sphingomonads from 4% of total 16S rRNA sequences on water and sediment to average of 50% and 27% in low and fully aerated samples, respectively, by the end of incubation. The reduction of concentration of all forms of CP (HCP and lower chlorinated forms) during the microcosm incubation suggested that much of the CP was fully dechlorinated.

The toxic potential of HCP and the long-term persistence in contaminated areas have demanded studies to address the removal of such compounds from the environment (Yang *et al.* 2021). In Finland, contamination levels have been monitored in HCP polluted areas, and cultivation-based studies of the microbial community allowed for the isolation of bacterial strains capable of aerobic HCP degradation (Männistö *et al.* 1999, 2001). Despite that, contamination levels of HCP have persisted. In this study, a PCR-based assay and DNA sequencing were employed to monitor the microbial community and the potential for aerobic biodegradation of HCP. The tools were employed to follow the ongoing experimental aeration treatment process in Kärkölä and suggested that the increase in *pcpB* gene copies and bacterial groups related to the degradation of HCP were connected to the lower levels of CP detected over time. These findings were verified by studying the process in a closed environment (microcosm). Hermon and collaborators employed microcosm experiments and functional genes and 16S rRNA enumeration to establish the correlation of environment conditions to processes encoded in the genes (Hermon *et al.* 2019). In our study, samples were collected from an area where the microbial biodegradation potential had been shown only by molecular methods. Although abiotic CP losses could not be verified in the microcosm, the changes in microbial community structure and *pcpB* counts corroborated the previous observations. Bio-stimulation with oxygen seems to support the growth of native CP degraders and facilitate the mineralization process of CP in the studied samples.

In this study, the success of the bioremediation induced by aeration was indicated by the lower CP concentration at the end of the experiment, coupled with an increased number of copies of *pcpB* gene (measured by qPCR) and increased abundance of Sphingomonads (observed by amplifying and sequencing 16S rRNA gene). The abundance of *pcpB* gene was measured by qPCR and suggested a favourable environment for the microbes carrying such gene. Meanwhile the abundance of Sphingomonads was detected by amplicon

sequencing. Thus, revealing the change in microbial community profile and supporting the connection between the Sphingomonads group and the HCP degrading gene. Although successful in showing that the bioremediation process was induced by the aeration of a natural microbial community, the methods employed here were insufficient to determine the specific microbes that were stimulated and could be responsible for the bioremediation. Moreover, the complete degradation of HCP is a process that depends on several enzymes and therefore a set of genes (Arora and Bae, 2014), and a wider group of microorganisms. As such, a more complete picture of the HCP removal process would need to reveal the microorganisms involved in other steps of the process.

4.2 Single-cell screening of microbes and specific functions

In this work, a novel method was developed with the aim to track the diversity of microorganisms that carried a specific function at single cell resolution. In this method, molecular screening of single cells is conducted from mixed microbial sample by simultaneously detecting two genes of interest, one being the phylogenetic marker 16S rRNA and the other encoding a functional trait (Fig. 2; IV, Fig. 1). The bacterial sample is partitioned into single cell PCR droplets using droplet-based microfluidics technology. Within each droplet, one-step fusion PCR is carried out to fuse the genes into a concatemer and amplify the product. The amplicon product is recovered from the droplets and used to construct sequencing libraries. Carriers of the functional gene and microbial community diversity are revealed by sequencing the full-length concatemer and 16S rRNA gene, respectively. Dispersal of the antibiotic resistance gene was chosen to establish the tracking system. However, any other trait of interest encoded in the genome may be studied as well. For instance, some of the tests were carried out to optimize the system for studying the dispersal of *pcpB* gene (data not shown). The development of the tool required a series of experiments to ensure that external DNA was not amplified, targeted genes were fused successfully and also to verify the method's success in detecting single cells (described further below).

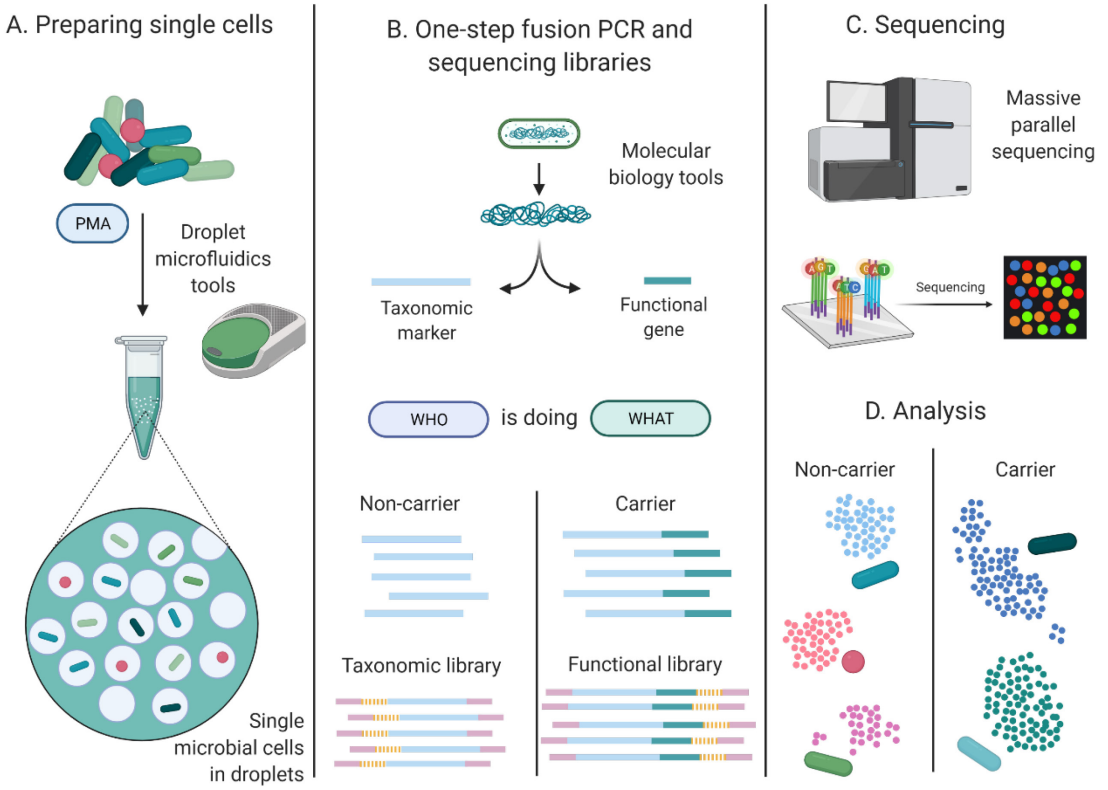


Figure 2. Overview of the method developed for single cell analysis. (A) Samples are treated with PMA, added to PCR reactions and partitioned by droplet microfluidics in order to have droplets with single cells. (B) Inside droplets, the taxonomic marker and the functional gene are amplified and fused. The amplicons are recovered and used to prepare sequencing libraries. (C) Sequencing of fused amplicons is done by any high-throughput sequencing method. (D) The analysis of the sequences obtained from 16S rRNA gene reveals the identity of the cells, while the functional gene sequence in the concatemer indicates the carriers.

4.2.1 Inhibiting the amplification of external DNA (exDNA) (IV)

External DNA (exDNA) is a known source of error in molecular detection of viable bacterial cells (Nocker and Camper 2006; Klein 2007; Emerson *et al.* 2017). Propidium monoazide (PMA) is a DNA-binding dye that attaches covalently to DNA after exposure to light. PMA does not enter intact cells and therefore it has been used for treatment of samples in order to immobilize exclusively the exDNA but not the intracellular DNA (Emerson *et al.* 2017). Once all total DNA is extracted from the sample, the PMA bound to exDNA prevents it from being amplified in a PCR reaction. Recently, PMA treatment has been used to prepare samples to be used in ddPCR assays. Gobert and collaborators optimized a PMA treatment of pig faecal samples to detect genomic DNA from viable bacterial cells (Gobert *et al.* 2018). In our study, intact and active individualized cells are the targets in single-cell analysis, and thus no DNA extraction is performed. Preliminary tests also revealed the presence of exDNA fragments in our bacterial

cultures. Therefore, to eliminate the bias caused by exDNA to single-cell analysis, we tested the efficiency of PMA treatment.

The potential inhibitory effect of PMA to ddPCR was evaluated, by adding increasing concentrations of photo-activated PMA to ddPCR reactions (EvaGreen fluorophore) (IV, Fig. 2). The RFU of positive droplets obtained from channel 1 (Ch1, FAM) decreased with increasing PMA concentrations, reaching the same level of negative droplets at PMA concentration of 2.5 µM. In channel 2 (Ch2, HEX/VIC), however, the RFU remained at similar levels at different PMA concentrations.

These results suggested that PMA can interact with the ddPCR chemistry by possibly shifting the emitted fluorescent spectra to longer wavelengths than can be detected in Ch1; nonetheless not affecting the wavelengths detected in Ch2. According to the manufacturer, the PMAxx™ exhibits light absorption and emission in the following range: λ_{Abs} = 464 nm (before photolysis); $\lambda_{Abs}/\lambda_{Em}$= ~510/~610 nm (with DNA/RNA, after photolysis) (Anon. 2019). However, the properties of PMA photolyzed in the absence of nucleic acid, as used here, are not described. Meanwhile, the absorption and emission wavelength ranges for EvaGreen (detected in ddPCR Ch1) and VIC (detected in ddPCR Ch2) are $\lambda_{Abs/Em}$ = 499/526 and 533/559 nm, respectively. Therefore, we suggest that the fluorescence emitted by EvaGreen fluorophore in the droplets after PCR is absorbed by the excess of free photolyzed PMA, which in turn emits it in another wavelength range. The fluorescence is detected by Ch2, and therefore the droplet can be assigned as positive or negative based on the fluorescence detected. The ddPCR and qPCR assays rely on the fluorescence detected from samples, therefore the interaction of PMA with these chemistries should be further studied.

Later, PMA was successfully used to treat cell suspensions and filtrate obtained from bacterial culture in order to prevent amplification of exDNA (IV, Fig. 3).

4.2.2 One-step fusion PCR (IV)

The antibiotic resistance gene (ARG) encoding beta-lactamase, *bla*CTX-M-14 (*bla*), found in the conjugative plasmid pEC13 (Mattila *et al.* 2017), was selected as the target trait for this project. A laboratory strain of *Escherichia coli* (*E. coli* K-12 BL21Gold) carrying the conjugative plasmid pEC13 served as an ARG-positive carrier. *Pseudomonas fluorescens* (ATCC 13525), *Salmonella enterica serovar* Typhimurium (SL5676) and *Klebsiella pneumoniae* (DSM681) were selected as non-carrier controls. Universal primers for bacteria, targeting the region V1-V2 of 16S rRNA gene (27F and 338R) were used as phylogenetic markers to identify species. The primer pair 293F and 320R was designed to amplify a fragment of 101 bp from *bla*, and an overhang encoding the reverse complement of 27F was added to the 5′ end of 320R. This overhang enables the assembly of the concatemer through the attachment to the amplified 16S rRNA fragment (IV, Fig. 4).

The optimization tests for the bla-16S concatemer (i.e., combination of fragments from 16S rRNA and beta-lactamase gene) production in ddPCR were

carried out by using dual asymmetric one-step fusion PCR (IV, Supplementary Fig. 3). Varying ratios and concentrations were tested for each primer pair. The results were analysed based on the clustering of fluorescence intensities of droplets. Initially, each pair of primer was tested separately. The reactions to amplify 16S rRNA were prepared using 1 to 100-fold difference of forward to reverse primers, and in primer concentrations varying between 100 to 400 nM. After the screening, the amplification could be verified when using up to a 10-fold difference in the ratio of primers. The reactions to amplify *bla* were also prepared using up to a 100-fold difference in forward to reverse ratio and the total primer concentrations tested were 100, 150 and 300 nM. Amplification of *bla* was verified in reactions using up to a 20-fold difference of forward to reverse. To continue the tests, 16S rRNA and *bla* primers were combined in the same ddPCR reaction to verify the efficiency to produce the concatemer. The concentration and ratio of 16S rRNA primers in the reaction were 225 nM, and 1 to 5 of forward to reverse. *bla* was tested in ratio from 10 to 30-fold difference, and concentrations ranging from 100 to 300 nM. The clustering of droplets in four possible groups (due to random distribution of DNA in the samples) – negative, for both targets; two single-positive, for 16S rRNA or for *bla*; and double positive – and cluster separation were taken into consideration to narrow the primer ratios and concentration options. In further single-cell experiments, the 16S rRNA pair was used with a ratio of 1 to 5 forward to reverse and final concentration of 225 nM, and *bla* used in ratio 10 to 1 of forward to reverse in 200 nM final concentration.

The concatemer production in the reaction is due to the depletion of primers to produce the intermediate fragments, therefore the complementary ends of the two fragments prime the concatemer polymerization. Thus, the use of a variable ratio of forward to reverse will interfere with the efficiency of concatemer production. The efficiency of concatemer production in bulk PCR has been shown to be higher when the outer primer concentration exceeds that of inner primers by 20-30 (Liu *et al.* 2018), however this has not been studied with single-cell experiments. Spencer and collaborators used ten times more outer primers in the epicPCR reaction (Spencer *et al.* 2016), nonetheless the dynamic of the reaction differs from here due to the employment of a single inner primer. Testing other targets and primer combinations will be useful to show whether the same ratios used here are applicable or if each primer combination requires case-specific optimization.

4.2.3 Sample partitioning (IV)

Once the errors stemming from free DNA could be eliminated by treating the cell suspension with PMA, the next step was to verify feasibility of the sample partitioning with ddPCR. Assuming that the distribution of cells into droplets would occur similarly to the distribution of DNA molecules, it was hypothesized that by using the right cell density, it would be possible to obtain a single cell within a droplet. Therefore, serial 2-fold dilutions prepared from PMA-treated cell suspension of *E. coli*, *P. fluorescens*, *S.* Typhimurium and *K. pneumoniae* were

used as a template in ddPCR assays targeting the 16S rRNA gene. In parallel, cell suspensions were also plated to count the cell density. The bacterial density obtained using 16S rRNA ddPCR was similar to the density obtained by plating (IV, Fig. 5). The decreasing density was consistent with the dilution factor using both methods. However, higher cell density was detected by ddPCR from *P. fluorescens* and *K. pneumoniae* cell suspension (IV, Fig. 5b and 5d). The detection of more cells by ddPCR than by plating could suggest that the PMA treatment was inefficient. However, in every batch of tests, PMA control samples were prepared to guarantee the efficacy of the treatment. Therefore, the discrepancy should be explained by other means. The CFU counting method is limited to enumerating the cells that are viable and in a replicative state, therefore outlier metabolic states can bias the quantification of total intact cells in a sample (Barer and Harwood 1999 and references therein). On the other hand, the use of molecular detection after PMA treatment ensures that intact cells can be detected, despite the metabolic state. To find out the reason for variation, other methods for cell quantification, for example microscopic cell counting using DAPI or Most Probable Number (MNP) assay could be employed together. In this study we tested the feasibility of CASY cell counter and flow cytometry using a live/dead cell double staining kit for counting the cells in our assays (data not shown). However, such methods were not appropriate for the comparison purpose. Enumeration of bacterial cells using with CASY varied between measurements, therefore it was considered unreliable. The use of flow cytometry required a prior staining step and optimization of both staining and reading. It proved to be slow and demanding protocol to be executed alongside the ddPCR and plating, thus not compatible with the proposed task.

4.2.4 Breaking ddPCR droplets (III)

Recovering amplicons from ddPCR droplets is a critical step for the method, as the sequencing of these DNA fragments will indicate which are the carriers of the studied trait. Therefore, chemical and physical methods to efficiently break ddPCR droplets were analysed. Three chemical methods: Chloroform, 1H, 1H, 2H, 2H-Perfluoro-1-Octanol (PFO) and n-Octane; and two physical methods: rapid cooling with liquid nitrogen (LN2) and silica columns (III, Table 1) were studied. The emulsion samples were produced by combining ddPCR reactions prepared with EvaGreen Supermix for ddPCR and Droplet Generation Oil for EvaGreen using the QX200™ ddPCR System (Bio-Rad). Efficiency was given by calculating the percentage of DNA recovered from droplets in comparison to what was initially added to the reaction mix. Among chemical methods, efficiency of DNA recovery using chloroform was up to 61 ± 4% of the initial DNA, higher than n-Octane and PFO, which provided recovery of approximately 59% (III, Fig. 2). Chloroform to oil ratio optimizations were tested but did not point to any value that could improve the outcome. Moreover, the use of n-Octane resulted in undesired reversal of phases, making it more difficult to collect the aqueous phase from the bottom (III, Fig. 3). The use of LN2 for rapid cooling and breaking the ddPCR droplets yielded the recovery of up to 70% of

initial DNA and was shown to work with both EvaGreen and Probe ddPCR kits (III, Supplementary Fig. 1). The thermal procedure based on rapid freezing using liquid nitrogen was shown to be very effective in coalescing the droplets and recovering the DNA. For not requiring chemical compounds with special disposal practices, and for being easy and reproductible, the LN2 method is a convenient procedure for recovering the amplicons from ddPCR droplets.

4.2.5 Identification of *bla* carriers in an artificial mixed sample (IV)

After revising the steps for single-cell analysis, sample preparation, partitioning using ddPCR system and the PCR for linking the target genes, we proceeded to apply the method for artificially generated bacterial samples. The feasibility of the single-cell method to identify microbial cells hosting the *bla* gene in a sample of different bacterial species was studied. The mixed samples were prepared by combining *E. coli* carrying *bla*CTX-M-14 with *S. typhimurium* or *K. pneumoniae* after treatment with PMA. Mixed samples and cell suspensions from each strain were used in the single-cell ddPCR assay. The bacterial cell count for each strain was estimated via the colony counting method and 16S rRNA ddPCR assay. After the amplification, the amplicons were recovered according to the methods established in study III. The concentration of the amplicon recovered from droplets is very small, therefore, to prepare the sequencing libraries, the material had to be reamplified. Spencer and collaborators found that blocking primers are needed during the reamplification step to prevent cross-talking between cells due to the fusion of unrelated intermediate products recovered from fusion PCR (Spencer *et al.* 2016). Here, the potential for generating such false concatemer fragments is even higher due to the use of all four primers and the recovery of amplicon products from different cells. 16S rRNA fragments amplified from non-carrier and 16S rRNA and *bla* fragments, produced during the step 1 of the fusion PCR and not fused, can be recovered from carrier-cells. Thus, the blocking system described by Spencer *et al.* was used for reamplification and NGS library preparation.

The abundance of different OTUs with *bla* was determined by sequencing the bla_16S concatemer (IV, Fig. 6). The bacterial cell count in ddPCR and plating were consistent for *E. coli* and *S. typhimurium*, and higher counts were obtained by ddPCR for *K. pneumoniae*. The observation was compatible with earlier assays for detection of single cells. The identification of *bla* host cells from mixed samples by OTU abundance revealed 0.47-1.15% of noncarrier (*S. typhimurium* or *K. pneumoniae*) host reads, suggesting that the developed single-cell ddPCR process was effective in diminishing the occurrence of false positive concatemers sequences. Without the use of blocking primers, the number of false positives was significantly higher (data not shown).

4.3 Applications of single cell methods

Single-cell technologies are promising tools for expanding even further the frontiers of microbiology. More than identification, single-cell sequencing can provide useful information for understanding the role of microbes in the environment and illuminating their interaction networks. Methods for linking specific functions to microbial taxa at single-cell resolution have been described earlier, hence providing promising tools for example to study functions that are disseminated via HGT (Brito 2021). In epicPCR, individual cells are encapsulated into acrylamide beads and then used as a template for a secondary emulsion where PCR amplification and fusion of targets is carried out (Spencer *et al.* 2016). This has been used to study functional genes (Spencer *et al.* 2016; Hultman *et al.* 2018; Qin *et al.* 2019) and virus-bacterial host interactions (Sakowski *et al.* 2021). Recently, OIL-PCR was introduced as an easier alternative by reducing the procedure into a single emulsion step (Diebold *et al.* 2021). Studies III and IV of this thesis introduce yet another alternative method to link genes at single-cell resolution. Here ddPCR, a commercially available droplet microfluidics system, is used to entrap individual cells. The presented workflow prepares the samples to prevent the amplification of exDNA. This is followed by the controlled partitioning of the sample into single cells within ddPCR droplets. In the PCR reaction, the targets are amplified and fused together. Amplicons are recovered efficiently from the emulsion and sequenced.

Obvious though it may be, single-cell studies depend on having partitions with single cells. Therefore, ensuring that cells are delivered alone, and no free DNA is trapped together, is of major concern for the workflow. Here we demonstrated that the use of PMA to treat samples is beneficial for preventing the amplification of the exDNA commonly found in the environment. Partitioning is another challenge. Successful partitioning is dependent on the utilized method but also on the type of the sample. The methods for delivering single cells can employ flow cytometry, microfluidics and manual emulsification (as show by Spencer *et al.* 2019 and Diebold *et al.* 2021). Here, we demonstrated that ddPCR can be employed to trap cells into PCR droplets. While different cell-enclosing tools can have benefits and disadvantages, all of them require having a bulk of cells detached from each other. We acknowledge that aggregates of cells and biofilms can clog the sample flow in sorting and microfluidic channels or result in several cells per partition in manual emulsification. Joyce and collaborators proposed that high ultrasonication (over 500 kHz) could be used to de-clump these cell aggregates (Joyce *et al.* 2011). Such a method could be considered in the initial preparation of samples for single-cell studies, if the sample type requires.

