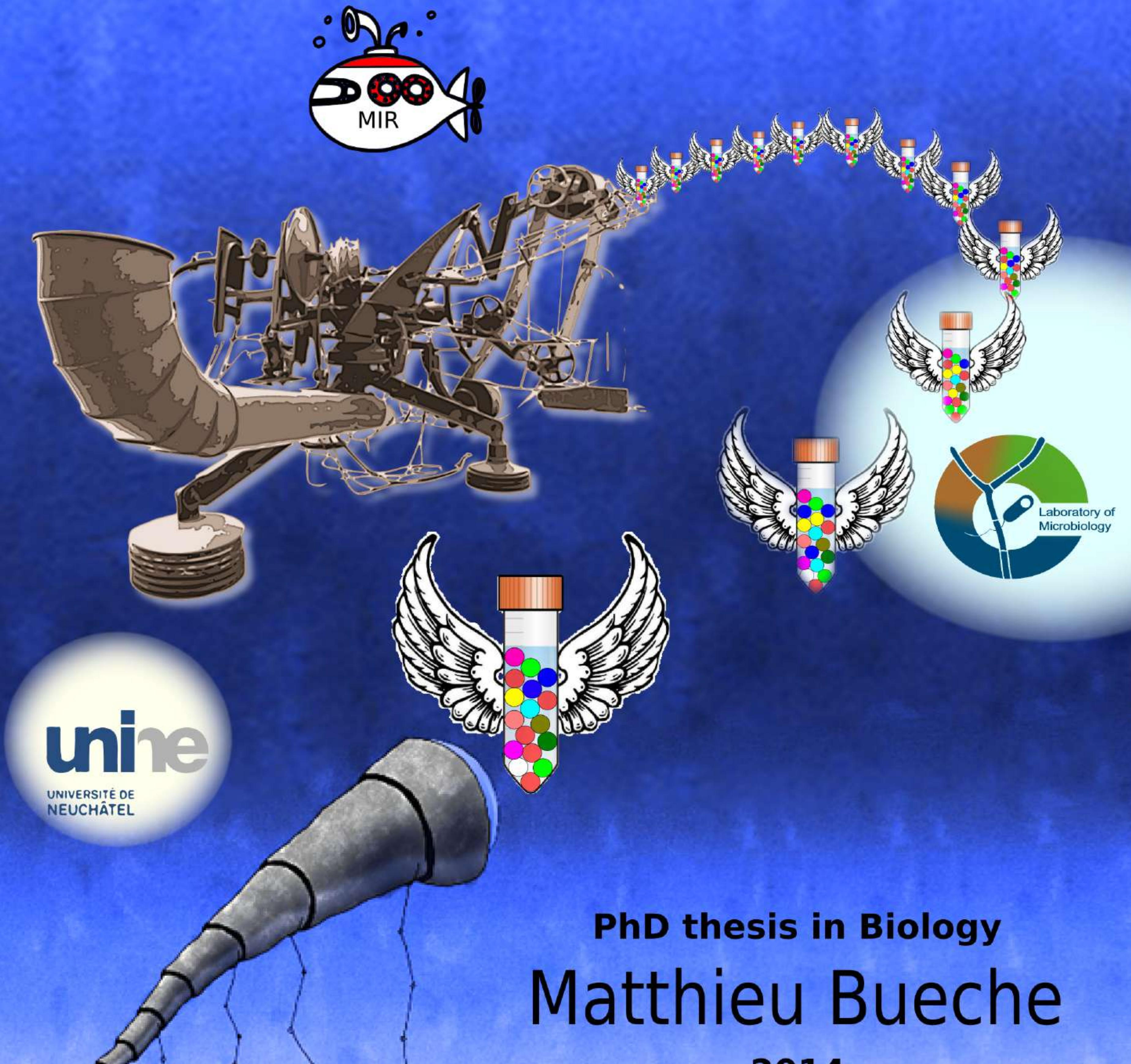
Spore-forming bacteria as indicators of pollution in sediments of Lake Geneva



PhD thesis in Biology

Matthieu Bueche

2014

UNIVERSITÉ DE NEUCHÂTEL

Faculté des sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel - Suisse
Tél: + 41 (0)32 718 2100
E-mail: secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Matthieu BUECHE

Titre:

"Spore-forming bacteria as indicators of pollution in sediments of Lake Geneva"

sur le rapport des membres du jury composé comme suit:

- Prof. ass. Pilar Junier, Université de Neuchâtel, directrice de thèse
- Prof. Daniel Hunkeler, Université de Neuchâtel
- Dr Jakob Zopfi, Université de Bâle
- Prof. Torsten W. Vennemann, Université de Lausanne

Neuchâtel, le 22 mai 2014

Le Doyen, Prof. P. Kropf

P. Kropf

Thesis direction:

Assistant Prof. Pilar Junier
Laboratory of Microbiology, University of Neuchâtel, Switzerland

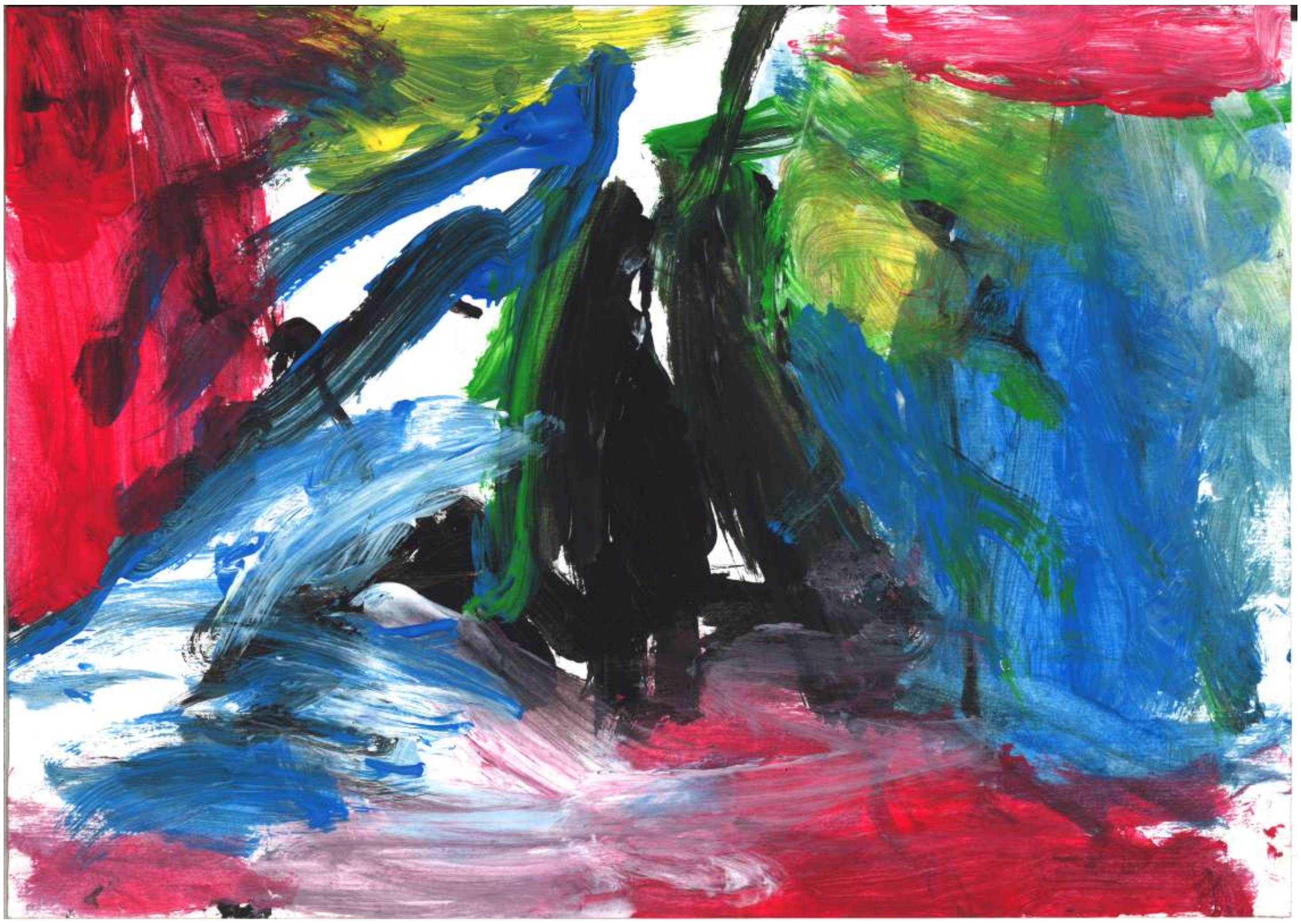
Assessment committee:

Prof. Daniel Hunkeler
Center for Hydrogeology and Geothermics, University of Neuchâtel, Switzerland

Dr. Jakob Zopfi
Department of Environmental Sciences, Environmental Geosciences / Biogeochemistry, University of Basel, Switzerland

Prof. Torsten Vennemann
Faculty of Geosciences and Environment, Stable Isotope Geochemistry, University of Lausanne, Switzerland

Private defense: 28th of April 2014



Dedicated to Zacharie, my little boy, who was born in the beginning of my thesis and has reoriented my priorities towards the essential things of life during these three years of hard work

Spore-forming bacteria as indicators of pollution in sediments of Lake Geneva

Short abstract:

Two components: Endospore-forming Firmicutes are well known for their capacity to survive under harsh environmental conditions through the formation of strong and persistent structures called endospores. Sediments of the Bay of Vidy (Lake Geneva), that are well known for their heavy metal pollution. The question was: Are endospore-forming Firmicutes predominant members of the bacterial communities in the most impacted sediments? The co-occurrence of these two components was effectively observed, and four particular strains of Clostridia were identified as dominant. However, their presence could be related to a combination of a dysfunction of the waste water treatment plant infrastructure and their survival in the state of an endospore. Episodic flooding events in the waste water treatment plant is effectively strongly suspected and supported by the data collected during this thesis. At the end of this work, the dominance of this particular group of bacteria does not seem to be uniquely related to their resistance to high concentrations of heavy metals, but also to the environmental changes occurring during the waste water treatment process and/or during the burial of bacterial communities in sediments.

Keywords: Endospore-forming Firmicutes, heavy metals pollution, waste water treatment plant, Lake Geneva sediments

Preface

This thesis was submitted to the Faculty of Science at the University of Neuchâtel (Switzerland), as a partial fulfilment of the requirements to obtain a degree of PhD. The work presented was carried out during three years, from 2011 to 2014, in the laboratory of microbiology headed by the assistant professor Pilar Junier. These three years were however not entirely dedicated to thesis work, but also to very interesting, but time consuming, academic duties with bachelor and master students. As a counterpart, the collaboration with a master student (as Master thesis supervisor) during the first year of the project was very helpful and stimulating.

The title of this thesis work is "Spore-forming bacteria as indicators of pollution in sediments of Lake Geneva". However, with the different findings and the inevitable re-orientation that occurred all along such a work, it should perhaps have been more correct to re-entitle it as follows: "Co-occurrence of trace metals and endospore-forming Firmicutes in sediments of Lake Geneva contaminated by human faeces". The project is part of the Elemo (exploration des eaux du Léman) project initiated thanks to Ferring Pharmaceutical and the EPFL (Ecole polytechnique fédérale de Lausanne). By grouping together scientists from different fields, but also from different Swiss and foreign universities, the Elemo project aimed at a better understanding of the bio-, geochemical functioning of Lake Geneva. Because of the increase of the population around its shore, Lake Geneva experienced in the recent decades an ever increasing anthropic pressure, which needs to be mastered to preserve the quality of its water. One and a half million people live on Lake Geneva's shore and about one half of them depend on it for drinking water resources. Furthermore, Lake Geneva is part of the Rhone river, which provides different water services for an important part of France. Preserving the quality of this water resource is therefore capital for the future.

During summer 2011, thanks to the collaboration with the "Consulat honoraire de la fédération de Russie à Lausanne", two Russian MIR submersibles were available for scientific field work in Lake Geneva. The presence of the submersibles in the lake was a large media event for EPFL and Ferring Pharmaceutical, but at the same time it provided to the scientists an incredible opportunity to examine more in detail different aspects of the water column, as well as the sediment floor of the lake.

As different scientists from different disciplines were involved in this project it was also, for each of us, an incredible opportunity to collaborate and to learn things from the others. This thesis project would certainly not have been the same without the helpful collaboration we had with people from the University of Geneva.

As said by Philipe Gillet (Vice President for Academic Affairs at the EPFL) during an interview, the Elemo project is a beautiful means to meet and to learn between different cultures, but also between different scientists.

Why is it so important to preserve the environment?
We often hear environmentalists saying that we should take more care of the environment, by preserving species and biodiversity. In most of the cases, I however think that people understand these recommendations as a means to preserve something that is not really useful for humans.
Many people see nature as a place, outside the cities, where some farmers are working to produce what people eat. For others, nature is a place where one can make some leisure activities or see beautiful animals or plants. For all of them, nature is something our species can exploit without limit, to meet our needs, as a kind of service specially dedicated to humans. This particular vision of the world is in fact probably due to the central place occupied by humans in the Christian religion. Effectively, as children of God, nature is perceived more as a theater set rather than the cradle of every life form on Earth.
I, however, think that the main problem of this point of view, is that people do not consider themselves as a part of this complex ecosystem, and therefore do not understand why it is so important to preserve the natural environment. They have not yet realized, that if nature is impaired, the human species will also be impaired, in a way or another. In their book, "What's life?", Lynn Margulis and Dorion Sagan gave an interesting but unusual definition of humans: "People, for example, redistribute and concentrate oxygen, hydrogen, nitrogen, carbon, sulfur, phosphorus, and other elements of Earth's crust into two-legged, upright forms that have an amazing propensity to wander across, dig into, and in countless other ways alter Earth's surface. We are walking, talking minerals!"[1]
In an evolutionary perspective, our species is on the top of the tree of life, far away from its trunk. One can see this position as an advantage giving us a kind of superiority on other "primitive" species. One must however not forget that this position is also the result of an extremely fragile equilibrium that depends on processes occurring thanks to organisms at the base of the tree of life. The life of "superior" organisms is effectively closely linked to the services provided by microbial communities in the ecosystems. A better understanding of their function is, therefore, necessary to understand how ecosystems work and how one can preserve them against deleterious human activities. Preserving the planet will first benefit ourselves! That is why research projects in environmental science are so important.

Thesis objectives

The global objective of this thesis was to achieve a better understanding of the relationship between endospore-forming Firmicutes and heavy and trace metal pollution accumulated in freshwater lake sediments. The thin red line followed all along this thesis could probably

be summarized as: "How to find a way to get the least biased vision of a particular portion of the bacterial community inhabiting lake sediments ?" If this condition seemed relatively easy to achieve in a first approach, this was, however, not always the case as the work progressed. During the three years at the University of Neuchâtel, my knowledge of this particular subject has effectively evolved, highlighting, step by step, different biases. Bypassing them was always the result of choices aiming at a minimization of the distortion of the reality. But what is exactly reality when we are talking about microbial communities? It's hard to get an unequivocal answer to this question!
In summary, if this awareness of the limits of science has greatly contributed to my personal education, I also hope that the different methodological and computational biases that were identified all along this thesis, as well as the way they were bypassed, will contribute, in the future, to an increase of our knowledge on endospore-forming Firmicutes, but also on the fate of heavy and trace metals in the environment.

Thesis outline

This thesis is in the form of a synopsis with attached publications. It is divided into three parts. Roughly, Part I is the core of the thesis and is composed of global informations on the subject, published papers and finally a conclusion. Part II is constituted by additional publications in which the author has contributed. Finally, part III is composed by an updated *Curriculum Vitae.*

In Part I, the first chapter introduces rapidly some particular properties of water. It gives an overview of the environmental issues linked to the pollution of fresh water resources with a particular emphasis on heavy metals collected by run-off water. The main remediation techniques used in heavy metal polluted soils and sediments are finally exposed. In a second part of chapter one, some basic information on bacteria is provided. This part defines, in particular, the kind of bacteria found in lake sediments, but also in waste-water treatment plants. This second part of chapter one ends with general informations on endospore-forming bacteria, as well as with a small description of the bacterial resistance mechanism for fighting against high environmental concentration of metals. Finally, in a third section of chapter one, a global presentation of Lake Geneva and the pollution issues in the Bay of Vidy is provided, with a particular emphasis on the rejection of the effluent of the waste water treatment plant of Lausanne. The functioning of the plant is then briefly described and this section ends with a small description of separation status of waste and run-off water for the city of Lausanne.

Chapter 2 tries to give a global overview of key methodological particularities used all along this work. The chapter starts with the description of the sampling procedure with MIR submersibles and the research boat from the Forel-Institute (University of Geneva, Switzerland). A quick description of the different methods available to study microbial communities is then provided. This chapter reports also the preliminary experiments performed to select and design the more efficient method to extract DNA from endospore-forming Firmicutes buried in Lake sediments. Finally, it describes also the pyrosequencing of the DNA samples and the bioinformatic treatment of the sequences.The chapter concludes with a quick description of the chemical analysis performed in this thesis.

Chapter 3 is composed of the publications of the author. It starts with one conference proceeding, "Endospore-forming bacteria as an indicator of pollution in sediments of Lake Geneva", reporting the participation to ICHMET 2012 (International conference on heavy metal in the environment), in Roma. This is followed by the *in press* article in Aquatic Sciences, "Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria"; the published article in Applied and Environmental Microbiology, "Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene *spo0A*"; and finally, the chapter ends with the article in preparation for Environmental Science and Technology, "Endosporulation as a resistance mechanism to metal-toxicity in sediments"

Chapter 4 describes laboratory experiments performed with four Clostridia strains and different heavy metals. Initially, it was planned to publish the results presented in this chapter. Experiments were however not sufficiently accomplished to allow this. These experiments explored the effect of heavy metal on sporulation rate and growth dynamic of the four Clostridia strains. During these experiments it was observed that heavy metals precipitated readily in the culture vials. Experiments to try to understand if this precipitation phenomena is related to physico-chemical conditions prevailing in the vials or to an immobilisation phenomena by bacterial biomass are finally reported.

Finally, Chapter 5 is composed by a synthesis of the work performed during this thesis, giving the main conclusions that can be drawn, as well as their outlook. A critical review of the different bias encountered all around this thesis is also provided in this chapter.

List of Papers

I Bueche, M., Sauvain, L., Wunderlin, T., Junier, T., Tercier, M., Loizeau, J.-L., Masson, M., Gascon, E., Junier, P.
Endospore-forming bacteria as an indicator of pollution in sediments of Lake Geneva
E3S Web of Conferences 1 (**2013**),doi: 10.1051/e3conf/20130133011

II Sauvain, L., Bueche, M., Junier, T., Masson, M., Wunderlin, T., Kohler-Milleret, R., Gascon D., E., Loizeau, J.-L., Tercier-W., M.-L., Junier, P.
Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria
Aquat. Sci. (**2013**), doi: 10.1007/s00027-013-0313-8

III Bueche, M., Wunderlin, T., Roussel-Delif, L., Junier, T., Sauvain, L., Jeanneret, N., Junier, P.
Quantification of Endospore-Forming Firmicutes by Quantitative PCR with the Functional Gene spo0A
Appl. Environ. Microbiol. (**2013**), doi: 10.1128/AEM.01376-13.

IV Bueche, M., Sauvain, L., Junier, P.
Endosporulation as a resistance mechanism to metal-toxicity in sediments
In preparation... To be published in Environmental Science and Technology.

V Wunderlin, T., Corella, J. P., Junier, T., Bueche, M., Loizeau, J.-L., Girardclos, S., Junier, P.
Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva (Switzerland–France)
Aquat. Sci. (**2013**), doi: 10.1007/s00027-013-0329-0

VI Wunderlin, T., Junier, T., Bueche, M., Corella, J. P., Loizeau, J.-L., Girardclos, S., Junier, P.
Endospores in sediment as a proxy for 20th century environmental lake history
In preparation... To be published in Nature communications.

VII Filippidou, S., Bueche, M., Wunderlin, T., Junier, T., Sauvain, L., Roussel- Delif, L., Jeanneret, N., Dorador, C., Molina, V., Ioannidou, A., Vargemezis, G., Johnson, D., Junier, P.
Survival strategy meets classic ecological theories: the case of endospore-forming bacteria diversity and abundance in extreme environments
Abstract for ISME2014. To be published in ISME journal.

Contribution report

Papers I, II were written in close collaboration with Loïc Sauvain, a master student in our laboratory who has chosen to do his master work on this subject. In these two publications, I performed 75% of the design of the experiments, 50% of the sampling procedure, 50% of the fluorescence microscopy and cell counting, 50% of the DNA extraction, and subsequent molecular characterisation of the bacterial communities, 100% the bioinformatic analysis, 75% of the statistical analysis, and, finally, 75% of the data representation.
Paper III was written by myself. In this publication, I performed 75% of the evaluation of target genes involved in endosporulation, 100% of the elaboration of a refined database of endospore-forming Firmicutes sequences, 100% of the primer development, 100% of the *in silico* specificity test by BLAST, 75% of the strains cultivation, 75% of the experimental accuracy tests, 100% of tuning of the PCR and qPCR conditions, 100% of the PCR and qPCR experiments with pure strains, 50% of the characterization of the environmental sequences, and, finally, 100% of the data analysis and representation.
Paper IV was written 75% by myself. In this publication, I performed 75% of the sampling, 50% of the chemical analysis, 75% of the DNA extraction, 100% of the bioinformatic analysis, 75% of the qPCR on 16S rRNA and *spo0A* genes, 100% of the DPA quantification, and, finally, 100% of the data analysis and representation.

Paper V and VI are the main work of T. Wunderlin where I contributed with the development of the DNA extraction methodology and the development of the qPCR quantification of endospore-forming bacteria. Finally, the seventh article, which is in fact not yet published, is the main work of S. Filippidou where I contributed with some DNA samples, as well as methodological, bioinformatic, and statistical analysis issues.

The first four papers are presented in the core of the thesis, whereas the fifth to the seventh are presented at the end.

Contents

Part I

Core of the Thesis

1

Background and Motivation

The aim of this first chapter is to highlight the motivation of the research conducted in this thesis. It begins with an introduction to aquatic environments and the pollution problems they are facing under anthropic pressure. Then an overview of the bacterial world is presented with a particular focus on endospore-forming bacteria. Finally, a short description of Lake Geneva is added to introduce the sampling environment used in this study.

Contents

1.1 Water

Water is an essential constituent of the biosphere in which it is present in each of its three phases: gaseous, liquid, and solid. Water is the main substrate for biological activities and also the main constituent of living organisms (up to 70% of their weight). The quantity of water on earth is estimated at around $1350*10^6$ km^3. 97% of this quantity is represented by the oceans and the remaining 3% is distributed between soft continental water (0.6%) and ice and snow mainly from polar caps (2.4%)(Figure 1.1). A very small fraction is also contained in the atmosphere (as clouds and vapours) and in living organisms (around $1*10^{-4}$% of the total).[2]

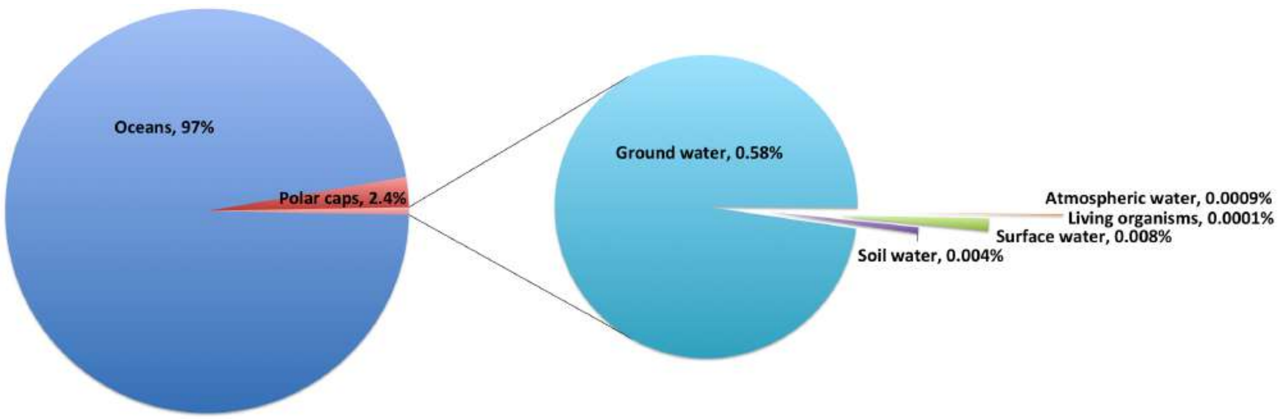


Figure 1.1: Repartition of water resources on earth.

Thanks to its physical and chemical properties, water plays a predominant role on climate regulation on the Earth. As universal solvent it allows the availability and the transport of nutrients in ecosystems.[3] Through dissolution and precipitation phenomena, water provides the main link between the geological cycles of elements from rocks to the uptake and temporal sequestration of them by the biosphere. This is the "coupling belt" between slow geological and fast biological cycling of elements on Earth.

Particular properties of water. As for other chemical compounds, molecules of water become closer when temperature deceases. This is, however, only true down to 4°C where fresh water reaches it is maximum density. At a lower temperature fresh water starts to freeze and molecules become more distant from each other.This phenomenon is called water thermal anomaly and has consequences on the functioning of aquatic ecosystems. For instance, this mechanism allows water stratification in lakes and oceans, leading to particular dispersion of nutrients and dissolved gases in the water column during different periods of the year. The fact that ice density is lower than the density of cold water is also interesting because it allows ice to float at the surface of water bodies. This ice sheet isolates the underlying water column from a cold atmosphere and allows living organisms to stay alive during the coldest seasons (Figure 1.2).

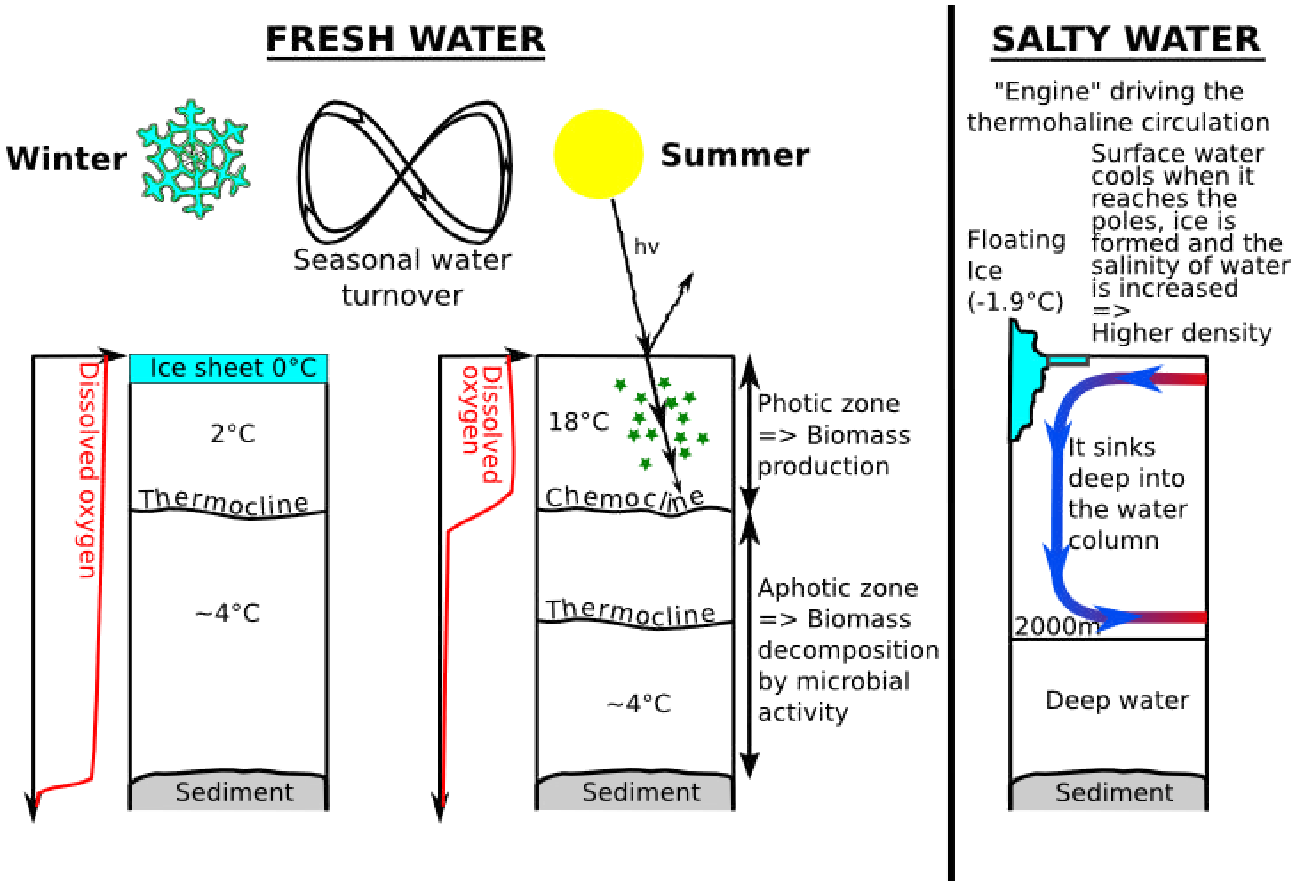


Figure 1.2: Illustration and implication of the water thermal anomaly in fresh and salty water.

The atoms in the water molecule are asymmetrically positioned giving to it polarity (negatively charged in the oxygen side, positively charged in the hydrogen side). This polarity allows the formation of hydrogen bonds between water molecules leading to some of its more interesting properties. In physics, the amount of energy needed to change a compound from one state to another (for instance from liquid to gaseous) is called latent heat. For water, latent heat is exceptionally high (only ammonia has higher values) and plays a capital role in the regulation of temperature on Earth. The presence of liquid water on Earth also derives from this high capacity to absorb energy without changing state.
The partial charge of the water molecule also explains why water is so efficient as a solvent for salts (ions) or other polar compounds. The dissolution of salts also has an influence on the properties of water. For instance, salty oceanic water, in contrast to fresh water, has a maximum density at a temperature below its freezing temperature. In the polar zones, where surface water is strongly cooled, this allows a large proportion of it to descend deep in the water column before the formation of ice. This phenomena is part of the principal engine driving the well known thermohaline circulation, which determines the oceanic currents all around the world. Atmospheric currents and the different climatic zones are greatly determined in response to this oceanic circulation.[4]
In the atmosphere, water vapour absorbs infra-red radiations emitted by Earth on reaction to solar irradiation. By itself this absorption accounts for about two thirds of the natural greenhouse effect. The formation of clouds and raining events are also a derivative from the polarity of water molecules. This comes from the high surface tension of water which allows for the formation of drops in the clouds.

From a biological perspective, the specific properties of water have also played a determinant role in the evolution of life on the planet. First its polar properties have permitted the formation of phospholipid droplets (which are amphiphile), which were probably the first forms of closed structures in which proteins have evolved, finally leading to the formation of protocells in the prebiotic soup.[5] The second important aspect lies in the oxidation state of the atoms in the water molecule. The oxygen atom in water molecules, is at its most reduced state. Chemically speaking, water is, therefore, an electron donor. During oxygenic photosynthesis these electrons are energized by light radiation and this energy is used to reduce molecules of carbon-dioxide (CO_2) into molecules of organic carbon ($<CH_2O>$), mainly sugar. As the the gap of energy between water and organic carbon is so high (redox potential of 1.14V between H_2O and $NADP^+$) this process requires the functioning of two serial photosystems.[6] Evolutionary speaking, the use of water as an electron source to fix carbon-dioxyde into biomass through oxygenic photosynthesis was a big step. Because of the availability of water in many places on Earth, living organisms gained the ability to colonize almost all the environments, and were not limited any more to places where reduced mineral compounds were excreted from hydrothermal sites, as previously. The presence of light radiation is of course also determinant for the production of biomass through photosynthesis. The transparency of water for the different wavelength of light is, therefore, also determinant for the zone in which the organisms responsible for primary production are established in the water column. Up to a given depth, photosynthesis will occur and water will be enriched in organic matter and dissolved oxygen. Below a certain level, there is not enough light to sustain photosynthetic activity and decomposition processes (mainly through microbial activity) will be dominant, consuming oxygen. Depending on different factors, such as the the quantity of biomass produced in the photic zone or the thermal stratification of the water column, a particular balance will occur between production, decomposition, and dispersion of molecules, leading to a given amount of dissolved oxygen at a particular depth of a lake (Figure 1.2). In eutrophic environments, where water is enriched by inorganic nutrients (like nitrogen or phosphorus), this can lead to anoxic conditions in the deeper layers of the water column. The use of dissolved oxygen by decomposition processes will also occur at a micro-scale in sediments.