Another challenge to address when preparing microbes for single cell studies is the cell lysis. For methods relying on amplification and detection of DNA, it is fundamental to have the cell lysed to facilitate the reactions to occur. While some microbial cells can be easily lysed by the high temperature achieved

in PCR, others can be very resistant and demand additional lysis methods. In epicPCR, the cells trapped in permeable acrylamide bead allows the employment of a variety of treatments for cell lysis, while the one-step emulsion PCR described in OIL-PCR requires the lysis reagents to be compatible with PCR.

Fusion PCR is the key for the linking of functional genes and microbial taxa discussed here. The fusion PCR employed in epicPCR and OIL-PCR methods allows the successful production of amplicons only if the functional gene is present because the additional tail in the fragment will act as the primer for the amplification of the phylogenetic marker. This differs from the strategy used in this work, where even in the absence of the desired genetic trait the phylogenetic marker can be amplified from any bacterial cell present in the droplet. This enables the projection of the whole community simultaneously, in addition to the detection of microbes with the gene of interest. The engineering of primers to detect more than one functional gene at a time is discussed by Diebold *et al.* (2021). Nonetheless, the methods for single-cell analysis of microbes are advancing and can help answer research questions that would otherwise be difficult to tackle.

4.4 Future considerations

While the common approaches for single cell studies, such as whole genome single cell sequencing, are a promising solution also for unravelling details of microbial life, some of its aspects can be a throwback. The challenge for reconstruction of complete genomes with good quality and the amount of data generated can hinder the advance in understanding specific processes. The method described in this thesis can be employed in transitioning steps from microbial communities to single cells, as the data generated is specific to chosen genetic trait. Therefore, the range of applications for this method is vast for processes carried out by microorganisms. These include for example questions explored in this thesis like the aerobic degradation of pentachlorophenol and distribution of antibiotic resistances. The method could also be utilized to study interactions between microbes, such as bacterial viruses and their hosts or the acquisition of novel traits by HGT. The method described here enables the study of genetic encoded functionalities that can be in the bacterial genome or in extrachromosomal nucleic acid inside. In the future, such methodology could be adapted to study transcripts by coupling cDNA from different genes inside a single cell, thus enabling not only the detection of carriers of specific DNA targets, but also the study of gene interactions upon certain stimuli.

5 CONCLUSIONS

Estimates of diversity and recent findings on uncultured microbes reveal that the vast world of microbes is still largely unknown. The challenges in understanding such a diversity and the complex interactions of microbes with their surrounding environment have prompted constant development and improvement of tools and study methods. Molecular biology was introduced in the past fifty years and has proven useful in studying microbes. Advances in other fields of sciences such as nanomaterials and microfluidics have provided technologies that can be employed in new methods and resources for microbiologists. Recently, single-cell technologies have been developed, hence allowing deeper investigation of microbial diversity and interactions. The combination of molecular and single-cell data with findings from classical methods (such as cultivation-based methods) is advancing our understanding of the microbial realm.

This thesis was dedicated to assess microbes (prokaryotes and viruses) using molecular techniques and explored contexts from communities to single cells. In studies I and II, cultivation and molecular techniques were used to detect viruses and bacteria in the context of communities. In study I, the methods were employed to look for clues that could link free-ranging capybaras to the maintenance of VACV in nature and the spark of outbreaks in farms. The work revealed the presence of viral DNA in animal faeces and indicated the possible silent circulation of the virus in a previously unknown area. In the second study (II), microcosm-based experiment was employed to observe the degradation of pentachlorophenol under controlled conditions. In this study, the response of a natural microbial community structure and metabolic capacity to degrade HCP under different cultivation conditions were observed with molecular methods.

Finally, studies III and IV focused on the development of a new method to facilitate the analysis of single bacterial cells. In this method, genetically encoded functionalities are linked to 16S rRNA fragments from the same cell by fusion PCR. The technical challenges were addressed by using the cultivation-based and molecular tools to provide a reliable method. The adoption of commercial platforms for the workflow – ddPCR system for cell partitioning and PCR; and

NGS for amplicon-based sequencing – makes the method accessible for many laboratories and research groups.

It can be concluded that molecular methods were sufficient to address the questions of studies I and II. Moreover, the combination of molecular methods and modern tools allowed the elaboration of an accessible method to study microbes at single-cell resolution. This shows that the field of development and optimization of novel methods is far from being saturated. And by keeping the mind open to ways of exploiting tools, methods and strategies, and maintaining rigorous analysis and fidelity to the scientific method, one can improve the resolution of findings.

Acknowledgements

The PhD journey hasn't been an easy ride. No one said it would, but no one said it wouldn't either. And for many times I thought about giving up. But I didn't. And here I am now: writing the last section of my book. In the first place, thanks to my resilience… but also thanks to the support in the most variable forms that I received from others.

I want to start by thanking to the funding bodies who supported my work and my attendance to conferences. My work was sponsored by the European Research Council, Academy of Finland and Jane and Aatos Erkko Foundation. Also, during my PhD I received funding from the Department of Biological and Environmental Science of the University of Jyväskylä and a personal grant from the Finnish Cultural Foundation.

I would like to thank my supervisors (and I was lucky to have many of them!). Thank you Marja Tiirola for giving me the opportunity to start the PhD and for trusting me to conduct my own research. To Matti Jalasvuori and Reetta Penttinen I thank you for accepting the challenge to supervise me half-way through the journey, and for all the inspiring discussions. I wouldn't have gone so far without your help. Reetta, you have always been so careful and dedicated to support me. I will never forget that! I also thank the supervisor who became a mentor in this journey, Anu Mikkonen.

I thank the reviewers of this thesis, who dedicated their time to read and comment and help me make it more understandable. Also, I thank those who somehow helped me in the writing process, either by inspiring me, or reading and giving their opinions. To Lotta-Riina Sundberg I thank for opening my eyes to the "big picture" of this work, your words made me re-think what my journey was all about. And to Emily Knott, who saved me many times during this final process. I am eternally grateful!

To Aaron Swartz, Alexandra Elbakyan and others that have fought for making the scientific knowledge freely available. And to all scientists whose work have been overlooked, in particular to all women in science who had - and those that still have - to fight to be accepted and recognized by their work.

In this journey a met a bunch of nice folks who became my support net and made my days easier and funnier. I had the pleasure to always share the office with the best office mates one could ever have. First Pauliina S. and Anita M., then Marjaana H. and Eveliina K. I wish I had more office time with our newcomer Samu (but I know he fits the profile of the office!). I also found support from the amazing people that shared the "coffee room", and before corona times we would meet twice a day for a cup of coffee and a relaxing conversation. I name Mervi K. and Emma P. but my heart goes with all of you who have been there. For a while I felt alone as a PhD student, but the "sauna and support" (old and new generation) came to show me that I am not alone. Thank you all for listening my cries, for sharing ideas, pizza and for having the strength to fight the system for better training conditions for the PhD students.

Coming to a new country was not an easy move, thus building a support net was fundamental for keeping me going. I thank the friends that Finland gave me and who made my ride cheerful and smooth. You brought joy to my life. To my dear Brazilian sisters in Finland, thanks for always being by my side and for supporting me with nice conversations, thanks for sharing advice and for baking delicious treats. To my Finnish friends, thanks for making me feel welcome in this land, thanks for showing me the "Finnish way" and thanks for accepting my "Brazilian way" of being.

I would like to thank my family and my extended family for all the support and dedication. Thanks to my brother and my sister (his partner) for sharing the joys and the sorrows of being a foreigner, and the love for science. To my parents and my parents in law, my gratitude for encouraging me to be who I am and to pursue my dreams no matter how far it would take me (and Gabriel & kids).

And finally, to my dear partner Gabriel my most gratitude for supporting me all the way through. You made a dream come true: to pursue my doctoral studies in an international environment. Moreover, you made the journey less heavy by helping with discussions, reading, and commenting manuscripts and (a bit of) this book. And for many times you reassured me that I could, even when I doubted myself you stood there for me. You also took good care of what are the most precious products of our existence, our little explorers Gregorio and Catarina. And to my kids, Gregorio and Catarina, I am grateful to you for keeping me on track. You pulled the strings that helped me to balance the work and my personal life, and you made sure I really balanced it! Love you!

References

Abrahão J.S., Guedes M.I.M., Trindade G.S., Fonseca F.G., Campos R.K., Mota B.F., Lobato Z.I.P., Silva-Fernandes A.T., Rodrigues G.O.L., Lima L.S., Ferreira P.C.P., Bonjardim C.A. & Kroon E.G. 2009. One More Piece in the VACV Ecological Puzzle: Could Peridomestic Rodents Be the Link between Wildlife and Bovine Vaccinia Outbreaks in Brazil? *PLoS ONE* 4: e7428.

Abrahão J.S., Drumond B.P., Souza Trindade G. de, Silva-Fernandes A.T. da, Ferreira J.M.S., Alves P.A., Campos R.K., Siqueira L., Bonjardim C.A., Ferreira P.C.P. & Kroon E.G. 2010. Rapid detection of Orthopoxvirus by semi-nested PCR directly from clinical specimens: A useful alternative for routine laboratories. Journal of Medical Virology 82: 692–699.

Abreu N.A. & Taga M.E. 2016. Decoding molecular interactions in microbial communities Shen A. (ed.). *FEMS Microbiology Reviews* 40: 648–663.

Albers S.-V., Forterre P., Prangishvili D. & Schleper C. 2013. The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nature Reviews Microbiology* 11: 713–719.

Amann R.I., Ludwig W. & Schleifer K.H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews* 59: 143–169.

Anantharaman K., Brown C.T., Hug L.A., Sharon I., Castelle C.J., Probst A.J., Thomas B.C., Singh A., Wilkins M.J., Karaoz U., Brodie E.L., Williams K.H., Hubbard S.S. & Banfield J.F. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* 7: 13219.

Anon. 2019. *PMAxx™ Dye, Product infromation*. Biotium. 2019.

Antunes J.M.A. de P., Borges I.A., Trindade G. de S., Kroon E.G., Cruvinel T.M. de A., Peres M.G. & Megid J. 2020. Exposure of free-ranging capybaras (Hydrochoerus hydrochaeris) to the vaccinia virus. *Transboundary and Emerging Diseases* 67: 481–485.

Arora P. & Bae H. 2014. Bacterial degradation of chlorophenols and their derivatives. *Microbial Cell Factories* 13: 31.

Avery O.T., MacLeod C.M. & McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine* 79: 137–158.

Barbosa A.V., Medaglia M.L.G., Soares H.S., Quixabeira-Santos J.C., Gennari S.M. & Damaso C.R. 2014. Presence of neutralizing antibodies to orthopoxvirus in capybaras (hydrochoerus hydrochaeris) in Brazil. *Journal of Infection in Developing Countries* 8: 1646–1649.

Barer M.R. & Harwood C.R. 1999. Bacterial Viability and Culturability. In: *Advances in Microbial Physiology*, pp. 93–137.

Bar-On Y.M., Phillips R. & Milo R. 2018. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* 115: 6506–6511.

Blevins S.M. & Bronze M.S. 2010. Robert Koch and the 'golden age' of bacteriology. *International Journal of Infectious Diseases* 14: e744–e751.

Böttger E. 1989. Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. *FEMS Microbiology Letters* 65: 171–176.

Brito I.L. 2021. Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology*: 1–12.

Chen K., Neimark H., Rumore P. & Steinman C.R. 1989. Broad range DNA probes for detecting and amplifying eubacterial nucleic acids. *FEMS Microbiology Letters* 57: 19–24.

Cobb M. 2015. Who discovered messenger RNA? *Current Biology* 25: R526–R532.

Costa G.B., Almeida L.R. de, Cerqueira A.G.R., Mesquita W.U., Oliveira J.S. de, Miranda J.B., Saraiva-Silva A.T., Abrahão J.S., Drumond B.P., Kroon E.G., Pereira P.L.L., Magalhães Soares D.F. de & Souza Trindade G. de. 2018. Vaccinia virus among domestic dogs and wild coatis, Brazil, 2013–2015. *Emerging Infectious Diseases* 24: 2338–2342.

Costa G.B., Miranda J.B., Almeida G.G., Oliveira J.S. de, Pinheiro M.S., Gonçalves S.A., Reis J.K.P. dos, Gonçalves R., Ferreira P.C.P., Bonjardim C.A., Abrahão J.S., Kroon E.G. & Trindade G. de S. 2017. Detection of vaccinia virus in urban domestic cats, Brazil. *Emerging Infectious Diseases* 23: 360–362.

Crick F.H.C., Barnett L., Brenner S. & Watts-Tobin R.J. 1961. General Nature of the Genetic Code for Proteins. *Nature* 192: 1227–1232.

Danovaro R., Canals M., Tangherlini M., Dell'Anno A., Gambi C., Lastras G., Amblas D., Sanchez-Vidal A., Frigola J., Calafat A.M., Pedrosa-Pàmies R., Rivera J., Rayo X. & Corinaldesi C. 2017. A submarine volcanic eruption leads to a novel microbial habitat. *Nature Ecology & Evolution* 1: 0144.

D'Anunciação L., Guedes M.I.M., Oliveira T.L., Rehfeld I., Bonjardim C.A., Ferreira P.P., Souza Trindade G. de, Lobato Z.P., Kroon E.G. & Abrahão J.S. 2012. Filling One More Gap: Experimental Evidence of Horizontal Transmission of Vaccinia Virus Between Bovines and Rodents. *Vector-Borne and Zoonotic Diseases* 12: 61–64.

Diebold P.J., New F.N., Hovan M., Satlin M.J. & Brito I.L. 2021. Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR 1 2. *bioRxiv*: 2021.01.22.427834.

Domingos I.J. da S., Oliveira J.S. de, Rocha K.L.S., Oliveira D.B. de, Kroon E.G., Costa G.B. & Trindade G. de S. 2021. Twenty Years after Bovine Vaccinia in Brazil: Where We Are and Where Are We Going? *Pathogens* 10: 406.

Doud D.F.R., Bowers R.M., Schulz F., Raad M. De, Deng K., Tarver A., Glasgow E., Meulen K. Vander, Fox B., Deutsch S., Yoshikuni Y., Northen T., Hedlund B.P., Singer S.W., Ivanova N. & Woyke T. 2020. Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. *The ISME Journal* 14: 659–675.

Doud D.F.R. & Woyke T. 2017. Novel approaches in function-driven single-cell genomics. *FEMS Microbiology Reviews* 41: 538–548.

Drews G. 2000. The roots of microbiology and the influence of Ferdinand Cohn on microbiology of the 19th century. *FEMS Microbiology Reviews* 24: 225–249.

Emerson J.B., Adams R.I., Román C.M.B., Brooks B., Coil D.A., Dahlhausen K., Ganz H.H., Hartmann E.M., Hsu T., Justice N.B., Paulino-Lima I.G., Luongo J.C., Lymperopoulou D.S., Gomez-Silvan C., Rothschild-Mancinelli B., Balk M., Huttenhower C., Nocker A., Vaishampayan P. & Rothschild L.J. 2017. Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* 5: 86.

Ferreira J.M.S., Abrahão J.S., Drumond B.P., Oliveira F.M., Alves P.A., Pascoal-Xavier M.A., Lobato Z.I.P., Bonjardim C.A., Ferreira P.C.P. & Kroon E.G. 2008. Vaccinia virus: shedding and horizontal transmission in a murine model. *Journal of General Virology* 89: 2986–2991.

Fox G.E., Pechman K.R. & Woese C.R. 1977. Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics. *International Journal of Systematic and Evolutionary Microbiology* 27: 44–57.

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

Gibson U.E.M., Heid C.A. & Williams P.M. 1996. A novel method for real time quantitative RT-PCR. *Genome Research* 6: 995–1001.

Giovannoni S.J., Britschgi T.B., Moyer C.L. & Field K.G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345: 60–63.

Gobert G., Cotillard A., Fourmestraux C., Pruvost L., Miguet J. & Boyer M. 2018. Droplet digital PCR improves absolute quantification of viable lactic acid bacteria in faecal samples. *Journal of Microbiological Methods* 148: 64–73.

Guatelli J.C., Gingeras T.R. & Richman D.D. 1989. Nucleic acid amplification in vitro: detection of sequences with low copy numbers and application to diagnosis of human immunodeficiency virus type 1 infection. *Clinical Microbiology Reviews* 2: 217–226.

Heather J.M. & Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1–8.

Heid C.A., Stevens J., Livak K.J. & Williams P.M. 1996. Real time quantitative PCR. *Genome Research* 6: 986–994.

Hermon L., Hellal J., Denonfoux J., Vuilleumier S., Imfeld G., Urien C., Ferreira S. & Joulian C. 2019. Functional genes and bacterial communities during organohalide respiration of chloroethenes in microcosms of multi-contaminated groundwater. *Frontiers in Microbiology* 10: 89.

Hershey A.D. & Chase M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 36: 39–56.

Higuchi R., Fockler C., Dollinger G. & Watson R. 1993. Kinetic PCR Analysis: Real-time Monitoring of DNA Amplification Reactions. *Biotechnology* 11: 1026–1030.

Hindson B.J., Ness K.D., Masquelier D.A., Belgrader P., Heredia N.J., Makarewicz A.J., Bright I.J., Lucero M.Y., Hiddessen A.L., Legler T.C., Kitano T.K., Hodel M.R., Petersen J.F., Wyatt P.W., Steenblock E.R., Shah P.H., Bousse L.J., Troup C.B., Mellen J.C., Wittmann D.K., Erndt N.G., Cauley T.H., Koehler R.T., So A.P., Dube S., Rose K.A., Montesclaros L.,

Wang S., Stumbo D.P., Hodges S.P., Romine S., Milanovich F.P., White H.E., Regan J.F., Karlin-Neumann G.A., Hindson C.M., Saxonov S. & Colston B.W. 2011. High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry* 83: 8604–8610.

Hinegardner R.T. & Engelberg J. 1963. Rationale for a Universal Genetic Code. *Science* 142: 1083–1085.

Holley R.W., Apgar J., Everett G.A., Madison J.T., Marquisee M., Merrill S.H., Penswick J.R. & Zamir A. 1965. Structure of a Ribonucleic Acid. *Science* 147: 1462–1465.

Hugenholtz P., Chuvochina M., Oren A., Parks D.H. & Soo R.M. 2021. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *The ISME Journal*.

Hultman J., Tamminen M., Pärnänen K., Cairns J., Karkman A. & Virta M. 2018. Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. *FEMS Microbiology Ecology* 94.

Johnson D. 1998. Biodiversity and ecology of acidophilic microorganisms. *FEMS Microbiology Ecology* 27: 307–317.

Joyce E., Al-Hashimi A. & Mason T.J. 2011. Assessing the effect of different ultrasonic frequencies on bacterial viability using flow cytometry. *Journal of Applied Microbiology* 110: 862–870.

Kallmeyer J., Pockalny R., Adhikari R.R., Smith D.C. & D'Hondt S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences* 109: 16213–16216.

Kaster A.-K. & Sobol M.S. 2020. Microbial single-cell omics: the crux of the matter. *Applied Microbiology and Biotechnology* 104: 8209–8220.

Klein D.A. 2007. Microbial communities in nature: A postgenomic perspective. *Microbe* 2: 591–595.

Kwok S., Mack D.H., Mullis K.B., Poiesz B., Ehrlich G., Blair D., Friedman-Kien A. & Sninsky J.J. 1987. Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. *Journal of Virology* 61: 1690–1694.

Labonté J.M., Field E.K., Lau M., Chivian D., Heerden E. Van, Wommack K.E., Kieft T.L., Onstott T.C. & Stepanauskas R. 2015. Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology* 6: 1–11.

Lane D.J., Pace B., Olsen G.J., Stahl D.A., Sogin M.L. & Pace N.R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences* 82: 6955–6959.

Lee C.K., Herbold C.W., Polson S.W., Wommack K.E., Williamson S.J., McDonald I.R. & Cary S.C. 2012. Groundtruthing Next-Gen Sequencing for Microbial Ecology–Biases and Errors in Community Structure Estimates from PCR Amplicon Pyrosequencing Gilbert J.A. (ed.). *PLoS ONE* 7: e44224.

Lee P.K.H., Men Y., Wang S., He J. & Alvarez-Cohen L. 2015. Development of a Fluorescence-Activated Cell Sorting Method Coupled with Whole Genome Amplification To Analyze Minority and Trace Dehalococcoides Genomes

in Microbial Communities. *Environmental Science & Technology* 49: 1585–1593.

Lehman I.R., Bessman M.J., Simms E.S. & Kornberg A. 1958. Enzymatic Synthesis of Deoxyribonucleic Acid. *Journal of Biological Chemistry* 233: 163–170.

Lima M.T., Oliveira G.P., Afonso J.A.B., Souto R.J.C., Mendonça C.L. de, Dantas A.F.M., Abrahao J.S. & Kroon E.G. 2019. An Update on the Known Host Range of the Brazilian Vaccinia Virus: An Outbreak in Buffalo Calves. *Frontiers in Microbiology* 9: 3327.

Lister J.J. 1830. XIII. On some properties in achromatic object-glasses applicable to the improvement of the microscope. *Philosophical Transactions of the Royal Society of London* 120: 187–200.

Liu Y., Chen J. & Thygesen A. 2018. Efficient One-Step Fusion PCR Based on Dual-Asymmetric Primers and Two-Step Annealing. *Molecular Biotechnology* 60: 92–99.

Männistö M., Salkinoja-Salonen M. & Puhakka J. 2001. In situ polychlorophenol bioremediation potential of the indigenous bacterial community of boreal groundwater. *Water Research* 35: 2496–2504.

Männistö M.K., Tiirola M.A., Salkinoja-Salonen M.S., Kulomaa M.S. & Puhakka J.A. 1999. Diversity of chlorophenol-degrading bacteria isolated from contaminated boreal groundwater. *Archives of Microbiology* 171: 189–197.

Marcy Y., Ouverney C., Bik E.M., Losekann T., Ivanova N., Martin H.G., Szeto E., Platt D., Hugenholtz P., Relman D.A. & Quake S.R. 2007. Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences* 104: 11889–11894.

Martínez Martínez J., Martinez-Hernandez F. & Martinez-Garcia M. 2020. Single-virus genomics and beyond. *Nature Reviews Microbiology* 18: 705–716.

Mattila S., Ruotsalainen P., Ojala V., Tuononen T., Hiltunen T. & Jalasvuori M. 2017. Conjugative ESBL plasmids differ in their potential to rescue susceptible bacteria via horizontal gene transfer in lethal antibiotic concentrations. *Journal of Antibiotics* 70: 805–808.

Maxam A.M. & Gilbert W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* 74: 560–564.

Meselson M. & Stahl F.W. 1958a. The replication of DNA in Escherichia coli. *Proceedings of the National Academy of Sciences* 44: 671–682.

Meselson M. & Stahl F.W. 1958b. The Replication of DNA. *Cold Spring Harbor Symposia on Quantitative Biology* 23: 9–12.

Miranda J.B., Borges I.A., Campos S.P.S., Vieira F.N., Ázara T.M.F. De, Marques F.A., Costa G.B., Luis A.P.M.F., Oliveira J.S. De, Ferreira P.C.P., Bonjardim C.A., Silva S.L.M. Da, Eiras Á.E., Abrahão J.S., Kroon E.G., Drumond B.P., Paglia A.P. & Trindade G.D.S. 2017. Serologic and molecular evidence of vaccinia virus circulation among small mammals from different biomes, Brazil. *Emerging Infectious Diseases* 23: 931–938.

Moreira C.A. 1956. Notas sôbre a evolução do virus vacínico em animais silvestres da fauna brasileira. *Memórias do Instituto Oswaldo Cruz* 54: 93–94.

Mullis K.B. & Faloona F.A. 1987. [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In: *Methods in enzymology*, pp. 335–350.

Nocker A. & Camper A.K. 2006. Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. *Applied and Environmental Microbiology* 72: 1997–2004.

Pinheiro L.B., Coleman V.A., Hindson C.M., Herrmann J., Hindson B.J., Bhat S. & Emslie K.R. 2012. Evaluation of a Droplet Digital Polymerase Chain Reaction Format for DNA Copy Number Quantification. *Analytical Chemistry* 84: 1003–1011.

Qin H., Wang S., Feng K., He Z., Virta M.P.J., Hou W., Dong H. & Deng Y. 2019. Unraveling the diversity of sedimentary sulfate-reducing prokaryotes (SRP) across Tibetan saline lakes using epicPCR. *Microbiome* 7: 1–12.

Quick J., Loman N.J., Duraffour S., Simpson J.T., Severi E., Cowley L., Bore J.A., Koundouno R., Dudas G., Mikhail A., Ouédraogo N., Afrough B., Bah A., Baum J.H.J., Becker-Ziaja B., Boettcher J.P., Cabeza-Cabrerizo M., Camino-Sánchez Á., Carter L.L., Doerrbecker J., Enkirch T., Dorival I.G.-, Hetzelt N., Hinzmann J., Holm T., Kafetzopoulou L.E., Koropogui M., Kosgey A., Kuisma E., Logue C.H., Mazzarelli A., Meisel S., Mertens M., Michel J., Ngabo D., Nitzsche K., Pallasch E., Patrono L.V., Portmann J., Repits J.G., Rickett N.Y., Sachse A., Singethan K., Vitoriano I., Yemanaberhan R.L., Zekeng E.G., Racine T., Bello A., Sall A.A., Faye O., Faye O., Magassouba N., Williams C. V., Amburgey V., Winona L., Davis E., Gerlach J., Washington F., Monteil V., Jourdain M., Bererd M., Camara A., Somlare H., Camara A., Gerard M., Bado G., Baillet B., Delaune D., Nebie K.Y., Diarra A., Savane Y., Pallawo R.B., Gutierrez G.J., Milhano N., Roger I., Williams C.J., Yattara F., Lewandowski K., Taylor J., Rachwal P., J. Turner D., Pollakis G., Hiscox J.A., Matthews D.A., Shea M.K.O., Johnston A.M., Wilson D., Hutley E., Smit E., Caro A. Di, Wölfel R., Stoecker K., Fleischmann E., Gabriel M., Weller S.A., Koivogui L., Diallo B., Keïta S., et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530: 228–232.

Raghunathan A., Ferguson H.R., Bornarth C.J., Song W., Driscoll M. & Lasken R.S. 2005. Genomic DNA Amplification from a Single Bacterium. *Applied and Environmental Microbiology* 71: 3342–3347.

Rojo D., Méndez-García C., Raczkowska B.A., Bargiela R., Moya A., Ferrer M. & Barbas C. 2017. Exploring the human microbiome from multiple perspectives: factors altering its composition and function. *FEMS Microbiology Reviews* 41: 453–478.

Rosenthal C.K. 2009. The beginning. *Nature Cell Biology* 11: S6–S6.

Roux S., Hawley A.K., Torres Beltran M., Scofield M., Schwientek P., Stepanauskas R., Woyke T., Hallam S.J. & Sullivan M.B. 2014. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* 3, e03125.

Saiki R., Gelfand D., Stoffel S., Scharf S., Higuchi R., Horn G., Mullis K. & Erlich H. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487–491.

Saiki R., Scharf S., Faloona F., Mullis K., Horn G., Erlich H. & Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350–1354.

Sakowski E.G., Arora-Williams K., Tian F., Zayed A.A., Zablocki O., Sullivan M.B. & Preheim S.P. 2021. Interaction dynamics and virus–host range for estuarine actinophages captured by epicPCR. *Nature Microbiology*: 1–13.

Sambrook. 2012. *[Molecular cloning].*

Sanderson J. 2019. Historical Background. In: *Understanding Light Microscopy*, John Wiley & Sons, Ltd, Chichester, UK, pp. 787–790.

Sanger F., Brownlee G.G. & Barrell B.G. 1965. A two-dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology* 13: 373-IN4.

Sanger F. & Coulson A.R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94: 441–448.

Sanger F., Donelson J.E., Coulson A.R., Kossel H. & Fischer D. 1973. Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage f1 DNA. *Proceedings of the National Academy of Sciences* 70: 1209–1213.

Sanger F., Nicklen S. & Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74: 5463–5467.

Schrader C., Schielke A., Ellerbroek L. & Johne R. 2012. PCR inhibitors - occurrence, properties and removal. Journal of Applied Microbiology 113: 1014–1026.

Shendure J., Balasubramanian S., Church G.M., Gilbert W., Rogers J., Schloss J.A. & Waterston R.H. 2017. DNA sequencing at 40: past, present and future. *Nature* 550: 345–353.

Spencer S.J., Tamminen M. V., Preheim S.P., Guo M.T., Briggs A.W., Brito I.L., A Weitz D., Pitkänen L.K., Vigneault F., Virta M.P. & Alm E.J. 2016. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal* 10: 427–436.

Straiton J., Free T., Sawyer A. & Martin J. 2019. From Sanger sequencing to genome databases and beyond. *BioTechniques* 66: 60–63.

Summers W.C. 2009. History of Microbiology. In: *Encyclopedia of Microbiology*, Elsevier, pp. 593–607.

Suttle C.A. 2007. Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* 5: 801–812.

Suzuki M.T. & Giovannoni S.J. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and environmental microbiology* 62: 625–630.

Sykes P.J., Neoh S.H., Brisco M.J., Hughes E., Condon J. & Morley A.A. 1992. Quantitation of targets for PCR by use of limiting dilution. *BioTechniques* 13: 444–449.

Thompson L.R., Sanders J.G., McDonald D., Amir A., Ladau J., Locey K.J., Prill R.J., Tripathi A., Gibbons S.M., Ackermann G., Navas-Molina J.A., Janssen S., Kopylova E., Vázquez-Baeza Y., González A., Morton J.T., Mirarab S., Zech Xu Z., Jiang L., Haroon M.F., Kanbar J., Zhu Q., Jin Song S., Kosciolek T., Bokulich N.A., Lefler J., Brislawn C.J., Humphrey G., Owens S.M., Hampton-Marcell J., Berg-Lyons D., McKenzie V., Fierer N., Fuhrman J.A., Clauset A., Stevens R.L., Shade A., Pollard K.S., Goodwin K.D., Jansson J.K., Gilbert J.A. & Knight R. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551: 457–463.

Trindade G.S., Emerson G.L., Carroll D.S., Kroon E.G. & Damon I.K. 2007. Brazilian Vaccinia Viruses and Their Origins. *Emerging Infectious Diseases* 13: 965–972.

Vogelstein B. & Kinzler K.W. 1999. Digital PCR. *Proceedings of the National Academy of Sciences* 96: 9236–9241.

Watson J.D. & Crick F.H.C. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171: 737–738.

Weisburg W.G., Barns S.M., Pelletier D.A. & Lane D.J. 1991. 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* 173: 697–703.

Whitman W.B., Coleman D.C. & Wiebe W.J. 1998. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* 95: 6578–6583.

Wilson K.H., Blitchington R.B. & Greene R.C. 1990. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. *Journal of Clinical Microbiology* 28: 1942–1946.

Woese C.R. & Fox G.E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* 74: 5088–5090.

Woyke T., Doud D.F.R. & Schulz F. 2017. The trajectory of microbial single-cell sequencing. *Nature Methods* 14: 1045–1054.

Woyke T., Doud D.F.R. & Eloe-Fadrosh E.A. 2019. Genomes From Uncultivated Microorganisms. In: *Reference Module in Life Sciences*, Elsevier, pp. 437–442.

Woyke T. & Jarett J. 2015. Function-driven single-cell genomics. *Microbial Biotechnology* 8: 38–39.

Wu R. 1972. Nucleotide Sequence Analysis of DNA. *Nature New Biology* 236: 198–200.

Wu R. & Kaiser A.D. 1968. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology* 35: 523–537.

Yang K., Zhao Y., Ji M., Li Z., Zhai S., Zhou X., Wang Q., Wang C. & Liang B. 2021. Challenges and opportunities for the biodegradation of chlorophenols: Aerobic, anaerobic and bioelectrochemical processes. *Water Research* 193: 116862.

Yanofsky C. 2007. Establishing the Triplet Nature of the Genetic Code. *Cell* 128: 815–818.

Zallen D.T. 2003. Despite Franklin's work, Wilkins earned his Nobel. *Nature* 425: 15–15.
Zhulin I.B. 2016. Classic Spotlight: 16S rRNA Redefines Microbiology. *Journal of Bacteriology* 198: 2764–2765.

ORIGINAL PAPERS

I

MOLECULAR EVIDENCE OF ORTHOPOXVIRUS DNA IN CAPYBARA (HYDROCHOERUS HYDROCHAERIS) STOOL SAMPLES

by

Lara Ambrosio Leal Dutra, Gabriel Magno de Freitas Almeida, Graziele Pereira Oliveira, Jonatas Santos Abrahao, Erna Geessien Kroon & Giliane de Souza Trindade 2017

Archives of Virology 162: 439–448.

https://doi.org/10.1007/s00705-016-3121-3

II

SUCCESSFUL AEROBIC BIOREMEDIATION OF GROUNDWATER CONTAMINATED WITH HIGHER CHLORINATED PHENOLS BY INDIGENOUS DEGRADER BACTERIA

by

Anu Mikkonen, Kati Yläranta, Marja Tiirola, Lara Ambrosio Leal Dutra, Pauliina Salmi, Martin Romantschuk, Shelley Copley, Jukka Ikäheimo & Aki Sinkkonen
2018

Water Research 138: 118–128.

https://doi.org/10.1016/j.watres.2018.03.033

Water Research 138 (2018) 118–128

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Successful aerobic bioremediation of groundwater contaminated with higher chlorinated phenols by indigenous degrader bacteria

Anu Mikkonen [a, b, *], Kati Yläranta [a], Marja Tiirola [b], Lara Ambrosio Leal Dutra [b], Pauliina Salmi [b], Martin Romantschuk [a, c], Shelley Copley [d], Jukka Ikäheimo [e], Aki Sinkkonen [a]

[a] Department of Environmental Sciences, University of Helsinki, Niemenkatu 73, 15140 Lahti, Finland
[b] Department of Biological and Environmental Science, University of Jyväskylä, Survontie 9 C, 40500 Jyväskylä, Finland
[c] Institute of Environmental Sciences, Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia
[d] Cooperative Institute for Research in Environmental Sciences and Molecular, Cellular and Developmental Biology, University of Colorado, CO 80309 Boulder, United States
[e] Pöyry Finland Ltd, PL 50, 01621 Vantaa, Finland

ARTICLE INFO

Article history:
Received 17 October 2017
Received in revised form
10 March 2018
Accepted 12 March 2018
Available online 13 March 2018

Keywords:
In situ bioremediation
pcpB gene
Pentachlorophenol hydroxylase
Quantitative PCR
Ion PGM amplicon sequencing
Sphingomonas sensu lato

ABSTRACT

The xenobiotic priority pollutant pentachlorophenol has been used as a timber preservative in a polychlorophenol bulk synthesis product containing also tetrachlorophenol and trichlorophenol. Highly soluble chlorophenol salts have leaked into groundwater, causing severe contamination of large aquifers. Natural attenuation of higher-chlorinated phenols (HCPs: pentachlorophenol + tetrachlorophenol) at historically polluted sites has been inefficient, but a 4-year full scale *in situ* biostimulation of a chlorophenol-contaminated aquifer by circulation and re-infiltration of aerated groundwater was remarkably successful: pentachlorophenol decreased from 400 $\mu g\,L^{-1}$ to <1 $\mu g\,L^{-1}$ and tetrachlorophenols from 4000 $\mu g\,L^{-1}$ to <10 $\mu g\,L^{-1}$. The *pcpB* gene, the gene encoding pentachlorophenol hydroxylase - the first and rate-limiting enzyme in the only fully characterised aerobic HCP degradation pathway - was present in up to 10% of the indigenous bacteria already 4 months after the start of aeration. The novel quantitative PCR assay detected the *pcpB* gene *in situ* also in the chlorophenol plume of another historically polluted aquifer with no remediation history. Hotspot groundwater HCPs from this site were degraded efficiently during a 3-week microcosm incubation with one-time aeration but no other additives: from 5400 $\mu g\,L^{-1}$ to 1200 $\mu g\,L^{-1}$ and to 200 $\mu g\,L^{-1}$ in lightly and fully aerated microcosms, respectively, coupled with up to 2400% enrichment of the *pcpB* gene. Accumulation of lower-chlorinated metabolites was observed in neither *in situ* remediation nor microcosms, supporting the assumption that HCP removal was due to the aerobic degradation pathway where the first step limits the mineralisation rate. Our results demonstrate that bacteria capable of aerobic mineralisation of xenobiotic pentachlorophenol and tetrachlorophenol can be present at long-term polluted groundwater sites, making bioremediation by simple aeration a viable and economically attractive alternative.

1. Introduction

Pentachlorophenol (PCP) is a xenobiotic with no known natural sources (Crawford et al., 2007). Due to its high toxicity and poor biodegradability PCP is a commonly used model pollutant in environmental research. Large-scale production and use as a timber preservative from the 1930s to the 1980s, typically in a chlorophenol mixture with 2,3,4,6-tetrachlorophenol (TeCP) as the main component, resulted in persistent environmental pollution (Bryant and Schultz, 1994; Männistö et al., 2001; Rautio, 2011). In the phenolic form PCP is very poorly soluble in water, but with pK_a of 4.74 the majority of it is present as the phenolate form in near-neutral environments, rendering the sodium and potassium salts over four orders of magnitude more soluble and prone to polluting large aquifers (Olaniran and Igbinosa, 2011). The same applies to 2,3,4,6-TeCP with pK_a of 5.38. Some 30 years after banning, these

* Corresponding author. Department of Biological and Environmental Science, University of Jyväskylä, Survontie 9 C, 40500 Jyväskylä, Finland.
E-mail address: anu.mikkonen@jyu.fi (A. Mikkonen).

https://doi.org/10.1016/j.watres.2018.03.033
0043-1354/

higher chlorinated phenols (HCPs) are still found to contaminate soil and groundwater at Finnish sawmill sites where chlorophenol impregnants were used (Sinkkonen et al., 2013).

One reason for the environmental persistence of PCP may be the lack of microbial degraders capable of catabolising the xenobiotic thermodynamically stable molecule at the prevailing conditions. PCP biodegradation is well possible in anaerobic conditions, but it leads to lower-chlorinated compounds (Bouchard et al., 1996; D'Angelo and Reddy, 2000). The long-term persistence of the compound in boreal oxygen-deficient groundwaters (constantly around +8 °C) and successful degradation after water aeration and fertilisation (e.g. Järvinen et al., 1994; Tiirola et al., 2001a) indicates that aerobic organisms may rule the process at low temperatures. On the other hand, aerobic degradation of chlorophenols may compete with ferrous iron oxidation for the supplied oxygen (see Langwaldt et al., 2007 and references therein). Even though PCP degradation has been studied actively already in the 1980s, the genes encoding a PCP mineralisation pathway have been fully described only in *Sphingobium chlorophenolicum* L-1 (Copley et al., 2012; Hlouchova et al., 2012, and references therein). Its aerobic oxidative dechlorination pathway initiated by pentachlorophenol hydroxylase (PcpB) is notably inefficient, with PCP turnover to tetrachlorobenzoquinone proceeding with a k_{cat} of 0.024 s^{-1}. Interestingly, the same enzyme initiates degradation of TeCP, hydroxylating it directly to tetrachlorohydroquinone, the second metabolite in the PCP mineralisation pathway (Hlouchova et al., 2012). TeCP may even be the preferred substrate: according to Tiirola et al. (2002b) PcpB is only expressed in *Novosphingobium lentum* MT1 when induced by 2,3,4,6-TeCP. Both degradation monitoring in bacterial liquid cultures (2,3,4,6-TeCP; *Novosphingobium lentum* MT1) (Tiirola et al., 2002a) as well as kinetic studies with the purified enzyme (2,3,5,6-TeCP; *Sphingobium chlorophenolicum* L-1) (Hlouchova et al., 2012) have confirmed PcpB to be more reactive on TeCP than PCP. We are not aware of any other pathways for TeCPs biodegradation that have been characterised on a genetic and enzymatic level. Also trichlorophenol can be hydroxylated by PcpB (Hlouchova et al., 2012), but other pathways exist including that of betaproteobacterial genus *Cupriavidus* initiated by trichlorophenol monooxygenase TcpA (Sánchez and González, 2007).

pcpB gene homologs have been detected in multiple *Sphingomonas sensu lato* (i.e. sphingomonad) isolates that degrade PCP in culture, both in the United States and Canada (Crawford et al., 2007) as well as Finland (Tiirola et al., 2002b). However, there are only two earlier works reporting cultivation-independent detection of the *pcpB* gene (PCR amplification, cloning and sequencing), in Canadian mixed slurry bioreactors (Beaulieu et al., 2000) and in Montana soil biopiles (Crawford et al., 2007). Quantitative polymerase chain reaction (qPCR) assays have greatly aided testing, justification and monitoring of bioremediation of various recalcitrant pollutants, perchloroethene-degrading *Dehalococcoides* with reductive dehalogenase genes as one of the benchmark examples (Ritalahti et al., 2006). For HCP degraders, no such assays have been available; to our knowledge, there are no earlier reports of PCR determination of the presence, abundance or diversity of any HCP degradation gene at original polluted sites or during HCP biodegradation. Investigation of HCP degrader communities have relied on likely biased and often cumbersome laboratory cultivation protocols (Bécaert et al., 2000; Männistö et al., 2001).

The aim of the current work was to establish protocols that enable cultivation-independent investigation of sphingomonads and *pcpB* gene carrying organisms. Novel assays were set up to cover the known diversity of the *pcpB* gene, amplifying shorter (qPCR) and longer (sequencing) fragments of the *pcpB* gene. We hypothesised the *pcpB* gene to be directly detectable at historically contaminated groundwater sites and its relative abundance in bacterial community to increase upon oxygenation of the water. Firstly, the novel assays were tested on plume microbial community samples from Pursiala site with long-term chlorophenol contamination but no bioremediation history. Secondly, groundwater from Pursiala was incubated aerobically in microcosms to quantify changes in chlorophenols and degrader communities. Thirdly, we report chlorophenol dissipation and *pcpB* gene abundance in Kärkölä, another long-term contaminated aquifer, that underwent 4-year *in situ* biostimulation by aeration. Our study sheds light on the potential for aerobic degradation of HCPs and the responsible *in situ* degrader communities.