Function and importance of water for life. What is the function of water in the ecosystem and organisms? Why is water so important for organisms? What are the water quality requirements to sustain life?

All known forms of life depend on water. In nature, rocks and other molecules are dissolved and transported by water to feed organisms at different scales and trophic levels. In evolved multicellular organisms water compose around 70% of the body mass and this amount is precisely regulated by homeostatic processes. In such organisms, as in natural ecosystems, water is involved in the transport of different chemical substances at different levels. Nutrients and metabolic waste, as well as communication molecules, are transported in aqueous solutions (cytosol inside the cells, blood and lymph outside the cells). From a biochemical point of view, water is at the center of catabolic and anabolic pathways. In heterotrophic cells, cellular energy is produced by the oxydation of small monomers (like glucose) by the cellular machinery. In nature, organic carbon is found in large quantities in polymers, such as starch, fatty-acids or proteins. The first step needed to use energy from these complex molecules is a hydrolysis (the addition of a water molecule) for the break-down into monomers. These monomers, as explained above can either be used to produce energy, or can be assembled to build new polymers through anabolic processes. Polymerisation occurs via a dehydratation step. Because life machinery is so complex, and because water is at the center of the main metabolic reactions, the need of high quality pure water is, therefore, essential for sustaining life on Earth. But what does exactly mean HIGH QUALITY PURE water?

1.1.1 Pollution

What is exactly pollution? Is it a measure of harmfulness for living organisms? In fact the answer to these questions is not easy to find. In nature a lot of substances, like cyanides or mycotoxins (like afflatoxins, for instance)[7] have harmful effects on life, but we do not normally consider them as pollution. The reason for this is that these substances originate from nature. But in this case, what about the poisoning of lakes by microcystins produced by some species of cyanobacteria?[8] Of course, their success is often the result of an eutrophication phenomenon; but microcystin-producing cyanobacteria have not been introduced in nature by humans. The notion of pollution has in fact greatly evolved during the last century in response to the increasing demography, the intensification of land-use, the industrial development, the progress in environmental sciences, and also in measuring technologies.[9] As a summary one can say that pollution occur when a substance with anthropic origin accumulates somewhere and generates a harmful effect directly on humans or on an ecosystem that provides a service to humans. Pollution can also be generated by the proliferation of a particular harmful organism under anthropic pressure. This focus on human benefits is important because it explains how the notion of pollution has evolved during the last decades. Ecology is the discipline of life sciences which tries to understand links between organisms and ecosystems. Notable advances occurred in this discipline thanks to the systems theory. To understand that almost all ecosystems are linked together it was effectively first necessary to understand natural environments in a systemic point of view. With such a comprehension of nature one can easily understand that every source of pollution can affect human health in a way or another through time.

Sources of pollution. What are the main sources of pollution known today? Pollution can be divided into two main categories depending on the way they enter natural ecosystems. Point-source pollution correspond to a large amount of a pollutant that is released in a particular point. Landfills and industrial effluents are two examples of this. In developed countries this kind of pollution source is generally under control because infrastructures to avoid pollutants dispersion are imposed by law.[10] In Switzerland, for instance, domestic as

well as industrial or agricultural waste water[11] need to be treated before reaching aquatic environments. The same is true for the effluents of landfills. Of course, even in developed countries, there is the emerging problem of micropollutants (traces of drug molecules and other chemicals, for instance), which are found in treated waste water and released into the environment. This source of pollution should, however, soon be limited by recent developments in the waste water treatment technologies. In developing countries (covering more than 3/4 of the entire world), due to their economic situation, point-source pollution is still a big environmental issue. In the majority of cases, urban and industrial effluents are released in the environment without any treatment.
The second category, the more difficult to control, correspond to diffuse pollution. Diffuse pollution is spread in larger areas in the environment and pollutant's concentrations are much lower than in the case of point sources. The typical example of diffuse pollution is the dispersion of pesticides and amending substances in agriculture. These substances are temporally stocked in soils but finally reach aquatic environment through leaching or erosion events. The second typical example is urban runoff water which collect all urban dusts and concentrate them in stormwater collection points. In some cases, even in developed countries, this stormwater is massively released into the environment without any treatments.

Runoff water. Aquatic environments are the most sensitive, depending on the size of the basin, because at one time or another toxic substances will reach them, finishing their travel into the oceans. In this study we have chosen to focus on the accumulation of urban originated pollutants in lake sediments. Urban runoff water is of particular interest because it will collect all the urban dust released by road infrastructure and usage, organic wastes from domestic / municipal wastes and animal droppings, but also from the normal degradation and usage of buildings and urban infrastructure (like public transportation services). The main substances released are mineral dust from different origins, but also nutrients, chlorides, hydrocarbons, pesticides, and finally trace-metals. On the list one can also add viruses and bacteria from faecal origin.[12] With the increasing of urban population, the amount of pollutants collected by urban runoff is poised to become one of the most problematic environmental issues in the upcoming years.
Because of their relatively low concentration in runoff, the most problematic substances from the list enumerated above are probably pesticides and trace-metals. Even if runoff water is properly treated in waste water treatment plants, such facilities often lack particular infrastructure for removing such micro-traces of pollutants. From different studies, levels of trace-metals measured in runoff water collectors are the following: 20 to 5000 $\mu g/L$ of zinc, 5 to 200 $\mu g/L$ of copper or lead and around 12 $\mu g/L$ of cadmium.[13] An evaluation of different sources for these four trace-metals (Cu,Cd,Pb, and Zn) has highlighted that contribution of each of them to runoff water pollution is highly dependent on the type of urban zone studied (commercial or residential). In summary, Cu is mainly coming from brake emissions of vehicles or from building siding. Cd is entering the system mainly by wet deposition (urban aerial dust or aerosols that are deposited during rain events), by leaching of building walls or by dry deposition (urban aerial dust deposition when weather is dry). Pb results from building walls or atmospheric depositions. Finally, Zn comes from

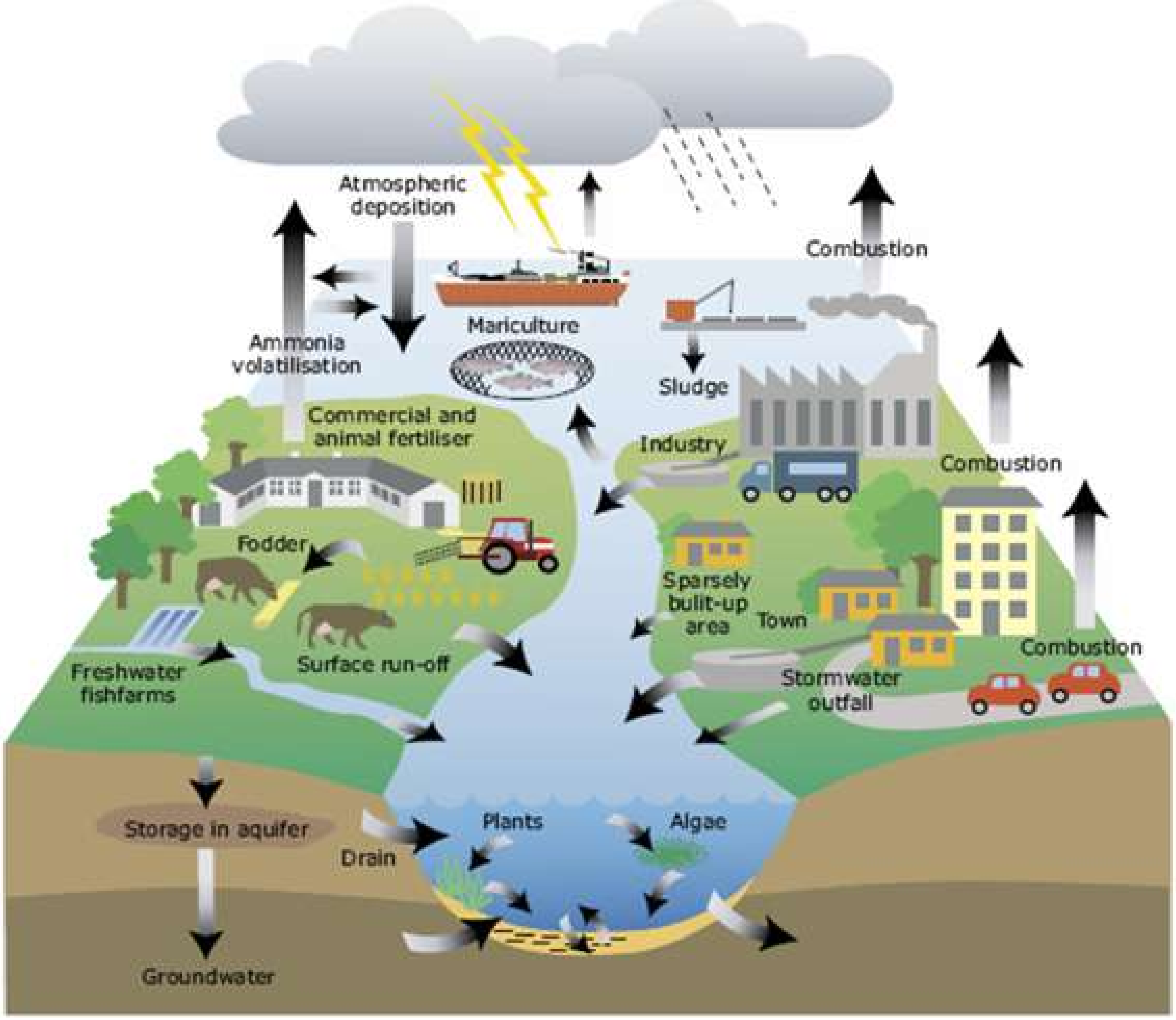


Figure 1.3: Water pollution sources (Figure from the web site of the Environmental Protection Agency from the Danish ministry of the environment) http://www.mst.dk (2014)

the leaching of brick walls and vehicle tire wear.[13]

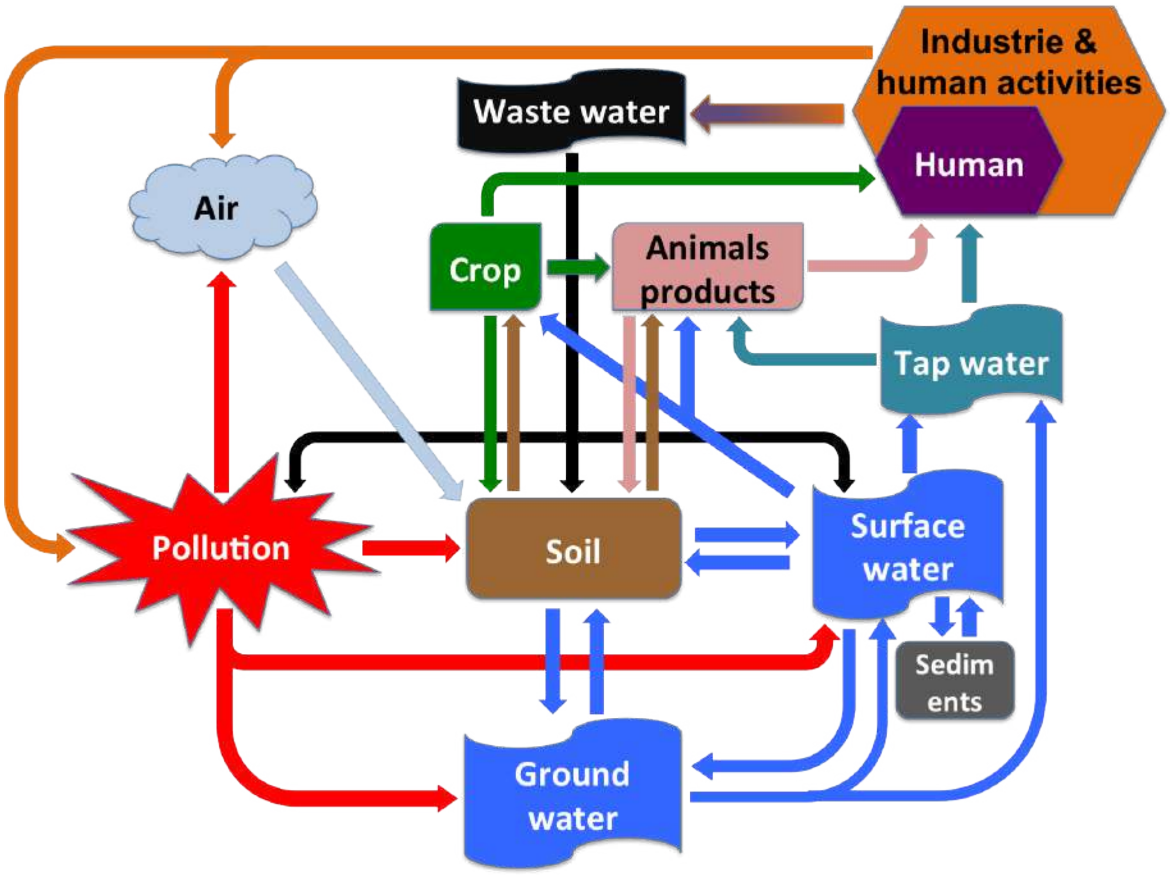


Figure 1.4: Pollutant fluxes in the environment are mainly mediated by water. Some ecosystem compartments could accumulate them for a relatively long period of time. Finally, humans are mainly exposed through food produced in polluted environments.

Accumulation of heavy metals and their release to the environment. What is the way pollutants accumulate in aquatic environments? What is the impact of these accumulations? Soils are particularly prone to accumulation of heavy metals, because they are porous media through which heavy metal polluted water can percolate. Soil is mainly composed by mineral particles of different sizes held together by decaying organic matter. Both of these two compounds, but particularly small size minerals (like clay) associated with humic acids (argilo-humique complex) confer a high cationic exchange capacity (CEC) to the structure, allowing complexation of any cation.[14] In some particular conditions (like over-fertilisation) the sequestration of heavy metals by soils could be reversible and metals could be leached, finally leading to a contamination of natural surrounding water. Concerning sediments, the way they accumulate heavy metals is quite the same, since sediments are often the result of soil erosion and deposition on the floor of water bodies. Furthermore, sediments are enriched by the deposition of decaying organic matter and waste from aquatic animals and plants that live in the overlying water column. Some anthropic actions, like the release of treated or untreated waste water in lakes and rivers, can also enrich sediment in decaying organic matter, increasing the number of negatively charged residues leading to an increase of CEC (and consequently the binding capacity for metal cations). When bound to organo-mineral complexes, heavy metals seem to be relatively stable in the environment. Laboratory experiments have demonstrated that it is not easy to leach heavy metal cations (Cu, Cd, Zn and Hg) from artificially contaminated sediments, even at a pH as low as 4 or 4.5. This suggests that heavy metals are bound in sediment by sorption, and by the precipitation of secondary minerals.[15] In natural environments conditions are, however, not as simple as in the laboratory. The chemical speciation of metals is influenced by many environmental factors, like salinity, sediments granulometry, hardness (pH buffering effect of calcite), Eh, and pH conditions. All these factors could finally have an impact on the metal binding and release, as well as their bio-availability and toxicity.[16] Microbial activity in sediments often consumes very rapidly dissolved available oxygen, leading to a decrease of redox potential. In such conditions, when sulphates are present, sulfides are produced and can immobilise divalent metal cations through precipitation of secondary metal sulfides. These minerals are stable as long as the redox potential is maintained at a low value. Aeration, pH decrease or organic matter mineralization could lead to a release of these temporarily trapped heavy metals in the environment.[17]

All around the world sediment pollution by heavy metals is reported by scientists, particularly in areas with the higher densities of population or particular human activities, like harbours or ore extraction sites. A search in web of science for titles containing "heavy metal" and "sediment" as keywords reports close to 500 publications in the last fifteen years, giving us a global idea of the extend of the issue of sediments pollution by heavy metals. In Switzerland, for instance, values higher than national guidelines on soils contamination are regularly reported for lake or river sediments sampled downstream from major urban areas.[18] In the last years, ICPR (International commission for the protection of Rhine) has defined some target values for the sediments of Rhine and has edicted rules to reduce inputs of heavy metals in the near future. The values retained were the following: 1 mg/kg for Cd, 50 mg/kg for Cu and Ni, 0.5 mg/kg for Hg, 100 mg/kg for Pb and 200 mg/kg for Zn.[19]

Plants are well known for their capacities to concentrate toxic metals in their tissues.[14] In many places of the world, heavy metal rich waste water is used to irrigate cultures, leading to an accumulation of these metals in fruit and vegetables.[20] Similarly, an accumulation of heavy metals is also reported in vegetables irrigated by polluted water[21] or grown on polluted soils. Cultures grown in Romania in heavily polluted sites have reported concentrations of Cd and Pb in fresh root vegetables that exceed 1.25 times and 11 times, respectively, the limit fixed by the world health organisation (WHO). In leafy vegetables contents of heavy metals were 7 times higher than normal concentrations for Cd and 17 times higher for Pb.[22]

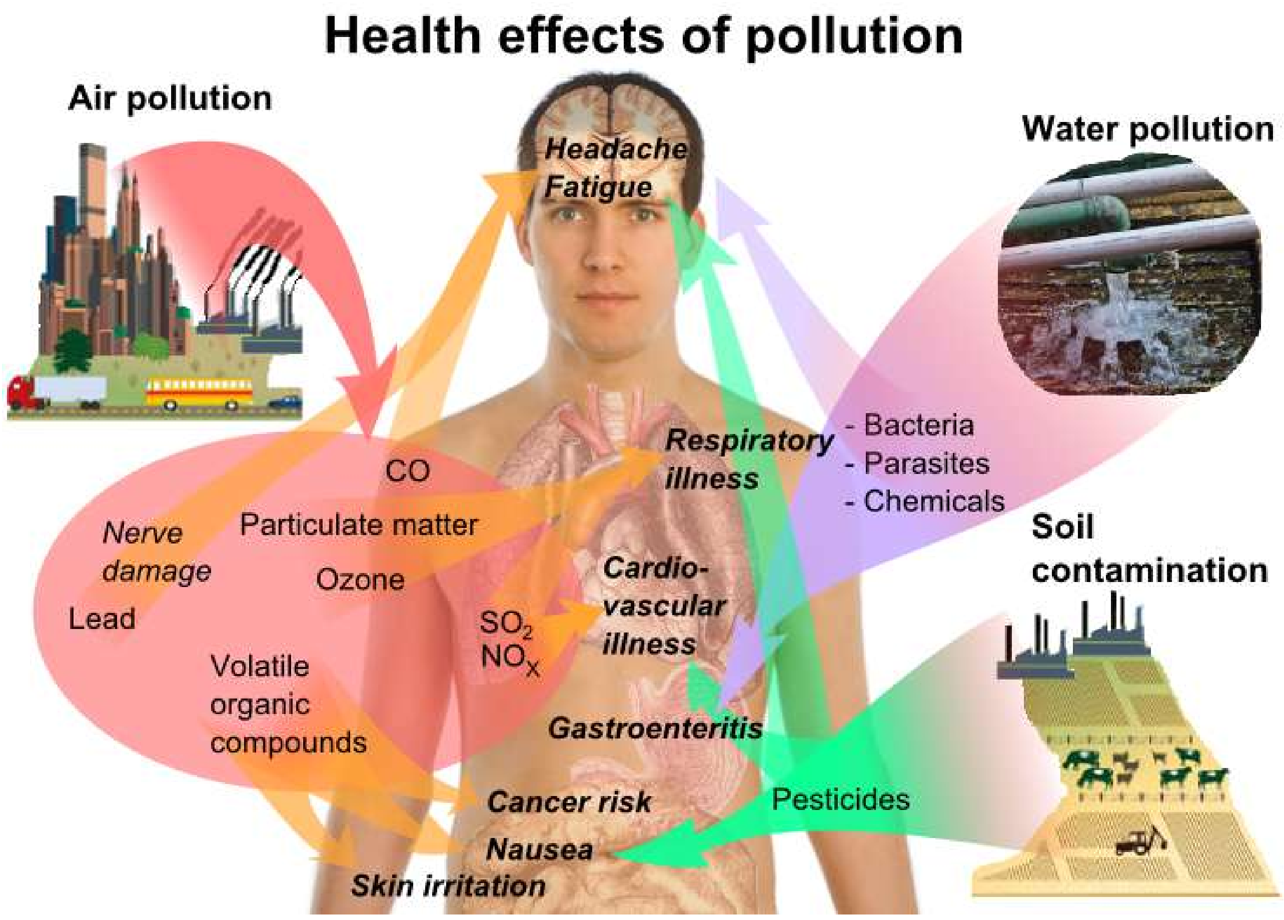


Figure 1.5: Overview of the main health effects on humans from some common types of pollution.[23]

Heavy metal toxicity. Metals play an important role as cofactors in many biochemical processes and are, therefore, necessary to living organisms as micronutrients. For instance, iron plays a central role in heme proteins like those involved in the respiratory chain in mitochondria (cytochrome c) or free-radical homeostasis (catalase). B_{12} vitamin (cobalamine) contains cobalt whereas the heterometal complex of nitrogenase from bacteria could contain iron, molybdenum or vanadium.[14] At higher concentrations, these metals, however, alter physiological functions in the cells leading to detrimental effects on the growth and reproduction of living organisms.[24]

Uptake, transport, utilization and release of metals by cells occurs mainly via metallothioneins and other metal binding proteins. In vertebrates, plasma proteins are the main systemic transporter for metals. In general the excretion of metals occurs via kidney filtration.[16] A focus on heavy metals shows that adverse effects on human health are mainly caused by an exposure to lead, cadmium, mercury and arsenic. Cadmium exposure occurs mainly through food and cigarettes. It has been demonstrated recently that an exposure to relatively low levels of cadmium could have a deleterious effect on kidney and bones on the long term, and that many European people are exposed daily to levels excessing these concentrations. For mercury, people are exposed through food (mainly fishes) and dental amalgam. Low risk of neurological troubles is encountered to a very small portion of the population that eat a large amount of fish. As foetus are much more sensitive, pregnant women must, however, avoid high fish consumption. Children are the most sensitive people to an exposure to lead, because of their "high gastrointestinal uptake"[i] and their "permeable blood–brain barrier"[i], causing a neurotoxic effect even at lower levels of exposure than previously thought. Exposure to arsenic occurs mainly via food and drinking water. Long-term exposure to arsenic will increase the risk of cancer and skin diseases.[25]

In sediments, biota is mainly made up of microorganisms. Exposure of microorganisms to high load of heavy metals has also detrimental effects on their metabolism.[26] In many cases a contamination of a microbial ecosystem by high levels of heavy metals leads to a dramatic drop of microbial diversity and a decrease of metabolic functions that are provided by the microbial community.[27,28] In bacteria, two mechanisms have been identified to regulate metal homeostasis. The first one involves metalloregulatory proteins that sense the concentration of a particular metal and act as a transcriptional regulator for the synthesis of ion uptake proteins. The second mechanism involves directly metal ion (mainly Mg^{2+} and K^{+}) which can act as a riboswitch to inhibit gene expression.[29] What is the long term effect of this reduction of bacterial metabolic diversity in sediments? Nobody knows!

1.1.2 Heavy metal removal techniques

Today, depending on the type of pollution, different removal technologies are available. The easiest way to avoid pollution of a natural ecosystem is of course to avoid the introduction of pollutants in nature. As the diversity of pollutants is very large, this chapter will focus on heavy metals, trying to draw the "big picture" of available techniques to remove those from the environment. Heavy metals, unlike organic contaminants, are not degradable and are persistent in polluted ecosystems. They can be displaced from one compartment to another of the ecosystem, but unless they are extracted they will remain there for a long time.

In the case of heavy metal polluted soils there are principally three types of remediation actions that can be used. The first one, probably the oldest, is the physical removal of contaminated layers and replacement by non-contaminated soil. This action could be very expensive depending on the area concerned. Furthermore, contaminated waste should be properly treated or deposited in a controlled landfill. One variation of this is the physical burying of the contaminated layers. Depending on the intended soil use, the surface could be covered by a uncontaminated fertile soil layer. This last solution is, however, not really a remediation technique as it only displace contaminants in deeper layers of the soil where they are not (or less) accessible to growing plants.[30] Different *ex situ* or *in situ* chemical treatment could also be used to leach heavy metals, but in many cases these chemicals induce either another source of pollution or diffuse it on a wider area.[31] Apparently, the two remediation methods that are most used actually are the chemical stabilisation or the phytoremediation. In the first case, some chemical non-toxic amendments are mixed with the soil to reduce the solubility and biodisponibility of heavy metals. Different materials can be used, such as lime, zeolite, clay minerals,

[i] Expressions gathered from Jarup et al (2003)[25]

compost, flying ashes, all involving complexation of metallic cations on negatively charged chemical groups at their surface.[32] This method, however, does not remove pollutants from the environment. Finally, phytoremediation is probably the more environmentally friendly technique available, but it takes time. Several variations of it exist, such as phytostabilisation, phytoextraction or rhizofiltration.[33]
Rhizofiltration is in fact principally used to remove contaminants from water which percolate through roots of bioaccumulating plants in rhizofiltration basins. As in other phytoremediation techniques, the plants are regularly collected and burned to concentrate pollutants. Ashes are finally confined and stocked in controlled landfills.[14] The "standard" operation of a waste water treatment plant, with coagulation-floculation technology, removes a high portion of heavy metals. In many cases, this is, however, not sufficiently effective and other treatments like chemical precipitation, membrane filtration, adsorption, ion exchange or electrochemical concentration should be added to remove a higher amount of heavy metals. From a review of several removal techniques the use of low-cost bio- or non-bioadsorbents seems to be a very promising technique for the treatment of heavy metal contaminated water. Among bioadsorbents, one can mention non-living biomass (potatoes peels, egg shells, chitin of crustaceans, rice husk), algal biomass (*Ulva lactuca* and *Cladophora fascicularis*) or microbial biomass (fungi and bacteria).[34] Heavy metal removal experiments with *Bacillus cereus*, *E. coli* or *Pseudomonas aeruginosa* have been reported in several studies. With their small sizes, bacteria provide the largest surface/volume ratio within living organisms, giving them excellent capacities to exchange molecules (nutrients and waste) with their environment. Depending on the strain and conditions, some bacteria can form biofilms to enclose cells together in a sticky matrix that can adhere to almost any kind of surface. The polymeric substances excreted to form biofilms are well known for their capacity to protect cells from adverse environmental conditions, but also for the adsorption of nutrients and other elements like heavy metals.[35] The main issue there is of course the recovery of contaminated microbial biomass after treatment. Mechanical immobilisation seems to be a good solution for various sorts of treatment processes.[36]

Today, the quality of discharges from sewage treatment plants continues to improve over time. With the ever-increasing volumes of water to be treated, the current problem that we must solve relies more on the accumulation of low-concentrated pollutants (like heavy metals and other micropollutants) on particular compartments of ecosystems. The well known biomagnification phenomenon in fish, leading to the accumulation of mercury in the food chain is an example of this. Sediments also act as a sink for different pollutants released in the environment. Since we do not know what are the long-term consequences of sediment pollution on aquatic ecosystems, it is, therefore, preferable to try to decontaminate them as much as possible. The removal of heavy metals from sediments is not a straightforward task. In fact, one can distinguish *in situ* and *ex situ* remediation of sediments. The first one principally aims at the stabilization of mobile and exchangeable metals. In the second, the sediments are dredged, the mobile metals are removed or immobilized and sediments are put back in place. This second solution is of course possible only for relatively small areas with high concentration of pollutants, such as harbours. In the *in situ* solution sediments can either be amended with substances increasing the sorption capacity (therefore, decreasing the mobility of heavy metals), like apatite or zeolite, or covered by an isolating substance (sediment capping) minimizing exchange with overlying water. These techniques do not remove pollutants, but transfer the problem to future generations.[17]

1.1.3 Personal view of today's challenges concerning water preservation

Clean water is essential for human life, but also for wildlife which maintain the "right" balance in ecosystem functioning. During its cycle, water evaporated from the oceans has many occasions to be contaminated by pollutants. First in the atmosphere where it can be contaminated by toxic molecules emitted from industrial plants or urban centres. During rain events it can also collect all the small particles (nano-size particles are the more persistant and problematic) of dust emitted by human activities. To fight against this atmospheric contamination, the first measure is, therefore, to limit as much as possible these harmful emissions. The major challenge is of course in developing countries where they do not have the financial resources to deal with this kind of problems. As there is only one single Earth supporting our common life, industrialized countries should, therefore, help them to produce things in a more environmentally friendly way.

Agriculture is the second main source of pollution in the water cycle. Both industrialized and developing countries should limit as much as possible the unwanted dispersion of chemicals in the environment. Of course, the "organic" revolution that developed countries experienced since relatively few years is more respectful for natural ecosystems. This is, however, a revolution of the "rich" and its impact on the entire Earth is probably not as significant as one can imagine. In fact, all around the world, erosion of soils is probably one of the major ways by which pollutants from agriculture reach aquatic environments. Agricultural practices should, therefore, be modified to maintain soil coverage all around the year. Such practices should furthermore significantly decrease the nutritive amendment needed because arable soil will stay in place rather than being removed by erosion.

Finally, due to population increase as well as the migration from rural to urban zones, urban centres are prone to a population increase in future. Satisfying the population needs for clean water as well as the treatment of the waste produced will become more and more challenging for big cities, even in developed countries. In order to be able to provide water of sufficient quality for many years, the rejection of pollutants around urban centres should be limited as much as possible. Here also, the help of developed countries is needed to sustain the efforts of developing countries in their struggle against the spreading of pollutants. People from industrialised countries often forget that a big proportion of what they eat comes from developing countries and is probably cultivated on polluted soils or irrigated with polluted water. If this is not yet true for all our food, it will probably become true in a near future.

1.2 Bacteria

Bacteria are the most ubiquitous forms of life on earth. In every place where liquid water is present, one can find active microbial life, coping sometimes with very harsh environmental conditions. Bacteria constitute the root of the tree of life and were present on Earth early in its history (around 4.3 Ga), playing an active role in the development of Earth's ecosystems. To deal with proto-Earth's conditions the first bacteria were thermophilic, anaerobic and also probably autotrophics (capable of building their biomass by the reduction of CO_2), using H_2 as electron donor. During a long period of time, bacteria have then experienced a metabolic diversification to take advantage of all the resources available on Earth at this time. Some kind of anaerobic photosynthesis also evolved. With the increase of biomass in the oceans, the amount of available organic carbon also increased, giving the opportunity for certain bacteria to evolve heterotrophy (the building of biomass by the recycling of biomass from other decaying cells). With the development of oxygenic photosynthesis, primary production is not any more established only around geothermal active zones exuding reduced sulphur compounds (typically H_2S), but can establish everywhere liquid water and light is present. As a consequence of this evolution, Earth's atmosphere became oxidant giving the opportunity to new more efficient metabolisms to evolve, leading to the development of larger cells (like eukaryotic cells) and multicellular organisms.[37]

The main particularity of bacterial cells stand in their lack of a nucleus. Bacterial chromosome, like other organelles is immersed in the cytosol and enclosed in the volume defined by the cell membrane. In bacteria, genetic information could be completed and enriched by the uptake of plasmids from the environment or by gene exchange. As growth occurs via cellular division, there is not a real genetic mixing that can occur through sexual reproduction, as in other organisms. Plasmid uptake and gene transfer along with mutations are, therefore, the only mechanisms that bacterial cells have at their disposal to evolve and adapt to new environmental conditions. In most bacterial cells, the membrane is surrounded by either a thick layer of a strong material (Gram-positives), or a thin layer of strong material and an outer membrane (Gram-negatives). In both cases, inner membrane play a determinant role in cellular energy production, as well as in the transfer of nutrients and waste in and out of the cytosol. Bacterial cells can have different shapes and sizes, but most of them have dimensions enclosed in a few micrometers, giving them excellent capacities for absorbing substances from the environment (greater surface/volume ratio in the biosphere). Some bacterial cells collect actively their nutrients through motility and chemotaxism mechanisms, while others stand in the same place, waiting for nutrients. Depending on the environment, both strategies have different trade-offs.