2. Materials and methods

2.1. Sites and sampling

We studied two former sawmill sites in Finland where KY5, a commercial fungicide consisting mostly of tetrachlorophenol (2,3,4,6-TeCP, 75–80%), pentachlorophenol (5–15%) and trichlorophenol (2,4,6-TriCP, 5–15%), had been used for timber treatment, resulting in persistent soil and groundwater pollution. In Pursiala, Mikkeli (N61.672, E27.290) KY5 was used for 32 years (1954–1986, with estimated 500–1500 kg a^{-1}) and in Kärkölä (N60.866, E25.268) for 54 years (1930–1984, with estimated 7000–10 000 kg a^{-1}) (Rautio, 2011). At both sites, the highest reported groundwater chlorophenol concentrations have been around 100 000 μg L^{-1}. Groundwater chemistry at both sites has been followed up regularly through an extensive system of sampling wells established to monitor the plume. The well casings extended 1 m above ground, they were capped and locked and only accessed by certified sampling personnel, who changed sampling hose between wells to minimize risk of cross-contamination.

In Pursiala, site investigation and follow-up has revealed no evidence of significant chlorophenol degradation or biotransformation *in situ* (Rautio, 2011). Five wells P1-P5, representing a 350-m long contamination gradient towards the CP hotspot, were sampled for microbial analyses in April 2013. In June 2016 Pursiala well P5 (hotspot) was sampled again for well bottom sediment slurry (1 L), as well as for groundwater (4 L) 1 m above the well bottom, to test the presence of aerobic PCP degradation potential. Sampling glass bottles were filled to the top and stored in the dark at 4 °C for 7 d before the start of the incubations. For both samplings, we received well-specific background data from the collaborating consultant companies who analysed groundwater samples taken minutes prior to the sediment sampling.

In Kärkölä, the hotspot (most contaminated soil) has been removed, preventing further CP emissions. Aerobic full-scale closed circle *in situ* bioremediation (biostimulation) was started in June 2012 to clean up the polluted aquifer, following the principle scheme in Fig. 1. Anaerobic contaminated water is pumped up from the recharge well, fully aerated by simple mechanical ejector (10–12 mg O_2 L^{-1}), and reinjected through an infiltration well located 200 m upstream in the CP plume. Aerated water infiltration rate has been approx. 65 m^3 d^{-1}. Changes in groundwater table and properties during the course of remediation were followed up with an extensive system of monitoring wells (Figure S1). In addition, two wells were sampled for microbial analyses in October 2012 and April 2013 (4 and 10 months of remediation, respectively): well K1 just next to infiltration well and well K2 halfway between the infiltration and recharge wells.

Standardised methods and accredited commercial laboratories were used in the groundwater analyses. CPs were quantified with GC-MS according to CSN EN 12 673, with a detection limit of 0.05 μg L^{-1} for each congener. Groundwater O_2 was measured

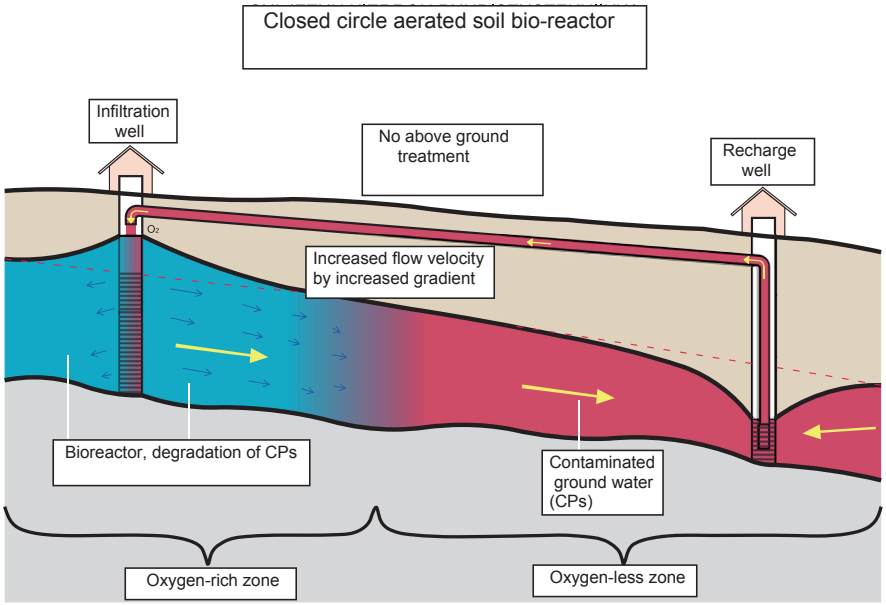


Fig. 1. Principle of Kärkölä *in situ* bioremediation of chlorophenols (CPs).

according to SFS-EN 25 813, alkalinity according to SFS 3005:1981 and soluble iron according to SFS 3027:1976. Groundwater humic substances were not analysed in this study, but according to earlier references little organic substrates other than chlorophenols are present at either of the sites (Rautio, 2011; Tiirola et al., 2002a).

Samples for microbial analyses (DNA extraction) were collected by pumping sedimented particles ('sediment slurry') from the bottom of the groundwater wells established to monitor the plume, at 13–41 m depth, to get more microbial biomass than that present in the groundwater alone. Another reason for sampling solids was to target sphingomonads, which were assumed to be sessile instead of planktonic (Pollock and Armentrout, 1999; Tiirola et al., 2002a). Well-bottom sediment slurry was collected by filling a sterile 50-ml plastic tube to the top, leaving minimal headspace. Chlorophenols are assumed to be highly soluble in these groundwaters of near-neutral pH, but if either limited solubility or biodegradation would affect the chlorophenol concentrations experienced by groundwater (planktonic) vs. sediment slurry (sessile) bacteria, this should show as different CP congener distributions in the two matrices. An in-house modification of the standard method (Sinkkonen et al., 2013) was used for CP test extractions from slurries from two Kärkölä wells sampled in October 2012; the relative abundances of the different chlorophenol congeners were very similar in water (μg L^{-1} - accredited commercial laboratory) and sediment slurries (mg kg^{-1} dwt - in-house method), with Pearson correlation above 0.99.

2.2. Nucleic acid extraction

The sampled slurries were allowed to settle in the dark at 4 °C for 4–6 d, after which the clear supernatant was decanted and discarded. Sediment DNA was extracted in duplicate with MO BIO PowerSoil DNA Isolation Kit (MO BIO, Carlsbad, CA, USA). A sufficient pellet for DNA extraction was collected by centrifuging multiple 2-mL volumes of the sedimented slurry at 10 000 g according to manufacturer's recommendations. To maximize DNA yield from clayey low-biomass subsurface sediments, we replaced the kit's PowerBead Tubes with 'G2 Beadbeating Tubes' purchased from GEUS, Copenhagen, Denmark. The G2 tubes are MO BIO PowerLyzer tubes with added G2 blocking agent (modified and fragmented salmon sperm DNA) to prevent immediate irreversible adsorption of released DNA on clay. The sample pellets and PowerSoil Bead Solution from the PowerLyzer Tubes were added to G2 tubes, vortexed briefly and allowed to stand for 5 min before addition of C1 solution and continuation of the extraction according to the manufacturer's protocol. Cells were disrupted by FastPrep FP120 Homogenizer (MP Biomedicals) for 40 s at 6.0 m s^{-1}. DNA was stored at −20 °C in aliquots.

DNA yield and quality were analysed with the Quant-IT PicoGreen dsDNA Assay Kit (Invitrogen, Thermo Fisher Scientific) and by agarose gel electrophoresis. The yields varied from 0.04 to 12.6 μg g^{-1} dry sediment but did not correlate with well depth, groundwater O_2, PCP or total CP concentrations. Because sediment sampling by pumping slurry from the groundwater well bottom is unlikely reproducible or representative in terms of particle size distribution, and because DNA extraction efficiency cannot be expected to be equal in sediments of different grain sizes and organic matter contents, we report the final sediment results not as absolute (per sediment dry weight) but as relative abundances.

2.3. pcpB gene primer design

The *pcpB* gene sequences were retrieved from GenBank, including sequences from pure cultures as reviewed by Crawford et al. (2007) as well as clones of Beaulieu et al. (2000) and Crawford et al. (2007), and aligned in ARB 5.2 (Ludwig et al., 2004). The alignment of the *pcpB* gene homologs detected by Saboo and Gealt (1998) in *Proteobacteria* not able to degrade PCP (*S. chlorophenolicum* L-1 *pcpB* gene positions 365–784) did not overlap with the majority of the other sequences and could thus not be used to exclude these non-sphingomonad non-degrader *pcpB* gene variants. As the forward primer pcpB-G by Beaulieu et al. (2000) (GGSTTCACSTTCAAYTTCGA, *S. chlorophenolicum* L-1 pos. 250–269) covered the full known diversity, only reverse primers for qPCR (short amplicon) and diversity analyses (longer amplicon) were designed (Table 1), taking into account both coverage as well

Table 1
Novel *pcpB* reverse primers. The degenerate positions are underlined. Positions refer to *S. chlorophenolicum* L-1 *pcpB* (total size 1617 bp).

Primer name	Application	Sequence (5'=>3')	Positions	Amplicon length with pcpB-G	Netprimer rating
pcpB_356r[a]	qPCR	TCGGTCTCATTCTGGTTGTAG TCGGTCTCATTCTGGTTATAG TCGGTCTCGTTCTGATTGTAG	336–356	107 bp	92–100
pcpB_512r	Sequencing	CCGATCACCCAGCGYGG	496–512	263 bp	75–88

[a] Equimolar mix of three primers - degeneracy forms not detected are not included.

as thermodynamic properties, and complementarity of the primer pair using NetPrimer (Premier Biosoft, 2013). PrimerBLAST (Ye et al., 2012) was used to test *in silico* the specificity of both primer pairs. Specificity of the *pcpB* gene primers was tested also *in vitro* with non-CP contaminated reference samples.

2.4. Gene enumeration by quantitative PCR: pcpB, sphingomonads, Bacteria

The *pcpB* gene was quantified with pcpB-G and pcpB_356r. qPCR was tested also with the longer amplicon, which showed separation of the samples comparable to the shorter amplicon (copy number Spearman correlation 0.87, $p = 6 \times 10^{-5}$, $n = 14$ Kärkölä DNA extracts), but lower amplification efficiency, as expected. Sphingomonads were quantified with the family-specific 16S rRNA gene primers of Zhou et al. (2012) SA429f (5′ TAAAGCTCTTTTACCCG3′) and SA933r (5′AAACCACATGCTCCACC3′). (Note that in the original reference, the reverse primer is given as in forward strand). Bacterial 16S rRNA genes were quantified with primers pE (5′AAACTCAAAGGAATTGACGG3′) and pF' (5′ACGAGCTGACGACAGCCATG3′) (Sinkkonen et al., 2014) with product length of approx. 170 bp, NetPrimer ratings of 85 and 84 with no cross-dimers, and coverage of 94% of *Bacteria* according to SILVA TestPrime (SILVA 117 RefNR database, one mismatch allowed).

All qPCR reactions were run with LightCycler 96 (Roche) in white LightCycler 8-Tube Strips (Roche) for increased sensitivity. Triplicate reactions of 20 μL consisted of 1 × FastStart Essential DNA Green Master (Roche) supplemented with 0.02% BSA (Fermentas, Thermo Fisher Scientific), 0.5 μM of both primers (Oligomer, Helsinki, Finland) and 2 μL of template (0.4–16 ng). The program consisted of preincubation at 95 °C for 10 min; cycling of melting at 95 °C for 10 s, annealing for 20 s and elongation at 72 °C for 20 s (30 s for the longer sphingomonad-16S rRNA amplicon); melting to 97 °C at 0.1 s^{-1} with 5 readings s^{-1}. The optimised annealing temperatures were 53 °C, 55 °C and 57 °C, and cycle numbers 40, 40 and 30, for the *pcpB* gene, sphingomonad-16S rRNA gene and bacterial 16S rRNA gene primers, respectively. Genomic DNA of fully sequenced strains purchased from DSMZ (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany), extracted with GeneJET Genomic DNA Purification Kit (Fermentas, Thermo Fisher Scientific) and quantified with Quant-IT PicoGreen dsDNA Assay Kit (Invitrogen, Thermo Fisher Scientific), were used as standards: *S. chlorophenolicum* L-1 (DSM 6824) for the *pcpB* gene (3.2×10^0–3.2×10^5 copies) and sphingomonad-16S rRNA gene (9.6×10^0–9.6×10^5 copies), *Cupriavidus necator* JMP134 (DSM 4058) for bacterial 16S rRNA gene (8.0×10^1–8.0×10^5 copies). Triplicate no-template-controls (with required threshold cycle >6.7 cycles higher than for the most dilute sample) and duplicate standard series ($R^2 > 0.985$) were included in each run using dilution aliquots of each standard, which was re-melted a maximum of three times to minimize degradation. Results were analysed with LightCycler 96 SW v.1.1 (Roche). The amplification efficiencies were ~90% for the *pcpB* gene and sphingomonad-16S rRNA gene and nearly 100% for bacterial 16S rRNA gene, with y-intercepts of the standard curves at 34, 38 and 34 cycles, respectively. Bacterial 16S rRNA gene copy numbers were found to correlate with extract DNA concentration (Spearman correlation 0.95, $p = 4 \times 10^{-9}$, $n = 18$ DNA extracts), supporting the reliability of this qPCR assay.

2.5. High throughput sequencing by Ion PGM: pcpB gene and bacterial 16S rRNA gene

The *pcpB* gene was amplified with pcpB-G and pcpB_512r in DNA Engine DYAD (MJ Research, St. Bruno, Canada) at 50 μL volume of 1 × Biotools buffer, 1 U Biotools Polymerase (Biotools Ultratools for 16S rRNA gene; Biotools, Spain), 0.2 mM of each dNTP (Fermentas, Thermo Fisher Scientific), 0.04% BSA (Fermentas, Thermo Fisher Scientific), and 1.0 μM of both primers (Oligomer, Helsinki, Finland). The PCR program consisted of preincubation at 94 °C for 5 min; 40 cycles (*pcpB* gene) or 35 cycles (sphingomonads) of melting at 94 °C for 45 s, annealing at 55 °C for 45 s and elongation at 72 °C for 30 s (40 s for longer sphingomonad-16S rRNA amplicon); final elongation at 72 °C for 5 min. The amplicons were cleaned up with High Pure PCR Product Purification Kit (Roche) for Ion Torrent PGM (Life Technologies, Thermo Fisher Scientific) sequencing. Barcoded sequencing adapters were ligated to the PCR products using the Ion Xpress Plus gDNA fragment library kit with Ion Xpress barcode adapters (Life Technologies, Thermo Fisher Scientific). Ion PGM Template OT2 400 (Life Technologies, Thermo Fisher Scientific) was used for the emulsion PCR and Ion PGM 400 Sequencing Kit (Life Technologies, Thermo Fisher Scientific) for final sequencing, which was done on the Ion 314 Chip (Life Technologies, Thermo Fisher Scientific) using the Ion Torrent PGM.

Bacterial 16S rRNA gene was amplified as in Mikkonen et al. (2011) with primers fD1 (5′AGAGTTTGATCCTGGCTCAG3′) and PRUN518r (5′ATTACCGCGGCTGCTGG3′) in final volume of 50 μL of 1 × Biotools buffer, 1 U Biotools Ultratools Polymerase (Biotools, Spain), 0.2 mM of each dNTP (Fermentas, Thermo Fisher Scientific), 0.05% BSA (Fermentas, Thermo Fisher Scientific), 0.6 μM of both primers (Oligomer, Helsinki, Finland) and 0.5–2 ng of template DNA. The cycling conditions were as for the *pcpB* gene except that the annealing and elongation steps were both 1 min and cycle number was limited to 28. To shorten the fragment size suitable for the sequencing chemistry, the products were re-amplified using barcoded adapter primers A_nn_fD1 (5′CCATCTCATCCCTGCGTGTCTCCGACTCAGnnAGAGTTTGATCMTGGCTCAG3′), where nn refers to a 10–12 bp long barcode. Shearing the product, Pippin prep purification of the 460–540 bp long constructs (Sage Science, Beverly, MA, USA) and ligation of the adapter P1 on the other side of the construct was done for pooled samples as previously described (Mäki et al., 2016) utilizing the chemistry of the Ion Xpress Plus gDNA fragment library kit (Life Technologies, Thermo Fisher Scientific). Downstream reactions were performed as described for the *pcpB gene* amplicons.

2.6. Ion Torrent amplicon sequence data analysis

High-throughput sequence data was analysed in mothur (Schloss et al., 2009). The *pcpB* gene sequence data were filtered based on quality and length (minimum quality window

average = 20 at window size = 10; maximum primer mismatch = 3 and barcode mismatch = 2; maximum homopolymer length = 6, no ambiguous bases, minimum length = 200), the primers and barcodes were trimmed out, and the data were rarefied to 175 sequences per sample. Unique sequences were aligned in ARB with the reference *pcpB* gene sequences used in primer design.

Raw bacterial 16S rRNA gene amplicon sequences were processed in mothur with the default methods (Schloss et al., 2011): filtered based on quality and length (minimum quality window average = 15 at window size = 10; maximum primer and barcode mismatch = 1, maximum homopolymer length = 8, no ambiguous bases, minimum length = 150) and screened to leave only well-aligned sequences (132–249 bp). These were preclustered with maximum difference = 2 (approx. 99% identity), chimeras were detected and removed, remaining sequences were classified against Greengenes taxonomy, based on which sequences of chloroplasts, mitochondria, *Archaea*, *Eukaryota* and unknown kingdom were removed. The remaining 60 000 sequences were classified as *Bacteria* and 94% could be identified to phylum level. ClearCut was used to calculate Neighbour Joining tree based on uncorrected pairwise distances between aligned sequences. Data was rarefied to 969 sequences per sample prior to calculation of phylogenetic α-diversity (command: phylo. diversity). Bacterial 16S rRNA gene sequences from *in situ* samples with MIMARKS details have been submitted to NCBI Sequence Read Archive under BioProject PRJNA349270.

2.7. Microcosm incubations for Pursiala groundwater

Three replicate incubation glass bottles of 162 mL, closed airtight with butyl septum, were filled to the top with partly or fully aerated well P5 groundwater, inoculated with 1% v/v settled sediment slurry of 14% dry matter aerated fully (vigorous bubbling and stirring at 300 rpm for 2 h to minimize chemical oxygen demand upon incubations), and amended with 10% ignited sand (w/v, grainsize 0.5–1.2 mm, Saint-Gobain Weber, Helsinki, Finland) to facilitate mixing. Preparations and incubations were carried out at 15 °C. The two initial aeration levels used were low, with 2 mg L^{-1} O_2 (16–20% air saturation at 15 °C), and high, with 10 mg L^{-1} O_2 (101–104% air saturation). Bottles were incubated for 21 d in the dark, mixing every other day by gently rotating each bottle 7 times. Oxygen concentration was monitored with non-invasive sensor spots and Microx transmitter (PreSens, Germany). After the last day mixing, sediment was allowed to settle for 1 h before decanting 60 mL water for chlorophenol analysis. The remaining contents of the bottle were shaken horizontally at 200 rpm for 5 min to gently detach microbes from the surfaces and homogenize flocs before sampling the water and suspended sediment for DNA-based analyses.

Concentrations of chlorophenols were analysed from the groundwater before and after incubations by SGS Inspection Services (Kotka, Finland) according to CSN EN 12 673. DNA was extracted from the initial sediment slurry (1.62 mL, matching the inoculation volume) with MO BIO PowerLyzer PowerSoil DNA Isolation Kit (MO BIO, Carlsbad, CA, USA), and from the initial groundwater (150 mL) and incubated water (40 mL) with MO BIO PowerWater DNA Isolation Kit (MO BIO, Carlsbad, CA, USA). The *pcpB* gene was quantified as described above, except for using Maxima SYBR Green/Fluorescein qPCR Master Mix (Thermo Scientific) and Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories, Inc.). Bacterial communities were characterised by Ion PGM high throughput sequencing of 16S rRNA gene V1-V2 region, amplified from 2.5 ng of template DNA using Maxima SYBR Green/Fluorescein qPCR Master Mix (Thermo Scientific) to amplify first with general bacterial

Table 2
Chlorophenol and oxygen concentrations, as well as bacterial, sphingomonad and *pcpB* gene copy numbers and bacterial α-diversity, up the chlorophenol plume in Pursiala site (no active remediation). Groundwater samples showing contamination level were taken as for the regular monitoring, typically 1 m above the well bottom (the wells were perforated from the bottom to a height of 10–30 m). DNA-based results are averages of duplicate DNA extractions from well-bottom sediment (both presented in Table S1).

Well code	Well depth (m)	Sampling time	Water O_2 (mg L^{-1})	Water total CPs (μg L^{-1})[a]	Water PCP (μg L^{-1})	Water TeCPs (μg L^{-1})	Water TriCPs (μg L^{-1})	Bacterial 16S rRNA gene copies (/ng DNA)	Sphingomonad 16S rRNA gene copies (/ng DNA)	*pcpB* copies (/ng DNA)	*pcpB*/bacterial 16S rRNA gene copies (%)	Bacterial Phylogenetic diversity
P1	41	Apr-2013	0.7	49	3	25.2	11	4.1×10^5	1.2×10^2	$<5.1 \times 10^0$	<0.001	47
P2	23	Apr-2013	0.6	1	<0.05	0.53	0.19	5.2×10^5	6.0×10^2	$<6.3 \times 10^1$	<0.010	31
P3	22	Apr-2013	0.5	841	74	591	140	4.1×10^5	1.2×10^3	$\leq 1.1 \times 10^1$	≤0.003	33
P4	32	Apr-2013	1.2	7380	390	5870	108	6.4×10^5	1.0×10^4	1.4×10^3	0.23	19
P5	20	Apr-2013	0.6	14 000	450	13 000	496	7.4×10^5	3.5×10^3	2.4×10^2	0.03	27

[a] For comparison, total chlorophenol concentration in drinking water must be < 10 μg L^{-1}.

primers 27 F (5′AGAGTTTGATCMTGGCTCAG3′) and 338R (5′TGCTGCCTCCCGTAGGAGT3′) for 30 cycles (anneal at 53 °C) and to append the sequencing primers and barcodes in additional 8 cycles. Products purified with Agencourt AMPure XP purification system (Beckman Coulter Life Sciences, Indianapolis, IN, USA) were pooled for 400 bp library sequencing, template prepared with Ion PGM Hi-Q View OT2 Kit and sequenced with Ion PGM Hi-Q View Sequencing Kit on Ion 316 Chip v2 chip (all Life Sciences, Thermo Fisher Scientific). The data was analysed in mothur as described above.

3. Results

3.1. Potential aerobic HCP degraders in chlorophenol plume

The qPCR amplification efficiency with the newly developed primers for pentachlorophenol hydroxylase gene (*pcpB* gene) was ~90% using the standard curve of the positive control strain (*Sphingobium chlorophenolicum* L-1). Both the novel primers produced amplicons of expected size with pcpB-G (Table 1). Product melting temperatures were the same for all the samples, and within 0.5 °C from the melting temperature of the positive control (i.e. standard). Specificity of the primers was successfully verified both *in silico* (no false positives from sequence databases) and *in vitro* (no PCR product detected from non-PCP contaminated reference samples: sandy and humus soils with and without hydrocarbon contamination).

The *pcpB* gene, however, was successfully detected and quantified by direct qPCR in the CP plume at the historically polluted Pursiala site. Relative abundance of the gene in the bacterial community was highest in the second-most contaminated sampling well (2‰ of bacterial 16S rRNA gene copy number), and approx. tenfold lower in the hotspot, possibly related to lower oxygen concentration (Table 2). The *pcpB* gene copy numbers were at the detection limit of the assay at the midpoint of the studied 350-m plume (slightly above the limit of 20 copies μL^{-1} in one duplicate DNA extract, below it in the other), and below detection limit at the two sampling wells furthest from the hotspot (HCP <30 $\mu g\ L^{-1}$ and $O_2 < 1\ mg\ L^{-1}$).

As expected, the number of bacterial 16S rRNA genes per quantity of extracted DNA was unchanged in the plume, but the abundance of sphingomonads varied hundredfold (Table 2). Significant positive correlation was observed between sphingomonad and *pcpB* gene assay results (Spearman rho = 0.73, p = 0.016, n = 10 extracts). *Proteobacteria* represented over 50% of the bacterial community characterised by high-throughput sequencing (Figure S2). *Betaproteobacteria* was the most abundant class (especially *Burkholderiales*), whereas the relative abundance of *Alphaproteobacteria* (especially *Sphingomonadales*) increased up the plume. The abundant *Sphingomonadaceae* OTUs that could be classified down to the genus level were identified as either *Sphingomonas* or *Novosphingobium*.

3.2. HCP degradation in aerated microcosms

Regardless of detection of the *pcpB* gene, there was no prior knowledge of the viability and HCP degradation potential of the *in situ* bacterial community at the Pursiala site, where no bioremediation had been studied or attempted. We returned to the site three years after the initial analysis and resampled the hotspot well P5 to test biotic effects of groundwater aeration in microcosms. PCP concentration was unchanged (440 $\mu g\ L^{-1}$), whereas TeCPs were reduced by more than half, still remaining at extreme concentration of 4900 $\mu g\ L^{-1}$. With proportions of 8% and 86% of the total chlorophenol concentration (5700 $\mu g\ L^{-1}$), respectively, this isomer distribution resembles relatively closely that of the original KY5 impregnant.

Vigorous 2 h aeration in the laboratory *per se*, without incubation, did not cause groundwater HCP loss (PCP concentration of fully aerated water was 450 $\mu g\ L^{-1}$ and TeCPs 4800 $\mu g\ L^{-1}$), whereas following 3-week incubation at 15 °C in the dark decreased both total PC and HCP concentrations by up to 96% (Fig. 2). The original groundwater PCP concentration decreased on average by 13% in the

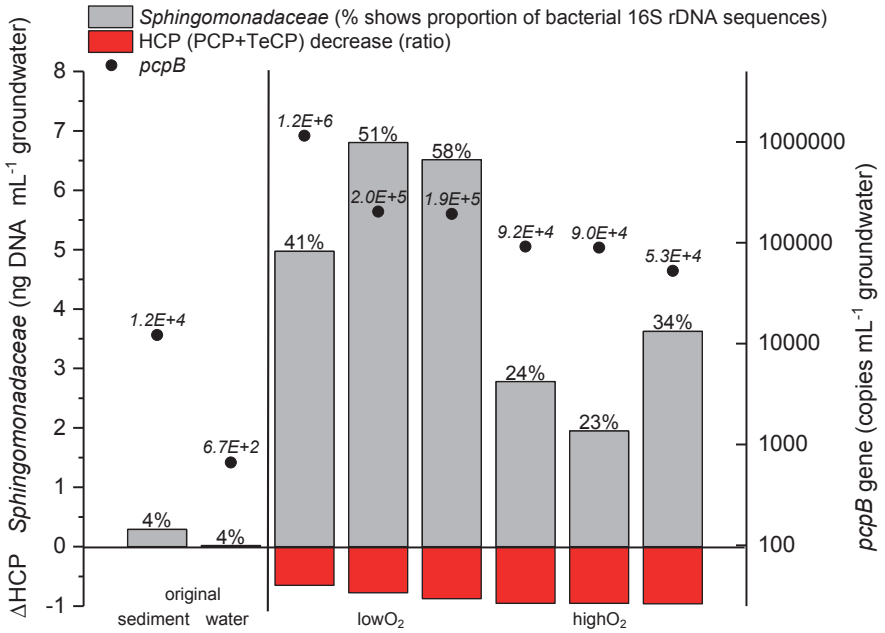


Fig. 2. Decrease in Pursiala groundwater HCPs coupled with increase in *Sphingomonadaceae* and the *pcpB* gene upon aerobic incubation in microcosms. Groundwater was aerated to low (2 mg L-1) or high (10 mg L-1) O2 concentration, amended with 1% sediment slurry from the same well and incubated in tightly sealed triplicate bottles for 21 d at 15 °C in the dark. The abundance of sphingomonads was calculated by multiplying the total microbial biomass (DNA yield) by the relative abundance of sphingomonads (proportion *Sphingomonadaceae* in bacterial communities analysed by 16S rRNA gene high throughput sequencing).

bottles of low initial aeration level (final range 330–440 μg L^{-1}) and TeCPs concentration by 83% (final range 220–1500 μg L^{-1}). In the bottles of high initial aeration level, PCP decreased on average by 51% (final range 200–240 μg L^{-1}) and TeCPs on average by 100% (final range 5–6 μg L^{-1}). Trichlorophenols decreased on average by 82% and 99% in the microcosms of high and low initial aeration level, respectively. Proportion of PCP enriched during incubation, being 17–61% of total CPs in the less and 95% in the fully aerated bottles.

Fully aerated bottles remained aerobic throughout the incubation (endpoint air saturation 64%), whereas oxygen was depleted from the less aerated bottles in the middle of the experiment. However, known anaerobic PCP degradation metabolites 2,3,4,5-TeCP and 2,4,5-TriCP (Field and Sierra-Alvarez, 2008) did not accumulate but reduced on average by 12% and 49%, respectively. Moreover, also total trichlorophenols decreased on average by 82%, dichlorophenols by 37% and monochlorophenols by 90% in the less aerated bottles, showing no indication anaerobic dehalogenation after oxygen depletion. pH varied from 7.8 to 8.1, at which chlorophenols are highly soluble, and losses due to sorption were likely minimal as no foreign organic matter was added to the glass bottles. 0.5% $AgNO_3$ used to kill abiotic control bottles, on the other hand, had dropped pH to 4–5, rendering PCP and TeCPS in poorly soluble protonated forms. This unfortunately prevented accurate quantification of possible abiotic losses.

The relative abundance of *Sphingomonadaceae* in both initial sediment and water upon start of the incubation was 4%, increasing on average to 50% and 27% in the microaerophilic and fully aerated bottles, respectively (each based on >15 000 bacterial 16S rDNA sequences) (Fig. 2). The *pcpB* gene copy number was 1.7×10^3 copies (ng extracted DNA)$^{-1}$ in the initial sediment and similar 1.5×10^3 copies (ng extracted DNA)$^{-1}$ (or 6.7×10^2 mL^{-1}) in the initial water, increasing on average by 2400% (to 5.2×10^5 copies mL^{-1}) and by 360% (to 7.8×10^4 copies mL^{-1}) in the microaerophilic and fully aerated bottles, respectively.

3.3. HCP degradation upon in situ aeration

Small-scale aeration tests similar to the one described above had earlier been performed at the Kärkölä site (data not shown), encouraging initiation of full-scale aquifer aeration by closed circle groundwater circulation in June 2012. After 47 months of *in situ* biostimulation, groundwater total CP concentration at the sampling well in the middle of the Infiltration and Recharge wells had decreased from 6000 to <100 μg L^{-1}, TeCPs from 4000 to <10 μg L^{-1}, and PCP from 400 to <1 μg L^{-1} (Fig. 3). Before aeration the percentage of PCP out of total CPs had been stably 4.5% for three years. After the start of the remediation its relative abundance increased slightly (8-month average 5.6%), but also this most chlorinated isomer was degraded rapidly with a dissipation curve resembling that of TeCPs. Also total chlorophenol concentration decreased (fourth year average in well K2 56 μg L^{-1}), indicating that chlorophenols were not only dechlorinated but mineralised. No major changes were observed in groundwater chloride concentrations (approx. 20 mg L^{-1} close to infiltration well and at recharge well, and 25 mg L^{-1} at K2 between them) were observed, possibly due to the comparatively low concentration of contaminants. However, a decreasing trend over time was seen in groundwater alkalinity in all the three monitoring wells, likely caused by HCl produced upon chlorophenol mineralisation (Figure S3).

Oxygen concentration in well K1 increased from 0.5 mg L^{-1} to 9.7 mg L^{-1} during the first month of remediation, and groundwater close to infiltration well remained well-aerated (>5 mg L^{-1}) throughout the 4 years. In well K2 oxygen concentrations were mostly above detection limit (0.2 mg L^{-1}, up to 2.7 mg L^{-1}) during the first 15 months of remediation, but undetectable after that. Differences in groundwater aeration from the infiltration well to the recharge well were evidenced by differences in groundwater iron content, which was <1 mg L^{-1} in well K1 but approximately 7 mg L^{-1} in the recharge well (Figure S3). Iron precipitation in and around the infiltration well gradually decreases infiltration capacity, and regular removal of the iron sludge and occasional acid wash of the well screen were required to maintain infiltration capacity stable for four years.

Microbial communities in Kärkölä were sampled 4 and 10 months after the start of the *in situ* aeration. The *pcpB* gene was detected in high abundance of up to several percent of bacteria both next to the infiltration well (K1, aerobic zone) and halfway between infiltration and recharge wells (K2, transition zone), tenfold higher abundance than in the Pursiala well with the highest *pcpB* gene abundance (Table 3). Again significant positive correlation was observed between the sphingomonad and *pcpB* gene copy numbers (Spearman rho = 0.76, p = 0.031, n = 8 extracts). *Alphaproteobacteria* (especially *Sphingomonadales*) were on average more abundant than in Pursiala (Figure S2), corresponding with higher sphingomonad abundance (Table 3). Well K1 October sample bacterial community resembled those of Pursiala, whereas in the next April the candidate phylum TM7 had become curiously abundant. In the transition zone well K2 Epsilonproteobacteria (specifically the sulfur-oxidizing genera *Sulfuricurvum* and *Sulfurimonas*)

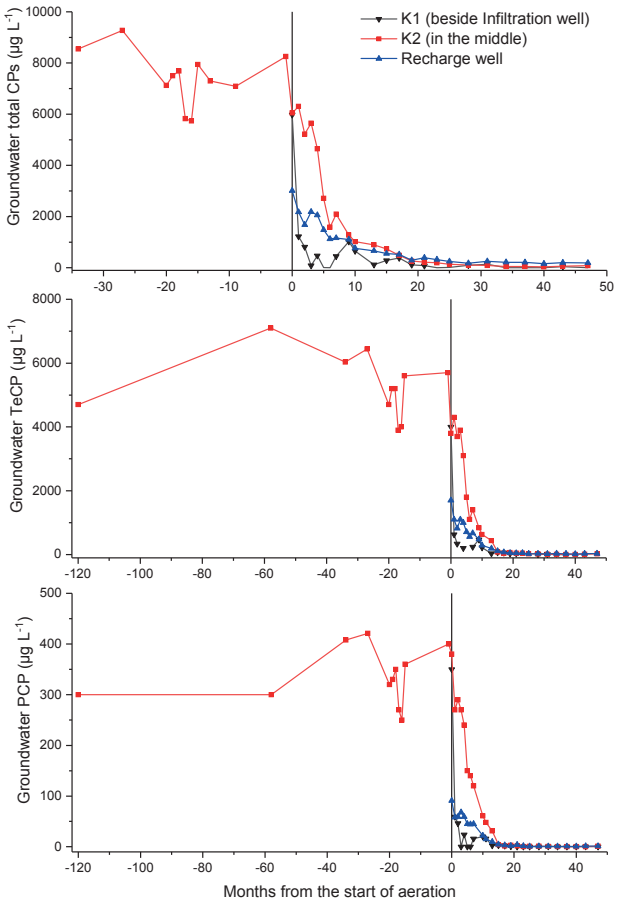


Fig. 3. Decrease of chlorophenol (CP) concentrations in groundwater at the Kärkölä site upon 4-year closed-circle aerobic *in situ* biostimulation. Note the different time scales on x-axes; before year 2009 only the main fractions of the polluting KY5 impregnant were quantified, after that all the 18 chlorophenol isomers.

Table 3
Chlorophenol and oxygen concentrations, as well as bacterial, sphingomonad and *pcpB* gene copy numbers and bacterial α-diversity, after 4 and 10 months of *in situ* aeration of Kärkölä site. Groundwater samples were taken as for the regular monitoring, 1.5–2.6 m above the well bottom. DNA-based results are averages of duplicate DNA extractions from well-bottom sediment (both presented in Table S1).

Well code	Well depth (m)	Sampling time	Water O_2 (mg L^{-1})	Water total CPs (μg L^{-1})[a]	Water PCP (μg L^{-1})	Water TeCPs (μg L^{-1})	Water TriCPs (μg L^{-1})	Bacterial 16S rRNA gene copies (/ng DNA)	Sphingomonad 16S rRNA gene copies (/ng DNA)	*pcpB* copies (/ng DNA)	*pcpB*/bacterial 16S rRNA gene copies (%)	Bacterial Phylogenetic diversity
K1	12	Oct-2012	7.2	470	24	210	112	4.5×10^5	1.3×10^4	1.2×10^4	2.7	20
		Apr-2013	8.9	657	16	232	301	3.2×10^5	8.1×10^3	1.4×10^4	4.2	32
K2	16	Oct-2012	0.4	4650	240	3130	1050	3.0×10^5	6.1×10^3	1.4×10^4	4.6	18
		Apr-2013	<0.2	1020	48	648	273	4.4×10^5	2.1×10^3	6.0×10^3	1.3	23

[a] For comparison, total chlorophenol concentration in drinking water must be < 10 μg L^{-1}.

dominated the bacterial community (Figure S2).

A longer *pcpB* gene amplicon was successfully amplified from all the Kärkölä samples, as well as from well P4 in Pursiala (the one with the highest *pcpB* gene abundance), enabling investigation of diversity of the gene with Ion PGM high-throughput sequencing. Except for obvious sequencing errors related to difficulties in Ion-Torrent homopolymer calling (indels resulting in frameshift that would make the enzyme dysfunctional), sequences from all the samples represented a single variant matching the *pcpB* gene of *Novosphingobium lentum* MT1. The same variant was recovered by direct Sanger sequencing of the purified *pcpB* gene amplicons from spring samples (without cloning; successful sequencing with both forward and reverse primers done at the DNA Sequencing and Genomics Laboratory core facility, University of Helsinki, Finland; Data S1).

4. Discussion

The use of chlorinated phenols as impregnants in the sawmill industry has caused wide-scale pollution of soil and groundwater. Decades after banning, these pollutants are still found in high concentrations at such sites. As the chlorophenol degradation potential of indigenous microbial community has been shown by cultivated groundwater isolates from the chlorophenol plume (Männistö et al., 1999, 2001), the explanation for the persistence of chlorophenols, at least at this site, must lie in unfavourable environmental conditions. Successful degradation of both higher chlorinated phenols (HCPs, i.e. pentachlorophenol and tetrachlorophenol) was observed in an aerated and fertilized bioreactor fed with groundwater from the historically polluted Kärkölä aquifer in the 1990s (reviewed by Langwaldt et al., 2007), but the efficiency of this pump-and-treat technology was insufficient to reduce aquifer chlorophenol content. Therefore full-scale *in situ* aeration of Kärkölä site by re-infiltration of aerated polluted groundwater upstream of the pumping (recharge) well was started in 2012. Neither additives such as fertilisers nor bioaugmentation was used. Impressive decrease, following first or second-order degradation kinetics, was observed in HCP concentrations during the four years of observation (2012–2016). PCP/TeCP ratio increased slightly in the beginning, but biostimulation by aeration brought down the concentration of also the fully chlorinated phenol remarkably efficiently: after few-months lag period, PCP concentration at the long-term monitoring well (halfway the 200 m line between infiltration and recharge wells) came down by >99.5% in four years. To our knowledge, this is the first report of successful aerobic full-scale *in situ* bioremediation of HCP-contaminated aquifer. There is at least one documented case of successful anaerobic (dechlorinative) PCP remediation *in situ* with permeable barrier technology, but this required both bioaugmentation as well as supplemental electron donor (Cole, 2000).

The key to successful aerobic *in situ* remediation in Kärkölä was assumedly the proliferation and competitiveness of native bacteria capable of HCP degradation once oxygen became available. Verified HCP degraders have been isolated from the Kärkölä groundwater and on-site fluidised-bed bioreactor before (Männistö et al., 2001; Tiirola et al., 2002b), but there was no prior cultivation-independent knowledge of their abundance or competitiveness in the aquifer *in situ* conditions. Iron oxidizing bacteria compete for the newly available oxygen, but are according to Langwaldt and Puhakka (2003) less competitive than chlorophenol oxidisers in the contaminated groundwater. Since little readily utilisable substrates other than chlorophenols are present in the Kärkölä groundwater (Tiirola et al., 2002a), heterotrophs able to utilize them as carbon and energy source are likely to get competitive advantage.

We developed novel primers for the *pcpB* gene, encoding the first and rate-limiting enzyme in the only well-known HCP mineralisation pathway, which enabled first ever cultivation-independent detection and quantification of HCP degradation potential *in situ*. Notably high relative abundance of the *pcpB* gene in the bacterial community was observed in the Kärkölä monitoring well-bottom samples taken during the fastest phase of the degradation, with 3–10% of bacterial cells possessing the pathway (assuming equal qPCR efficiency for the 16S rRNA gene and *pcpB* gene, and that each cell has two copies of the former and one of the latter). In groundwater samples the relative abundance might have been either lower (due to preference for growth attached to surfaces), higher (due to preference for assumedly higher oxygen level in the water), or similar. Anyhow, such a percentage of potential degraders of any persistent pollutant in an *in situ* community is exceptionally high. Considering that the gene was not found in non-chlorophenol polluted reference samples, and even laboratory pure cultures are known to readily discard it in the absence of chlorophenol selection pressure, such high abundance can only result from proliferation of the degraders at the *in situ* conditions, matching well with the first or second-order decrease in HCPs upon aeration. Unfortunately the final proof of *in situ* HCP decrease by PcpB-initiated biodegradation remains to lack, as we cannot conclusively verify growth or increased metabolic activity of the degraders during the remediation; no samples were taken for microbial analyses before the aeration started, and no RNA was extracted for transcript analyses. Moreover, the PcpB-initiated aerobic mineralisation pathway, where the first reaction is the rate-limiting step, is not known to accumulate any metabolites that could be observed in the groundwater.