Phylosophical claim:
In terms of cell number, complex multicellular organisms harbour more prokaryotic cells than eukaryotic cells. In humans, for instance, around 90% of cells are bacteria mainly in the intestinal tract. Firmicutes are the dominant members of the community in our digestive system. In terms of genetic information, our microbiome harbour probably much more diversity than our own cells, allowing bacteria to be more versatile in terms of metabolic functions. This allows for the question of our real identity. Are we a kind of "tool" in the "hands" of bacteria, or do we have our own identity? From a bacterial perspective, complex organisms, like humans, are probably considered as a kind of elaborated biofilm in which they are heated, fed, and protected from the environment... In summary, a sort of bacterial paradise!

1.2.1 "Standard" microbiota from lake sediments

The main aim of this chapter is to try to define a sort of core microbiota one can find in freshwater lake sediments. Today, with the advances in next generation sequencing (NGS), the reflex is to try to define this by a comparison between NGS based studies on sediment microbiota from similar lakes. In fact, it makes not really sense to tackle this issue in such a way, because the answer will depend on many environmental factors. A theoretical consideration of the potential functioning of the benthic ecosystem is a better approach. The first factor to take into account is probably the depth at which the sediment floor stands, determining whether solar irradiation can reach the sediment floor or not, and this if phototrophs could establish at this depth. The scenario of phototrophic activity at the surface of sediment can, however, be forgotten in most of the cases since sediment floor often stands below the photic zone. In the same order of ideas, the kind of chemical stratification is important to determine if dissolved oxygen can reach the sediment surface (or the thin overlying water layer) or not. This factor will define a "starting point" in the electron tower (in fact, the chemical reagents available to produce energy via RedOx reactions) determining the kind of microbial metabolisms that can take place in the sediments. Prior to do the list of all the environmental factors that can be determinant, it is, therefore, important to define more precisely the context of this study. As described in the following chapters, this study focuses on Lake Geneva, which can be considered as a monomictic lake most of the time. The water column is, therefore, oxygenated down to the sediment floor. The lake is considered as mesotrophic, except in zones which suffer from the release of high quantities of nutrients, as described later. The sampling of sediment cores has been made below the photic zone and the temperature of the overlying water was approximately 10°C. In "conventional" lowland Swiss lakes, sediments are composed from runoff material brought by rivers, mainly eroded agricultural soils. In addition to the organic matter of terrestrial origin, sediments are regularly amended by the settling of dead organisms living in the upper water column (primary producers and heterotrophs feeding on them). A contribution of dissolved nutrients (like nitrate, sulphate, and phosphate) is also carried out

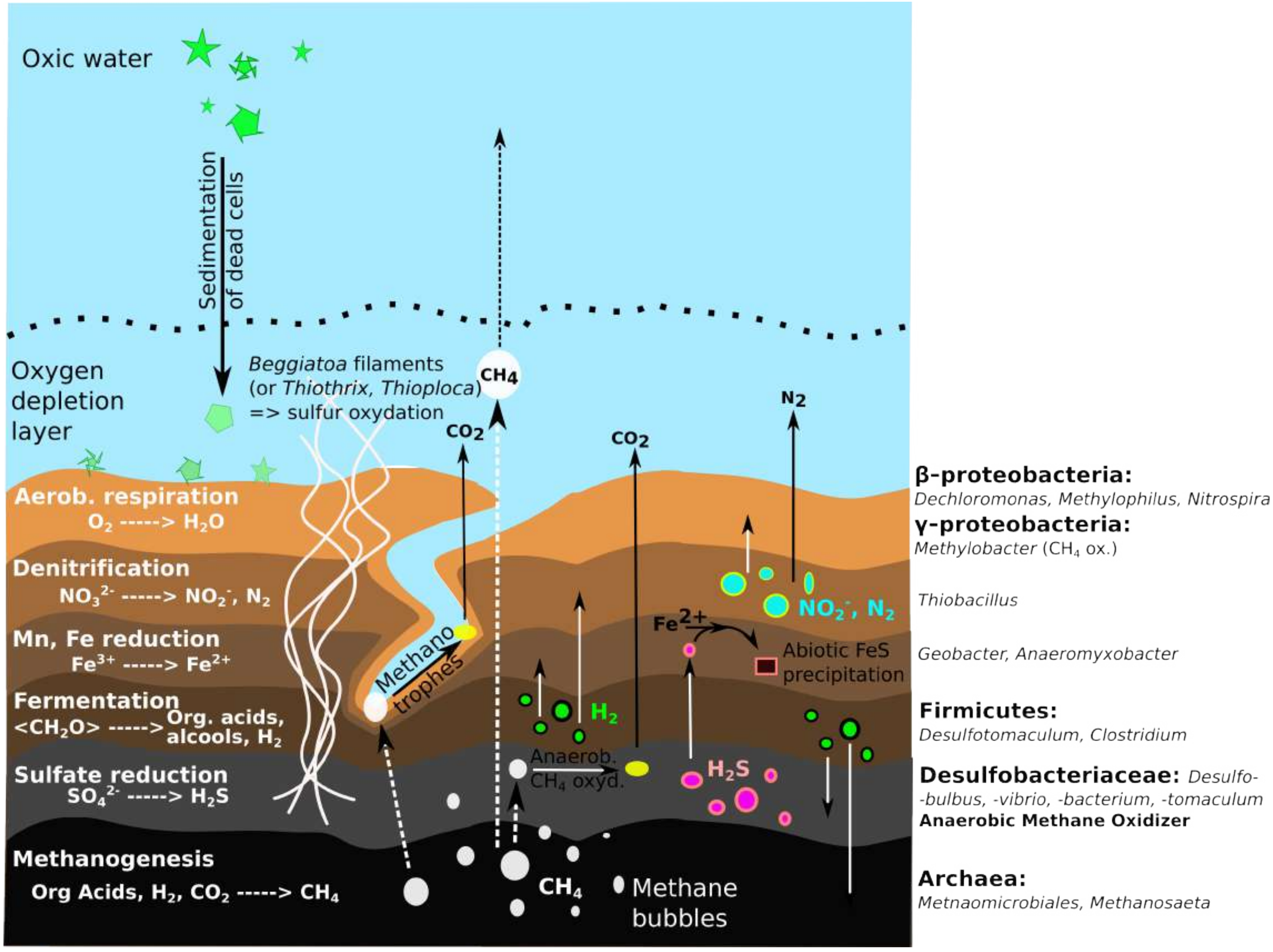


Figure 1.6: Overview of some microbial metabolisms and "Core microbial community" that is likely to be found in sediments from Lake Geneva

via the rejections from anthropic activities. From this short description of the ecosystem of concern one can start the description of possible microbial metabolism that can take place in sediments, therefore the bacterial groups that could be encountered.
As a starting point, a general shift in the communities could be observed while going down into the sediments. Starting from the dominance of Proteobacteria at the surface, one can find more Firmicutes with fermentative metabolism in deeper sediments, and finally methanogenic Archaea.[38] The information gathered from many authors[39–43] is summarized in the Figure 1.6.
The redox conditions become more and more reducing when going deep into the sediment. The steepness of this gradient is mainly a function of the availability and quality of organic carbon deposited on the top of sediment. For instance, in highly eutrophic environments or deep thermally/chemically stratified lakes, organic carbon deposition rate is high needing a large quantity of oxygen for its mineralisation. Transition towards anoxia is, therefore, very abrupt and occurs in the few first millimetres, leading to a dominance of anaerobic species.[40]

1.2.2 Microbiota of waste water treatment plant

As described in the following chapter concerning Lake Geneva, the site we have chosen to study (Bay of Vidy) is under the influence of the release from the waste water treatment plant of Lausanne. To have a comparison point, it is, therefore, interesting to have a global idea of what kind of microorganisms one can encounter in waste water treatment plants. All around the world, different kinds of waste water treatment plants exist, involving different technologies, but also different types of microorganisms. The main differences is depending on the quantity of organic matter to be treated. For high load, anaerobic digestion is preferentially used whereas aerobic treatment is preferred for medium to low organic matter concentrations.[37] In the case of the waste water treatment plant of Lausanne, the technology used to remove organic matter is mainly composed by a conventional aerobic secondary treatment involving activated sludge process. This process relies on the formation of flocs (large aggregates of matter) by the growth of slime-forming bacteria, like *Zoogloea ramigera*. Other organisms, like protozoa, fungi, filamentous (or not) bacteria can then attach to these flocs and take part in the biomineralisation process. The question explored in this chapter is whether one can find a common core in the active bacterial population found in these plants. In fact this question has already been addressed by several researchers. It is a kind of "Holy-Grail quest" for specialists of waste water treatment plant. Between 2009 and 2011, a Danish team has regularly monitored bacterial population of 28 waste water treatment plants of a very wide range of sizes using quantitative FISH and specific probes to target different groups of bacteria as well as metabolic functions. The result is summarized in the Figure 1.7.
Of course, this repartition is a little bit biased because it is based on a non-exhaustive selection of FISH probes. In the words of the authors, a parallel metagenomic study seems to confirm these conclusions. Another study that was also conducted on full-scale waste water treatment plants as well as on pilot and laboratory scale equipment led almost to the same repartition of the phyla (Figure 1.8), except that Actinobacteria where not dominant members of the communities.[44] In both studies, Firmicutes are present, but seem to play a very minor role in the degradation of organic matter during the treatment of waste water.

1.2.3 Endospore-forming bacteria

Endospore-forming bacteria are found as a non-homogenous group within the phylum Firmicutes.[46] This phylum has been historically separated into three classes: Bacilli, Clostridia and Molicutes, but only the first two contain endospore-forming species. In the new version of the Bergey's Manual of Systematic Bacteriology they, however, removed Molicutes from Firmicutes, keeping only Bacilli and Clostridia classes and a new one called Erysipelotrichia. All these three classes contain exclusively bacteria with rigid cell walls constituted mainly by muramic acid. Some of them contain also teichoic acid. Most of

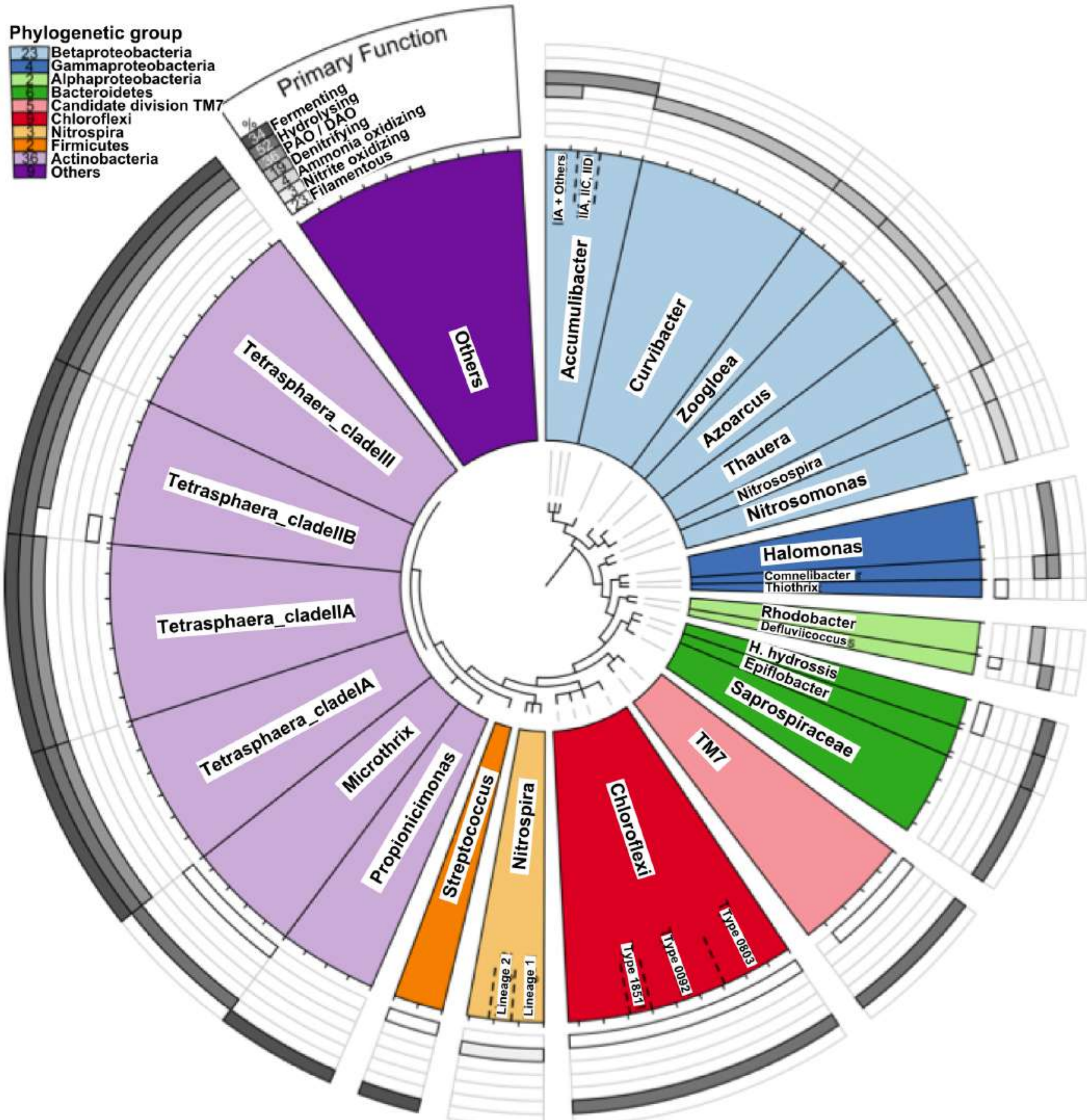


Figure 1.7: Core microbial community found in activated sludge from 28 Danish waste water treatment plants.[45] The abundance of each group was determine by quantitative FISH using specific probes

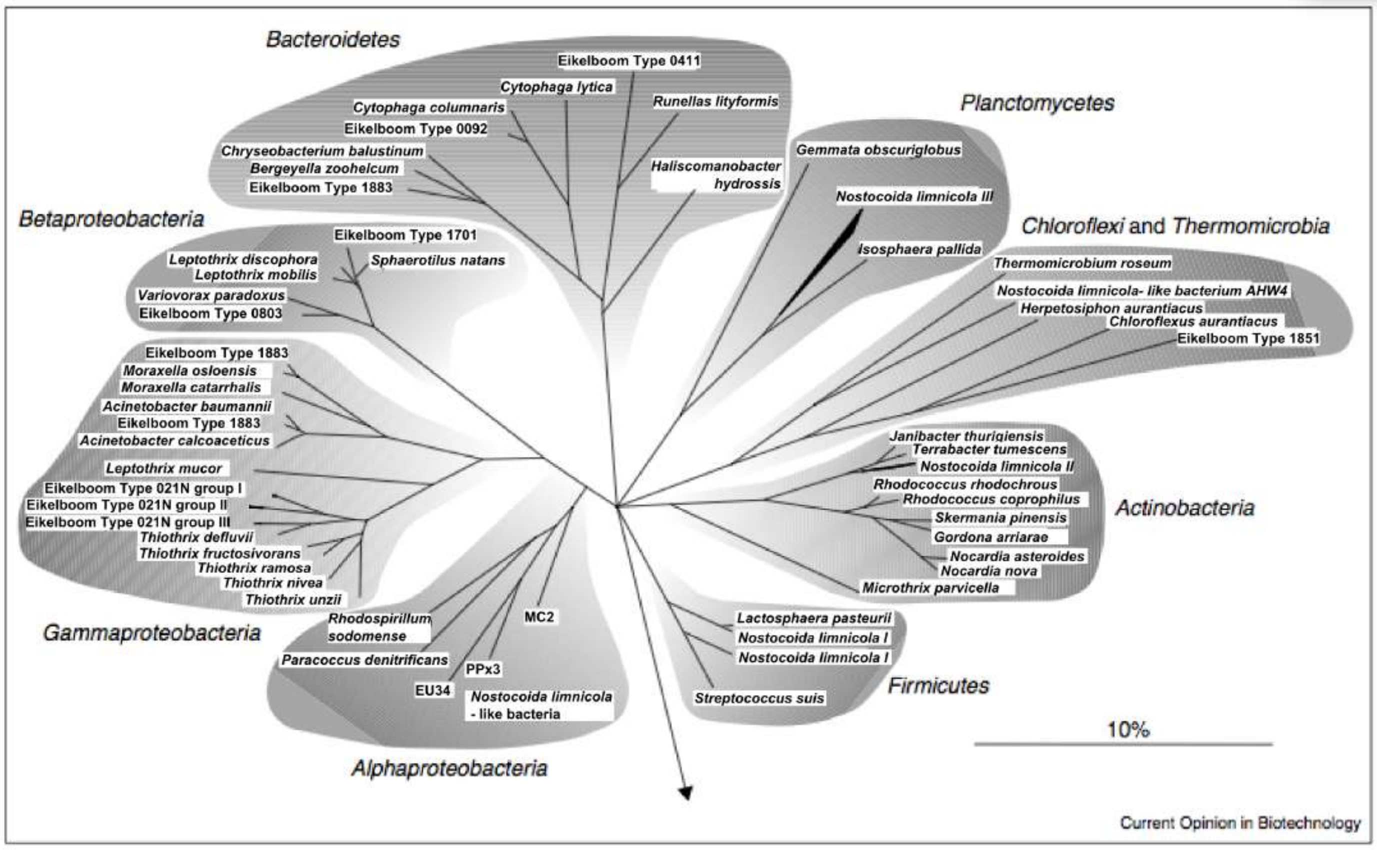


Figure 1.8: Core microbial community found in activated sludge from two full scale waste water treatment plants (industrial and domestic) as well as in different pilot and lab scale reactors. This study was conducted using a cloning-sequencing approach.[44]

Firmicutes stain Gram-positive, but few of them stain as Gram-negative (*Veillonellaceae*, *Syntrophomonadaceae*). Most of the Firmicutes are chemo-organotrophes and grow at neutral pH and ambient temperature. The phylum contain also some thermo-, halo-, acido- and alkaliphiles. All of them contain low DNA mol% G + C.[47] Classically, Bacilli are considered as aerobic, whereas Clostridia assemble anaerobic strains. Figure 1.9 gives the main characteristics permitting the identification of most of the endospore-forming Firmicutes.[37]

Since their discovery in the late 19th century, the knowledge gathered so far on the biology of endospore-forming Firmicutes has been mainly gained by laboratory studies on cultivable strains.[48]
The survival advantage conferred by endosporulation is one of the factors leading to the ubiquitous distribution of them in natural environments. Endospore-forming Firmicutes are common inhabitants of soils. As endospores can last for a relatively long period of time under unfavorable conditions they can survive long distance transportation by climatic or biological agents, leading to a better dispersion rate compared to non-endospore-forming taxa.[49,50] In addition, the large metabolic diversity found within this group is probably another major factor explaining their ubiquity.[51]
In laboratory experiments, endosporulation is triggered by starvation.[52] However, in natural environments cells frequently face starvation events. Other strategies like motility, chemotaxis, DNA uptake, or transformation, are used first to overcome environmental stress. In fact, endosporulation is considered as the ultimate mechanism triggered when all other survival strategies have failed.[53] Figure 1.10 shows the main steps of the life cycle of an endospore-forming Firmicute.

Characteristics	*Genus*
Rods	
Aerobic or facultative, catalase produced	*Bacillus* *Paenibacillus*
Microaerophilic, no catalase; homofermentative lactic acid producer	*Sporolactobacillus*
Anaerobic:	
Sulfate-reducing	*Desulfotomaculum*
Does not reduce sulfate, fermentative	*Clostridium*
Thermophilic, temperature optimum 65–70°C, fermentative	*Thermoanaerobacter*
Gram-negative; can grow as homoacetogen on $H_2 + CO_2$	*Sporomusa*
Halophile, isolated from the Dead Sea	*Sporohalobacter*
Produces up to five spores per cell; fixes N_2	*Anaerobacter*
Other rod-shaped endospore formers	
Acidophile, pH optimum 3	*Alicyclobacillus*
Alkaliphile, pH optimum 9	*Amphibacillus*
Phototrophic	*Heliobacterium*, *Heliophilum*, *Heliorestis*
Syntrophic, degrades fatty acids but only in coculture with a H_2-utilizing bacterium (Section 14.5)	*Syntrophospora*
Reductively dechlorinates chlorophenols (Section 14.12)	*Desulfitobacterium*
Cocci (usually arranged in tetrads or packets), aerobic	*Sporosarcina* (Figure 18.7)

Figure 1.9: Characteristics of the different groups of endospore-forming Firmicutes[37]

From an ecological point of view, one might suppose that endospore-forming Firmicutes should play an important role in the recovery of natural ecosystem after a perturbation event. As many natural environments are subjected to cyclic variations, in terms of nutrient flow, UV-irradiation, desiccation, freezing, boiling, and pressure, among others, the characteristics of endospore-forming Firmicutes seem to increase their ecological fitness in such situations.[54,55] However, even if these cause the admiration of microbial ecologist, up to now, the ecological signification of endospore formation in nature is not completely understood.

In addition to natural habitats, endospore-forming Firmicutes are often the cause of problems in anthropogenic environments, such as industrial production plants or hospitals. Food processing[56] or pharmaceutical synthesis,[57] are both industrial fields requiring perfectly cleaned production chains. If this condition seems relatively simple to achieve at a small laboratory scale, the same is not true at an industrial level, where efficiency and profitability need to be taken into account. As endospore-forming Firmicutes are able to resist many disinfection treatments (chemical, heat, and UV sterilization), it is difficult to remove them without a reduction on the quality of the end product. Furthermore, due to their adhesive capabilities that facilitate attachment to smooth surfaces (like stainless steel), the formation of biofilms is often encountered in production plants, leading to very challenging issues for the industry.[58,59] For the food industry, in addition to the well-known health concerns

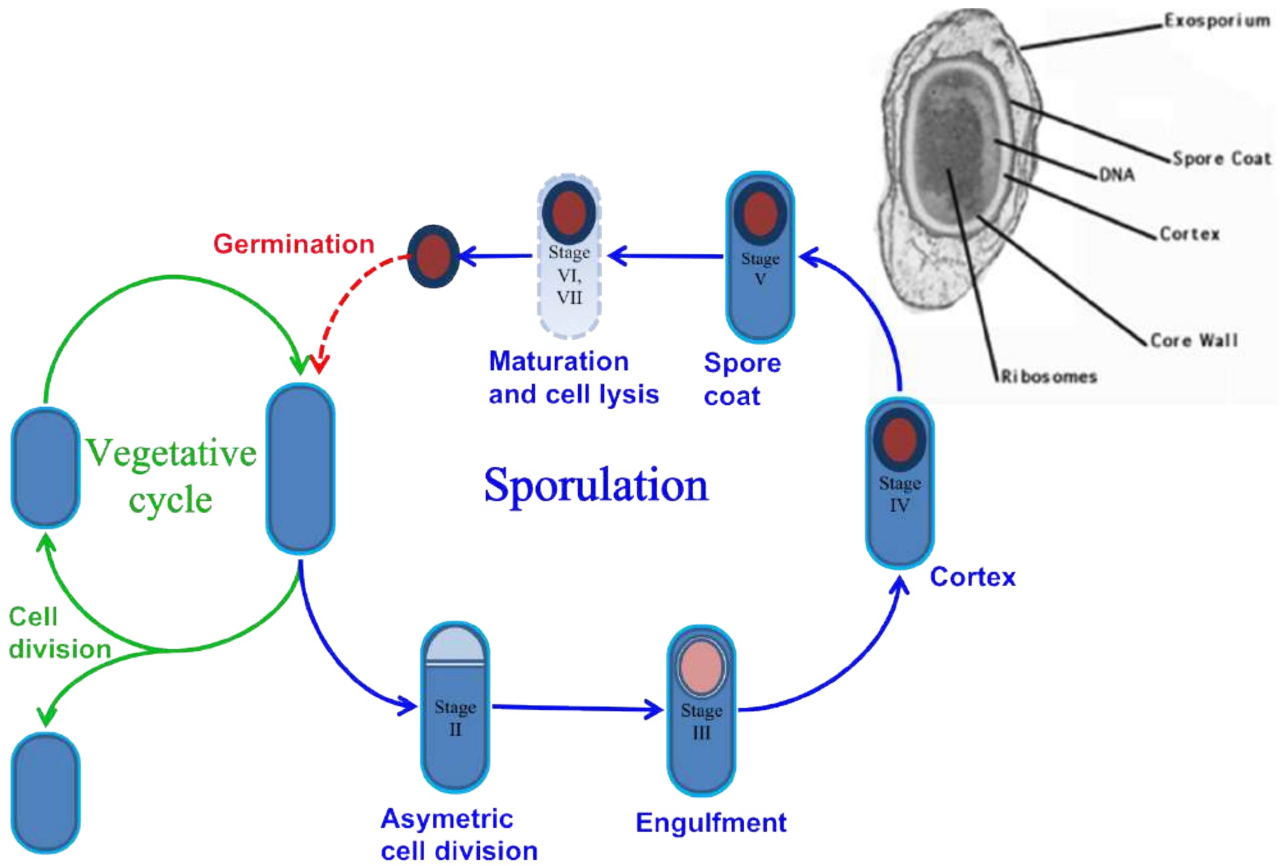


Figure 1.10: Vegetative cycle (left) and the different phases of an endospore formation (right) in endospore-forming bacteria. Figure modified from Erington et al (2003).[53] Above-right: Aspect and different layers of a bacterial endospore.

related to endospore-forming Firmicutes contamination by pathogenic species like *Bacillus cereus* or *Clostridium botulinum*, there is also the aspect of spoilage by non-pathogenic species, leading to a shorter shelf life of the product. All these issues represent significant economic losses, especially for the food industry.[60] Besides all these negative aspects, one must also mention the positive role of endospore-forming Firmicutes in many domains related to human life, such as their role as plant-growth promoting rhizobacteria (PGPR) or as pathogen biocontrol agents in agriculture,[61] or their use for the production of fermented food and probiotics by the healthcare industry.[62]
Taking into account their particular skills to resist to a lot of detrimental treatments, one can ask if endospore-forming Firmicutes harbour some particular mechanism leading to a better resistance (or simply tolerance) to high content of heavy metals in natural environments (particularly in sediments, for this study). If yes, what are their impact on the sequestration and release of these metals in aquatic environment? This constitute the essential driving question of this research project.

1.2.4 Mechanism of tolerance and resistance to heavy metal

Metals enter bacterial cells mainly by unspecific fast transporters, such as CorA (Magesnium uptake system) for Ni^{2+}, Co^{2+}, Zn^{2+}, and Mn^{2+}; Pit (Phosphate inorganic transport) for arsenate; or fast sulfate-uptake system for chromate. As these unspecific transporters are expressed constitutively in cells, their role in the regulation of internal concentrations of heavy metal is minimal and can only be efficient through mutation.[63] Once inside the cells, heavy metals can easily link to SH groups in proteins, strongly perturbing enzymatic activities. Some heavy metal could also compete with other essential metallic ions needed to perform metabolic activities.[24] For instance, Cd^{2+} could replace Zn^{2+} in DNA-binding proteins (like zinc-finger proteins), or Ca^{2+}, Ni^{2+} and Co^{2+} could replace Fe^{2+} in heme proteins or in different siderophores. Finally, elevated concentrations of non-essential heavy metals generate oxydative stress inside the cells.[24] The solution to regulate an excess of undesirable intracellular heavy metals is to use "metal-ion homeostasis factors" and "metal-resistance determinants".[63] As heavy metals cannot be degraded, three main solutions are available to cope with these intracellular excesses.

1. Active efflux of specific ion (via specific inducible transporters)

2. Segregation / complexation (immobilisation) of undesirable metallic cations into complex compounds containing thiol-groups. This solution, however, requires large quantities of energy (ATP) to synthesize sequestration proteins.

3. Diminution of the toxicity by reducing heavy metals into a less toxic form (reduction of Hg^{2+} to Hg^0, for instance).

Active efflux mechanisms can be mediated by P-type ATPases (for Mg^{2+}, Mn^{2+}, Ca^{2+}, K^+, Cu^{2+}, Zn^{2+}, Cd^{2+}, Pb^{2+}, and Ag^+), A-Type ATPases (arsenite), proton antiporters (Co^{2+}, Zn^{2+}, Cd^{2+}, Ni^{2+},Cu^{2+}, and Ag^+), or by specific slow chemiosmotic transporters (Zn^{2+}, Cd^{2+}, Co^{2+}, and Fe^{2+}). Active influx could also be mediated by ABC-binding cassettes (Mn^{2+}, Zn^{2+}, Ni^{2+}, and Fe^{2+}). This last transport system is, however, not active for the efflux of these metallic ions.[63]

Extracellular complexation mechanisms could also play an important role in the bio-availability of heavy metal ions. Complexation by exopolymeric substances,[64] or S-layers[65,66] is effectively reported by many authors. This extracellular immobilisation system is, however, not specific, and lead also to a diminution of the bio-availability of essential metals and nutrients.

Finally, the last mechanism to escape elevated concentrations of heavy metals could be endosporulation. This is the purpose of this work!

1.3 Lake Geneva

Lake Geneva is the largest freshwater lake in Western Europe with a volume of 89 km^3, a maximum depth of around 300 m, a length of 73 km and a maximum width of 14 km. It is located in a depression in the Molasse basin in the western Swiss Plateau. This depression has been formed 1 million years ago in reaction to the erection of the massif of Mont Blanc.[67] Lake Geneva could be divided into three distinct basins. To the east, towards the mouth of the Rhone, one can found the « Haut-Lac ». Further east, in front of Lausanne, stands the biggest and deepest part, the « Grand-Lac », with a maximum depth of 309 m. Finally, in the western part, the « Petit-Lac » begins with the narrowing in front of the cities of Nyon and Yvoire. The volume of this basin is much smaller and the maximum depth does not exceed 79 m. At the western end of Lake Geneva, the Rhone exits the lake and continues its travel towards the Mediterranean Sea. The altitude of the lake is 372 m above sea level and is artificially regulated since the late 19th century. This level regulation was necessary to avoid floods when snow and ice melt in the surrounding mountains. The two main tributaries of the lake, the rivers Rhone and Dranse, provide 80% of the water that flow into the lake. They are characterized by a water regime that corresponds to an alpine glacio-nival and alpine nival type. The sediments that are carried by them have an Alpine to pre-Alpine signature.[68]

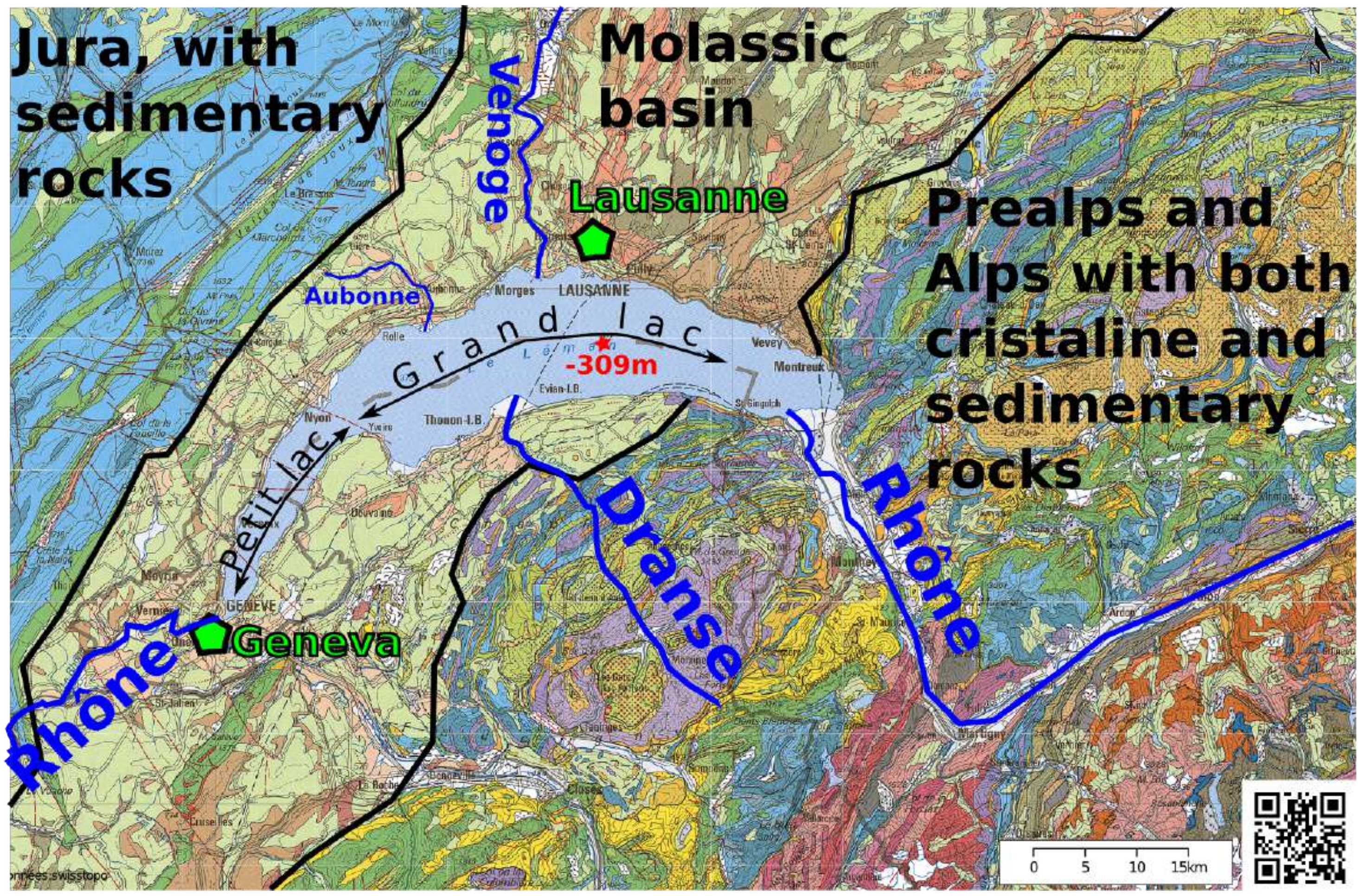


Figure 1.11: Lake Geneva and surrounding geology

Because of their characteristics, the two main basins of Lake Geneva, "Grand-Lac" and

"Petit-Lac" can be considered as two separate lakes having different hydrological functioning. For instance, the summer thermal stratification of the "Petit-Lac" disappears every winter allowing water turnover and oxygenation of deep layers of the water column (monomictic). This is, however, not always the case for the "Grand-Lac" that becomes meromictic during hot winters when surface water does not cool enough to descend down to the floor of the lake. During these particular years, oxygenation of the deepest water does not occur.[69]

For decades, the "health" of Lake Geneva has been followed by the CIPEL (Commission internationale pour la protection des eaux du Léman); A Franco-Swiss Commission that monitors different aspects of Lake Geneva ecosystem and its main tributaries (Rivers Rhone, Dranse, Aubonne, and Venoge) The commission publishes annual scientific reports that are the basis for decision making concerning the maintenance and the amelioration of the health status of the lake. The industrialization of the "50s", a more intensive agriculture, and an increase of the population around the lake, have largely contributed to an increase of the pollution during the "Glorious Thirty", leading to an awareness of public and authorities concerning environmental issues. As a reaction, the first waste water treatment plants have been constructed and operated at the end of the "60s". Since these years, the regulations have become much stricter concerning the agricultural inputs and the release of industrial effluents in the environment. In the "80s", the ban of phosphates in detergents has played a major role in restauration of aquatic ecosystems of Swiss lakes which were prone to eutrophication. Actually, if the situation may appear as satisfactory, we, however, need to take into account the densification of the population on the shores of Lake Geneva, which can lead to pollution problems linked to higher amounts of wastewater and runoff to be treated and released into the lake.This is particularly true for larger cities like Lausanne, which release treated waste water directly into the lake, in the Bay of Vidy.[70]

1.3.1 Bay of Vidy

Bay of Vidy is the portion of the lake bordering the city of Lausanne. Natural sediments are released into the bay by the "Chamberonne" river which collects water from the "Mèbre" and the "Sorge" rivers both having drainage basins through agricultural zones to the north of Lausanne. Sediments in the bay should, therefore, mainly be composed by erosion material from these croplands, such as sand from Marine Molasse (sedimentary rocks), organic matter form soils, chemical amending substances from agriculture, as well as pesticides. Apart from these constituents, the bay suffers for many years from significant pollution problems due to the release of effluents from the waste water treatment plant and the runoff of the city of Lausanne. Despite treatment, water rejected into the bay contains a large amount of organic matter and bacteria that accumulate around the outlet of the rejection pipe. Until the late "90s", a first pipe was in operation, releasing treated water into the bay relatively close (around 300 m) to the shore and at a shallower depth. This was, however, problematic because effluents were rejected above the thermocline and dispersed in the upper layers of the water column up to the surface of the lake. At this

time it was, therefore, strongly recommended to avoid swimming in this polluted water (in fact a kind of "bacterial soup"). The problem was finally solved in 2000 by the extension of the pipe further from the shore (around 700 m) and deeper into the lake (around 30 m) for releasing water below the thermocline. With these modifications a better removal of waste by deep currents was planned. If surface water of the bay is again safe for swimming, what are the consequences for sediments of deeper zones of the lake, where treated water is rejected?
As seen in the previous chapter on pollution, heavy metals and other micropollutants accumulates easily in sediments. Several recent studies, mainly conducted at the Forel-Institute, showed that a strong accumulation of heavy metals (around 9 mg/kg of total mercury) can be measured in the sediments close to the outlet pipe.[71–73]

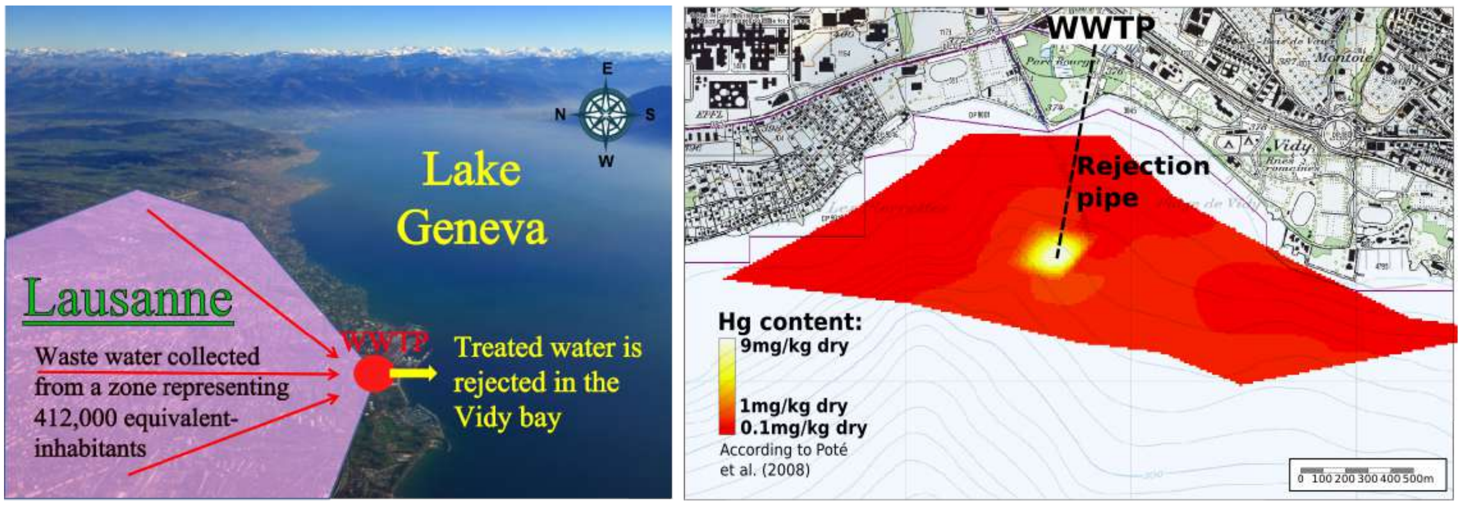


Figure 1.12: Bay of Vidy

The Elemo project brought together people from different universities and particularly some from the Forel-Institut who took the opportunity offered by MIR submersibles to visualize the sediments of the Bay of Vidy and continue their research projects on the pollution induced by the waste water treatment plant effluents. During preliminary discussions on the global Elemo project it quickly became evident to us that we must take advantage of the knowledge already available on the Bay of Vidy, and that a collaboration could occur with the teams of Jean-Luc Loizeau (Forel-Institut) and Marylou Tercier-Waeber (UniGE), both having a strong background on this polluted site.
In summary, Bay of Vidy and the collaborations that could be considered, constitute the ideal conditions for our project on endospore-forming bacteria as biological indicators of pollution in the environment.

1.3.2 Waste water treatment plant of Lausanne

This chapter is mainly a summary of the informations found in the report concerning micropollutant removal pilot experiments conducted recently in the waste water treatment plant of Lausanne.[74]

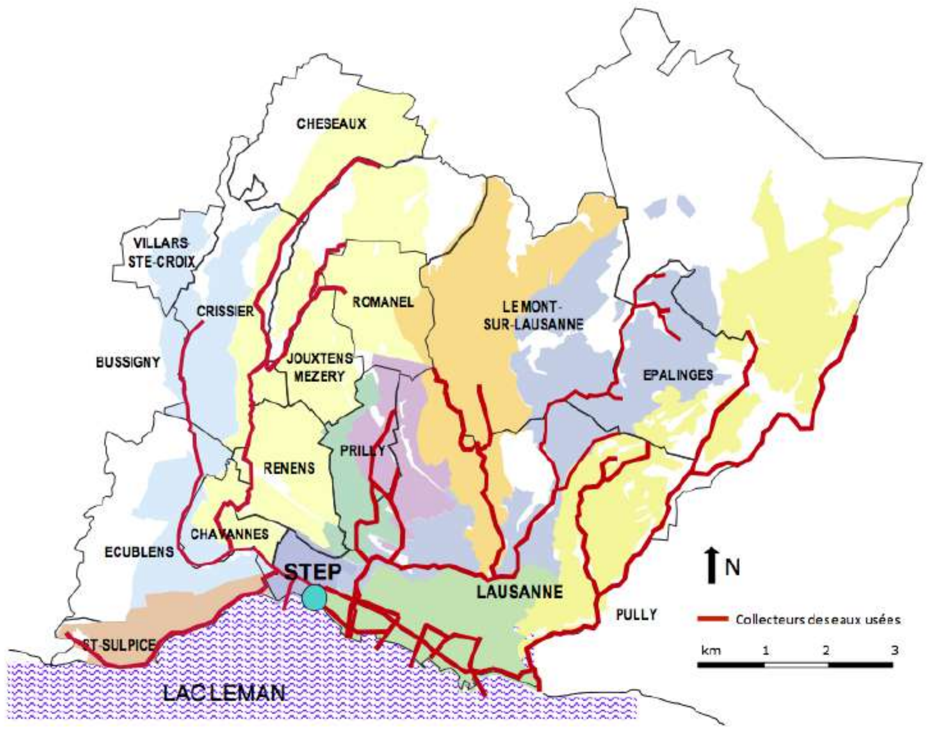


Figure 1.13: Origin of waste water treated in the waste water treatment plant of Lausanne.[74]

The waste water treatment plant of Lausanne operates since 1964 and was built next to the lake in Vidy. It has been designed to treat waste water ($45 * 10^6 m^3/yr$) from a catchment of 74 km^2 including 15 municipalities (Figure 1.13).

Initially it was constituted of a degritting/ screening device, followed by a primary sedimentation basin, an aerobic biological treatment and a final sedimentation, all devices allowing to treat up to 1.2 m^3/s. In 1971 a phosphorus removal device by flocculation (addition of $FeCl_3$) was added to fight against eutrophisation of the lake. Finally, a physico-chemical treatment plant was added in 1976 in parallel to the existing to increase the capacity of an additional 1 m^3/s (Figure 1.14). With up to 2.5 m^3/sec (if we take also into account runoff) released into the lake, a sedimentation rate of approximately 3 $cm/year$ has been estimated for the area close to the rejection pipe.[75]

The 2000s were mainly marked by the addition of deodorization units as well as an automatic system for harvesting and dewatering sludge. Approximately 110 tons of sludge is harvested daily, dehydrated and incinerated in the Vidy plant providing energy to the district heating.

In conclusion, this plant is the biggest of Canton Vaud and also one of the oldest with very basic technology removing only organic carbon and phosphorus from waste water. No nitrification module is installed for oxidizing ammonium into nitrate, and also no anaerobic basin allows for denitrification. Furthermore, the removal efficiency is one of the lowest of Vaud, perhaps due to the design of the aeration basin, which does not allow a sufficiently high residence time (around 2 hours), or because of the high amount of clear water that enter the treatment plant also affecting the residence time. The separation of water is effectively not yet entirely carried out in the city of Lausanne. All of this is clearly shown by the results published in 2011 (Figure 1.15).[76] Finally, there is approximately 12% of the water that enters the plant that is rejected into the lake without any treatment.

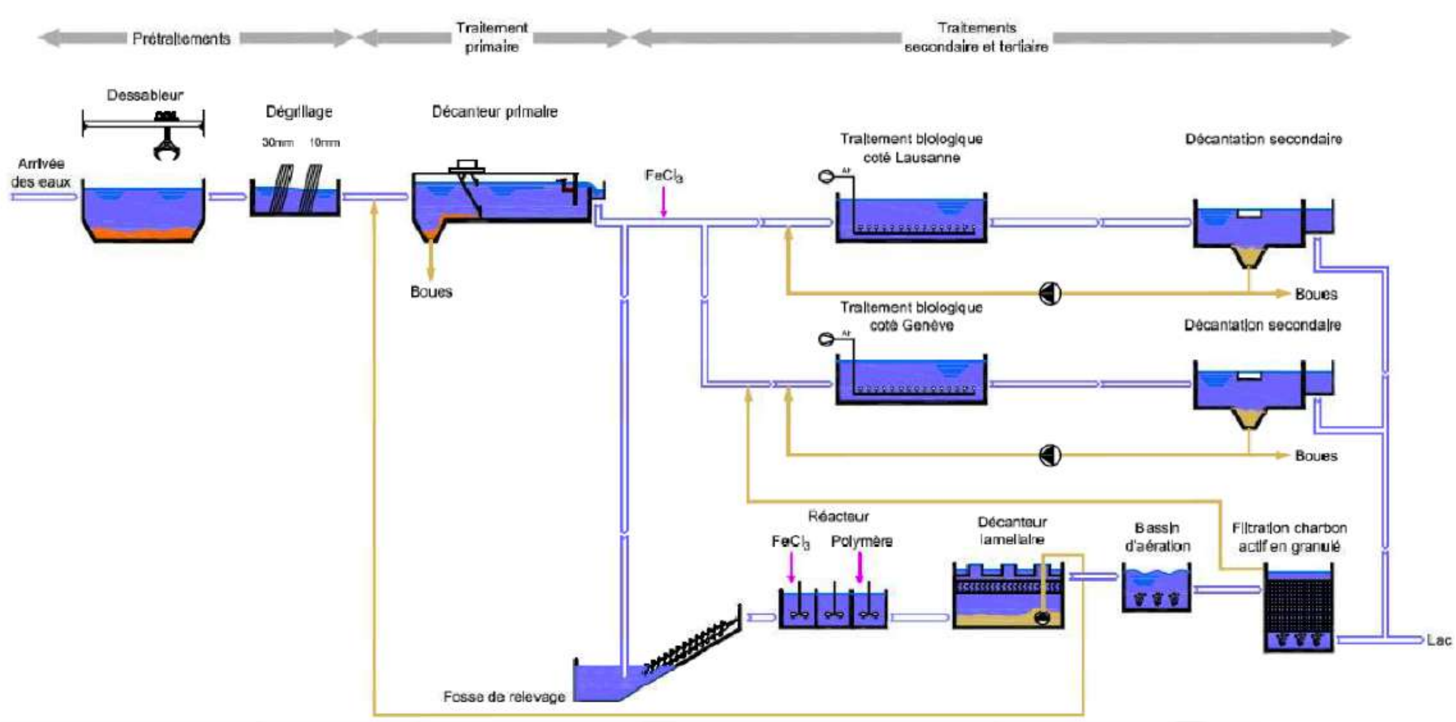


Figure 1.14: Diagram of the installation of the waste water treatment plant of Lausanne.[74]

General informations

Flow (m^3/Day) Mean In	Flow (m^3/Day) Mean Out	Flow (m^3/Day) Overflowed	Specific flow l/EH*day Mean	Specific flow l/EH*day Dry weath.	Receiving environ.	Energy used MWh/yr	DBO5 Treatment Yield %	Corg Treatment Yield %	Portho mg/l	Ptot Treatment Yield	Ammonium mg/l In	Ammonium mg/l Out	Nitrate mg/l Out
94409	83182	11227	391	320	L. Geneva	12869	74.1	89.4	0.17	78.9	19.78	17.88	2.87

Sludge composition

Dry mater %	Morg %	Ntot (% of Dry matter)	N-NH3 (% of Dry matter)	Ndisp (% of Dry matter)	P2O5 (% of Dry matter)	Hg (ppm)	Cd (ppm)	Ni (ppm)	Cr (ppm)	Cu (ppm)	Pb (ppm)	Zn (ppm)	AOX (ppm)
34.1	77.5	5	1.8	2.5	4.6	3.8	0.8	11	53	259	24	520	0.69

Figure 1.15: Measurement results for the waste water treatment plant of Vidy. Year 2011.[76]

Pilot installations During the last years, the waste water treatment plant of Lausanne has participated to a Swiss pilot project for testing new technologies for the removal of micropollutants in waste water. A pilot scale plant (below 7% of the actual capacity of the plant) has been installed in parallel to the existing infrastructure, first for testing the efficiency of moving bed bioreactor to transform ammonium in nitrate (Figure 1.16).

The second part of the pilot scale installation is intended to test the removal efficiency of two systems for treating organic micropollutants (like pharmaceuticals, pesticides, hormonally active substances). The first one, ozonation with sand filtration is presented in Figure 1.17. The second one, powdered activated carbon with ultrafiltration is presented in Figure 1.18. On a selection of more than 50 organic micropollutants the moving bed technology improve the removal of micropollutants by a factor 2 (removal of 40%) when compared to a conventional plant. The ozonation treatment and the powdered activated carbon increase also the efficiency by a factor 2, reaching an overall elimination of 80% of organic micropollutants. A verification by 25 ecotoxicological assay has also showed a clear reduction of the toxicity of the water released after the treatment.
During this experimentation, no evaluation has, however, been made concerning heavy metals. In principle, heavy metals should be removed by complexation on flocs during conventional treatment. After sedimentation, heavy metals are, therefore, eliminated with

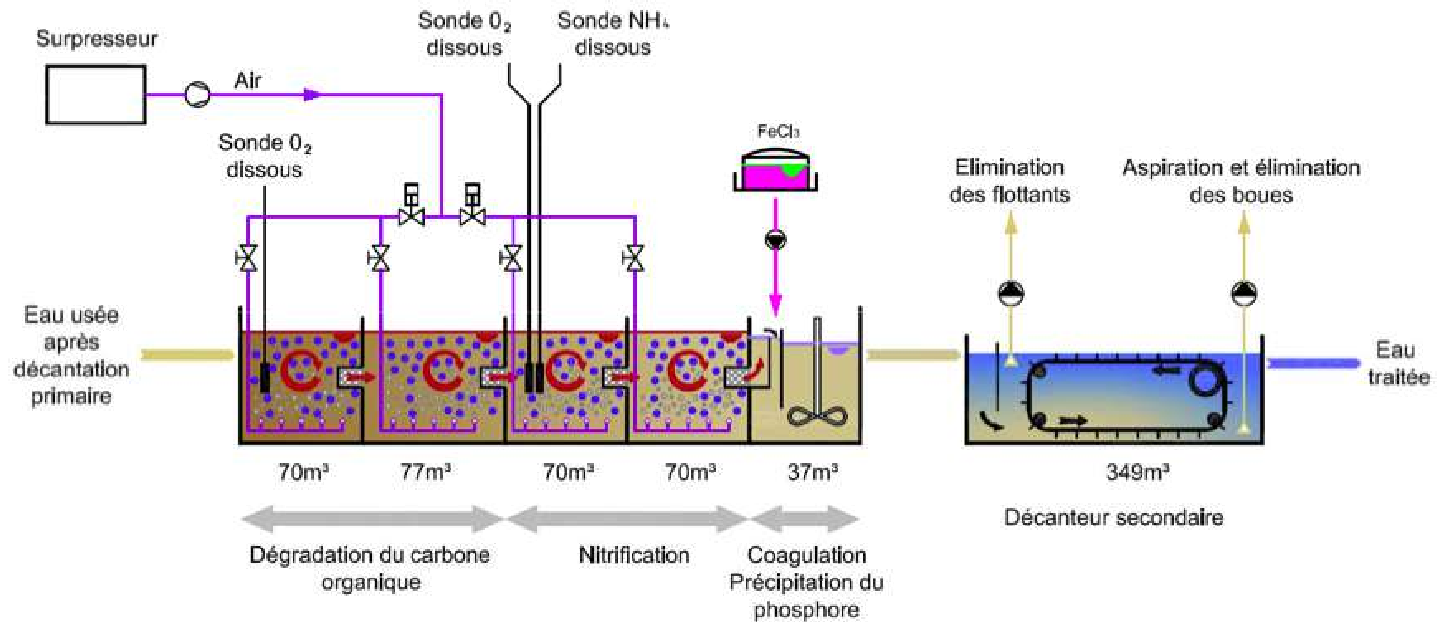


Figure 1.16: Pilot plant for the removing of ammonium. In operation for more than one year, treating 7% of the total flow of the waste water treatment plant of Vidy.[74]

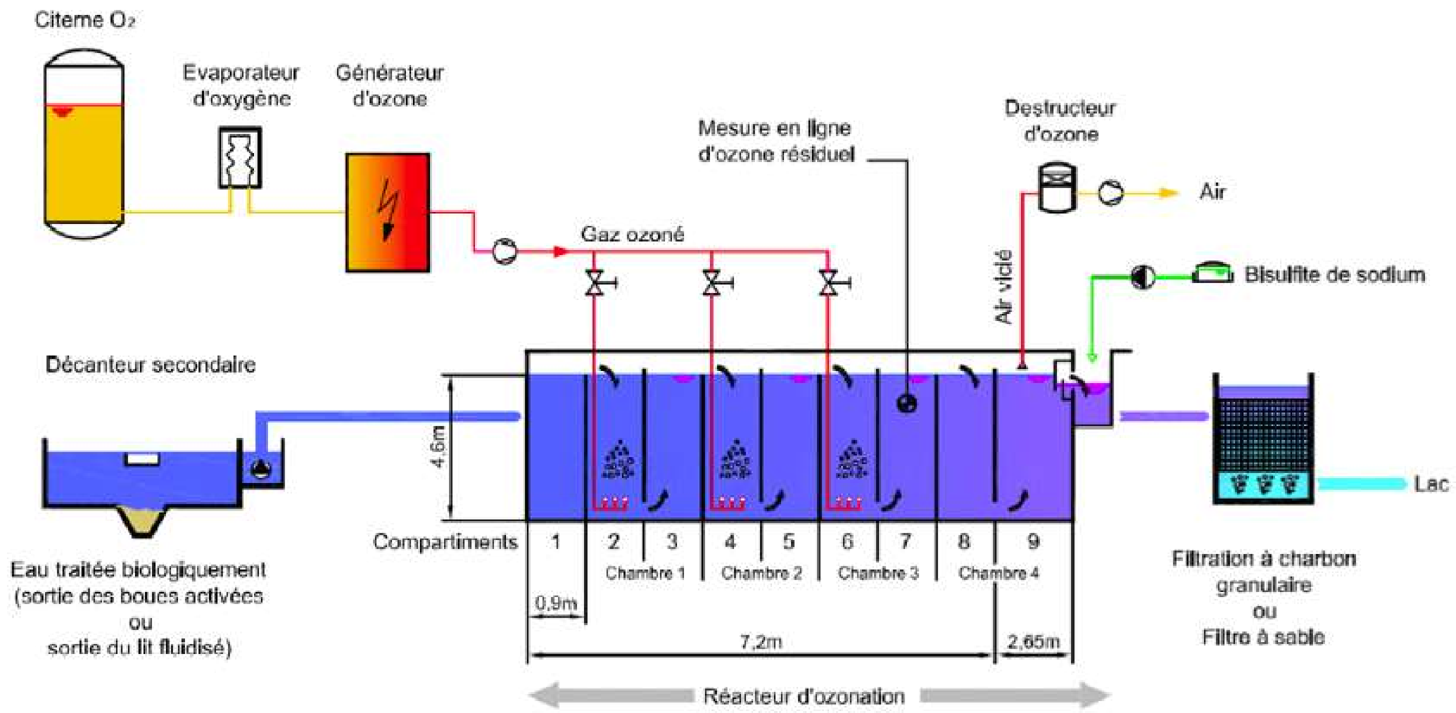


Figure 1.17: Pilot plant for the removing of micropollutants by ozonation.[74]

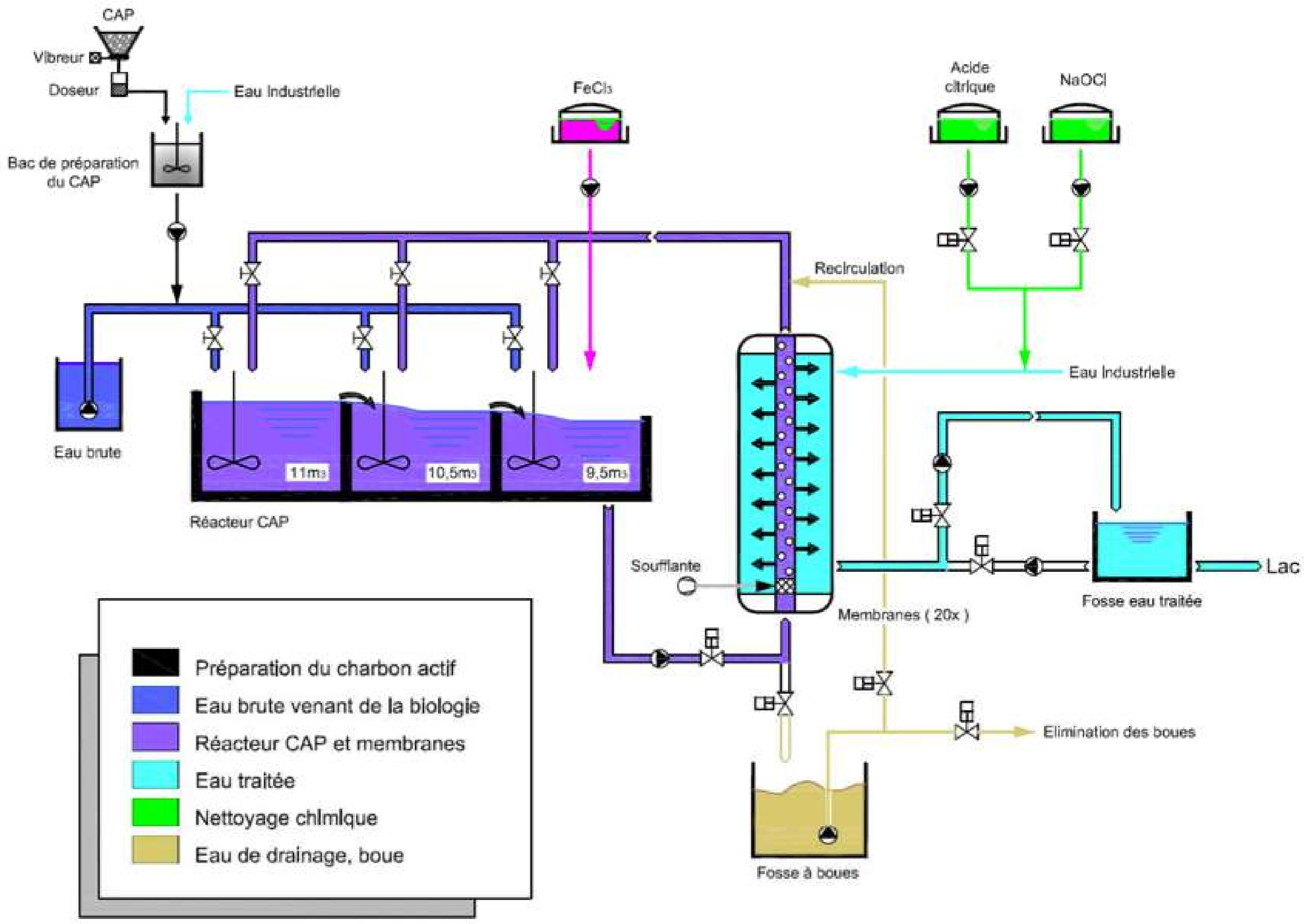


Figure 1.18: Pilot plant for the removing of micropollutants by powdered activated carbon and ultrafiltration.[74]

the sludge. The results shown in Figure 1.15 effectively confirm that a lot of heavy metals are found in sludge. In 2011, a value of 3.8 mg/kg of mercury was measured in sludge from the waste water treatment plant of Lausanne. This level is in the same range as the concentrations that were measured in the sediments collected in 2005 by Poté et al[73] around the outlet of the rejection pipe in the Bay of Vidy (Figure 1.12). A regular overflow of the treatment plant could perhaps explain why this amount of heavy metals are found in the sediments from the Bay of Vidy. This hypothetical overflow is also highlighted by the results published in 2011 by the authorities of Canton Vaud (Figure 1.15), showing that approximately one tenth of waste water is escaping from a complete treatment in the plant. To avoid the release of heavy metals in the bay, the complete separation of waste water and clear water should be achieved and the waste water treatment plant of Vidy should be modernized. Both projects are under realisation by the city of Lausanne. Modernisation is effectively planned in a close future and is reported in the document concerning the review of waste water treatment in Vaud for 2011.[76] Ammonium and organic micropollutants removal units should be added to the plant in the future.

1.3.3 Runoff water of Lausanne

The separation of runoff and waste water is not yet completely achieved in Lausanne and an amount of clear water is still going through the waste water treatment plant of Vidy. This input of clear water in the waste water treatment plant is not desirable for two reasons. First, the induced dilution of waste water reduces the efficiency of the treatment. Second, it is an energetic, economic, and ecological nonsense to sully this surface runoff water with faecal contaminants to treat it thereafter in installations which are not designed for this kind of water. In few years the city of Lausanne has invest a lot to separate clear water from waste water. Storm overflow of "Le Capelard", operated in 2012 is a good example of this. Now a big proportion of clear water is discharged directly into the lake.
But what about heavy metals contained in this runoff water? Of course, they are probably not those we found associated with the sediments sampled around the rejection pipe, but they still induce pollution of the aquatic ecosystem of the lake! What is, therefore, their long term impact on the health of the ecosystem and our health? Nobody knows! A precautionary approach should dictate to us to avoid this kind of rejection into the environment and to found a way to eliminate heavy metal pollution from clear water before their rejection in nature. This research project should gives some clues for solving this problem.

2

Concepts and Methods

This chapter introduces some key aspects of the theoretical, experimental, and computational concepts used throughout the thesis

Contents

2.1 Sampling

The sampling was done in the Bay of Vidy, in front of the city of Lausanne. This area was selected because of the well known contamination of the sediments by heavy metals released together with treated waste water from the municipal treatment plant of Lausanne. A first sampling campaign took place during summer 2011, right after the beginning of my thesis, taking the opportunity to collect sediment samples with the two Russian research MIR submersibles (Figure 2.1). In the context of my project, the main advantage given by the submersible rely on the visualisation of the sediment's surface giving some hints to select more precisely a particular sampling location. The criteria used will be described later. Eight cores were retrieved during this first campaign (red dots in the Figure 2.2).