Another aquifer historically polluted with chlorophenol timber preservatives, Pursiala, was sampled to investigate the presence and abundance of the *pcpB* gene in HCP plume at a site with no remediation history. The *pcpB* gene was detected in three sampling wells closest to the hotspot. No product was amplified from the two wells further down the plume (groundwater PCP $\leq 3\ \mu g\ L^{-1}$, TeCPs $\leq 25\ \mu g\ L^{-1}$), supporting the specificity of the PCR assay for chlorophenol-contaminated environments. However, this detection of the *pcpB* gene *in situ*, even if only in contaminated soils, does not mean the organisms present produce functional PcpB at the *in situ* chemical and physical conditions. Moreover, detection of the only well-characterised pathway does not mean that the bacteria carrying it account for any HCP decrease; in fact, in Pursiala the persistence of the pollution suggests the opposite. We thus aerated the Pursiala hotspot groundwater, inoculated it with sediment from the same well (where the *pcpB* gene was earlier detected), and observed changes in chlorophenols, the *pcpB* gene and bacterial community in microcosms. After 21-day incubation at 15 °C, HCP concentration decreased by 96% and the *pcpB* gene increased by 360% in the fully aerated microcosms, whereas HCP concentration decreased by 77% and the *pcpB* gene increased by 2400% in the lightly aerated microcosms. These results provide strong evidence that catabolically capable degraders (carrying the *pcpB* gene) were present at the contaminated aquifer and could be activated by aeration to efficiently degrade both TeCPs and PCP. Lower HCP degradation rate in the less-aerated microcosms was likely caused by lack of measurable oxygen after day 10, ceasing aerobic biodegradation - however, no increase in anaerobic dehalogenation products was observed either. Lower counts of the *pcpB* gene (and lower relative and absolute abundance of sphingomonads, the assumed host taxon) in the fully than the lightly aerated microcosms is more curious. This may be explained by the fact that PcpB itself causes oxidative stress by releasing H_2O_2 in uncoupled, futile cycles - the enzyme is notably inefficient and apparently represents "evolution in progress" (Hlouchova et al., 2012). Indeed, *Novosphingobium lentum* MT1, the dominant strain from the Kärkölä chlorophenol-treating bioreactor ran in the late 1990s, was found to be microaerophilic and to degrade HCPs more efficiently at 8 °C than at room temperature (Tiirola et al., 2002a).

Other suggested routes for bacterial PCP degradation include the actinobacterial (formerly rhodococcal but renamed mycobacterial) pathway initiated by cytochrome P-450 (Uotila et al., 1992) and strictly anaerobic dehalogenases (Field and Sierra-Alvarez, 2008). Unfortunately, the genes are not well-characterised and specific primers to study either of these pathways are, to our knowledge, not available; shotgun metagenomics or metaproteomics on controlled study systems could shed light on their relevance *in situ*. However, *Mycobacterium* was not abundant in the bacterial communities of the aerated chlorophenol-degrading systems, *Mycobacteriaceaea* relative abundances being <0.5% in the Kärkölä wells and <0.03% in the Pursiala microcosms. In poorly aerated contaminated environments especially the anaerobic degradation pathways would be highly beneficial, and indeed some of the lower-chlorinated phenols observed in the aquifer may originate from reductive dechlorination. Unfortunately there is no indication of such natural metabolism having markedly decreased contaminant concentrations in Pursiala (Rautio, 2011); one reason for this may be that anaerobic biodegradation of chlorophenols generally requires addition of organic electron donors (Field and Sierra-Alvarez, 2008). Former suggested anaerobic degraders did not seem to be major members of the indigenous communities; for example *Clostridium* and *Desulfitobacterium* were absent or formed max 1‰ of the bacterial communities in the Kärkölä and Pursiala *in situ* samples. Accumulation of lower-chlorinated phenols, typical of anaerobic dechlorination, was not observed upon bioremediation treatments either, even after running out of oxygen (Kärkölä Recharge well or Pursiala less-aerated microcosms). Also fungi are known to have several pathways for PCP modification and at least partial degradation, but the associated enzymes do not seem to be chlorophenol-specific but cometabolic (Field and Sierra-Alvarez, 2008). Our findings suggest that the curious aerobic HCP mineralisation pathway initiated by pentachlorophenol hydroxylase PcpB is present and provides some competitive advantage even in Pursiala non-aerated plume. Also in Pursiala groundwaters with little other carbon sources (Rautio, 2011), HCPs can serve as an abundant source of carbon and energy, if any oxygen becomes available. PcpB-initiated pathway grants the host cell access to this rich resource, but causes marked metabolic stress to the host itself (Hlouchova et al., 2012). The *pcpB* gene was both transferred horizontally and discarded from the genomes of previously studied isolates surprisingly readily, especially considering that the gene is likely present as a single copy situated in a chromosome, not in a plasmid (Tiirola et al. 2002a, 2002b; Copley et al., 2012). The presence of the *pcpB* gene might thus serve as a "biosensor", indicating the current situation - bioavailability and/or ecological effect of HCPs - in the *in situ* bacterial community.

Earlier studies on culturable isolates have established that the *pcpB* gene is found exclusively in sphingomonads (Tiirola et al., 2002b; Copley et al., 2012). Amplification of fragments somewhat homologous to the *pcpB* gene has been reported also from non-degrader *Gammaproteobacteria* and *Betaproteobacteria* (Saboo and Gealt, 1998), but there are no reports of non-sphingomonads being able to utilize the gene for HCP degradation. Our cultivation-independent results are in agreement with the *pcpB* gene being present exclusively in sphingomonads, as positive correlation was observed between the two at both sites. Slightly higher (max. twofold) counts for the *pcpB* gene than sphingomonad 16S rRNA genes were detected in some extracts, but this could be explained by the slight variability in the amplification efficiencies of the differently sized amplicons (107 vs. 504 bp, respectively).

Betaproteobacteria, especially *Burkholderiales*, were even more abundant in the bacterial communities than *Sphingomonadales*. This order contains many well-characterised degrader taxa, for example genus *Cupriavidus* known for trichlorophenol degradation potential (Sánchez and González, 2007). As no correlation was observed between their and the *pcpB* gene relative abundance (Spearman rho = −0.25, p = 0.31, n = 18 extracts), *Betaproteobacteria* were unlikely the class carrying this gene.

Sequencing (both Sanger and next generation) revealed the presence of only one variant of the *pcpB* gene in Kärkölä and Pursiala. The Sanger sequences of the amplified 263 bp fragment showed 99% identity to the *pcpB* gene of *Novosphingobium lentum* MT1 (two mismatches, synonymous, other of which at the degenerated base of the reverse primer). The same variant of the gene has been earlier detected in multiple sphingomonad isolates, representing several genera, cultivated from the on-site aerobic bioreactor that processed contaminated Kärkölä groundwaters (Tiirola et al., 2002b), as well as in some of the Canadian mixed slurry bioreactor clones reported by Beaulieu et al. (2000). The complete lack of diversity in such an abundant degradation gene at two long-term contaminated sites 150 km apart seems surprising. Three possible explanations include limited distribution of the other *pcpB* gene variants, high positive selection for the observed variant and/or limited divergence/evolution. We never detected the *pcpB* gene in soils that were pristine or oil-polluted, and both *S. chlorophenolicum* L-1 and *N. lentum* MT1 are notorious for their readiness to discard the gene in cultivation (Tiirola et al., 2000a; Copley et al., 2012). PCP-polluted sites may thus represent the only "islands" where the pathway is maintained and multiplied (via reproduction and horizontal gene transfer). Out of the several *pcpB* gene variants detected in North America - or other thus far unknown center of origin and diversity - perhaps only one has made its way to the studied sites in Finland. It is also possible that the MT1-type *pcpB* gene is selected for in aquatic environments, such as Kärkölä and Pursiala groundwaters and the Canadian mixed slurry bioreactors of Beaulieu et al. (2000). However, we detected the same variant of the gene, and only that, also in chlorophenol-contaminated surface soils at a third site in Finland over 200 km away from the sites of this study (unpublished data).

5. Conclusions

- Novel PCR-based assays were developed that enable investigation of the presence, abundance and diversity of all known variants of the gene encoding pentachlorophenol hydroxylase (*pcpB*).
- Applying these assays, we were for the first time able to detect and quantify a gene specific for degradation of higher chlorinated phenols (HCPs: pentachlorophenol and tetrachlorophenol) *in situ* at polluted sites.
- Detection of *pcpB* in chlorophenol plume and its high abundance during aerobic HCP degradation both *ex situ* and *in situ* indicate that the specific degraders are present and can be activated by groundwater aeration.
- Cultivation-independent analysis supported the earlier cultivation-dependent results of *pcpB* being present only in sphingomonads, and of only one of the *pcpB* variants being found in Finland.
- Bacteria capable of aerobic mineralisation of xenobiotic HCPs present at long-term polluted groundwater sites make bioremediation by simple aeration a viable and economically attractive alternative.

Funding

This work was supported by the Academy of Finland [grant numbers 139847, 260797]; the European Research Council under the European Union's Seventh Framework Programme [FP/2007-450 2013, grant number 615146]; and Koskinen Oy (Kärkölä remediation and groundwater analyses, including consultancy by Jukka Ikäheimo). The other authors declare no conflict of interest. Funding sources had no involvement in the interpretation of data, writing of the report or the decision to submit the article for publication.

Acknowledgements

The three anonymous reviewers are acknowledged for their insightful and constructive criticism that helped to significantly improve the manuscript. We thank the consultants working at the Pursiala site, Iiro Kiukas/Ramboll and Arto Itkonen/FCG, for helpful discussions, sampling and providing the groundwater data. Anita Mäki is thanked for help with Ion PGM sequencing and Ivonne De La Garcia/Pöyry for drawing the Kärkölä remediation schematic graph.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.watres.2018.03.033.

References

Beaulieu, M., Becaert, V., Deschenes, L., Villemur, R., 2000. Evolution of bacterial diversity during enrichment of PCP-degrading activated soils. Microb. Ecol. 40, 345–355.
Bécaert, V., Deschênes, L., Samson, R., 2000. A simple method to evaluate the concentration of pentachlorophenol degraders in contaminated soils. FEMS Microbiol. Lett. 184, 261–264.
Bouchard, B., Beaudet, R., Villemur, R., McSween, G., Lépine, F., Bisaillon, J.G., 1996. Isolation and characterization of *Desulfitobacterium frappieri* sp. nov., an anaerobic bacterium which reductively dechlorinates pentachlorophenol to 3-chlorophenol. Int. J. Syst. Bacteriol. 46, 1010–1015.
Bryant, S.E., Schultz, T.W., 1994. Toxicological assessment of biotransformation products of pentachlorophenol: Tetrahymena population growth impairment. Arch. Environ. Contam. Toxicol. 26, 299–303.
Cole, J.D., 2000. Demonstration of a Permeable Barrier Technology for the In-situ Bioremediation of Pentachlorophenol Contaminated Groundwater. Dissertation for the Degree of PhD (Civil Engineering). Oregon State University, Oregon, United States.
Copley, S.D., Rokicki, J., Turner, P., Daligault, H., Nolan, M., Land, M., 2012. The whole genome sequence of *Sphingobium chlorophenolicum* L-1: insights into the evolution of the pentachlorophenol degradation pathway. Genome Biol. Evol. 4, 184–198.
Crawford, R.L., Jung, C.M., Strap, J.L., 2007. The recent evolution of pentachlorophenol (PCP)-4-monooxygenase (PcpB) and associated pathways for bacterial degradation of PCP. Biodegradation 18, 525–539.
D'Angelo, E.M.D., Reddy, K.R., 2000. Aerobic and anaerobic transformations of pentachlorophenol in wetland soils. Soil Sci. Soc. Am. J. 64, 933–943.
Field, J.A., Sierra-Alvarez, R., 2008. Microbial degradation of chlorinated phenols. Rev. Environ. Sci. Biotechnol. 7, 211–241.
Hlouchova, K., Rudolph, J., Pietari, J.M.H., Behlen, L.S., Copley, S.D., 2012. Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by *Sphingobium chlorophenolicum*. Biochemistry 51, 3848–3860.
Järvinen, K.T., Melin, E.S., Puhakka, J.A., 1994. High-rate bioremediation of chlorophenol-contaminated groundwater at low temperatures. Environ. Sci. Technol. 28, 2387–2392.
Langwaldt, J.H., Puhakka, J.A., 2003. Competition for oxygen by iron and 2,4,6-trichlorophenol oxidizing bacteria in boreal groundwater. Water Res. 37, 1378–1384.
Langwaldt, J.H., Tiirola, M., Puhakka, J.A., 2007. Microbial adaptation to boreal subsurface environment: implications in bioremediation of polychiorophenol contaminated groundwater. In: Margesin, R., Schinner, F., Marx, J.-C., Gerday, C. (Eds.), Psychrophiles: from Biodiversity to Biotechnology. Springer Verlag, pp. 409–428.
Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al., 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32,

1363–1371.
Mikkonen, A., Lappi, K., Wallenius, K., Lindström, K., Suominen, L., 2011. Ecological inference on bacterial succession using curve-based community fingerprint data analysis, demonstrated with rhizoremediation experiment. FEMS Microbiol. Ecol. 78, 604–616.
Mäki, A., Rissanen, A.J., Tiirola, M., 2016. A practical method for barcoding and size-trimming PCR templates for amplicon sequencing. Biotechniques 60, 88–90.
Männistö, M.K., Salkinoja-Salonen, M.S., Puhakka, J.A., 2001. *In situ* polychlorophenol bioremediation potential of the indigenous bacterial community of boreal groundwater. Water Res. 35, 2496–2504.
Männistö, M.K., Tiirola, M.A., Salkinoja-Salonen, M.S., Kulomaa, M.S., Puhakka, J.A., 1999. Diversity of chlorophenol-degrading bacteria isolated from contaminated boreal groundwater. Arch. Microbiol. 171, 189–197.
Olaniran, A.O., Igbinosa, E.O., 2011. Chlorophenols and other related derivatives of environmental concern: properties, distribution and microbial degradation processes. Chemosphere 83, 1297–1306.
Pollock, T.J., Armentrout, R.W., 1999. Planktonic/sessile dimorphism of polysaccharide-encapsulated sphingomonads. J. Ind. Microbiol. Biotechnol. 23, 436–441.
Rautio, J., 2011. Detrimental Substances in the Pursiala Groundwater Area, Mikkeli. Thesis for the degree of M.Sc. (Tech.). Lappeenranta University of Technology, Faculty of Environmental Technology, p. 113 (in Finnish).
Ritalahti, K.M., Amos, B.K., Sung, Y., Wu, Q., Koenigsberg, S.S., Löffler, F.E., 2006. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. Appl. Environ. Microbiol. 72, 2765–2774.
Saboo, V.M., Gealt, M.A., 1998. Gene sequences of the *pcpB* gene of pentachlorophenol-degrading *Sphingomonas chlorophenolica* found in non-degrading bacteria. Can. J. Microbiol. 44, 667–675.
Sánchez, M.A., González, B., 2007. Genetic characterization of 2,4,6-trichlorophenol degradation in *Cupriavidus necator* JMP134. Appl. Environ. Microbiol. 73, 2769–2776.
Schloss, P.D., Gevers, D., Westcott, S.L., 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 6, e27310.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.
Sinkkonen, A., Kauppi, S., Simpanen, S., Rantalainen, A.-L., Strömmer, R., Romantschuk, R., 2013. Layer of organic pine forest soil on top of chlorophenol contaminated mineral soil enhances contaminant degradation. Environ. Sci. Pollut. Res. 20, 1737–1745.
Sinkkonen, A., Ollila, S., Romantschuk, M., 2014. Changes in *TcpA* gene frequency explain 2,4,6-trichlorophenol degradation in mesocosms. J. Environ. Sci. Health, Part B 49, 756–759.
Tiirola, M.A., Männistö, M.K., Puhakka, J.A., Kulomaa, M.S., 2002a. Isolation and characterization of *Novosphingobium* sp. strain MT1, a dominant polychlorophenol-degrading strain in a groundwater bioremediation system. Appl. Environ. Microbiol. 2002, 173–180.
Tiirola, M.A., Wang, H., Paulin, L., Kulomaa, M.S., 2002b. Evidence for natural horizontal transfer of the *pcpB* gene in the evolution of polychlorophenol-degrading sphingomonads. Appl. Environ. Microbiol. 68, 4495–4501.
Uotila, J.S., Kitunen, V.H., Saastamoinen, T., Coote, T., Häggblom, M.M., Salkinoja-Salonen, M.S., 1992. Characterization of aromatic dehalogenases of *Mycobacterium fortuitum* CG-2. J. Bacteriol. 174, 5669–5675.
Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinf. 13, 134.
Zhou, L., Li, H., Zhang, Y., Wang, Y., Han, S., Xu, H., 2012. Abundance and diversity of *Sphingomonas* in Shenfu petroleum-wastewater irrigation zone, China. Environ. Sci. Pollut. Res. 19, 282–294.

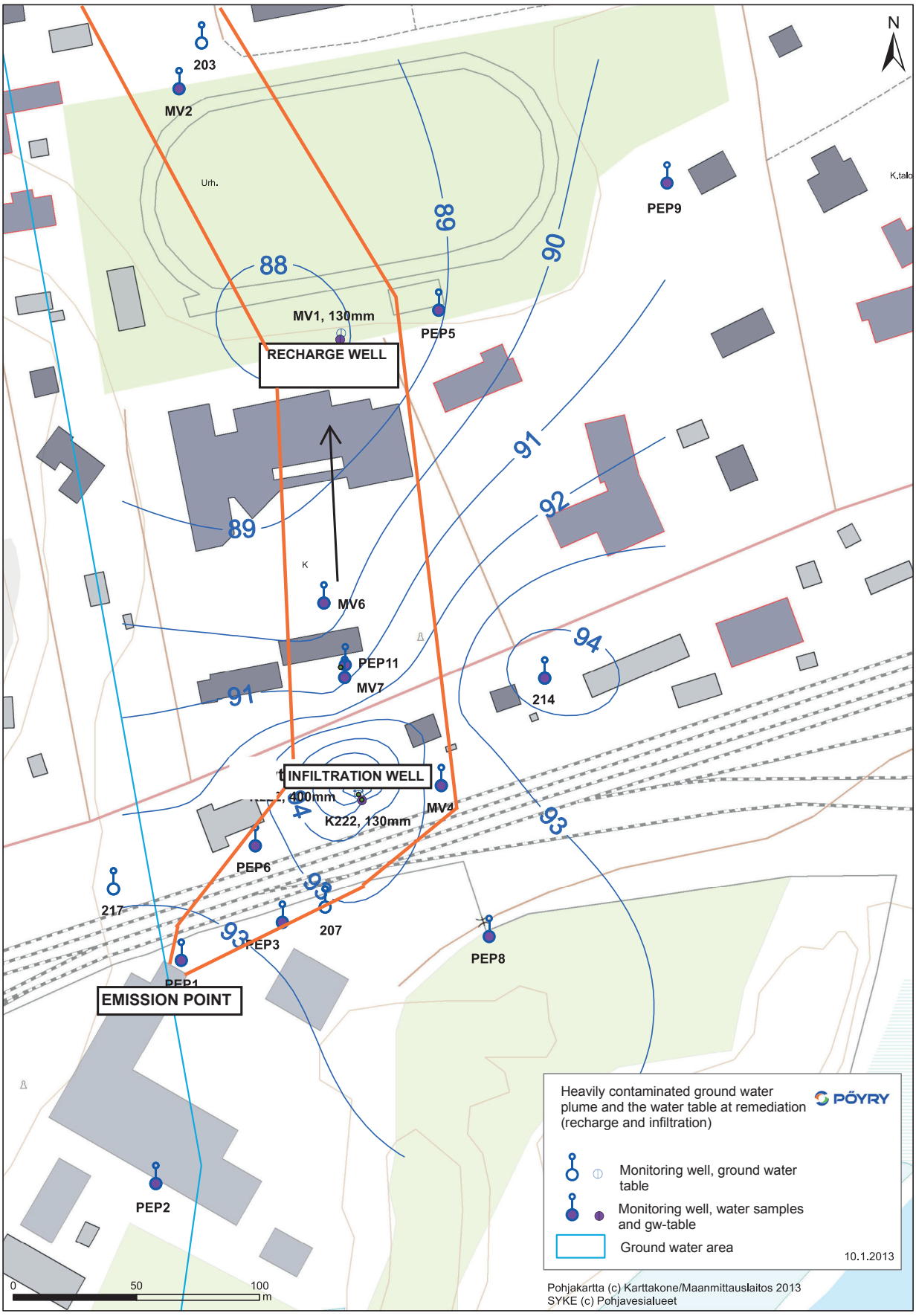


Figure S1. Map of Kärkölä site. K222/2 in the map is well K1 and MV6 in the map is well K2.

Table S1. Original microbiological data (duplicate DNA extracts).