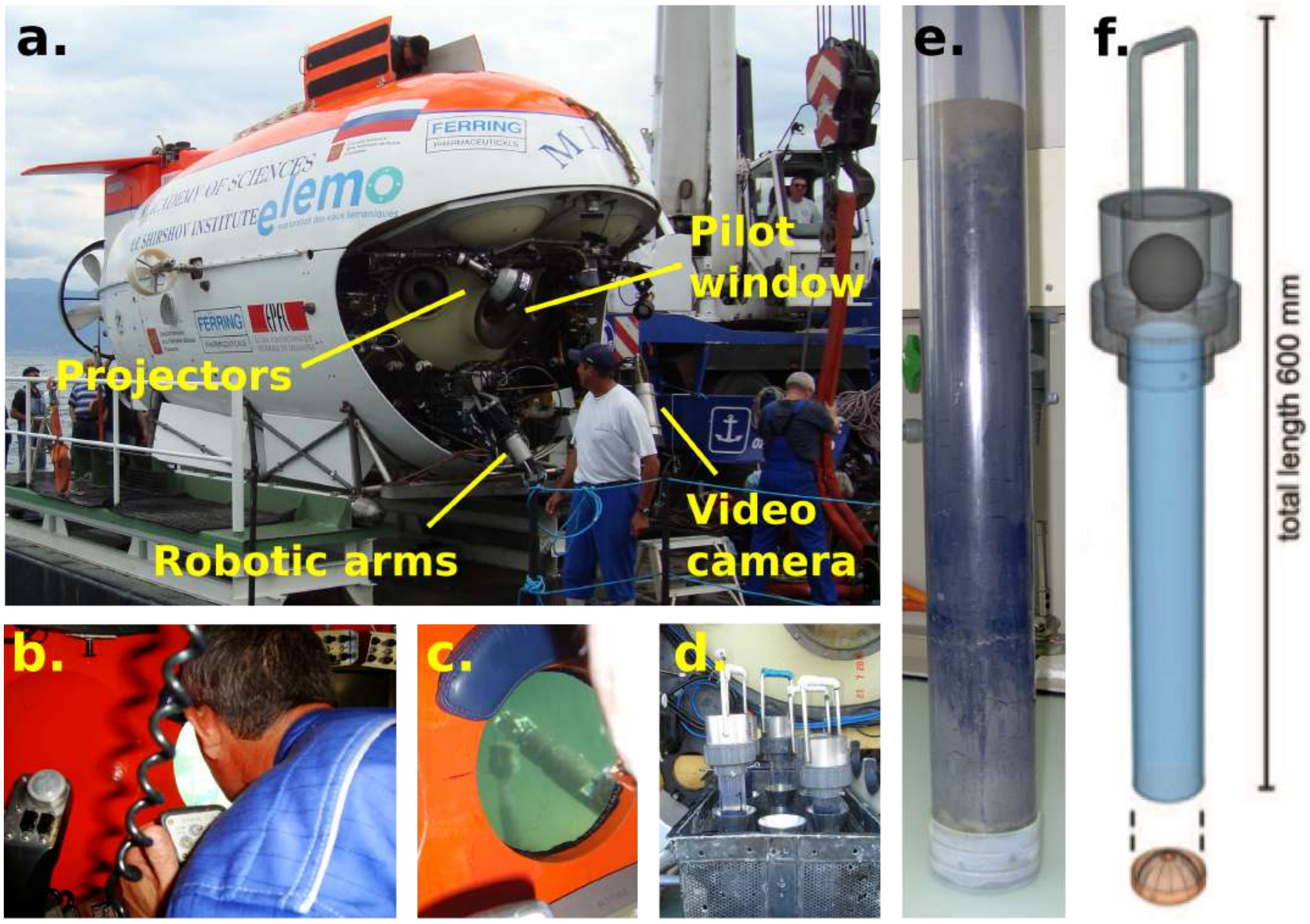
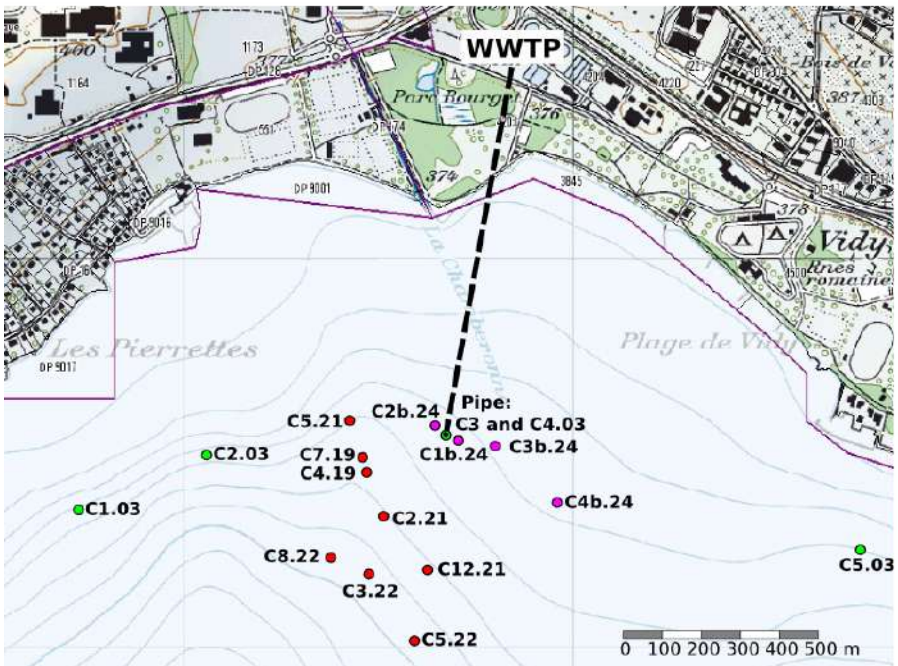


Figure 2.1: MIR submersible with equipment for collecting sediment cores and doing *in situ* physico-chemical measurements (a.). Pilot manipulating robotic arm(b.). View of an arm with a sediment corer from inside the sub (c.) Basket in front of the sub with sediment corer inside (d.) Sediment core (e.) Layout of the sediment corer used (f.).

As the Elemo project brought together a large number of researchers with different projects focusing on the polluted sediments of the Bay of Vidy, the exact sampling locations were the result of a compromise between all the groups interested by sediment samples from this area. To complete the sediment core collection with samples less impacted by the effluent from the waste water treatment plant (environmental blanks), as well as samples from the direct area of the releasing pipe, a second sampling campaign was conducted from the research boat "La Licorne" of the Forel Institute, University of Geneva (Figure 2.3). Five cores were retrieved in the beginning of October 2011 with an Uwitec (Mondstein, AU) gravity corer (green dots in Figure 2.2). Finally, 4 additional cores were retrieved at the end of May 2012 with the same methodology in order to collect more samples from areas covered by white filamentous bacteria (see description below).



Campaign	Core Name	Depth (m)	CH East coord.	North coord.
19.07.2011 MIR1	C4.19	41	534461	151484
	C7.19	42	534471	151446
21.07.2011 MIR1	C5.21	46	534435	151602
	C2.21	47	534515	151335
	C12.21	55	534627	151197
22.07.2011 MIR1	C8.22	60	534379	151229
	C3.22	60	534476	151186
	C5.22	70	534594	151014
03.10.2011 Boat of the Forel-Institute	C1.03	36	533735	151351
	C2.03	35	534063	151491
	C3.03	32	534676	151543
	C4.03	32	534676	151543
	C5.03	31	535739	151248
24.05.2012 Boat of the Forel-Institute	C1b.24	32	534709	151526
	C2b.24	31.5	534648	151567
	C3b.24	31	534803	151513
	C4b.24	31	534961	151369

Figure 2.2: Sampling locations in the Bay of Vidy. Points in red correspond to cores collected during MIR campaign, green points correspond to the first campaign with "La Licorne", purple points to the second sampling with the boat.

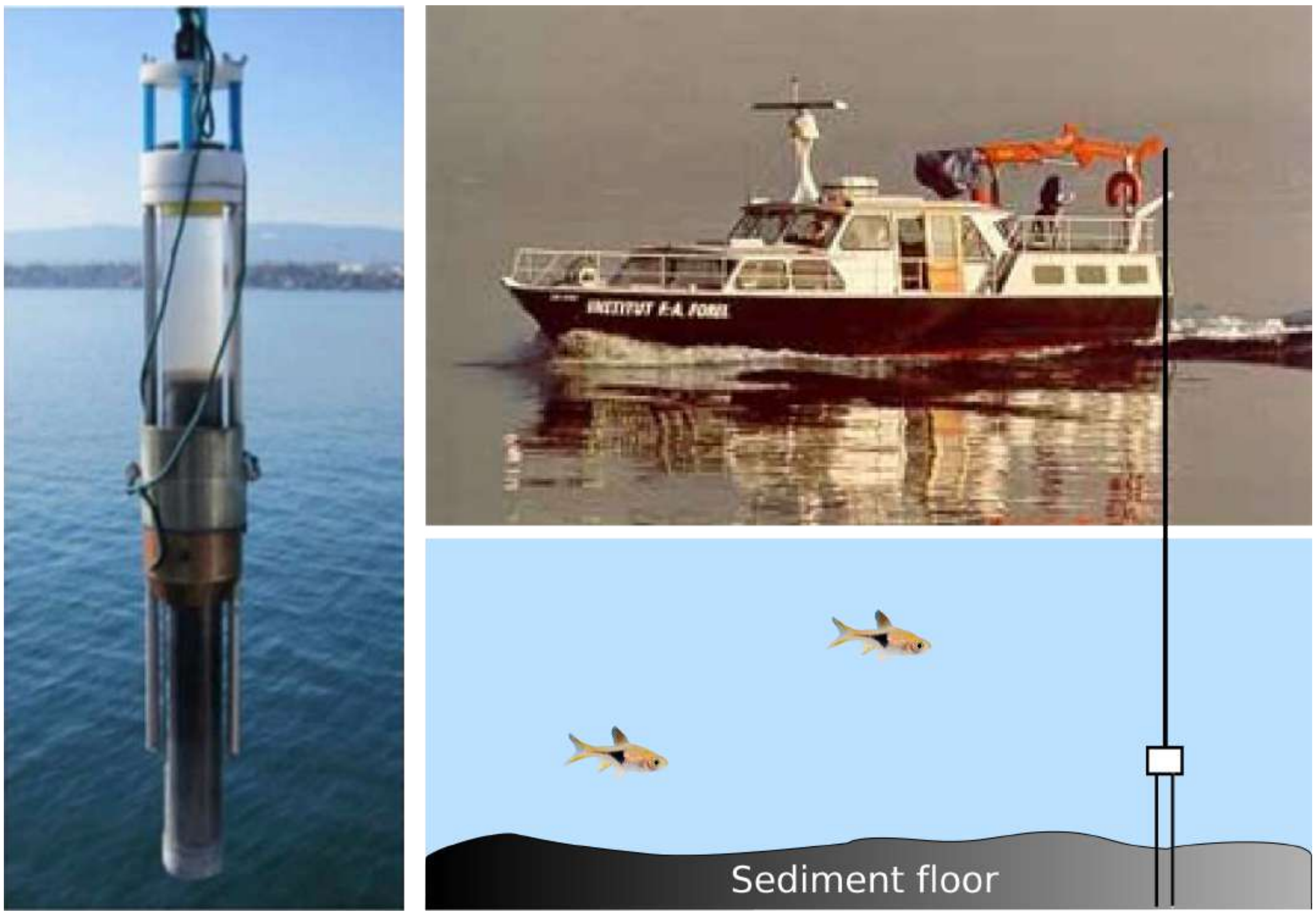


Figure 2.3: Sediment cores collected from "La Licorne", research boat from the Forel Institute (UniGE). Cores were retrieved with an Uwitec (Mondstein, AU) gravity corer (image to the left).

During MIR sampling campaign (summer 2011), different aspects of sediment surface were reported depending of the sampling zone (Figure 2.4). Close to the rejection pipe sediments were covered by a white layer of filamentous bacteria that look like *Beggiatoa* sp. When coring in this area, a great amount of gas bubbles were also emitted by sediments, reporting an intense mineralisation of a high load of organic matter in anaerobic conditions. Gas produced could be a proxy for methanogenesis (CH_4), as well as sulphate reduction (H_2S), fermentation (H_2) or nitrate reduction (N_2). All these process could take place in anoxic sediments in different micro-niches depending on the local availability of the substrates as well as the local redox conditions at a microscale. However, H_2S is highly soluble, H_2 is consumed by different metabolisms (like methanogenesis), and the low concentrations of nitrate in this aquatic environment[?] does not allow the formation of large quantities of N_2. At the end, only methane (CH_4) can be degassed by sediments.

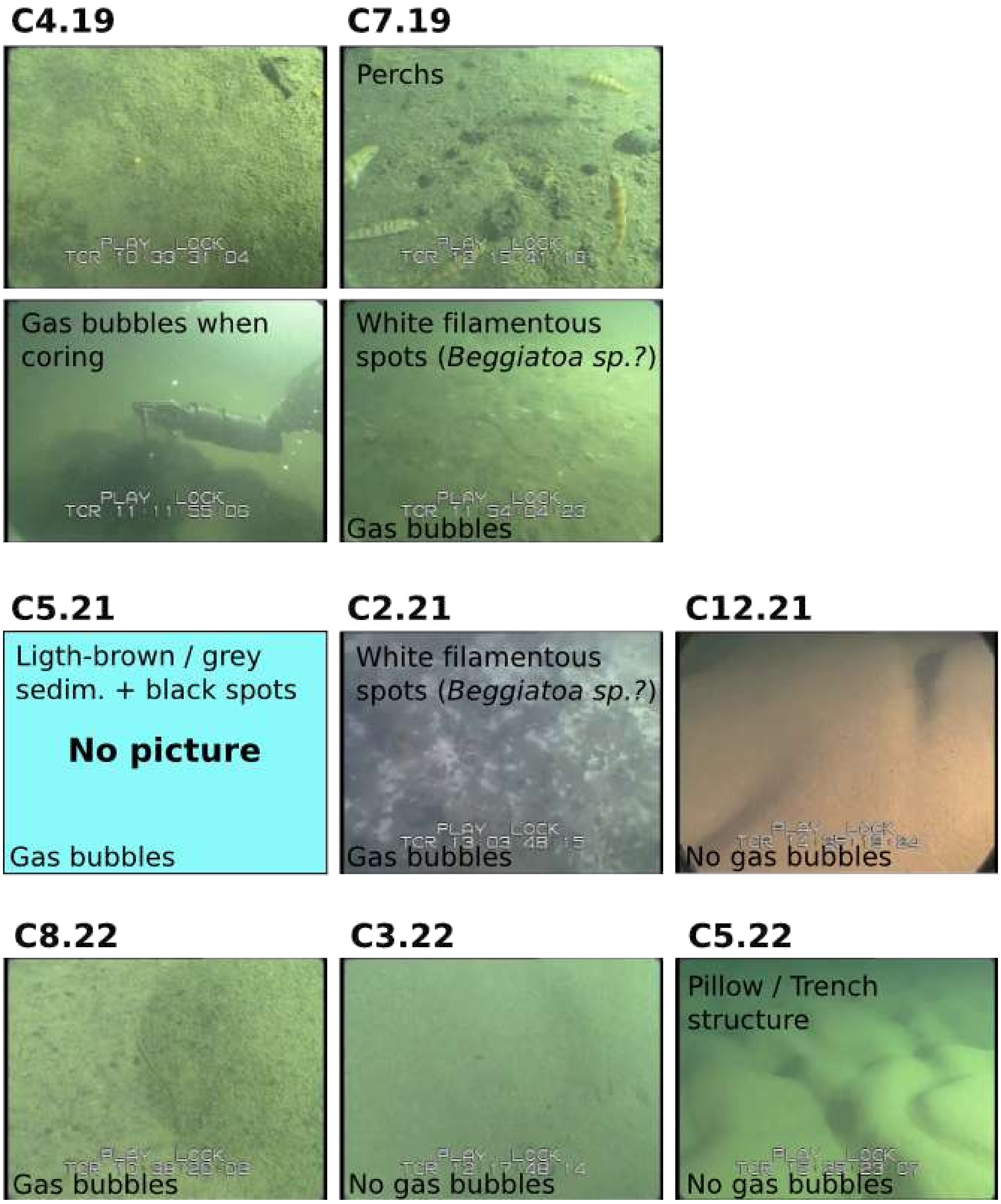


Figure 2.4: Aspect of the sediment surface encountered in the different sampling locations during the three days of MIR sampling campaign.

No pictures of the aspect of the sediment surface were available for the second sampling campaign from the boat. For the third sampling campaign, a GoPro camera was attached to the gravity corer allowing afterwards visualization of the coring process. This reveals that sediments collected during this last campaign (Cxb.24 series, purple points in Figure 2.2) were all covered by a patchy white mat of filamentous bacteria, except the one retrieved further away from the pipe (C4b.24). When observing this white mat with phase-contrast microscope one could see filaments full of white elemental-sulfur granules. These inclusions are typically found in *Beggiatoa* sp. which produces filaments of the same shape and size than those observed in this study (Figure 2.5).

As demonstrated by Winogradsky in 1949, the amount of sulfur granules inside *Beggiatoa*'s filaments is a function of H_2S concentration in the surrounding water.[37] The *Beggiatoa* sp. coverage of the sediment around the rejection pipe, as well as the big amount of sulphur granules observed inside the cells provide evidence to infer an important sulphate reduction activity in the anoxic layers of the sediments. These bacteria are well known for their capacity to oxidize sulphide anaerobically by the use of nitrate. Some authors claims that *Beggiatoa* sp. is responsible for the suboxic zone in sediments (zone between a strongly reduced layer where sulphate reduction takes place and an oxic layer at the top of the sediments) and that this biological oxidation of sulphide compete with chemical oxidation by Fe(III) as electron acceptor.[77]

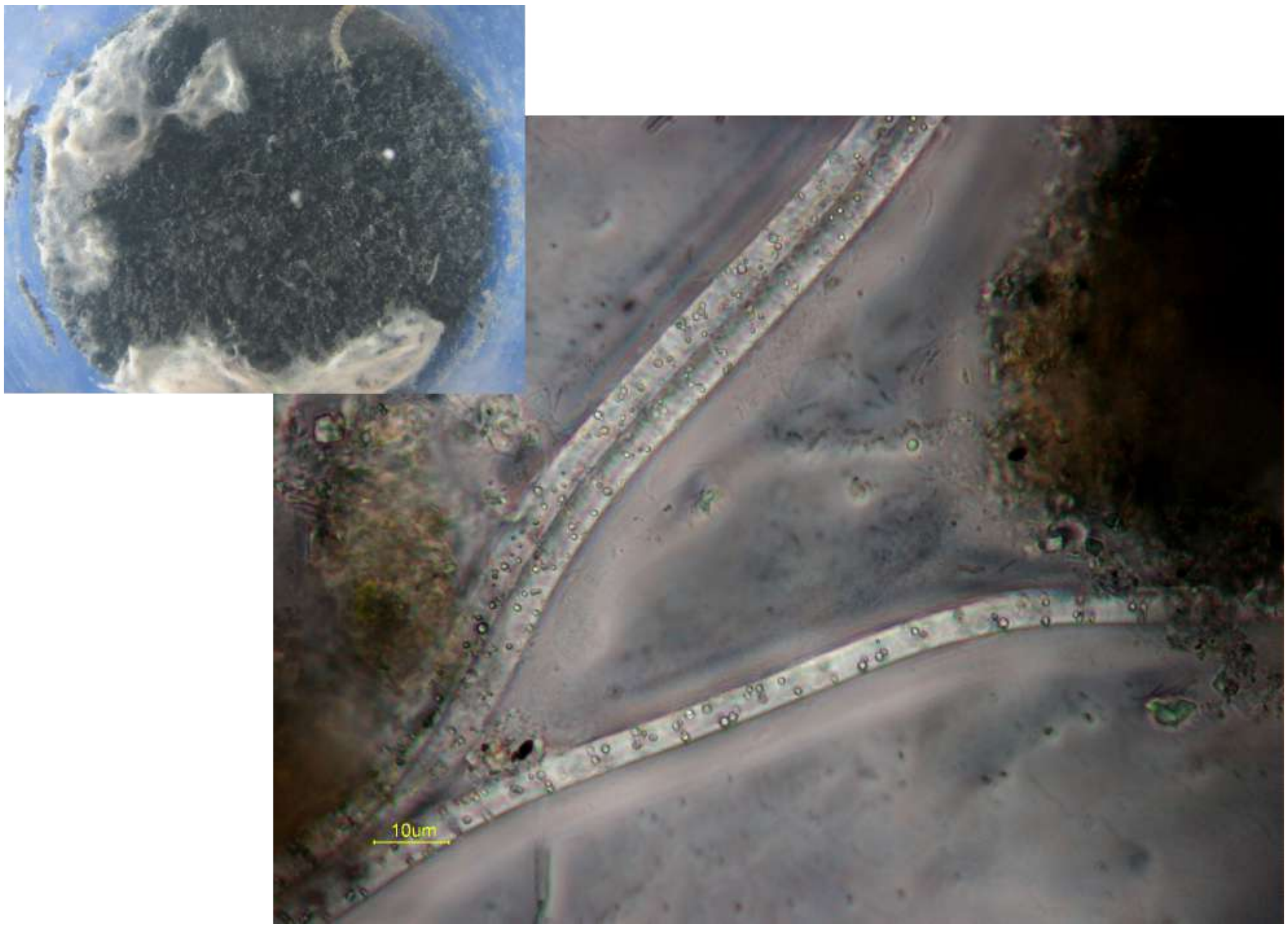


Figure 2.5: White filaments at the surface of some cores sampled in the Bay of Vidy and microscopic observation with phase contrast and a magnification of 1000x. White spots inside filaments are typically found in *Beggiatoa* sp. and is composed by elementary sulphur. Some other strains like *Thiothrix* sp. can also have this aspect and these type of inclusions.

Furthermore, *Beggiatoa* sp. have the ability to glide into the sediment to locate its optimal positioning regarding redox conditions, sulphide, nitrate and oxygen availability.

Microelectrode measurements of dissolved oxygen (Figure 2.6) show a very abrupt transition towards anoxia in all the sediments gathered from the Bdyay of Vidy. Logically, the most abrupt transition has been measured in cores collected close to the pipe where the input of organic matter is the most important. This high load of nutrients increases the mineralisation process leading to a rapid consumption of available oxygen. In these cores

(C3.03 in Figure 2.6) a transition towards anoxia was recorded in the first millimetre deep inside the sediment. For comparison, the transition in the cores retrieved further away from the rejection pipe was recorded in the first 2 or 3 millimetres. An attempt at measuring H_2S in parallel to O_2 has resulted in a failure since it has given only low concentrations of sulphides in the first centimetre investigated. As sulfur is necessarily provided by decaying organic matter released into the bay, this failure can be explained either by the presence of *Beggiatoa* sp. which consumes the sulphides as soon as they are produced, or by the presence of iron(III) which consumes sulphide via chemical oxidation. Furthermore, once iron is reduced as Fe(II), it can easily bind to sulfur anions, forming Fe-sulphide minerals, such as pyrrhotite.[78] In all these cases, sulphides are consumed and remain at low concentrations in sediments. In the Bay of Vidy, the iron needed to feed this process is probably provided by the waste water treatment plant effluents, since iron-chloride is used as floculant to remove phosphorus in the waste water treatment process (Figure 1.14).

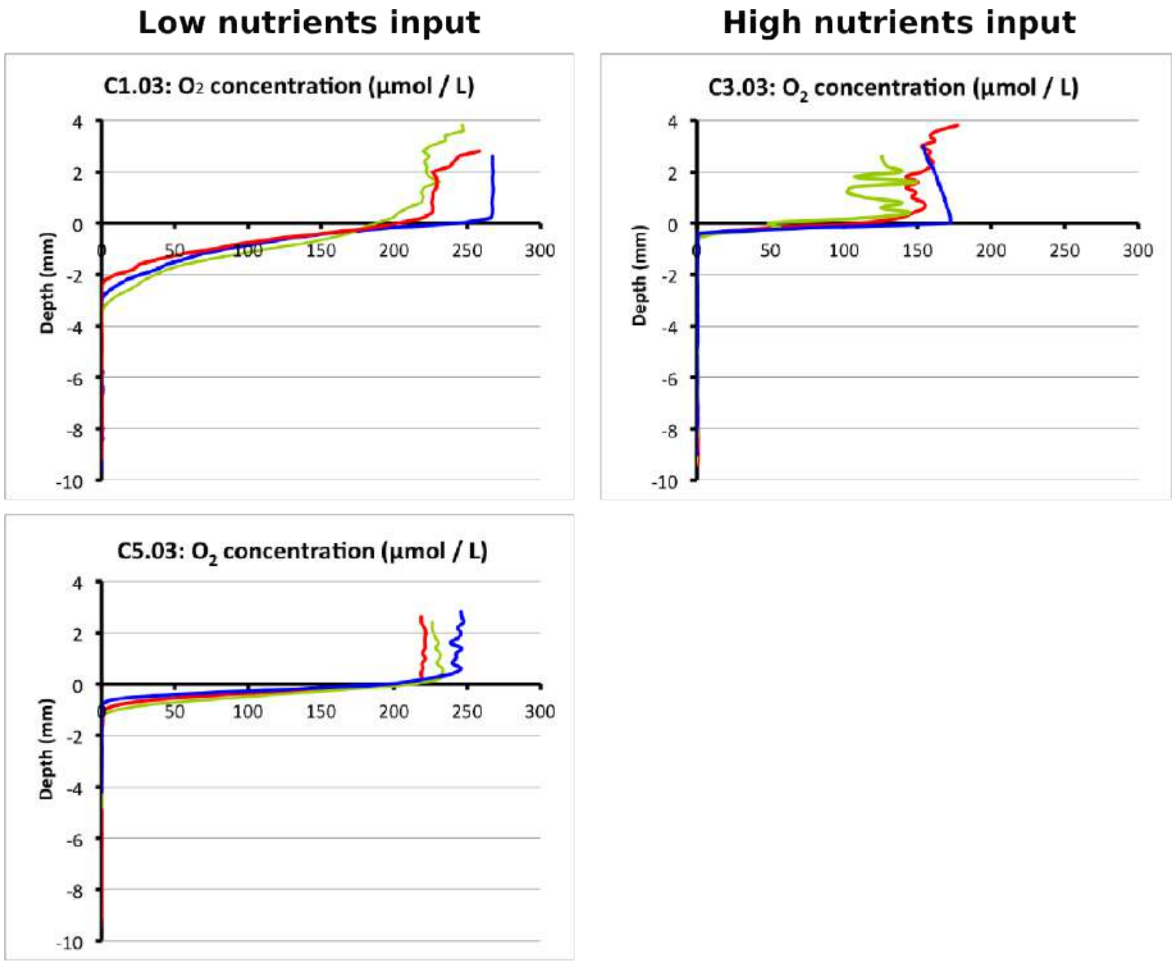


Figure 2.6: Oxygen profiles measured in sediments from three different locations of the Bay of Vidy. Measurements were done with microelectrodes. Even in sediments far from the pipe, with low nutrients input (C1.03 or C5.3), the transition toward anoxia is observed within only few millimeters. For comparison, transition occur within less than one millimetre in the sediment core collected near the pipe (C3.03)

2.2 Methods to study microbial diversity

Today, two main options are available to study microbial communities in natural as well as anthropic environments. The oldest, isolation and cultivation based method, has the advantage of providing easily information on metabolic capabilities of an isolated strain, without requiring expensive and complex laboratory equipment. This method suffers, however, from the well-known great plate count anomaly, which expresses the fact that approximately 99% of bacteria from the environment are uncultivable and remain, therefore, unseen by scientists.[37] Today, cultivation-based methods do not constitute any longer a mean for exploring the diversity of microbial communities from the environment. This kind of methods are helpful in combination with molecular techniques and "provides a context in which to corroborate the theoretical findings of metagenomics and much more direct access to genomes of environmental isolates".[79] In this study, as we are interested in the big picture of heavy metal resistant members of bacterial communities in sediments, we have, therefore, chosen to use the second type of methods based on molecular techniques. Molecular techniques rely on the study of DNA extracted from a given environment. In order to have a holistic view of the different communities of organisms, as well as to understand their function in the ecosystem, DNA extraction is, therefore, a critical step.[80]

Once extracted, DNA could be used for different downstream analyses. Polymerase chain reaction (PCR) techniques permit an amplification of a given portion of the DNA molecule enclosed between a forward and a reverse primer. When applied to DNA extracted from environmental samples, it allows to highlight, depending on the primer sequence, if a particular organism or a particular function is present or not. A variation of this is the real-time PCR (or qPCR) which allow the quantification of the occurrence of a given portion of DNA in addition to its detection in the environment. This technique has been used in this work for quantifying endospore-forming Firmicutes in the sediment samples. The design of the needed forward and reverse primers is described in the third published manuscript presented in this thesis.[81]
Fingerprinting techniques, like DGGE (denaturing-gradient gel electrophoresis), TGGE (thermal-gradient gel electrophoresis), SSCP (single strand conformation polymorphism), RISA (ribosomal intergenic spacer analysis), RFLP (restriction fragment length polymorphism) or one of their derivatives, permit to have a rapid overview of the structure of a bacterial community, or part of it depending on the target gene that is involved. For overall bacterial community analysis, 16S rRNA gene is the most currently used. As said before, it is also possible to target genes involved in a particular metabolic function, like *amoA*, a gene encoding an enzyme responsible for the oxidation of ammonia to nitrite. For this work, DGGE profiling has been performed with the 16S rRNA gene to have a global overview of the structure of the different bacterial communities in sediment samples. Information gathered this way, was mainly used to determine if communities were different from one another and if it was relevant to send all the samples for 454-pyrosequencing.

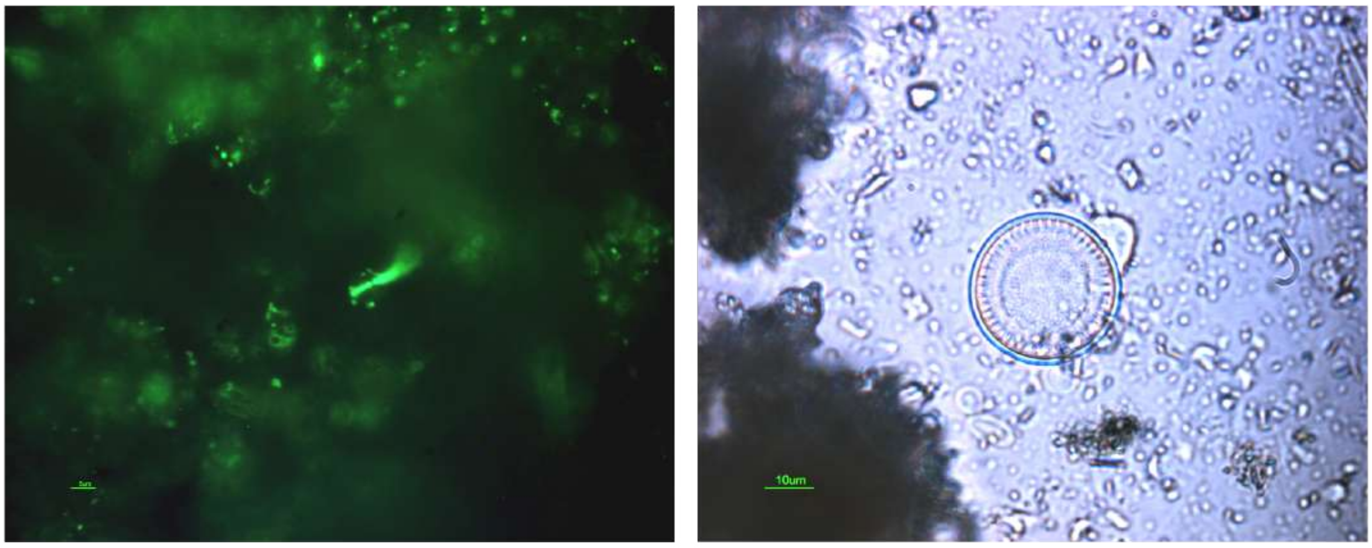
This approach is described in the second manuscript presented in this document.[82]
Finally, sequencing allows to have access to the succession of gnucleotides either of a given portion of the DNA molecule (amplicon-based sequencing, first step is a PCR amplification) or of all the DNA molecules that are present in the environment (metagenomic sequencing). Depending on the sequencing technology used, the number of sequences recovered, as well as their length, may be subject to large variation. For this study, 24 samples were selected for amplicon-based 454-pyrosequencing of a portion of the 16S rRNA gene. A complete description of the different steps involved in this analyses is presented in further chapters.

Storage issues:
Microbial communities are very sensitive to environmental conditions and prone to a rapid modification of their species composition and abundance when a change occurs in environmental conditions. To have the more realistic view of the communities in the sediment cores that were collected, their preservation until DNA extraction procedure is, therefore, capital. The first major determinant factor is of course the time elapsed between sampling and DNA extraction. Some authors claims that this factor does not have a big impact on bacterial community structure of small portions of soils even after 14 days of preservation at room temperature.[83] This, however, seems to me a little strange and as a precautionary approach I prefer to believe authors that claims that storing delay can deeply alter non-fixed bacterial communities after only 1 to 3 days, even when stored at the same temperature as in the environment.[84,85]
Other factors like the oxygenation of the surface water, as well as the temperature and the light exposition, could also play an important role in maintaining of the initial bacterial community during storage. In fact, when thinking about all this it became quite clear to us that the most important factor to preserve bacterial communities as close as possible as they were in the environment is to avoid any shock in the environmental conditions during storage. To remain as close as possible to the conditions that were present in the sediment floor of the Bay of Vidy, it has, therefore, been decided to store the cores in the dark at $4°C$ while oxygenating the surface of overlaying water by bubbling air to preserve the redox gradient at the surface of the sediments. Fully oxygenated water was effectively observed by MIR instruments during the first sampling campaign.
The sub-sampling of the cores and the fixation of the bacterial communities (procedure described in the further chapters) were performed as soon as possible, but in any cases within a maximum delay of one week. This delay is due to the time needed for the cell separation and DNA extraction procedures chosen to be used in this study (see description below). In contrast, since material for bacterial community analysis was collected (sub-sampled) in the middle of the core one can assume that the surrounding core material will act as a buffer to preserve environmental conditions. Regarding this aspect as well as the storage conditions, the delay of one week should not have induced a big change in the bacterial communities.

2.2.1 Cell separation and DNA extraction

When speaking about DNA extraction, one can distinguish between two options, the direct and the indirect DNA extraction methods. The first one relies on the direct lysis of cells in the environmental matrix, followed by a DNA purification step. The big advantage of this method rely mainly on its speed, because it requires less laborious steps to gather DNA from the environment. It is also recognized to give better extraction yields, as well as a less biased overview of the *in-situ* communities. Its major drawbacks are that DNA is often fragmented in small pieces (depending on the lysis method used), that it also gathers unwanted extracellular DNA, that it gives less purified products often contaminated by humic acids (which can significantly alter following downstream molecular procedures), and that it is inefficient to extract DNA from hard structures, like endospores, Gram-positive bacteria, or from cells strongly attached and protected by the environmental matrix.[86]
In the case of the indirect methods, those rely on a first separation of cells from the environmental matrix, followed by a full DNA extraction procedure. These kind of methods are, therefore, much more time consuming. DNA recovery yield is often lower, but its quality is much higher, allowing more sensitive downstream molecular analysis on longer DNA fragments (less fragmentation of DNA). The selection of one rather than another method require several parameters to be taken into account. As there is no optimal method that give the best results in every case, one should make a choice based on criteria like the target organisms studied, their resistance to lysis process, the strength of their attachment to the environmental matrix, the physico-chemical composition of samples, the bias induced by extracellular DNA, the quality and purity of DNA needed for downstream analysis.[80] In fact, rather than the type of DNA extraction (direct or indirect) used, the lysis protocol seems to be the major factor leading to bias, favouring a particular group of bacteria rather than another. As a consequence, the lysis protocol should be adapted to the hardness of the cell wall of the targeted organisms.[87,88] As we were not interested in eukaryotic or extracellular DNA, and as we planned to do some precise description of bacterial communities thanks to molecular techniques, an indirect methodology seemed to be the most suitable for this study. The first part of this work was, therefore, devoted to the selection of the best way to do indirect DNA extraction from lake sediments with the aim to study endospore-forming Firmicutes. To achieve this, several preliminary tests were made and are described below.

Cells separation: The first separation assays were performed on a test core retrieved from the Bay of Vidy in June 2011. The methodology was inspired by a development made in our laboratory by another PhD student (Tina Wunderlin) and completed by some parts of a protocol suggested in a review on cell extraction methodology for deep subsurface sediments (Jens Kallmeyer et al, graduate school of oceanography, University of Rhode Island, USA).[89] In summary, the method tested consisted of three distinct steps at the end of which microscopic observation and counting were performed after SYBR® Green I staining. The methodology and pictures from microscopic observations are presented in appendix A.1. The aspect of raw sediments is presented in Figure 2.7.



1000x, SYBR green staining, layer 0-3 cm   1000x, without staining, layer 3-9 cm

Figure 2.7: Microscopic observation of raw sediments from the Bay of Vidy with SYBR® Green I staining (left) and with phase contrast (right). Bacteria are clearly encapsulated in the sediment matrix.

The steps of the first methodology that was tested could be summarized in the three following points:

1. To extract cells from the sediment matrix a first homogenization step was performed by the use of an ultra-turrax (IKA, Staufen, Germany) dispersing tool and a dispersing solution made of 1% sodium hexametaphosphate, as suggested in another study aiming at extracting microbial biomass from biofilters.[90] In our lab, this protocol was supplemented by a low speed centrifugation to precipitate heaviest mineral particles. After two rounds of the treatment, supernatant was filtered through $0.2\mu m$ cellulose-acetate filter and kept frozen until further DNA extraction (Protocol developed by Tina Wunderlin)

2. The second part of the protocol consisted of an acidic dissolution step to extract cells enclosed in secondary carbonates. This dissolution step was followed by a detergent treatment to release cells encapsulated in exopolymeric substances.

3. Finally, a sonication step was added to break down the remaining strongest particles aiming at extraction of bacterial cells entrapped in these structures.

The two last steps was suggested by the work of Kallmeyer et al.[89] As their method also involves a density gradient centrifugation through a layer of Histodenz^{TM}, this step was also added to this first evaluation work. Results summarized in Figure 2.8 show that the simplest treatment with ultra-turrax and low speed centrifugation gather only one third of the cells entrapped in the sediments when comparing to the full treatment with the two further steps and density gradient centrifugation.

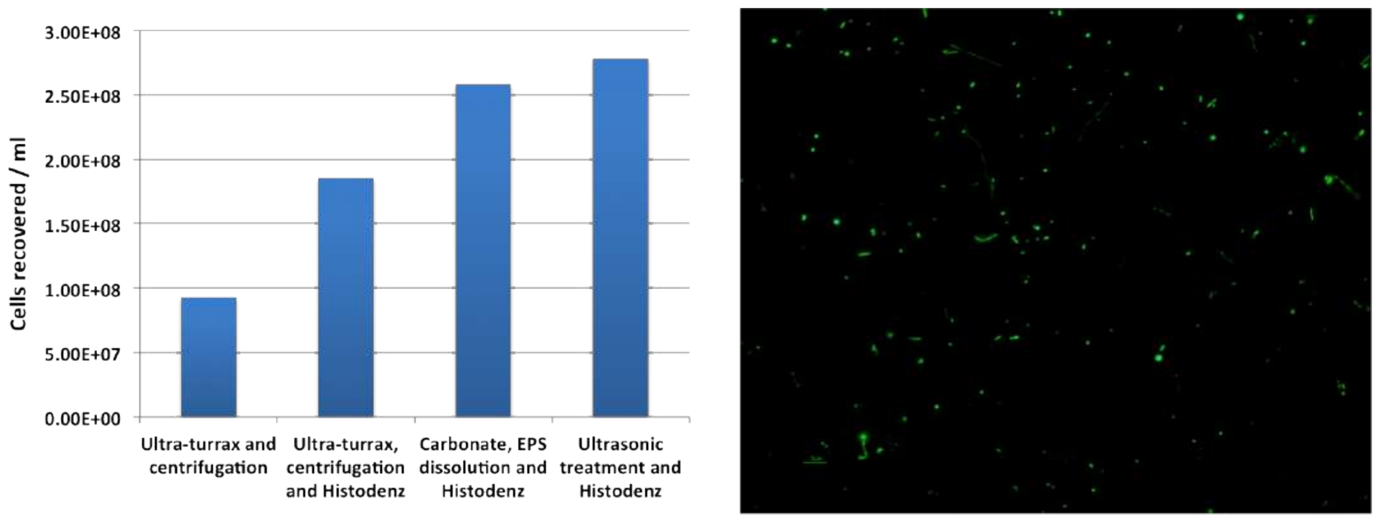


Figure 2.8: Results of the first tests aiming at finding the best method for separating bacterial cells from the sediments of the Bay of Vidy. Left part: number of cells recovered after the different steps of the separation treatment. Right part: Epifluorescence microscopic picture of the cells after the last step of the treatment. As seen in the image, cells are well individualized.

Although the full treatment seems to be the more efficient, it suffered, however, from a big limitation when extracting endospore-forming bacteria. It is effectively not clear if endospores stay entrapped in the Histodenz^{TM} layer, like other cells, or if they are pelleted with mineral particles at the bottom of the centrifugation tubes.[91] This hypothesis has, however, been verified by preliminary test performed in our laboratory by Tina Wunderlin, a colleague of mine. For the furthest tests, Histodenz^{TM} centrifugation was then removed from the separation methodology. Figure 2.9 presents the layout of the last separation tests that were performed in the beginning of this study.

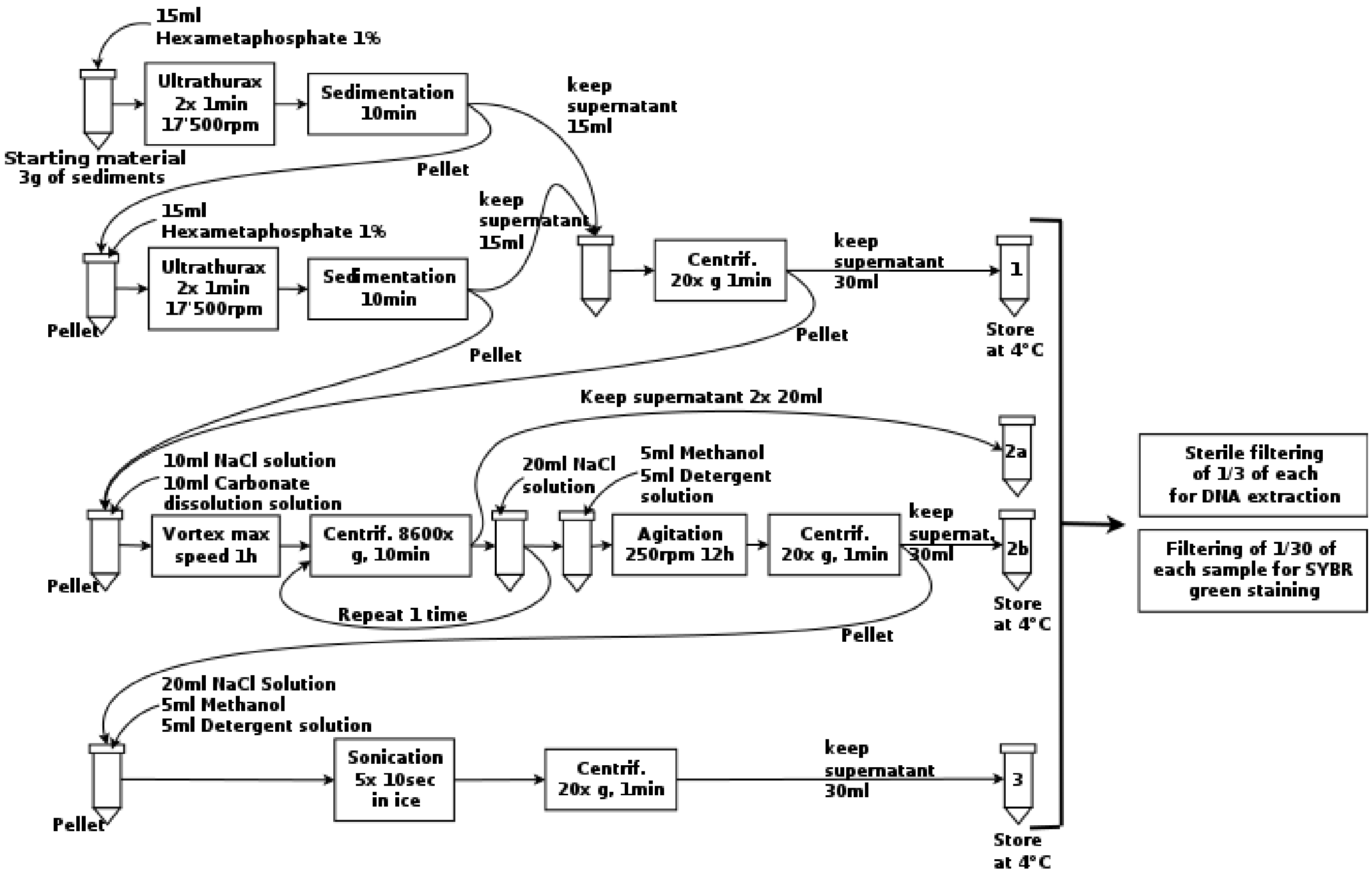


Figure 2.9: Diagram of the second methodology tested for the separation of bacterial cells from sediments. At the end of each step, DNA was extracted with FastDNA® SPIN Kit for soil (MP biomedicals, Santa Ana, USA) with up to four sequential bead-beatings.

At the end of each step, a third of the remaining supernatant was filtered through $0.2\mu m$ acetate-cellulose filter. DNA extraction was performed with the FastDNA® SPIN Kit for soil (MP biomedicals, Santa Ana, USA) including a variable number of sequential bead-beating rounds. Community profiles were established by denaturing-gradient gel electrophoresis (DGGE) on the V3 region of the 16S rRNA gene, giving us an idea of the microbial community extracted at each step. In parallel, a quantification of bacteria was performed by real-time PCR (qPCR) on the 16S rRNA gene. For these two last steps, standard protocols are available in our lab since it constitutes a routine analysis. The results of these analysis are presented in Figure 2.10.

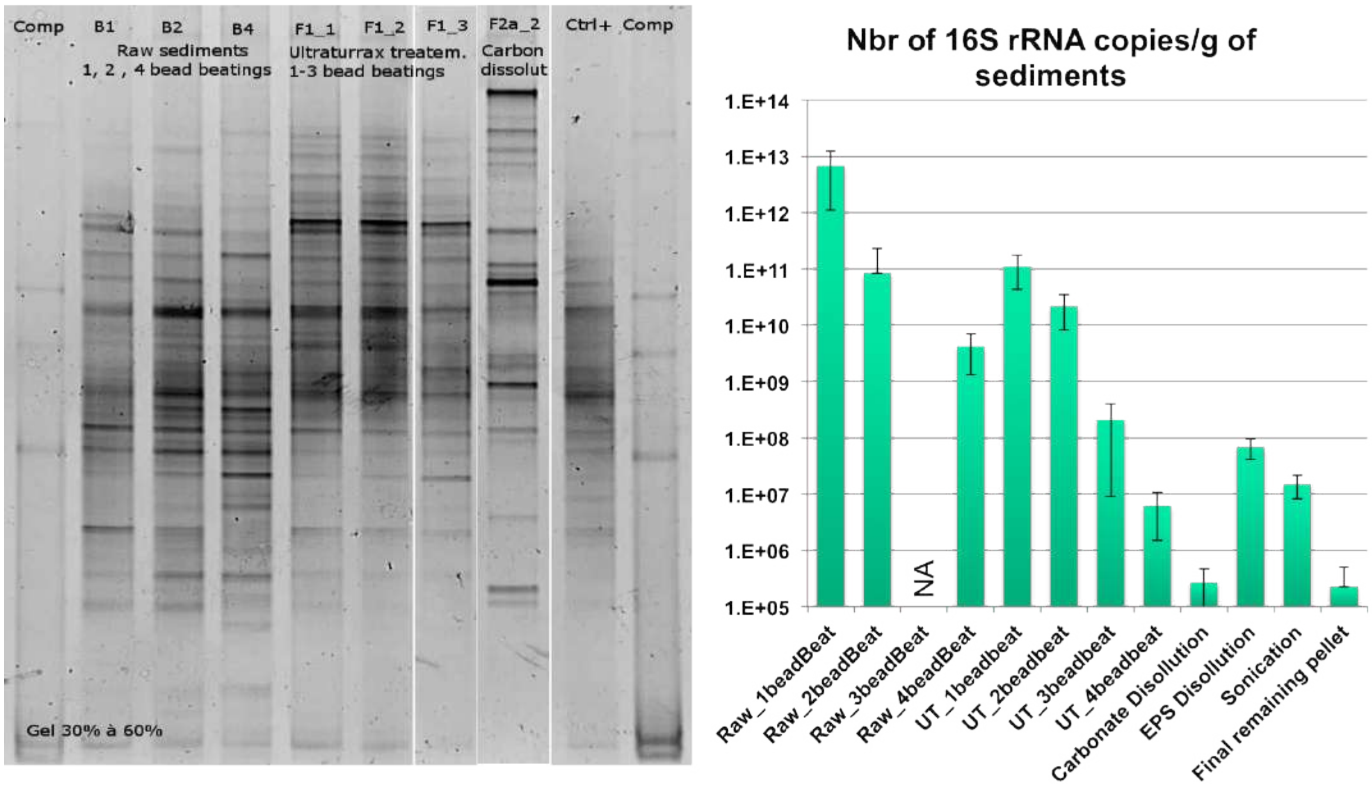


Figure 2.10: Results of the second separation tests performed at the beginning of this study. DGGE profile (left) shows clearly that the composition of bacterial communities may vary depending on the methodology used for the extraction. Effectively, direct DNA extraction (B1, B2 and B4) does not gives the same overview of the bacterial communities when compared to indirect methodology (F1_1 to F1_3 and F2a_2). qPCR quantification (right) shows that direct method gives much better yield (2 order of magnitudes more DNA) than indirect method. Sequential steps of bead beating extract a decreasing amount of DNA.

Because of a lack of time to complete these very interesting experiments before going to the field for sampling, some data are missing from these results. For example it should has been very interesting to re-do the DGGE analysis for samples that do not give results (too few DNA recovered to perform DGGE analysis), like those corresponding to the end of steps 2 and 3 of the protocol. In the same order of ideas it could also have been interesting to find a way to avoid the co-precipitation of endospores with mineral particles during the density gradient centrifugation. An optimisation of the densities of Histodenz^{TM} layers should perhaps help to separate endospores form minerals.

As it was not possible to finish the development of our own method prior to the first sampling campaign with the MIR submerssible, we finally decided to use the simplest protocol involving ultra-turrax homogenization of sediments suspended in 1% sodium hexametaphosphate. As demonstrated in the different preliminary tests performed (Figures 2.8 and 2.10) the separation of cells from sediments is not complete with this solution. As part of this drawback, this method is, however, not so bad, because it is much less time consuming and allows a more realistic treatment of sediments prior the degradation of the bacterial communities during the storage of the cores. Furthermore, even if the members of the community gathered by additional separation steps are different, as shown by DGGE lane F2a_2 when comparing to simple ultra-turrax treatment (lanes F1_1, _2, _3), the additional steps will only allow the recovery of a portion of the community representing less than 0.1% of the total abundance as demonstrated by qPCR measurements (Figure 2.10, right part). Finally, as the final product is intended to DNA extraction, it is not necessary to have well individualized cells as it should be the case for other kind of analysis, like flow-cytometry counting of bacterial abundance.

DNA extraction: DNA extraction could also be subject to intense investigation to find the method that gives the least biased image of the community. Effectively, depending mainly on the cell lysis technology employed, one kind of bacteria could be favoured rather than another. In fact, no single lysis protocol could extract DNA simultaneously from all kind of bacterial cells that are present in complex environments, such as soil or sediments. To achieve better results, several lysis protocols with different stringencies should be used and final products should be pooled together.[88]
The selection of the more suitable extraction kit was based on an evaluation of the efficiency of commercial kits to extract endospores from different soils.[92] FastDNA® SPIN Kit for soil (MP biomedicals, Santa Ana, USA) gave the best results in this study. We, therefore, decided to use this kit in our following work.
As part of the selection of the best separation method, the preliminary tests gave us precious results concerning the number of bead beating steps needed to extract the highest proportion of DNA from the samples. From Figure 2.10 it appears that three sequential bead beating steps are sufficient to extract most of the DNA. For further work we decide to use four sequential rounds.
Finally, in order to have high quality DNA, but also to remove an excess of guanidine thiocyanate in the final DNA product delivered by the extraction kit, we decided to add an extra ethanol precipitation step at the end of the procedure.

The overall protocol involving cell separation, DNA extraction and ethanol precipitation is presented in appendix A.2.

2.2.2 Pyrosequencing

During the last 10 years sequencing technologies have improved at a remarkable speed. Today next generation sequencing technologies allows to obtain up to 3 billions of small sequences (length around 150 base-pairs) within a single run (Illumina sequencing). Thanks to such technologies the microbial diversity of complex environments like soils, sediments or human intestinal tracts is finally reachable. The big limitation of these new generation high throughput sequencing technologies, like Illumina or SOLiD, relies on the limited size of the sequences that are recovered at the end of the process. This limit of size often requires a tedious assembling step before the assignation of the sequences to a particular gene in a database. For exploring microbial community structure based on 16S rRNA gene sequencing, the 454-pyrosequencing technology offer a good compromise since it allows to recover up to 1 million of sequences with a length of around 500 to 700 base-pairs (approximately one third of the entire 16S rRNA gene). The main drawback of this technology is, however, its cost (around 8000 US dollars for an entire run). To limit spending, a multiplexing technique was developed allowing several samples to be treated in the same run. Multiplexing involve first a PCR amplification of the region of interest of the genome, then the attachment of a particular sequence tag at the beginning of sequences corresponding to each sample. For this project, two series of 12 samples were sent for 454-pyrosequencing of the 16S rRNA gene to Eurofins MWG GmbH (Ebersberg, Germany). Each serie was run on 1/4 of plate, allowing the recovery of up to 200'000 sequences in total. As the length of the sequences is limited (around 500 base-pairs) by the sequencing technology and as the entire 16S rRNA gene is three times longer it was necessary to select the more interesting zone to sequence. The selection of the more suitable primers for this study is described below.

Choice of primers for 454-pyrosequencing Depending on their design, primers can amplify preferentially a particular group of organisms leading to an over-representation of particular taxa in the dataset of sequences. Since modern molecular taxonomy is based on ribosomal small-subunit sequences (16S in bacteria and Archaea, 18S in eukarya) its, therefore, normal to select it for diversity assessment of a particular environment. Different "universal" primers have been designed for amplifying bacterial 16S rRNA genes, but relatively few of them are suitable for 454-pyrosequencing due to the limitation of sequence length. The 16S rRNA gene is composed by 4 domains in which conserved portions of the gene are separated by nine hypervariable regions. It has been demonstrated that the best molecular differentiation between bacterial taxa is achieved by sequence contained within the two first conserved domains of the gene.[93] Figure 2.11 shows the different taxa amplified by four sets of primers suggested for 454-pyrosequencing studies[94] and containing the V3 hypervariable region we use routinely in our laboratory for DGGE anylysis. Results shown in this figure are based on the sequences recovered when runing NCBI Primer-BLAST[95] with the different sets of primers.

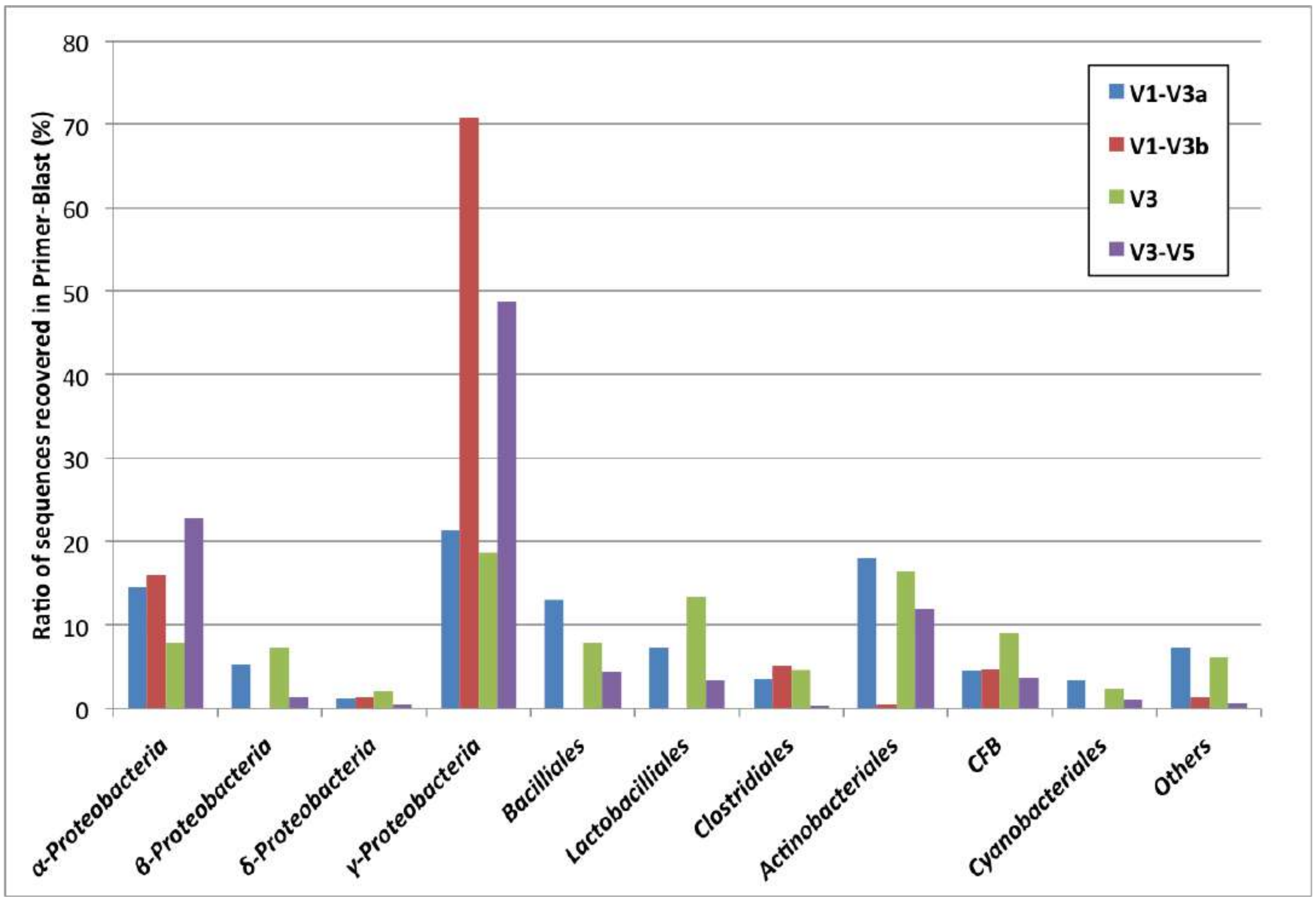


Figure 2.11: Results of Primer-BLAST[95] query for four different sets of primers suitable for 454-pyrosequencing of the 16S rRNA gene. V1-V3a is the couple of primers that seems to be the more universal.

The Figure 2.11 clearly shows that the best set of primers, giving the most holistic view of the community, correspond to the V1-V3a primer pair. Their sequences are given in Figure 2.12 as well as the amplicon length. With 529 base-pairs amplified, amplicons generated by this set of primers have the ideal length for 454-pyrosequencing. We, therefore, choose this primer set for further work.

Primer name	Sequence	Position (E. coli)
Eub8f	AGA GTT TGA TCC TGG CTC AG	8–27
Eub519r	GTA TTA CCG CGG CTG CTG G	519–536

PCR conditions : Touchdown 65->55°C, -1°C/cycle
Amplicon length: 529bp

Figure 2.12: Sequences of both reverse and forward primers used for 454-pyrosequencing. PCR conditions are also given as well as amplicon length for *E. coli*

Bioinformatic analysis of genetic data In a first approach, 454-pyrosequencing data were analysed with NCBI-BLAST[96] tool run locally on a desktop computer. The appendix A.6 presents the commands and the scripts used for this purpose.
However, a COST (European cooperation in science and technology) workshop on NGS

(next generation sequencing) and metagenomics followed at the national biotechnology center of Madrid (Spain) in July 2012, showed us the existence of several, more clever ways, to treat 454-pyrosequencing data. Some tools, like Mothur[97] or Qiime[98] integrate several pieces of free software allowing to run a whole pipeline for analysing of NGS data, and more particularly 16S rRNA gene sequences. These tools are based on a first clustering of the sequences by operational taxonomic units (OTUs) picking. OTUs picking is a way to group similar sequences together based on their level of identity. In the case of the standard pipeline of Qiime used in this thesis, an identity level of over than 97% was used to group sequences together. This threshold level has been defined as the "bioinformatic standard" within which we are faced to an unique specie.[99] As NGS sequencing technologies are relatively prone to sequencing errors, the first step when dealing with this kind of data is, therefore, to filter sequences and remove those that do not reach the quality requirements. Qiime software integrates the needed tools to perform this step. Chimera detection and elimination is also another big part of NGS data analysis. Qiime software also integrates some functions to check for chimeras in the dataset. The main steps of "standard" Qiime workflow for 454-pyrosequencing data are the following:

1. Demultiplexing of the sequences and quality checking
2. OTU picking
3. Taxonomic assignment of the OTUs
4. Computation of rarefaction curves
5. Beta-diversity analyses
6. Bootstrap analyses of the beta-diversity
7. Multivariate representation of the beta-diversity data, using either weighted or unweighted distances computation (this last aspect will be discussed in the conclusion of this manuscript)

The command used to run these different steps, as well as their outputs, are presented in appendix A.3.

As it was difficult to integrate environmental data into Qiime to run multivariate analyses, especially written R scripts were used to perform these steps. Starting from the table resulting form Qiime OTUs picking step and integrating chemical measurements done on the sediments (see description below) these scripts performed multivariate analyses to visualize links between microbial community composition and environmental factors. The R scripts used are documented in appendix A.4
Finally, as the assignation of OTU in the Qiime worflow was not entirely satisfactory a taxonomical reassignment of each OTU by BLAST against a more complete 16S rRNA database was performed. The commands and the scripts used for this purpose are described in appendix A.5

2.3 Chemical measurements

Chemical measurements that were performed during this study are out of the scope of my competences and are briefly described in the different articles. All these measurements were possible thanks to the help of various geochemists or soil scientists. In summary, ICP-MS (inductively coupled plasma mass spectrometry) measurements of total particulate heavy metals in the sediments were done at the university of Geneva in the department of analytical chemistry thanks to the help of Marie-Louise Tercier-Waeber and Matthieu Masson (postdoc). The mercury measurements were done at the Forel-Institute (University of Geneva) with an atomic absorption spectrometer (AAS) thanks to the help of Elena Gascon, PhD student in the group of Jean-Luc Loizeau. CHN analysis, as well as phosphorus content and loss on ignition were performed at the University of Neuchâtel thanks to the help of Roxanne Kohler-Milleret from the soil and vegetation laboratory headed by Jean-Michel Gobat. Total sulphur and some other molecules (Na_2O, MgO, Al_2O_3, SiO_2, MnO, Fe_2O_3, P_2O_5, K_2O, CaO, TiO_2) and elements (V, Cr, Ni, Cu, Zn, Rb, Sr, Y, Zr, Ag, Sn, Ba, and Pb) were measured by XRF (X-ray fluorescence spectrometry) thanks to the help of Jean-Claude Lavanchy from the Centre of Mineral Analysis (CAM) at the University of Lausanne. Finally, dissolved heavy metals in cultivation laboratory experiments with Clostridia strains were performed thanks to the help of Jasqueline Peña and Anna Simanova (first assistant) from the laboratory of environmental geochemistry at the University of Lausanne.

3

Published papers

This chapter is intended to show the papers that were written by the author.

Contents

Paper I

In which we found a link between trace-metal pollution and endospore-forming bacteria...