Site	Sampling day	Groundwater tube	CP (µg/L)	O_2 (mg/L)	dw for DNA extraction (g)	DNA (ng/µL)	DNA yield (µg/g dw)	Bacterial 16S rRNA gene copies (/µL)	Sphingomonad 16S rRNA gene copies (/µL)	*pcpB* copies (/µL)	Bacterial PD
Pursiala	23/04/2013	P1	49	0.7	0.184	4.18	2.27	1.9E+06	4.7E+02	<2.0E+01	50.8
	23/04/2013	P1	49	0.7	0.184	3.63	1.97	1.3E+06	4.4E+02	<2.0E+01	43.2
	23/04/2013	P2	1	0.6	0.193	0.31	0.16	1.4E+05	1.4E+02	<2.0E+01	29.1
	23/04/2013	P2	1	0.6	0.193	0.32	0.17	1.9E+05	2.4E+02	<2.0E+01	32.9
	23/04/2013	P3	841	0.5	0.174	1.51	0.87	6.3E+05	2.9E+03	<2.0E+01	31.9
	23/04/2013	P3	841	0.5	0.174	2.56	1.47	1.1E+06	1.4E+03	2.5E+01	33.6
	23/04/2013	P4	7380	1.2	0.317	6.53	2.06	3.4E+06	6.3E+04	9.1E+03	18.5
	23/04/2013	P4	7380	1.2	0.317	5.73	1.81	4.4E+06	6.4E+04	8.5E+03	19.8
	23/04/2013	P5	14000	0.6	0.255	0.19	0.07	1.6E+05	6.7E+02	4.5E+01	25.2
	23/04/2013	P5	14000	0.6	0.255	0.18	0.07	1.1E+05	6.2E+02	4.3E+01	29.7
Kärkölä	24/10/2012	K1	470	7.2	0.118	5.20	4.41	2.2E+06	7.2E+04	5.8E+04	19.9
	24/10/2012	K1	470	7.2	0.118	5.46	4.62	2.7E+06	6.5E+04	7.3E+04	19.4
	29/04/2013	K1	657	8.9	0.091	6.46	7.06	2.1E+06	5.2E+04	8.9E+04	33.4
	29/04/2013	K1	657	8.9	0.063	7.92	12.56	2.5E+06	6.4E+04	1.1E+05	31.4
	24/10/2012	K2	4650	0.4	0.077	3.44	4.48	1.0E+06	2.0E+04	4.8E+04	19.0
	24/10/2012	K2	4650	0.4	0.077	3.35	4.37	1.0E+06	2.1E+04	4.4E+04	16.3
	29/04/2013	K2	1020	<0.2	0.065	4.57	7.08	2.1E+06	1.0E+04	2.7E+04	23.3
	29/04/2013	K2	1020	<0.2	0.043	5.03	11.68	2.2E+06	1.0E+04	3.0E+04	22.5

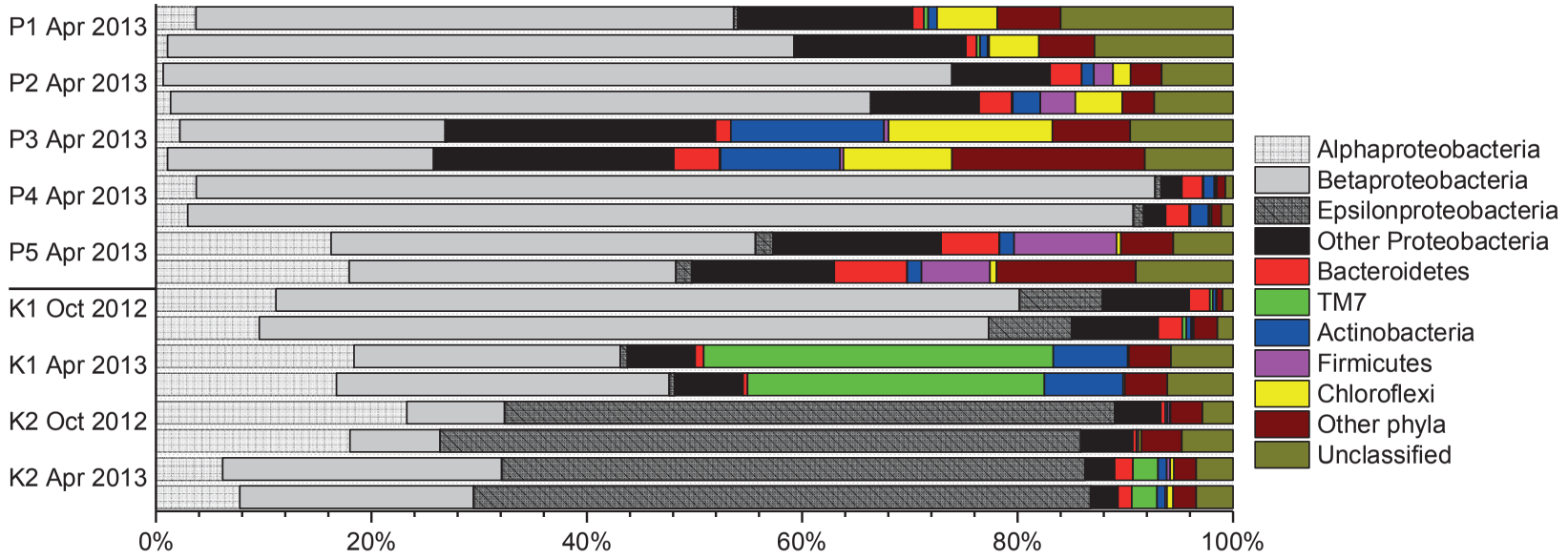


Figure S2. Bacterial community composition. Pursiala wells (P1-P5) above, Kärkölä wells (K1-2) below. The two bars for each sample represent the duplicate DNA extractions, showing the technical reproducibility of the analysis.

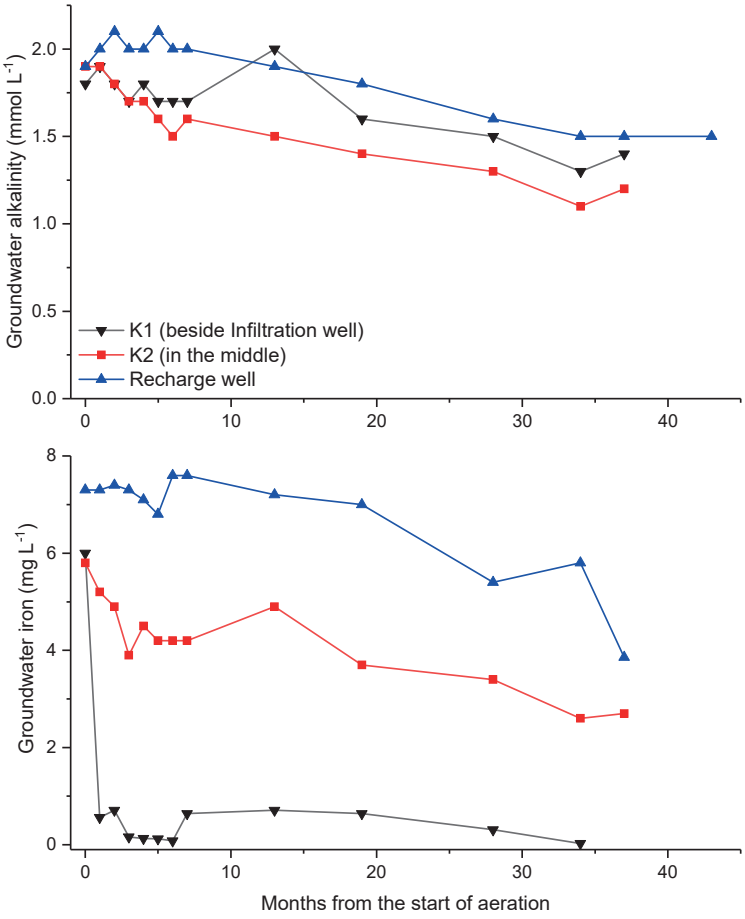


Figure S3. Differences and changes in groundwater alkalinity and iron concentration at the Kärkölä site upon 4-year closed-circle aerobic *in situ* biostimulation.

***pcpB* sequences (Sanger, consensus) from the three wells where a 263 bp PCR product was successfully amplified**

>11_Karkola1_spring_250bp
caatttcgagaacaccgacgcgaagcccatcctggacttctcgacccttcccgggcgttttccgttcatcatgatctaca
atcagaacgagaccgaacgcgtgcttcggcagcatctcgatgcaacgttcaacttccgtccggaatggggaacgcaactg
cttacgctgaaacaaggcgagagcggcatcgaagtgggacttcgccttgcagacggcagcaaggaaacgatccgcccgcg
ctgggtgatc

>13_Karkola2_spring_169bp
tctcgacccttcccgggcgttttccgttcatcatgatctacaatcagaacgagaccgaacgcgtgcttcggcagcatctc
gatgcaacgttcaacttccgtccggaatggggaacgcaactgcttacgctgaaacaaggcgagagcggcatcgaagtggg
acttcgcct

>17_Pursiala4_spring_256bp
cacgttcaatttcgagaacaccgacgcgaagcccatcctggacttctcgacccttcccgggcgttttccgttcatcatga
tctacaatcagaacgagaccgaacgcgtgcttcggcagcatctcgatgcaacgttcaacttccgtccggaatggggaacg
caactgcttacgctgaaacaaggcgagagcggcatcgaagtgggacttcgccttgcagacggcagcaaggaaacgatccg
cccgcgctgggtgatc

*all 3 show 99% identity to *Novosphingobium lentum* MT1 *pcpB* gene for pentachlorophenol-4-monooxygenase (2 mismatches, synonymous) and even higher identity with *Novosphingobium* sp. K1 (1 mismatch, synonymous - the degenerated base of the reverse primer)*

III

DNA RECOVERY FROM DROPLET DIGITAL™ PCR EMULSIONS USING LIQUID NITROGEN

by

Lara Dutra, Ole Franz, Veli-Mikko Puupponen & Marja Tiirola 2020

BioTechniques 69: 451-454

https://doi.org/10.2144/btn-2020-0076

Benchmark

BioTechniques

DNA recovery from Droplet Digital™ PCR emulsions using liquid nitrogen

Lara Dutra*,[1], Ole Franz[1], Veli-Mikko Puupponen[1] & Marja Tiirola[1]

[1]*Department of Biological & Environmental Science, Nanoscience Center, University of Jyväskylä, Finland; *Author for correspondence: Laambros@student.jyu.fi*

BioTechniques 69: 00–00 (December 2020) 10.2144/btn-2020-0076

First draft submitted: 27 May 2020; Accepted for publication: 22 September 2020; Published online: 26 October 2020

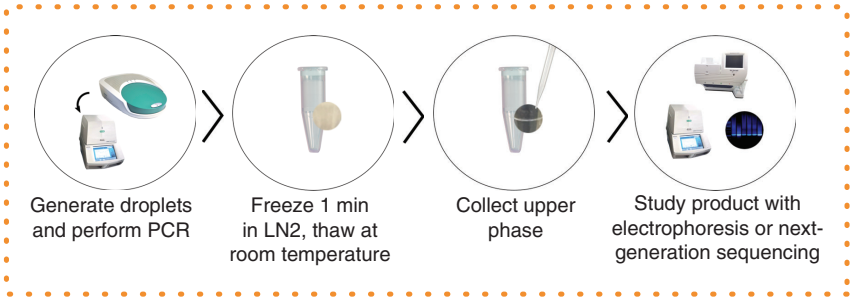
ABSTRACT

Droplet microfluidics is a technology that enables the production and manipulation of small volumes. In biosciences, the most popular application of this technology is Droplet Digital™ PCR (ddPCR™), where parallel nanoliter-scale PCR assays are used to provide a high sensitivity and specificity for DNA detection. However, the recovery of PCR products for downstream applications such as sequencing can be challenging due to the droplets' stability. Here we compared five methods for disrupting the droplets to recover DNA. We found that rapid freezing in liquid nitrogen results in a clear phase separation and recovery of up to 70% of the DNA content. Liquid nitrogen freezing can thus offer a simple and environmentally friendly protocol for recovering DNA from ddPCR.

METHOD SUMMARY

The coalescence of the Droplet Digital™ PCR emulsion droplets is achieved by rapid freezing in liquid nitrogen. Thawing of the frozen emulsion forms separate oil and aqueous phases; the latter contains the DNA, which is readily available for secondary PCR reactions and other downstream applications (see Graphical abstract).

GRAPHICAL ABSTRACT



KEYWORDS:

amplicon recovery ● breaking droplets ● Droplet Digital™ PCR (ddPCR™) ● droplet microfluidics ● emulsion PCR

Droplet microfluidics (DM) technology has proved to be a unique and versatile tool for a broad range of biological assays. It relies on the physicochemical properties of two immiscible liquids and their manipulation through interconnected microfluidic channels, enabling the generation and manipulation of small and separated volumes [1–3]. The properties of the droplets, such as size and stability, can be adjusted by altering the microfluidic channel geometry, flow rate and reagent composition [1–3]. As a result, droplets can be manipulated in a controlled manner for different ends.

DM has been used for various purposes ranging from molecular analyses to the cultivation of individual microbes [3,4]. The replication of microbes inside droplets has revolutionized the study of metabolism, antibiotic resistance and enzyme activity [4,5]. Methods have also been developed to select and cultivate combinations of organisms for studying cell–cell interactions [6]. Lately, DM has even been explored for studying single cells [7,8]. Nonetheless, the most popular and well-established use of DM has been for nucleic acid detection and quantitation.

The Droplet Digital™ PCR (ddPCR™) system is a droplet-based PCR platform commercialized by Bio-Rad (CA, USA) and has been widely used for pathogen detection [9,10], food quality analysis [11], environmental studies [12] and medical research [13]. ddPCR relies on DM's generation of numerous nanoliter-sized droplets, inside which independent DNA amplification reactions are carried out. As with quantitative PCR, the increase in DNA copy number due to specific amplification is revealed via fluorescent dye. However, ddPCR analysis

Table 1. Summary of methods used for breaking the ddPCR emulsion.

Method	Protocol	Procedure	Ref.
Chemical	Standard chloroform protocol (Bio-Rad)	i) Remove excess oil ii) Add 20 μl of TE buffer iii) Add 70 μl of chloroform iv) Homogenize by vortexing at top speed for 1 min v) Centrifuge at 16,000 × *g* for 10 min vi) Collect the aqueous phase	[16]
	n-Octane protocol	i) Remove excess oil ii) Add 20 μl of TE buffer iii) Add 23.3 μl of n-Octane iv) Homogenize by vortexing at top speed for 1 min v) Centrifuge at 16,000 × *g* for 10 min vi) Collect the aqueous phase	
	1H,1H,2H,2H-Perfluoro-1-octanol protocol	i) Remove excess oil ii) Add 20 μl of TE buffer iii) Add 23.3 μl of PFO iv) Homogenize by vortexing at top speed for 1 min v) Centrifuge at 16,000 × *g* for 10 min vi) Collect the aqueous phase	
Physical	Silica column (GeneJET PCR purification kit, Thermo Scientific)	i) Add 70 μl of binding buffer ii) Add 70 μl of isopropanol iii) Homogenize by vortexing at top speed for 1 min iv) Transfer to silica column and proceed according to manual	[17]
	Thermal	i) Remove excess oil† ii) Add 20 μl of TE buffer† iii) Freeze for 1 min iv) Thaw at room temperature v) Collect the aqueous phase	

Volumes are adjusted per single emulsions (20 μl of PCR reaction and 70 μl of Droplet Generation Oil) processed by the QX200 Droplet Generator.
†Optional step.
PFO: 1H,1H,2H,2H-perfluoro-1-octanol; TE buffer: Tris-EDTA buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

is based on the end-point fluorescence intensity measured for each separate droplet, thereby increasing method sensitivity. After the ddPCR process, the droplets are usually discarded, despite their potential as material for downstream genetic studies. For instance, DNA recovered from droplets has been used for production of next-generation sequencing libraries for single-cell genetics [8,14] and transcriptomics studies [15]. The lack of efficient methods for breaking the ddPCR emulsion forms a major challenge to using the PCR products in downstream analysis.

In this study we addressed the challenge of recovering DNA from droplets generated by the QX200™ ddPCR System (Bio-Rad, CA, USA). Different chemical and physical methods were compared: chloroform, n-Octane and 1H,1H,2H,2H-perfluoro-1-octanol (PFO) as chemical treatments, and freeze-thawing and silica columns as physical methods (Table 1).

The ddPCR emulsion samples were prepared with QX200 ddPCR EvaGreen Supermix and QX200 Droplet Generation Oil for EvaGreen according to the manufacturer's instructions. To assess the efficiency of DNA recovery, DNA fragments up to 1 kb were added to the ddPCR master mix as input. Initially, GeneRuler 100-bp DNA ladder (Thermo Fisher Scientific, MA, USA) was used as the DNA control, containing fragment sizes ranging from 100 to 1000 bp. Later, purified PCR amplicons or Qubit® dsDNA HS Standard #2, of sizes 263 bp and 1 kb, respectively, were chosen. Up to three independent emulsion samples were combined into one tube to form a single replicate; the number of replicates per protocol in each experiment varied from one to seven. The initial and recovered DNA was quantified using the Qubit dsDNA high-sensitivity assay (Thermo Fisher Scientific, MA, USA). The method efficiency was calculated as the percentage of DNA recovered against the initial DNA input for each sample (Supplementary Table 1).

Our reference method for DNA retrieval was the chloroform protocol described in the Bio-Rad ddPCR application guide [16]. Briefly, this protocol consists of combining the emulsion replicates, removing excess oil and homogenizing the sample with Tris-EDTA buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and chloroform (Sigma-Aldrich, MO, USA). These steps were followed by a high-speed centrifugation to separate the phases into a lower oil phase and an upper aqueous phase, with the latter containing the DNA.

We tested the efficiency of the standard protocol with several chloroform-to-oil ratios (Chl:O), which were achieved by increasing or decreasing the chloroform volume and keeping other conditions identical. A Chl:O of close to 2 was the protocol recommendation, for which 70 μl of chloroform was added to the combined emulsion sample after removing the excess oil. Higher ratios (Chl:O >2) were achieved by adding twice the volume and lower ratios (Chl:O <2) were achieved by adding a third of the chloroform volume. The average recovery of DNA using the standard protocol (Chl:O ratio ≈ 2) was 45 $\pm$ 4%. Chl:O >2 yielded 34 $\pm$ 18% of the initial DNA, and Chl:O <2 yielded 61 $\pm$ 4% recovery (Figure 1). Despite the differences in the DNA recovery, the results presented here do not demonstrate which chloroform-to-oil ratio could improve the outcome. Whereas Chl:O <2 provided higher apparent recovery, in some samples Chl:O >2 was necessary for a clear phase separation.

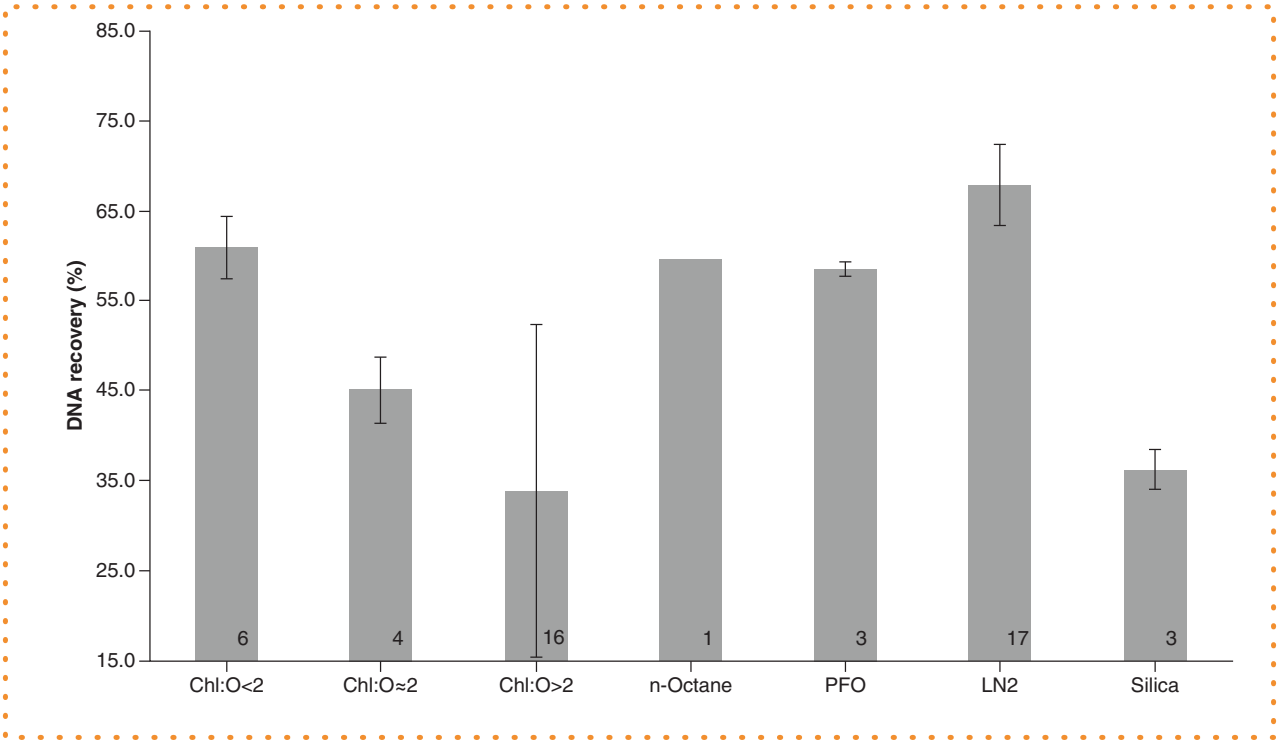


Figure 1. Efficiency of DNA recovery from ddPCR droplets. Three solvents were tested to break the droplets in ddPCR emulsion: chloroform, n-Octane and 1H,1H,2H,2H-perfluoro-1-octanol. Adjustments of chloroform-to-oil ratio (Chl:O) were tested. In addition, two physical methods were tested: freezing-thawing in liquid nitrogen and purification with silica columns. The number of replicates used for each method is shown at the bottom of the respective column and the bars represent the standard deviation of the recovery percentage.
LN2: Liquid nitrogen; PFO: 1H,1H,2H,2H-perfluoro-1-octanol.