E3S Web of Conferences 1, 33011 (2013)
DOI: 10.1051/e3sconf/20130133011

Endospore-forming bacteria as an indicator of pollution in sediments of Lake Geneva

M. Bueche[1], L. Sauvain[1], T. Wunderlin[1], T. Junier[1], M. Tercier[2], J. -L. Loizeau[3], M. Masson[3], E. Gascon[3] and P. Junier[1]

[1] Laboratory of Microbiology, University of Neuchâtel, CH-2000 Neuchâtel, www.unine.ch/biol
[2] Analytical and Biophysical Environmental Chemistry Group (CABE), University of Geneva, 1200 Geneva (Switzerland)
[3] Institute F.-A. Forel, University of Geneva,1200 Geneva (Switzerland)

Abstract. Treated wastewater and runoff-water is released by the outlet of the sewage treatment plant of Vidy (Lausanne) directly into the Lake of Geneva via a pipe located 300m from the shore. Even if this water is properly treated with modern technologies, we can observe an accumulation of micro pollutants into the sediments, and particularly heavy-metals. The main objective of this project is to investigate how these elevated concentrations of heavy metals affect both abundance and diversity of prokaryotes in the sediments. A special emphasis was given to endospore-forming bacteria, which could use sporulation as a survival strategy to resist in highly contaminated areas. This study could have implications both for understanding the role of endospore-forming bacteria in the environment as well as in terms of improving the bioremediation processes.

Key words: Endospore-forming bacteria, Heavy metals, Metagenomic analysis

Introduction

Lake Geneva is the largest freshwater lake in Western Europe. Since several decades it has been the subject of close environmental monitoring. Different studies from the Forel Institute(Pote et al., 2008; J-L Loizeau et al., 2004) indicate that the sediments of the Vidy Bay contain high level of heavy-metals (HM). This is mainly due to the rejection into the bay of the runoff surface water, but also the treated wastewater from the city of Lausanne and its suburbs (412'000 equivalent inhabitant). Even if the concentration of HM in the rejected water is low, the surrounding sediments accumulate them over time. Endospore-forming bacteria (EFB) are well known for their ability to resist to harsh environmental condition over a long period of time(Vreeland et al., 2000). When conditions are not optimal, endospore formation is triggered. Spores are resistant structures intended to protect and conserve the genetic material of the organisms until the conditions become suitable for vegetative growth (Nicholson, 2002). For a long time, spores had been considered as a dormant state in which no metabolic activity takes place. Recent studies(Junier et al., 2009; Rosson and Nealson, 1982; Tebo and Obraztsova, 1998) have however demonstrated that some redox activities are active at the surface of endospores.

This study aims at evaluating an eventual link between endospore-forming species and the load of heavy-metals in the environment. A better understanding of the role of this special group of bacteria in polluted areas could lead to improved bioremediation processes in the future.

Materials and Methods

This study is a part of the "Elemo" project that aims to gain a better understanding of the biogeochemical processes taking place in the water column but also in the sediments of Lake Geneva. Thanks to the sponsors, two MIR submersibles could be used to monitor the surface of the sediments and to select the more interesting coring zones around the rejection pipe of the wastewater treatment plant of Lausanne. A first set of eight cores was retrieved from precisely selected areas by using the robotic arm of the submersible. In order to have more contrasted heavy-metal concentrations, five additional cores were retrieved from a boat during a second sampling campaign.

Cores were transported in our cold room and conserved there for no more than 4 days. During these days, air has been bubbled at the surface of supernatant water in order to preserve the gradient of oxygenation into the sediments as well as the structure of the microbial community at the oxic/ anoxic interface.

According to the expertise we have developed in our laboratory, and in order to have a more holistic view of the microbial community, we have then opted for an indirect DNA extraction method. Sediments were split into two separated layers (0 to 3cm and 3 to 9cm deep). They were first homogenized in 1% Hexametaphosphate solution with an Ultra Turrax (IKA, Germany). Mineral particles were then separated from cells by sedimentation and gentle centrifugation. Recovered supernatants, harboring microbial cells, were filtered on 0.2µm cellulose-nitrate filters. One half of each filter was cut in pie pieces and frozen into a bead-beating tube from the FastDNA® Spin Kit for Soil (MP Biomedicals) until processing. In order to harvest DNA from very resistant structures, like spores, a modified protocol, including 4 sequential bead-beating steps, has been used. Inhibitor free DNA has then been obtained by an extra ethanol purification step at the end of the extraction procedure.

A first overview of the bacterial community has been obtained by DGGE fingerprinting on the V3 region of 16S rRNA gene. Based on this, both layers in six cores have been selected for further analyses. The quantification of the overall bacterial population has been done by qPCR on the 16S rRNA gene. Quantification of the endospore- forming subpopulation has been done by qPCR on an endosporulation specific gene (*spo0A*). Specific primers have been developed for this. 454 pyrosequencing was carried out by “Eurofins MWG GmbH” (Germany) on V1-V3 region of 16S rRNA gene. Bioinformatic identification of the approximately 20’000 sequences per sample has been done in our lab by running NCBI blastn locally.

Air dried/agate grinded sediments were used for physico-chemical analyses. A CHN analysis preceded or not by an acidic digestion step, has been used to retrieve both inorganic and organic fraction of C and N in the sediments. The total amount of HM was determined by an acidic digestion (aqua regia) followed by ICP-MS analysis.

A first set of multivariate exploratory analyses has been done with the open source R software and the library vegan. Cluster analysis and nonmetric multidimensional scaling have been used to achieve a better understanding of the main differences among the characteristic bacterial groups that shape the communities in the samples. An overview of the effect of physico-chemical factors on the prevalence of the different species has been plotted by the environmental fitting procedure described in the vegan tutorial(Oksanen, 2011). Further statistical analyses, like multiple factor analysis (MFA) and the hypothesis-driven redundancy analysis (RDA) are in progress in order to demonstrate more accurately the link between HM and specific bacterial groups.

Finally, wetlab experiments with laboratory strains are planned to achieve a better understanding of the physiological mechanisms that allow endospore-forming bacteria to cope with HM pollution. For these experiments, flow-cytometry will be used to quantify the proportion of endospore vs. vegetative cells in the different cultures amended with increasing amount of HM.

Results and Discussion

The physico-chemical analyses, as well as the composition of the bacterial community assessed by pyrosequencing, show that four out of 12 samples analyzed are clearly separated from the others among the 12 analyzed. Furthermore, the first multivariate analyses tend to demonstrate a strong link between the prevalence of endospore-forming species and the HM content of the sediments (Fig.1). As a good correlation is also observed between HM and nutrients (represented by C.org and N.total vectors) it is not clear if the first set of physico-chemical factors is sufficient to explain the prevalence of EFB in the community of these 4 particular samples. In order to answer to this uncertainty, a hypothesis-driven multivariate analysis has to be computed. RDA analysis has been chosen but results are not yet available.

Finally, quantification of EFB with the qPCR approach on *spo0A* gene gives reliable results when compared with the pyrosequencing data. Higher abundance of endospore-formers are reported in the four samples with high HM concentrations.

Conclusion

According to these first results, a clear link could be established between EFB and the load of HM in the sediments. At this point of the work we still don’t know if this selection is due to the capacity of EFB to form resistance structures and therefore to escape harsh environmental conditions, or if it is due to other special features of Clostridiales. The further wetlab experiments we have planned should answer this. Additionally, we hope that we will isolate some important species that display such resistance properties. This is particularly important from an ecological point of view, because resistant species are also those that can re-colonize the empty niches after the perturbation caused by high HM loads. In other words, EFB could be involved in the resilience of ecosystems after HM pollution events. Furthermore, if we are successful with these isolation steps, we could also make a characterization of these species, and eventually discover some new depollution mechanisms that take place in these environments. This could be relevant for further bioremediation processes.

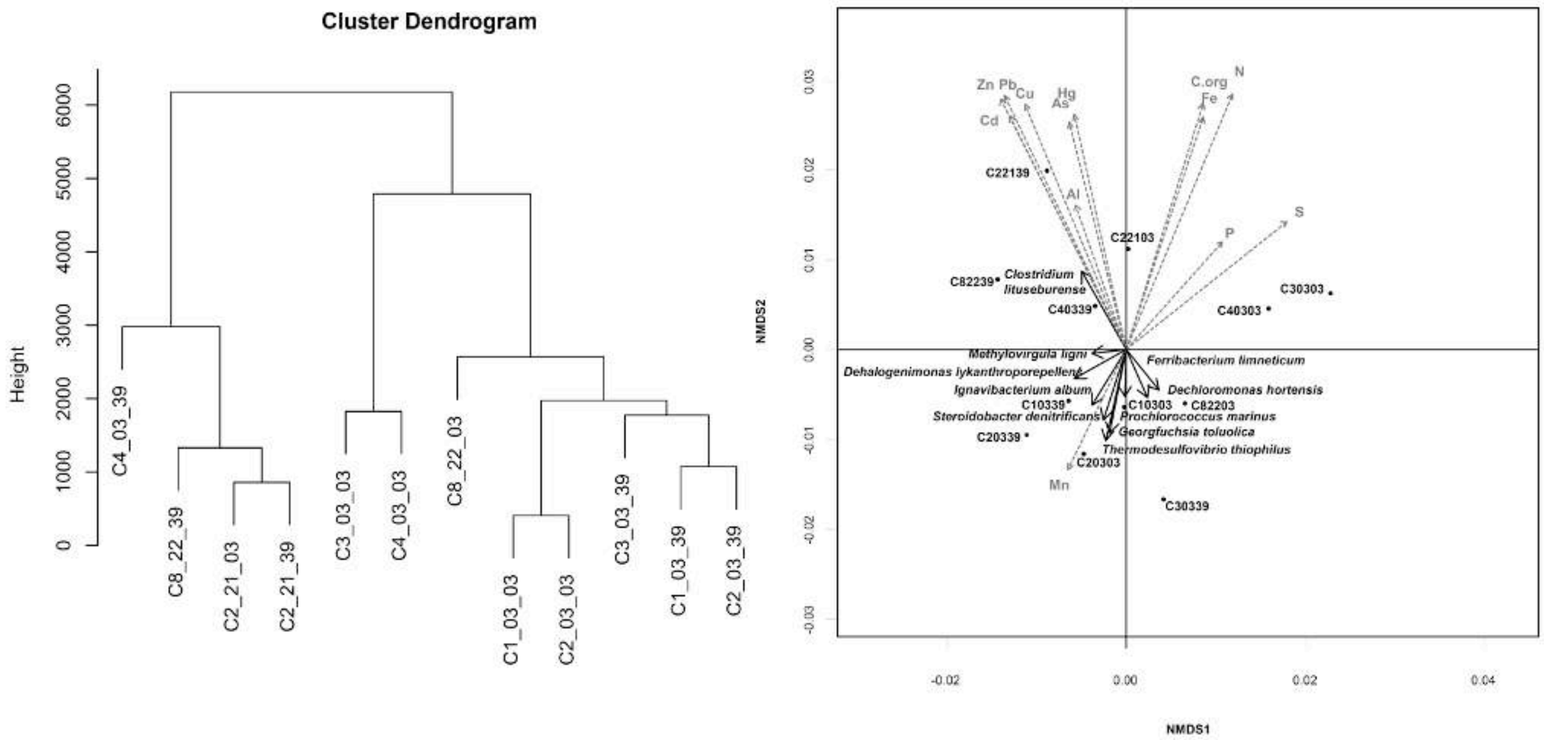


Fig. 1. Left: Cluster analyses (euclidean distance, average clustering) of the bacterial community assessed by 16S rRNA gene pyrosequencing with DNA obtained from 12 selected sediment samples. The four samples to the left form a clear separate group. Right: Nonmetric multidimensional scaling ordination (euclidean distance) of the 12 samples and environmental fitting of measured physico-chemical factors. This representation demonstrates that heavy-metals are affecting bacterial communities of these four particular samples by favoring endospore-forming species like *Clostridium lituseburense*. To refine this conclusion, more sensitive statistical methods (like redundancy analysis - RDA) are still needed.

Acknowledgements

The authors thank all the members of the Elemo project and particularly the financial support provided from Ferring pharmaceuticals (St-Prex) and the “Consulat honoraire de Russie” (Lausanne). We also thank the logistic support provided by the EPFL (Ecole Polytechnique Fédérale de Lausanne).

References

Junier P, Frutschi M, Wigginton NS, Schofield EJ, Bargar JR, Bernier-Latmani R. Metal reduction by spores of Desulfotomaculum reducens. Environ Microbiol 2009;11:3007–17.

Loizeau J-L, Pardos M, Monna F, Peytremann C, Haller L, Dominik J. The impact of a sewage treatment plant’s effluent on sediment quality in a small bay in Lake Geneva (Switzerland-France). Part 2: Temporal evolution of heavy metals. Lakes & Reservoirs: Research and Management 2004; 9:53–63.

Nicholson W. Roles of Bacillus endospores in the environment. Cellular and Molecular Life Sciences 2002;59:410–6.

Oksanen J. Multivariate Analysis of Ecological Communities in R: vegan tutorial. R tutorial; October 2011. http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf.

Pote J, Haller L, Loizeau JL, Bravo AG, Sastre V, Wildi W. Effects of a sewage treatment plant outlet pipe extension on the distribution of contaminants in the sediments of the Bay of Vidy, Lake Geneva, Switzerland. Bioresource Technology 2008; 99: 7122–31.

Rosson R, Nealson K. Manganese binding and oxidation by spores of a marine Bacillus. Journal of Bacteriology 1982;151:1027–34.

Tebo B, Obraztsova A. Sulfate-reducing bacterium grows with Cr(VI), U(VI), Mn(IV), and Fe(III) as electron acceptors. Fems Microbiology Letters 1998;162:193–8.

Vreeland R, Rosenzweig W, Powers D. Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. Nature 2000;407:897–900.

Paper II

In which we found a link between trace-metal pollution and endospore-forming bacteria...

Published in Aquatic sciences, online October 26th 2013, p 1-14
which should be used for any reference to this work

Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria

Loïc Sauvain · Matthieu Bueche · Thomas Junier · Matthieu Masson · Tina Wunderlin · Roxane Kohler-Milleret · Elena Gascon Diez · Jean-Luc Loizeau · Mary-Lou Tercier-Waeber · Pilar Junier

Abstract Lake sediments in areas close to the outlet of wastewater treatment plants are sinks for pollutants. Bacterial communities in sediments are likely affected by the released effluents, but in turn they might modify the distribution and bioavailability of pollutants. On the shore of Lake Geneva, Switzerland, wastewater from the City of Lausanne is treated and discharged into the lake via an outlet pipe in the Vidy Bay. The objectives of this study were to assess (1) the impact of the treated wastewater release on the bacterial communities in the Vidy Bay sediments and (2) the potential link between bacterial communities and trace metal sediment content. Bacterial community composition and abundance were assessed in sediments collected in three areas with different levels of contamination. The main factors affecting bacterial communities were inferred by linking biological data with chemical analyses on these sediments. Near to the outlet pipe, large quantities of bacterial cells were detected in the three upper most cm (3.2×10^9 cells assessed by microscopy and 1.7×10^{10} copies of the 16S rRNA gene assessed by quantitative PCR, per gram of wet sediment), and the dominant bacterial groups were those typically found in activated sludge (e.g. *Acidovorax defluivii* and *Hydrogenophaga caeni*). Three samples in an area further away from the outlet and one sample close to it were characterized by 50 % of endospore-forming Firmicutes (*Clostridium* spp.) and a clear enrichment in trace metal content. These results highlight the potential role of endospore-forming Firmicutes on transport and deposition of trace metals in sediments.

Keywords Sediments · Bacterial communities · Firmicutes · Trace metals · High-throughput DNA sequencing · Lake Geneva

This article is part of the special issue "éLEMO – investigations using MIR submersibles in Lake Geneva".

Electronic supplementary material The online version of this article (doi:10.1007/s00027-013-0313-8) contains supplementary material, which is available to authorized users.

L. Sauvain · M. Bueche · T. Junier · T. Wunderlin · P. Junier (✉)
Laboratory of Microbiology, University of Neuchâtel, Rue Emile Argand 11, 2000 Neuchâtel, Switzerland
e-mail: pilar.junier@unine.ch

M. Masson · M.-L. Tercier-Waeber
Analytical and Environmental Chemistry, Department of Inorganic and Analytical Chemistry, University of Geneva, Geneva, Switzerland

R. Kohler-Milleret
Laboratory Soil and Vegetation, University of Neuchâtel, Neuchâtel, Switzerland

E. Gascon Diez · J.-L. Loizeau
Earth and Environmental Science Section, Institute F.-A. Forel, University of Geneva, Geneva, Switzerland

Introduction

Access to clean drinking water is undoubtedly one of the major concerns for humans. However, agricultural, industrial, and domestic activities lead to increasing chemical contamination of freshwater via runoff or wastewater discharge (Schwarzenbach et al. 2006; F ̈o rstner and Wittmann 1981). In industrialized countries, wastewater treatment plants (WWTPs) remove a large fraction of contaminants. Nevertheless a significant quantity of chemicals are still not efficiently retained or broken down during the process and their release into aquatic environments remain a significant concern for environmental and public health (Schwarzenbach et al. 2006). Therefore, we still need to understand how to deal not only with the sources, but also with the fate of these contaminants that may accumulate over-time. Among the inherently persistent contaminants are the trace metals (TMs), which are involved in biogeochemical cycles in aquatic ecosystems and distributed in various chemical forms, including inorganic and organic colloids as well as particulate matter (Tercier-Waeber et al. 2012). Colloids and particulate matter (through coagulation, aggregation and sedimentation) play a key role in the transport of TMs from the water column to the sediment. Thus, sediment is one of the major sinks of TM contaminants (Schwarzenbach et al. 2006; P o t ´e et al. 2008; Thevenon et al. 2011a).

Bacteria, through their diverse metabolic activities and large combined biomass, play a key role in biochemical cycles in aquatic sediments (Nealson 1997). In particular, bacterial metabolic activity can facilitate the remobilization of TMs from the sediment into the water column (Fo¨rstner and Wittmann 1981; Zhang et al. 1996). An example of this is the methylation of mercury occurring in anaerobic sediments, which is accelerated by in situ sulfate- and iron-reducing bacteria (such as *Geobacter* sp.) (Kerin et al. 2006; Benoit et al. 2003; Yu et al. 2012). Methyl-mercury is a neurotoxin that is transferred through the food chain by bioaccumulation, representing a major risk of intoxication in humans via consumption of contaminated fish (Selin 2009). Hence, remobilization of TMs adsorbed in the sediment increases their potential (eco)toxicity impact, creating a threat to the aquatic ecosystem, as well as human health (Fo¨rstner and Wittmann 1981; Schwarzenbach et al. 2006). Therefore, it is essential to improve our knowledge of the effect of TMs on microbial communities in contaminated sediments, as well as the effect of microbial activities on TM immobilization and remobilization.

Various approaches have been taken to study the link between microbial communities and metal pollution in sediments. One approach is the concept of pollution-induced community tolerance (PICT), developed as a sensitive tool to detect minor effects of anthropogenic pollution on biological communities (Blanck et al. 1988; Blanck 2002). A recent study shows that PICT could be evaluated to test the effect of chronic metal pollution on estuarine sediment microbial communities (Ogilvie and Grant 2008). A link between pore water copper concentrations and PICT was observed. However, no correlation between community structure and development of tolerance to metals was found. Another study investigating the community structure in anoxic freshwater sediments along a gradient of contamination with zinc also failed to find a correlation between community structure and various levels of zinc and arsenic (Gough and Stahl 2011). One of the reasons for this is that a broad approach consisting in the evaluation of the whole microbial community may miss more subtle changes in specific functional groups. For example, a study investigating the effect of mercury associated to tailings from the Idrija Mine showed the presence of mercury detoxification genes solely downstream from the discharge into the river with the same name, compared to upstream (Hines et al. 2000). In other naturally enriched environments such as the Rio Tinto in Spain, sulfate-reducing bacteria have been shown to be associated with the attenuation of the effect of acid mine drainage (AMD) thanks to the immobilization of metals in insoluble sulfides (Sa´nchez-Andrea et al. 2012). In other cases, sulfate-reducers have been associated with the production of methyl-mercury in AMD impacted sediments (Batten and Scow 2003; Shipp and Zierenberg 2008). Those examples are not exclusive of aquatic ecosystems as similar targeted effects have been observed in Cu-con-taminated soils (Berg et al. 2012).

Another microbial group that can be of particular interest is the group of endospore-forming bacteria (EFB), belonging to the phylum Firmicutes. Endospores are highly specialized structures that allow bacteria to remain dormant for long periods of time under unfavorable environmental conditions (Errington 2003) and could be an efficient strategy to survive in polluted environments. Recent reports support this hypothesis, especially with respect to trace metals. Field studies of microbial communities present in U-contaminated sites (Anderson et al. 2003; Chang et al. 2001), as well as in drainage water contami-nated with Pb, Cu and Zn (Nakagawa et al. 2002) suggested that EFB are important and might tolerate pol-lution better than non-spore formers. However, in the case of more common and widely spread anthropogenic sources of pollution, such as effluents of urban or industrial wastewater treatment plants, there is a gap of knowledge on the prevalence and diversity of EFB.

The aim of this project is evaluating the abundance and diversity of bacterial communities in lake sediments and the potential influence of TM contamination on these parameters. To achieve this, sediment samples were collected in three areas differentially impacted by the WWTP effluent of the City of Lausanne according to previous studies (Thevenon et al. 2011a, b; P o t ´e et al. 2008; Wildi et al. 2004; Pardos et al. 2004; Loizeau et al. 2004; Bravo et al. 2011). Bacterial abundance and diversity were assessed using the 16S rRNA gene as a universal bacterial molecular marker. High-throughput DNA sequencing was conducted to obtain an overview of the bacterial community composition. This type of analysis allows the recovery not only of the dominant groups, but also rare members of the community. In the sequence analyses, particular emphasis was placed on assessing the prevalence of EFB. Finally, a characterization of the sediments was conducted to infer the influence of chemical factors on the changes in the microbial community composition.

Materials and methods

Study site

Lake Geneva (580 km^2, 8 9 k m ^{3}, maximum depth 310 m), nowadays considered mesotrophic (Dorioz et al. 1998), is one of the largest freshwater reservoirs of Europe and currently provides drinking water for approximately 700,000 inhabitants (Pote´ et al. 2008). In the WWTP of Lausanne, Switzerland, serving one of the largest cities along the shore of Lake Geneva (236,000 population equivalents), 92,400 m^3/day of water were released into the lake, from which 9,900 m^3/day were untreated (release from combined sewage overflow; SESA 2012). The outlet of the WWTP is located within the Vidy Bay at 700 m from the lakeshore at a depth of 35 m. A previous study showed that the highest TMs levels in the sediment of the Vidy Bay were found around the outlet of the WWTP as shown in Fig. 1a (Pote´ et al. 2008). Total mercury con-centrations of up to 8.7 mg/kg (dry weight sediment) were measured in this area (Pote´ et al. 2008), and it has been reported that methylmercury follows a similar trend (Bravo et al. 2011). Mercury values are about 18 times higher than the probable effect levels (PELs) indicated by the Canadian Sediment Quality Guidelines (CCME 1999); and more than four times higher than the limit concentra-tion for relevant sediment contamination, accordingly to the International Commission for the Protection of the Rhine (ICPR) (ICPR 2009).

Sampling

In order to obtain a large variation in particulate TM concentrations, sampling sites were determined on the basis of measurements published previously (Pote´ et al. 2008) (Fig. 1a). Six cores were collected during two sampling campaigns. Two initial cores were collected with the MIR submersible during the sampling campaign of the ELEMO project (July 2011), in a supposedly moderately contaminated area (about 300–400 m from the outlet pipe), which were coded with the letter M (M1 and M2). Four additional cores were retrieved in October 2011 from a boat; two of the cores near the outlet pipe of the WWTP (i.e. supposed as the highest contaminated area), hereafter indicated by the letter N (N1 and N2); and two from an area with supposedly low particulate TM, distal to the outlet pipe (indicated by the letter D as D1 and D2). During both campaigns Plexiglas tubes of 6 cm (diameter) were used as push-cores for the sediments. At the laboratory, the cores were sub-sampled under sterile conditions. If immediate sub sampling was not possible, the cores were stored at 4 °C and the overlaying water was bubbled with air in order to preserve the oxygen gradient at the sediment–water interface. Two top layers were analyzed separately in triplicates: 0–3 cm, indicated as upper; and 3-9 cm, indicated as lower.

Chemical analysis

Water content of the sediment was estimated weighting the sediment before and after freeze-drying. Another fraction of the sample was used for chemical analysis. Prior to the analysis of chemical compounds adsorbed onto the sediments, they were air-dried and agate ground. Total particulate organic carbon ($C_{org\text{-}part}$) and nitrogen (N_{part}) were measured with a CHN-analyzer (CHNEA1108-Elemental analyser, CE Instruments Ltd, Wigan, UK) following a carbonate fumigation exposing the samples to concentrated HCl vapors for 6 h (Harris et al. 2001). After Kjeldahl oxidation, total phosphorus was determined colorimetrically at 880 nm using the molybdate procedure (Murphy and Riley 1962). X-ray fluorescence (XRF) was performed with the Uniquant® method (Thermo Fisher Scientific, Vantaa, Finland) to quantify sulfur. Total particulate TM concentrations were measured by quadrupole-based Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) (HP 4500, Agilent Technologies, Santa Clara, CA, USA) following an aqua regia digestion (Fo¨rstner and Wittmann 1981). The standard ''LKSD-4'' was used as control to validate the metal quantification. Additional measurements of particulate metal concentration were obtained by XRF. These values were used for the extrapolation of particulate metal concentration shown in Fig. 1b, c. Finally, total mercury was quantified by atomic absorption spectrometry (AAS) (Advanced Mercury Analyser; AMA 254, Altec, Dvůr Králové nad Labem, Czech Rep.).

Extraction of cells from sediments

Three grams of wet sediment were mixed with a dispersing solution (1 % Na-hexametaphosphate) in a ratio of 1/5 and homogenized 2 × 1 min at 15,500 rpm using an Ultra-Turrax® homogenizer (T18 basic ULTRA-TURRAX®, IKA, Staufen, Germany). The heaviest mineral particles were removed by sedimentation (10 min) allowing the recovery of the cells contained in the supernatant. The extraction was repeated once more with an equal volume of the dispersion solution. The two supernatants were pooled together and centrifuged at low speed (20 g for 1 min) in

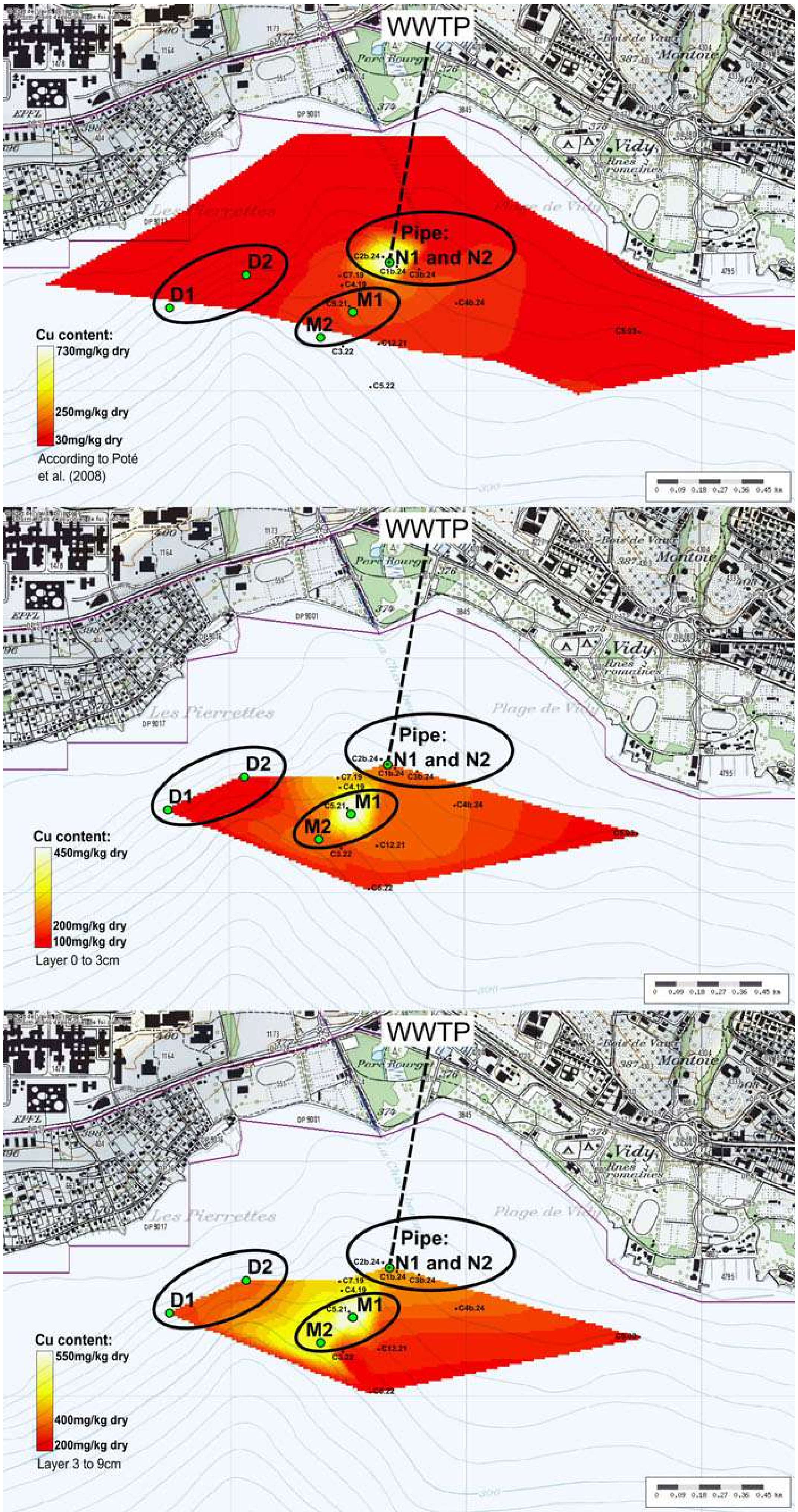


Fig. 1 Map of the Vidy Bay indicating the positions of the cores retrieved with the MIR submersible or with the boat from the F-A Forel Institute (*green dots*). The pipe (*dashed line*) rejecting the wastewater treated by the WWTP from Lausanne is also shown. The cores were retrieved near (N1 and N2) the pipe, or at middle (M1 and M2) and distal (D1 and D2) positions from it. For this study three areas were selected based on previously reported Hg concentrations (Pote´ et al. 2008) as shown in **a**, and are indicated by *black circles*. Two cores per area were considered for the analyses. Current Cu concentrations measured by XRF in the upper layers (**b**) and lower layer (**c**) are presented for comparison with the previously published Hg values (color figure online)

order to remove the remaining sediment particles and to recover a final supernatant containing cells and spores. From this final supernatant, 30 μL were kept to perform fluorescence microscopy, while the rest (about 24 mL) was divided into two equal fractions, and each one was filtered onto two different 0.2 μm pore-size nitrocellulose filters (11407-47-CAN, Sartorius, Goettingen, Germany) and stored at −80 °C. One of the filters was kept as a back-up.

Fluorescence microscopy (FM) and cell counting

The 30 μL of sediment extract (recovered cells from the sediment) were diluted 50× in a saline solution (0.015 % NaCl) containing 2 % formalin for fixation. Then 500 μL of the diluted cell extract were filtered through black 0.2 μm pore-size nitrocellulose filters (GTBP02500, Mil-lipore, Billerica, MA, USA). Filters were placed on microscopy slides with 50 % glycerol and 50 % Tris–EDTA and stained for 15 min with a SYBR® Green I 20× solution (adapted from (Weinbauer et al. 1998). The stained filters were observed at a magnification of 1,000× with an epifluorescence microscope using filter cube L5 (Leica DMR, Leica Microsystems, Wetzlar, Germany). Ten random images were taken per filter and automatic counting of the cells was performed using the program *daime* (Daims et al. 2006).