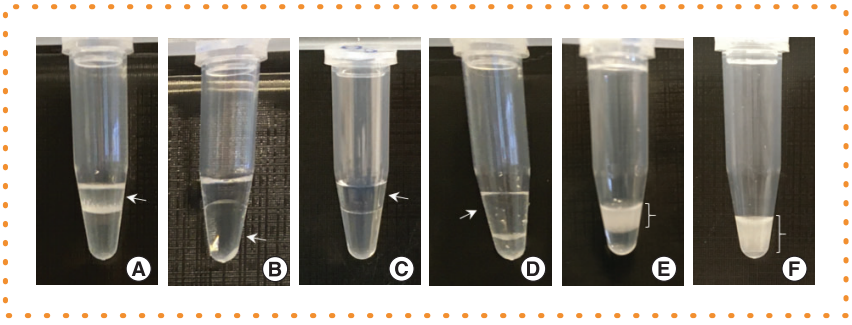


Figure 2. Phase separation after breaking the ddPCR emulsion. The photo panel shows the phase separation by using **(A)** chloroform, **(B)** n-Octane, **(C)** 1H,1H,2H,2H-perfluoro-1-octanol and **(D)** liquid nitrogen. The arrows indicate the aqueous phase. In the tubes on the right, the intact emulsion sample is shown before **(E)** and after **(F)** removing the excess oil, with the emulsion layer indicated by the white brackets.

In addition to chloroform, two other chemical compounds, n-Octane and PFO (both Sigma-Aldrich), were compared. The use of low-density n-Octane resulted in an undesirable reversal of the phases when compared with other treatments (Figure 2A–D), making the now lower aqueous part hard to retrieve. DNA was successfully recovered from only one of the three replicates using n-Octane. For PFO, the average recovery of DNA was 59 $\pm$ 1% (Figure 1). A comparison of the DNA yields between the alternative chemical methods showed that PFO and n-Octane were as good as Chl:O $<$2. Nonetheless, with regard to phase separation, PFO appeared to perform better than n-Octane or chloroform; PFO resulted in a clear upper aqueous phase, while n-Octane resulted in inverted phases and chloroform left a cloudy layer between the two phases (Figure 2A).

Two physical methods were also investigated to explore alternative means for recovering DNA from the emulsion. The GeneJET PCR Purification Kit (Thermo Fisher Scientific, MA, USA) was tested as a physical silica column extraction method for recovering DNA

from the droplets. The emulsion was homogenized with the binding buffer and added to the silica columns for binding, washing and recovery, according to the manufacturer's instructions (including isopropanol addition, as recommended for short fragments) [17]. The silica columns from the GeneJET PCR Purification Kit were able to recover 36 ± 2% of the DNA from the emulsion.

As a second physical method, we developed a temperature-based protocol for breaking the emulsion samples. Initially, the freezing temperature of the oil was investigated by storing 250-µl aliquots at -20°C (freezer), -80°C (dry ice/deep freezer) or -195°C (liquid nitrogen [LN2]) for up to 24 h. Incubation at -20 and -80°C did not result in state change, while samples in LN2 were quickly frozen. In further experiments, samples were cooled in LN2 for 1 min, followed by thawing at room temperature. Initial tests showed that freeze-thawing was enough to coalesce the droplets (Figure 2D–F). The average DNA recovery from the aqueous phase of LN2-frozen emulsions was up to 68 ± 5% of the total DNA. Minor changes to the procedure – for example, keeping or removing the excess oil or adding Tris-EDTA buffer to the emulsion – did not have a significant effect on the yield. Phase separation was successfully achieved with emulsions prepared with both kits commercialized by Bio-Rad: ddPCR for EvaGreen® and ddPCR for Probes (Supplementary Figure 1). When compared with the other methods, the LN2 protocol was able to provide the highest recovery rate with minimal sample manipulation steps. This approach differs from the rest likely due to two simultaneous processes: first, the rapid expansion of the water droplets increases the liquid–liquid boundary surface area, lowering the concentration of adsorbed, stabilizing surfactant; and second, the buildup of additional surfactant may be limited by the low temperature [18].

In conclusion, we show that our novel LN2-based methodology is a simple and an efficient way to recover DNA from ddPCR emulsions. Chloroform and fluorinated solvents pose a safety risk and generate hazardous waste; LN2 is available in most research institutes and can provide a safer and greener alternative. In addition, the recovery rate achieved using LN2 presents less variation when compared with other protocols and prevents contamination risk by minimizing the need for manipulation and addition of reagents in the sample. This method can support the use of the ddPCR system, not only as a tool for quantitative analysis, but also as a sample preparation process for downstream applications such as secondary PCR and sequencing. The suggested method also provides an alternative way to handle fluorinated oil-based emulsions in general and to collect the encapsulated contents for many types of applications.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2020-0076

Author contributions

L Dutra and M Tiirola conceived the idea. L Dutra and O Franz carried out experiments. V Puupponen helped in interpreting the temperature results. L Dutra analyzed results and wrote the first version of the manuscript; all authors provided critical comments. The final version of the manuscript was reviewed by all authors.

Acknowledgments

The authors thank P Salmi (University of Jyväskylä, Finland) for the help in laboratory work during the initial development of this study. We also thank G Almeida and M Jalasvuori (University of Jyväskylä, Finland) for kindly reading and commenting on the manuscript.

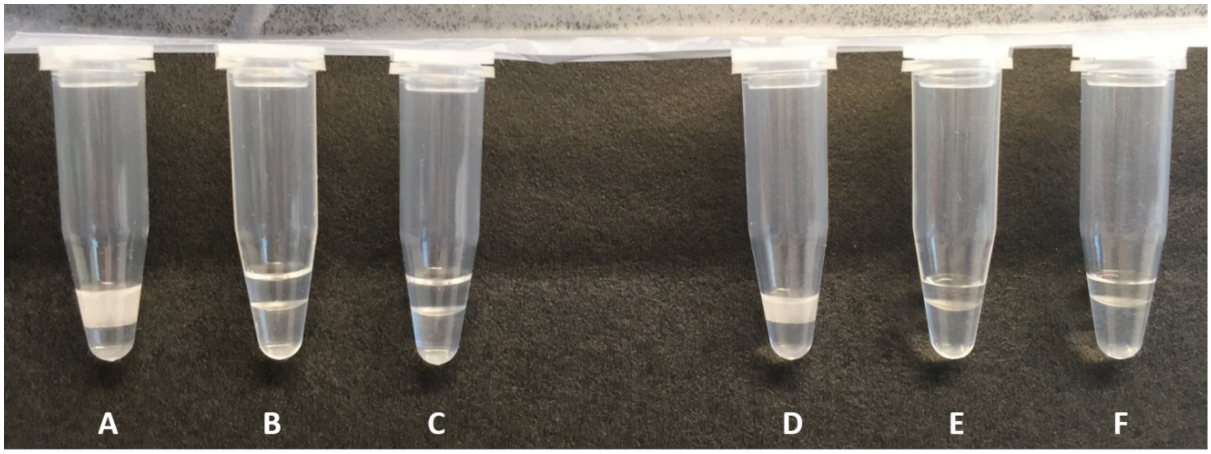
Financial & competing interests disclosure

This work was funded by the Academy of Finland (grant 323063) and European Research Council CoG grant (615146) and PoC grant (812729; EU FP7 and H2020 programs, respectively) to M Tiirola. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

The English language has been checked and corrected by professional editing service Scribendi.

References

1. Baret J-C. Surfactants in droplet-based microfluidics. *Lab. Chip* 12(3), 422–433 (2012).
2. Gruner P, Riechers B, Chacòn Orellana LA *et al.* Stabilisers for water-in-fluorinated-oil dispersions: key properties for microfluidic applications. *Curr. Opin. Colloid Interface Sci.* 20(3), 183–191 (2015).
3. Shang L, Cheng Y, Zhao Y. Emerging droplet microfluidics. *Chem. Rev.* 117(12), 7964–8040 (2017).
4. Kaminski TS, Garstecki P. Controlled droplet microfluidic systems for multistep chemical and biological assays. *Chem. Soc. Rev.* 46(20), 6210–6226 (2017).
5. Leygeber M, Lindemann D, Sachs CC *et al.* Analyzing microbial population heterogeneity – expanding the toolbox of microfluidic single-cell cultivations. *J. Mol. Biol.* 431(23), 4569–4588 (2019).
6. Ohan J, Pelle B, Nath P *et al.* High-throughput phenotyping of cell-to-cell interactions in gel microdroplet pico-cultures. *BioTechniques* 66(5), 218–224 (2019).
7. Thibault D, Jensen PA, Wood S *et al.* Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. *Nat. Commun.* 10(1), 5729 (2019).
8. Pellegrino M, Sciambi A, Treusch S *et al.* High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res.* 28(9), 1345–1352 (2018).
9. McMahon TC, Blais BW, Wong A, Carrillo CD. Multiplexed single intact cell Droplet Digital PCR (MuSIC ddPCR) method for specific detection of enterohemorrhagic *E. coli* (EHEC) in food enrichment cultures. *Front. Microbiol.* 8, 1–14 (2017).
10. Yu F, Yan L, Wang N *et al.* Quantitative detection and viral load analysis of SARS-CoV-2 in infected patients. *Clin. Infect. Dis.* 71 15 793–798 (2020).
11. Porcellato D, Narvhus J, Skeie SB. Detection and quantification of *Bacillus cereus* group in milk by droplet digital PCR. *J. Microbiol. Methods* 127, 1–6 (2016).
12. Mauvisseau Q, Davy-Bowker J, Bulling M *et al.* Combining ddPCR and environmental DNA to improve detection capabilities of a critically endangered freshwater invertebrate. *Sci. Rep.* 9(1), 14064 (2019).
13. Oscorbin I, Kechin A, Boyarskikh U, Filipenko M. Multiplex ddPCR assay for screening copy number variations in *BRCA1* gene. *Breast Cancer Res. Treat.* 178, 545–555 (2019).
14. Spencer SJ, Tamminen MV, Preheim SP *et al.* Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J.* 10(2), 427–436 (2016).
15. Klein AM, Mazutis L, Akartuna I *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161(5), 1187–1201 (2015).
16. Biorad. Droplet Digital™ PCR Droplet Digital™ PCR Applications Guide. Bulletin 6407, 1–145 (2018).
17. Thermo Scientific. GeneJET PCR Purification Kit. 1–7 (2015).
18. Swenson H, Stadie NP. Langmuir's theory of adsorption: a centennial review. *Langmuir* 35(16), 5409–5426 (2019).



Supplementary figure 1 – Applicability of LN2 method to break ddPCR emulsion. ddPCR reactions were prepared with ddPCR™ Supermix for Probes (no dUTP) and Droplet Generation Oil for Probes (three tubes on the left) or ddPCR™ Supermix for EvaGreen® and Droplet Generation Oil for EvaGreen® (three tubes on the right). The tubes (A) and (D) show the emulsion before using LN2 method, and tubes (B), (C), (E) and (F) shows the oil and water layers after breaking the emulsion.

Supplementary table - Raw data from experiments studying DNA recovery from droplets using chemical and physical procedures.

Date	Experiment	Sample	Protocol	emulsion per replicate	DNA fragment lenght	Remove oil	Chloroform to oil ratio	N-Octane	Perfluoro-octanol	Isopropanol	Add TE	CIAA wash	Liquid Nitrogen (LN2)	Initial DNA (ng)	Total DNA recovered (ng)	mean Total DNA recovered (ng)	ev Total DNA recov	CV= StDev/mean	% recovered (individual sample)	% recovered (average)	% of StDevp
19.9.2018	III	S1A1	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	yes	NA	516,00	186,20	285,63	92,88	0,33	36,08	55,35	18,00
19.9.2018	III	S1A2	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	yes	NA	516,00	300,52	285,63	92,88	0,33	58,24	55,35	18,00
19.9.2018	III	S1A3	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	yes	NA	516,00	370,16	285,63	92,88	0,33	71,74	55,35	18,00
19.9.2018	III	S2A1	Silica	1	variable	NA	NA	NA	no	yes	yes	NA	NA	516,00	174,97	186,97	11,45	0,06	33,91	36,23	2,22
19.9.2018	III	S2A2	Silica	1	variable	NA	NA	NA	no	yes	yes	NA	NA	516,00	197,77	186,97	11,45	0,06	38,33	36,23	2,22
19.9.2018	III	S2A3	Silica	1	variable	NA	NA	NA	no	yes	yes	NA	NA	516,00	188,17	186,97	11,45	0,06	36,47	36,23	2,22
21.9.2018	A	1a1	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	120,42	120,62	63,63	0,53	27,12	27,17	14,33
21.9.2018	A	1a2	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	280,80	120,62	63,63	0,53	63,24	27,17	14,33
21.9.2018	A	1a3	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	84,78	120,62	63,63	0,53	19,09	27,17	14,33
21.9.2018	A	1a4	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	86,40	120,62	63,63	0,53	19,46	27,17	14,33
21.9.2018	A	1a5	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	85,00	120,62	63,63	0,53	19,14	27,17	14,33
21.9.2018	A	1a6	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	135,54	120,62	63,63	0,53	30,53	27,17	14,33
21.9.2018	A	1a7	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	87,36	120,62	63,63	0,53	19,68	27,17	14,33
21.9.2018	A	1a8	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	79,92	120,62	63,63	0,53	18,00	27,17	14,33
21.9.2018	A	1a9	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	444,00	125,32	120,62	63,63	0,53	28,23	27,17	14,33
21.9.2018	A	1c1	Chloroform (Chl:Oil ratio = 2)	1	variable	yes	2	NA	NA	NA	yes	no	NA	444,00	181,20	196,60	21,78	0,11	40,81	44,28	4,91
21.9.2018	A	1c2	Chloroform (Chl:Oil ratio = 2)	1	variable	yes	2	NA	NA	NA	yes	no	NA	444,00	212,00	196,60	21,78	0,11	47,75	44,28	4,91
25.9.2018	V	1a1	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	310,00	170,91	101,79	56,81	0,56	55,13	32,83	18,33
25.9.2018	V	1a2	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	310,00	62,15	101,79	56,81	0,56	20,05	32,83	18,33
25.9.2018	V	1a3	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	310,00	48,97	101,79	56,81	0,56	15,80	32,83	18,33
25.9.2018	V	1a4	Chloroform (Chl:Oil ratio > 2)	1	variable	yes	>2	NA	NA	NA	yes	no	NA	310,00	125,12	101,79	56,81	0,56	40,36	32,83	18,33
27.3.2019	VI	LN1	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	11,42	13,42	1,03	0,08	57,12	67,12	5,13
27.3.2019	VI	LN2	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	13,83	13,42	1,03	0,08	69,14	67,12	5,13
27.3.2019	VI	LN3	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	13,84	13,42	1,03	0,08	69,18	67,12	5,13
27.3.2019	VI	LN4	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	12,84	13,42	1,03	0,08	64,19	67,12	5,13
27.3.2019	VI	LN5	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	12,96	13,42	1,03	0,08	64,79	67,12	5,13
27.3.2019	VI	LN6	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	14,15	13,42	1,03	0,08	70,76	67,12	5,13
27.3.2019	VI	LN7	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	13,55	13,42	1,03	0,08	67,74	67,12	5,13
27.3.2019	VI	LN8	LN2	2	long	no	NA	NA	NA	NA	yes	NA	yes	20,00	14,82	13,42	1,03	0,08	74,09	67,12	5,13
10.10.2019	VII	Chloroform1	Chloroform (Chl:Oil ratio < 2)	3	short	yes	<2	NA	NA	NA	yes	no	NA	95,40	60,40	60,59	1,06	0,02	63,31	63,52	1,11
10.10.2019	VII	Chloroform2	Chloroform (Chl:Oil ratio < 2)	3	short	yes	<2	NA	NA	NA	yes	no	NA	95,40	59,64	60,59	1,06	0,02	62,52	63,52	1,11
10.10.2019	VII	Chloroform3	Chloroform (Chl:Oil ratio < 2)	3	short	yes	<2	NA	NA	NA	yes	no	NA	95,40	61,74	60,59	1,06	0,02	64,72	63,52	1,11
10.10.2019	VII	LN1	LN2 (TE)	3	short	yes	NA	NA	NA	NA	yes	NA	yes	95,40	65,12	66,99	1,64	0,02	68,26	70,22	1,72
10.10.2019	VII	LN2	LN2 (TE)	3	short	yes	NA	NA	NA	NA	yes	NA	yes	95,40	68,20	66,99	1,64	0,02	71,49	70,22	1,72
10.10.2019	VII	LN3	LN2 (TE)	3	short	yes	NA	NA	NA	NA	yes	NA	yes	95,40	67,65	66,99	1,64	0,02	70,91	70,22	1,72
10.10.2019	VII	LN4	LN2 (oil)	3	short	yes	NA	NA	NA	NA	no	NA	yes	95,40	68,40	62,93	4,74	0,08	71,70	65,97	4,97
10.10.2019	VII	LN5	LN2 (oil)	3	short	yes	NA	NA	NA	NA	no	NA	yes	95,40	60,00	62,93	4,74	0,08	62,89	65,97	4,97
10.10.2019	VII	LN6	LN2 (oil)	3	short	yes	NA	NA	NA	NA	no	NA	yes	95,40	60,40	62,93	4,74	0,08	63,31	65,97	4,97
10.10.2019	VII	N-oct1	n-Octane	3	short	yes	NA	yes	NA	NA	yes	NA	NA	95,40	NA	NA	NA	NA	NA	NA	NA
10.10.2019	VII	N-oct2	n-Octane	3	short	yes	NA	yes	NA	NA	yes	NA	NA	95,40	NA	NA	NA	NA	NA	NA	NA
10.10.2019	VII	N-oct3	n-Octane	3	short	yes	NA	yes	NA	NA	yes	NA	NA	95,40	56,88	NA	NA	NA	59,62	NA	NA
10.10.2019	VII	Octanol1	PFO	3	short	yes	NA	NA	yes	NA	yes	NA	NA	95,40	56,64	55,84	0,78	0,01	59,37	58,53	0,82
10.10.2019	VII	Octanol2	PFO	3	short	yes	NA	NA	yes	NA	yes	NA	NA	95,40	55,80	55,84	0,78	0,01	58,49	58,53	0,82
10.10.2019	VII	Octanol3	PFO	3	short	yes	NA	NA	yes	NA	yes	NA	NA	95,40	55,08	55,84	0,78	0,01	57,74	58,53	0,82
17.10.2019	B	C4	Chloroform (Chl:Oil ratio < 2)	3	short	yes	<2	NA	NA	NA	yes	no	NA	430,20	238,00	251,43	13,37	0,05	55,32	58,44	3,11
17.10.2019	B	C5	Chloroform (Chl:Oil ratio < 2)	3	short	yes	<2	NA	NA	NA	yes	no	NA	430,20	251,55	251,43	13,37	0,05	58,47	58,44	3,11
17.10.2019	B	C6	Chloroform (Chl:Oil ratio < 2)	3	short	yes	<2	NA	NA	NA	yes	no	NA	430,20	264,74	251,43	13,37	0,05	61,54	58,44	3,11
17.10.2019	B	C7	Chloroform (Chl:Oil ratio = 2)	3	short	yes	2	NA	NA	NA	yes	no	NA	430,20	208,95	197,56	16,11	0,08	48,57	45,92	3,75
17.10.2019	B	C8	Chloroform (Chl:Oil ratio = 2)	3	short	yes	2	NA	NA	NA	yes	no	NA	430,20	186,16	197,56	16,11	0,08	43,27	45,92	3,75
17.10.2019	B	LN1	LN2 (TE)	3	short	yes	NA	NA	NA	NA	yes	NA	yes	430,20	279,60	299,88	20,87	0,07	64,99	69,71	4,85
17.10.2019	B	LN2	LN2 (TE)	3	short	yes	NA	NA	NA	NA	yes	NA	yes	430,20	321,30	299,88	20,87	0,07	74,69	69,71	4,85
17.10.2019	B	LN3	LN2 (TE)	3	short	yes	NA	NA	NA	NA	yes	NA	yes	430,20	298,75	299,88	20,87	0,07	69,44	69,71	4,85

IV

SINGLE-CELL RESOLUTION GENETIC ASSOCIATION ANALYSIS FOR BACTERIA IN HETEROGENEOUS COMMUNITIES BY UTILIZING DROPLET DIGITAL PCR

by

Lara Ambrosio Leal Dutra, Reetta Penttinen, Ole Franz, Paulina Salmi, Kimi Nurminen, Matti Jalasvuori & Marja Tiirola 2021

Manuscript

Request a copy from author.