DNA extraction and precipitation

DNA was extracted using the MP FastDNA® SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA, USA). This kit has been chosen because of its ability to extract DNA from endospores compared to other kits (Dineen et al. 2010). A modified protocol was applied in order to better extract DNA from hard-to-lyse cells and endospores. Briefly, the modified protocol consisted of three consecutive bead-beating steps treated separately following the manufacturer instruction before being pooled together at the end of the procedure. Total DNA was then precipitated from the aqueous phase by adding sodium acetate [final concentration of 0.3 M (pH 7.6)] and 2 volumes of ice-cold pure ethanol. DNA was incubated in ice for 1 h and collected by centrifugation (1 h at 21,460×g and 4 °C). The supernatant was removed and the precipitate washed with 1 mL of cold 70 % ethanol and centrifuged again (30 min at 21,460×g and 4 °C). A maximum of supernatant was removed and the remaining droplets were evaporated by vacuum centrifugation (Univapo 150H, Uniequip, Planegg, Germany) for 5 min. DNA was resuspended in 50 μL of PCR-grade water. Total DNA was quantified before and after ethanol precipitation with Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) to calculate the recovery rate of DNA.

Real-time PCR on the 16S rRNA gene

Real-time PCR on the V3 region of the 16S rRNA gene was performed with a Corbett Rotor-Gene 3000 (Qiagen, Hilden, Germany) using the primers 338f (5′-ACT-CCTACGGGAGGCAGCAG-3′) and 520r (5′-ATTACCG CGGCTGCTGG-3′) (Bakke et al. 2011; Muyzer et al. 1993), which are routinely used to asses bacterial abundance or to carry out profiling of bacterial communities. Reaction mix consisted of DNA template (ranging between 0.5 and 20 ng), 0.3 μM of each primer and 1× Rotor-Gene SYBR Green PCR Master Mix (Qiagen). Total reaction volume of 10 μL was reached with PCR-grade water. The program started with a hold at 95 °C for 15 min to activate the polymerase (HotStarTaq *Plus* DNA Polymerase) followed by 40 cycles of denaturation at 95 °C for 10 s, annealing at 55 °C for 15 s and elongation at 72 °C for 20 s. The number of gene copies was calculated by com-parison with a standard curve using dilutions from 10^8 to 10^2 copies/μL of a plasmid containing the V3 sequence from an environmental clone prepared with the Invitro-gen™ Zero Blunt® TOPO® PCR Cloning kit for sequencing and inserted in competent *E. coli* cells (Invit-rogen™ One Shot® TOP10). Plasmid extraction was performed with the Wizard® Plus SV Miniprep DNA Purification System (Promega, Madison, WI, USA) fol-lowing manufacturer instructions. The DNA concentration was deduced from the quantification with Qubit® 2.0 Fluorometer (Invitrogen) of the final plasmid product.

PCR amplification of the 16S rRNA gene and denaturing gradient gel electrophoresis (DGGE)

The V3 variable region of the 16S rRNA gene was amplified using the primers GC 338f (5′-GC-clamp-ACTCCTACGGGAGGCAGCAG-3’) and 520r (Bakke et al. 2011; Muyzer et al. 1993). PCR reactions were performed in a total volume of 50 μL containing: 1 ng of DNA template, 1× ThermoPol Reaction Buffer (New England Biolabs, Ipswich, MA, USA), 1.25 U of Taq DNA Polymerase (New England Biolabs), 3 mM of $MgCl_2$, 0.25 μM of each primer (Microsynth, Balgach, Switzerland) and 0.2 μM of each dNTP (Promega). PCR amplifications were carried out with a Thermo Scientific Arktik Thermal Cycler (Thermo Fisher Scientific). PCR program consisted of an initial denaturation at 94 °C for 10 min, followed by 10 cycles with denaturation at 94 °C for 30 s, annealing at 65 °C (touchdown of −1 °C/cycle) for 30 s and elongation at 68 °C for 30 s, the following 25 cycles consisted of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s and elongation at 68 °C for 30 s. The amplification was completed by a final elongation at 68 °C for 5 min. PCR products were verified by standard 1.2 % agarose gel electrophoresis. To obtain enough DNA to conduct the DGGE analysis (minimum 600 ng), two PCR products of each sample were pooled together and concentrated by vacuum centrifugation (Univapo 150H, Uniequip) during 45 min.

DGGE was carried out in a Bio-Rad D-Code Universal Detection Mutation system (Bio-Rad Laboratories, Hercules, CA, USA). 8 % acrylamide gels with a 30–60 % gradient of denaturants (with 100 % of denaturants consisting of 40 % (vol/vol) formamide and 7 M urea) (Muyzer et al. 1993) were cast between two glass plates (16 by 16 cm), 1.5 mm spacers and 1 cm-wide loading wells. A minimum of 600 ng of DNA in 20 µL was mixed with 5 µL of loading buffer (60 % sucrose, 0.25 % bromophenol blue and 0.25 % xylencyanol FF) and loaded into the wells (Ferris et al. 1996). Runs were conducted at 60 °C with a constant voltage of 150 V during 4 h 30 min in 1× TAE buffer (0.04 M Tris base, 0.02 M sodium acetate and 1 mM EDTA; pH adjusted to 7.4). Then the gels were stained with 1X SYBR® Gold (Invitrogen) during 30 min and visualized using the GenoPlex system (VWR, Radnor, PA, USA). GelCompar II (Applied Maths, Sint-Martens-Latem, Belgium) and the open source R software, (www.r-project.org/) was used for statistical analysis of the gels. To allow the comparison of different gel images, a ladder containing PCR products from ten strains (50 ng per strain) having different G-C contents and resulting in a pattern of bands distributed throughout the gel, was loaded on each side of the gel.

Pyrosequencing and bioinformatic analysis

High-throughput DNA amplicon sequencing on the V1-V3 variable regions of the 16S rRNA gene was conducted by Eurofins MWG Operon (Ebersberg, Germany) using GS FLX+ technology (454 Life Sciences, Branford, CT, USA). The primers Eub8f (5′-AGAGTTTGATCCTGG CTCAG-3′) and Eub519r (5′-GTATTACCGCGGCTGCT GG-3′) were used. These primers were selected based on recent results indicating that this region is the best performing for profiling and identification of diverse members of bacterial communities (Li et al. 2009). The sequences (about 20,000 per sample) sent by Eurofins MWG Operon, were compared by BLASTN (Altschul et al. 1990) (http://www.ncbi.nih.gov) to a 16S rRNA gene database of 8911 sequences (the most complete database available, containing only good quality sequences of type strains downloaded from the Ribosomal database project; http://rdp.cme.msu.edu/), and containing well-described groups of iron- and sulfate-reducing bacteria that could be expected in metal-contaminated sediments. For each ana-lyzed sequence only the closest hit in the reference database was conserved. In addition the sequences were clustered into putative OTUs (identity of >95 %) and identified with the pick_otus_through_otu_table.py script from the QIIME package (Caporaso et al. 2010).

Statistical analysis

A first set of multivariate exploratory analyses was conducted with the R software and the package *vegan*. Pearson correlations were calculated between chemical elements and/or between bacterial abundance measured by FM and real-time quantitative PCR. Canonical correspondence analysis (CCA) was selected to analyze the relationship between the environmental factors and the composition of the bacterial communities. In this analysis the bacterial groups that constitute the communities of the different samples were constrained with the data of environmental variables. This analysis was selected because it allows establishing a correspondence between changes in abundance and variation of environmental conditions, and takes into consideration rare species, which are obtained during high-throughput DNA sequencing (Ramette 2007). The CCA was carried out following the procedure described in the vegan tutorial. Diversity indexes (Shannon and Evenness) were calculated using R. Finally the chao1 index was used as a statistical estimator of species richness in a rarefaction analysis (Colwell and Coddington 1994). Rarefaction curves (Chao1) were calculated using the script alpha_rarefaction.py in QIIME.

Results

Chemical analysis

Analyses of the 12 samples (six cores and two depths) showed a large variation of the sediment chemical composition (Table 1). For half of the cores (N1, D1, and D2), the upper sediment layer (0–3 cm) contained higher $C_{org\text{-}part}$ than the lower layer (3–9 cm). The highest $C_{org\text{-}part}$ contents were measured in core M1 (both layers), respectively with 7 and 7.72 % (of total dry sediment weight). The lowest $C_{org\text{-}part}$ levels were found in the sediments sampled further away from the outlet pipe (D1 and D2). A similar trend for the N_{part} content was found but with a range varying between 0.17 and 0.65 %. Total particulate phosphorous ranged from 0.08 to 0.38 %, measured in the upper layer of core D1 and the upper layer of core N1, respectively. The highest particulate concentration of sulfur was measured in the upper layer of core N2, with 1.4 % while the lowest concentration was measured in the upper layer of core D2 with 0.43 %. The water content of the sediments varied from 57 to 72 % (Table 1). Particulate manganese showed small variability with values ranging

Table 1 Location of the sampling sites and chemical data of six sediment cores collected near the outlet pipe of the WWTP of Lausanne, in the Vidy Bay, Switzerland

Expected level of HM (according to Poté et al. 2008)	Core	Sampling site locations (Swiss coordinates)		Layer[a] (cm)	ω (%)	C_{org} (%)	N (%)	P (%)	S (%)	Mn (mg/kg)	Fe (g/kg)	Hg (mg/kg)	Cu (mg/kg)	Zn (mg/kg)	Cd (mg/kg)	Pb (mg/kg)	As[b] (mg/kg)	Al[b] (g/kg)
		X (m)	Y (m)															
Low	D1	533,735	151,351	0–3	69.61	2.67	0.23	0.08	0.48	408	18.2	0.24	75	170	0.74	43	8	16.3
				3–9	62.98	2.64	0.21	0.11	0.76	387	19.9	0.84	114	322	1.90	124	10	13.9
	D2	534,063	151,491	0–3	69.64	2.78	0.23	0.10	0.43	405	24.8	0.27	81	205	0.81	53	9	15.3
				3–9	57.14	2.36	0.17	0.29	0.94	396	22.5	0.80	137	542	2.76	154	9	9.2
Medium	M1	534,515	151,335	0–3	71.67	7.00	0.53	0.24	1.19	390	43.5	*7.65*	*292*	*1,032*	*6.27*	*311*	*15*	*18.6*
				3–9	73.10	7.72	0.65	0.17	0.98	413	47.8	*8.98*	*362*	*1,564*	*10.13*	*498*	*22*	*20.8*
	M2	534,379	151,229	0–3	71.75	4.01	0.38	0.14	0.69	415	33.2	0.87	165	399	1.66	100	13	17.7
				3–9	64.84	5.08	0.40	0.34	0.86	400	34.3	*1.25*	*293*	*974*	*4.38*	*274*	*16*	*16.3*
High	N1	534,676	151,543	0–3	79.27	5.96	0.62	0.38	ND[c]	392	40.0	1.42	155	416	1.88	115	13	13.4
				3–9	66.32	3.79	0.31	0.23	1.20	443	33.9	1.15	123	349	1.76	97	11	16.2
	N2	534,676	151,543	0–3	75.38	4.69	0.46	0.32	1.40	369	35.8	1.30	129	346	1.59	127	11	13.0
				3–9	71.35	6.30	0.58	0.20	1.19	401	36.0	*2.85*	*341*	*902*	*8.06*	*294*	*22*	*21.4*

In each core two independent depths (upper 0–3 cm; lower 3–9 cm, column layer) were analyzed. *ω* water content. All values measured in dry sediments. Metal content measured in mg/kg, except for Fe and Al. *ND* not determined. Samples containing elevated heavy and trace metal concentrations (Hg, Cu, Zn, Cd, Pb, As, and Al) are indicated in italic

[a] Layer analyzed (in centimeter from the top of the core)

[b] Values of As and Al only considered as indicatives (quantification not accurate)

[c] Not determined due to a lack of sediment remaining for the analysis

from 369 to 443 mg/kg. The maximum particulate iron content was found in the lower layer of core M1 and the minimum in the upper layer of core D1, with 4.8 and 1.8 g/kg, respectively.

A general overview of TM_{part} concentration obtained by XRF on 11 cores (including the 6 cores of the present study) showed that for the upper layer, the highest values are not found in the vicinity of the outlet pipe (comparison between Fig 1a, b). For the lower layer, the TM_{part} con-taminated area is more diffuse and included sediment cores in close vicinity to the outlet pipe (N2) as well as those retrieved in the area supposedly containing intermediate TM_{part} concentrations (M1 and M2). Accordingly, samples that contained elevated TM_{part} concentrations measured by ICP-MS corresponded to those areas identified by XRF. Four samples contained relatively elevated TMs concen-trations (Table 1): the lower layer of the cores N2 and M2 and both layers of the core M1. The highest levels of each TM_{part} were found in the lower layer of core M1, with 8.98 mg/kg of Hg, 362 mg/kg of Cu, 1,564 mg/kg of Zn, 10.13 mg/kg of Cd, and 498 mg/kg of Pb. The lowest TM_{part} contents were measured in the upper layer of the cores D1 and D2, both sampled in the distal area, sup-posedly less influenced by the outlet pipe of the WWTP. Measurements of particulate As and Al were only consid-ered as indicative because the measurement of the standard

Table 2 Bacterial abundance and diversity in sediment cores of the Vidy Bay

Core	Layer (cm)	Fluorescence microscopy (cells/g)[a]	Real-time PCR (copies of gene/g)[a]	Shannon index (H′)	Evenness (J)
D1	0–3	1.3×10^9	3.3×10^9	5.62	0.77
	3–9	2.2×10^9	4.3×10^8	5.36	0.74
D2	0–3	1.2×10^9	1.5×10^9	5.48	0.76
	3–9	1.3×10^9	6.2×10^7	5.31	0.74
M1	0–3	6.9×10^8	3.2×10^8	4.57	0.64
	3–9	8.0×10^8	5.3×10^8	4.25	0.61
M2	0–3	5.1×10^8	1.7×10^9	5.1	0.72
	3–9	7.1×10^8	2.8×10^8	4.25	0.61
N1	0–3	2.6×10^9	1.7×10^{10}	4.69	0.66
	3–9	1.1×10^9	6.1×10^8	5.45	0.77
N2	0–3	3.2×10^9	1.2×10^{10}	5.08	0.71
	3–9	1.5×10^9	4.7×10^8	4.39	0.61

Shannon index and evenness were strongly correlated ($R^2 > 0.98$)

[a] Cells and copies of gene measured per gram of wet sediment

LKSD-4 was not accurate for these two elements. Pearson correlation matrix (Supplementary Fig. 1) revealed signif-icant positive correlations between $C_{org\text{-}part}$, N_{part}, Fe_{part} and all TM_{part} concentrations. The S_{part} concentration was also positively correlated with P_{part} and Fe_{part} content.

Quantification of bacterial communities

Quantification of bacterial abundance resulted on average in 1.4×10^9 cells (fluorescent microscopy, FM-) and 3.2×10^9 16S rRNA gene copies (real-time quantitative PCR) per gram of wet sediment. This represented a mean of 2.3 copies of 16S rRNA gene per cell. Both analyses revealed the largest abundance of bacteria in the upper layer of cores N1 and N2 sampled near the outlet pipe of the WWTP, with 2.6 and 3.2×10^9 cells/g counted by FM, and 1.7×10^{10} and 1.2×10^9 copies of gene/g measured by real-time PCR, respectively (Table 2). The lowest abundance of cells was found in both layers of cores M1 and M2, whereas the lowest gene copy numbers were measured in the lower layer of core D2. A significant correlation (p value of 0.05; Supplementary Fig. 1) was found between the two methods used to measure bacterial abundance.

Bacterial community fingerprinting

Fingerprinting analysis by DGGE showed high reproducibility between triplicates of each sample (data not shown). Thus, in order to compare the bacterial community composition between the 12 samples, a gel containing the three replicates pooled together was generated (Fig. 2). This allowed an improved analysis by avoiding differences between gels (e.g. migration or staining). All profiles showed a large number of bands (more than 70 bands per lane), attesting an important bacterial diversity in both layers and in each core. Fingerprints clearly differed between the cores corresponding to the three areas of sampling and between the two layers of a same core. A similar band pattern was observed in cores D1 and D2 (both layers), upper layer of cores N1 and N2, and lower layer of cores M1 and M2. In contrast, this was not the case for fingerprints corresponding to the upper layer of cores from the intermediary area (M1 and M2) and to the lower layer of cores sampled near the outlet pipe (N1 and N2). A pattern of two strong bands located in the region of high denaturant concentration (black squares in Fig. 2; lane 4 of layer 0–3 cm and lanes 3, 4 and 6 of layer 3–9 cm) was detected in the four samples of highest TM_{part} levels (M1upper, M1lower, M2lower and N2lower).

Pyrosequencing and bioinformatics analyses

Close to 22,000 sequences per sample were obtained by pyrosequencing. The sequences matched 3,388 different type strains (out of 8,911) from the reference database in a BLAST search (data not shown). The eight most abundant species identified in each sample are presented in Supplementary Fig. 2. Sequences were assigned to an OTU at the species level with >97 % identity, at the genus level with >95 % identity and at the phylum level with >80 %identity (Schloss and Handelsman 2005; Stackebrandt and Goebel 1994). In the four samples containing the highest TM_{part} levels (M1upper, M1lower, M2lower and N2lower) the most abundant OTU found was identified at the species level as the endospore-forming bacterium *Clostridium lituseburense*, which represented 19–21 % of the total of sequences. Two other endospore-forming bacteria, *Clostridium bartletti* and *Clostridium disporicum,* corresponded to the second or third most abundant species detected in TM-contaminated sediments. All together *Clostridium* spp. represented up to 49 % of the total sequences in the four samples with high TM_{part} content, but reached a maximum of 10 % in the other sediments. *C. lituseburense* was also found among the eight most abundant species identified (except for D1upper and D2upper) in other samples, but represented about 4 % of the community.

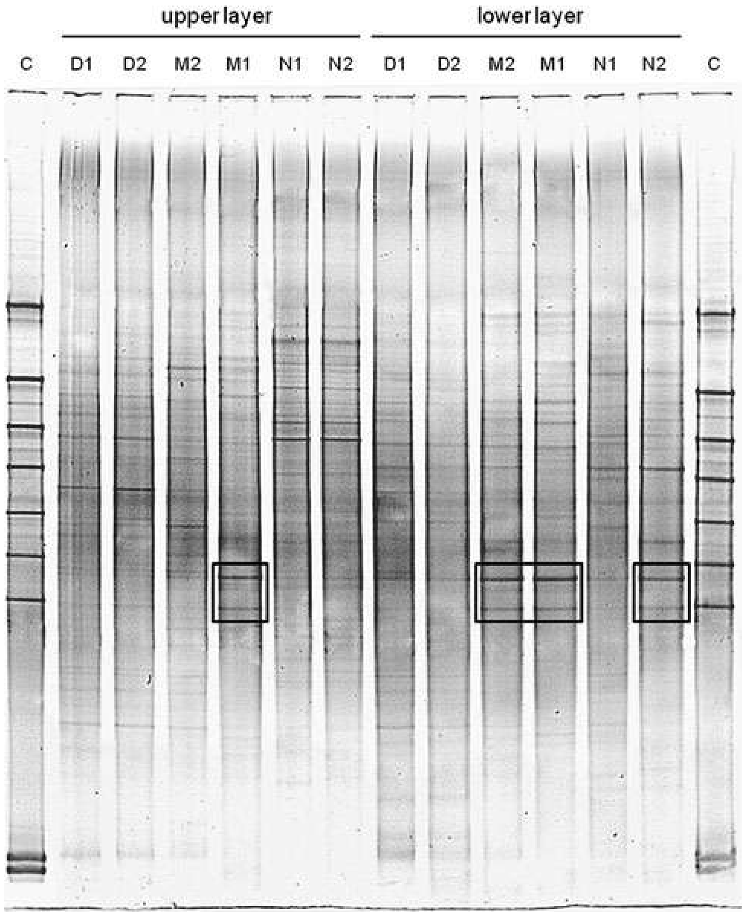


Fig. 2 DGGE profiles for the V_3 region of the 16S rRNA gene of the bacterial communities in sediment cores collected in the Vidy Bay. For labeling of the samples see Fig. 1. The two *lanes* C correspond to the ladder. A pattern composed by two strong bands located in the region with high denaturant concentration was found in the lanes corresponding to sediments with high particulate trace metal content (Table 1) and is indicated by *black squares*

In the upper layer of the two cores sampled near the outlet pipe, the dominant species was *Hydrogenophaga caeni* (identified at the species level), representing 22 % and 14 % of the sequences in samples N1upper and

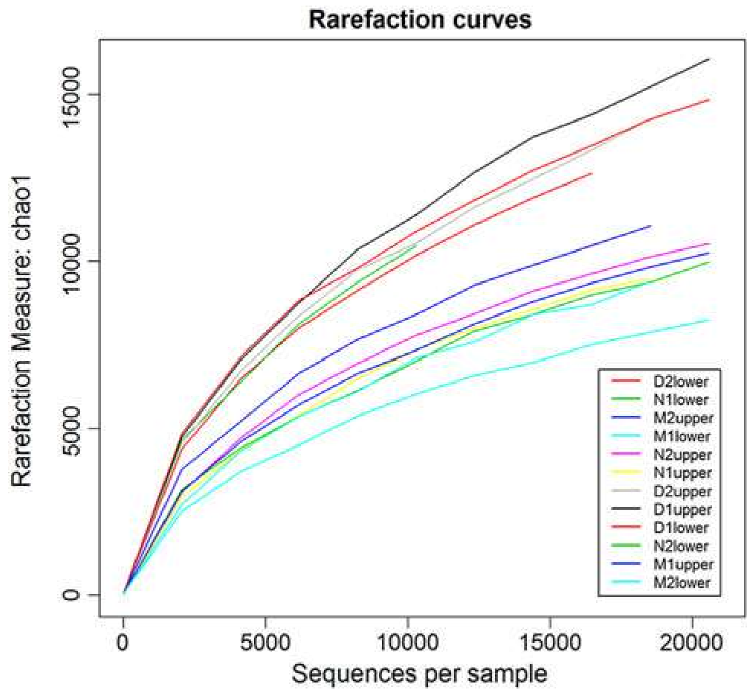


Fig. 3 Rarefaction curves for the 12 samples showing the diversity detected compared with the predicted total diversity. The x axis represents the number of sequences sampled while the y axis represents a measures of the species richness detected, estimated with the Chao1 index. The *legend* on the *bottom right* shows the correspondence between the curves and the samples. For labeling of the samples see Fig. 1

N2upper, respectively. In the same samples the second most abundant species assigned (6 and 7 % of the total sequences) was *Acidovorax defluvii*. In sample N1lower the two most abundant assigned OTUs (over 5 % of the sequences) were related to the phylum Chloroflexi and *C. lituseburense.* For both layers of core D1 and in the upper layer of core D2 retrieved from the distal area, the majority of the sequences were affiliated to the (per)chlorate-reducing bacteria related to *Dechloromonas* spp. The highest proportion of the sequences (6 %) from lower layer of the sample D2 was most closely related to the Gammaproteobacteria. In the upper layer of the core M2, which did not show a high TM_{part} concentration (Table 1), the two most abundant assigned OTUs were the Fe(III)-reducing *Geobacter* spp. (Nevin et al. 2005) and *D. hort-ensis*, with 9 and 6 % of total sequences, respectively.

Based on the pyrosequencing data, the Shannon index (H′) and evenness (J) (Table 2) were calculated to obtain a global overview of the diversity of bacterial communities in the sediment samples. The highest diversity was found in the upper layer of cores D1 and D2, and in the lower layer in core N1. The lowest values were found in the lower layer of the cores M1, M2 and N2, i.e. sediment core layers characterized by the highest TM_{part} concentrations (Table 1). The logarithmic shapes of the rarefaction curves obtained for the sequence data (Fig. 3) indicated that most of the species richness occurring in the sediments was detected. Indeed as the identification of sequences progressed (x axis), the detection rate for new species leveled off (y axis).

Canonical correspondence analysis (CCA)

An ordination plot of the samples was constructed using the results of pyrosequencing after OTU picking and identification of species using QIIME (Fig. 4). Three principal groups of samples were observed. The groups corresponded to the clustering based on the DGGE analysis (data not shown). Based on the community composition the four samples with high TM_{part} levels (M1upper, M1lower, M2lower and N2lower) formed a group in the bottom right part of the plot. Accordingly, the vectors Cd, Pb, Cu, Hg, and As are the environmental factors that corresponded more significantly the community composition in this group of samples. This group is also characterized by the prevalence of endospores-forming bacteria. As previously outlined (Supplementary Fig. 2) the prevalence of endospore-forming bacteria belonging to the Firmicutes (mainly Clostridiaceae, *Clostridium* spp., *Sarcina* spp., *Syntrophomonas* spp.) in these samples is clearly highlighted in the CCA.

The samples from the upper layer of the cores N1 and N2, near the outlet pipe of the WWTP, formed a second group (in the top right part of the plot). The community composition was characterized by non-endospore-forming Firmicutes (e.g. *Streptococcus* spp.) and non-Firmicutes (e.g. *Acidovorax* spp. and Comamonadaceae, corresponding to the identification of *H. caeni* by BLAST in Supplementary Fig. 2. However, endospore-forming bacteria were not enriched in these samples.

The remaining samples (D1upper, D1lower, D2upper, D2lower, M2upper and N1lower) represented a third group anti-correlated with all environmental factors, except Mn, but correlated with a large diversity of bacteria phyla such as Betaproteobacteria (*Dechloromonas* spp, *Thiobacillus* spp, and *Desulfobacterium* spp.), Gammaproteobacteria, and Bacteroidetes. However no Firmicutes, including endospore-forming bacteria, were enriched in this group.

Discussion

The sampling sites in this study were selected with the objective to retrieve sediments differently impacted by the TMs released with the effluent from a WWTP. In the model system studied here, the WWTP of the City of Lausanne, previous studies suggested that the release of treated wastewater was correlated with high TM concen-tration in the sediments clustered around the outlet pipe (Thevenon et al. 2011a, b; P o t ´e et al. 2008; Wildi et al. 2004; Pardos et al. 2004; Loizeau et al. 2004; Bravo et al.

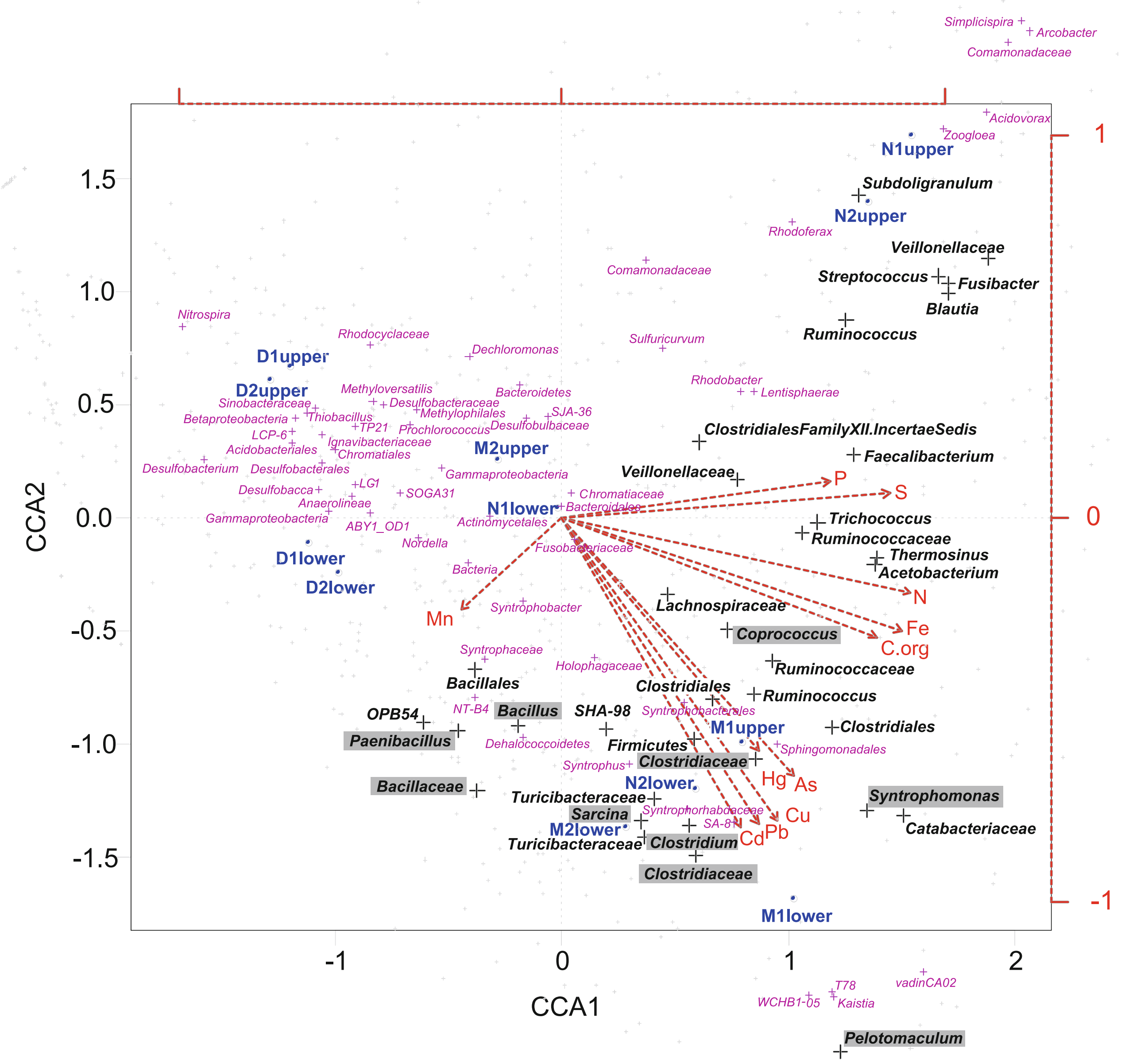


Fig. 4 Canonical correspondence analysis (CCA) triplot of the samples (*blue text* and *dots*) based on the bacterial groups identified in the sediment of the Vidy Bay and constrained by environmental variables. The ordination plot of the samples was constructed with results of OTUs picking and identification with the program QIIME from pyrosequencing data. Each OTU is represented by a cross in the plot. OTUs from sequences corresponding to less than 1 % (non Firmicutes) or 0.1 % (Firmicutes) of the community are shown in *light grey* crosses. OTUs representing non-Firmicutes are shown in *purple* and with *letters* in a *small font size*. OTUs corresponding to Firmicutes are shown in *black* and with a *larger font size*. Those assigned to endospore-forming Firmicutes are highlighted in *grey*. Additionally the centroids corresponding to the different environmental variables are represented by *dashed red arrows* (color figure online)

2011). However, this was not observed in the present study, in which contaminated sediments were also found relatively far from the outlet pipe. In addition, TM_{part} concentrations measured near the outlet pipe were much lower compared to those previously published (Pote´ et al. 2008), but similar to values published more recently (Haller et al. 2011), both reporting measurements from sediments collected in 2005, which corresponded to 16.8 cm deep according to a sedimentation rate of 2.8 cm/year (Loizeau et al. 2003). A reason for the differences observed could be that the deposition of organic matter and TM close to the outlet pipe is highly heterogeneous (Masson and Tercier-Waeber 2013). This is supported by the chemical characterization of cores N1 and N2, which were obtained from the area near the outlet pipe. For example, in core N1, a relatively low level of TM_{part} was detected for both layers. On the contrary, TM_{part} concentrations in the two layers of the core N2 were different, with low TM_{part} values for the upper layer and very high values for the lower layer. Probably the most surprising result was that the other samples with the highest TM_{part} levels (M1upper, M1lower and M2lower) were not retrieved near the outlet pipe, but rather in the intermediate area (Fig. 1b, c), which was supposedly less influenced by the direct release of treated wastewater. This could be directly linked to the transport of TM_{part} associated to sediment particles or colloids with different size, which will influence sedimentation rate. However this cannot be evaluated based on the results since only bulk measurements were made. In addition, another element that needs to be considered in the future is the TM_{part} speciation linked to redox gradients or different mineral or organic fractions in the sediments. This is important since other studies have shown microzonation and redox gradients occurring at a millimeter scale can influence the avail-ability of metals, as well as microbial communities or the biogeochemical processes affected by those (Huerta-Diaz et al. 2012).

A correlation was observed between $C_{org-part}$, $N_{org-part}$ and TM concentrations. In all the TM-contaminated sediments an increased $C_{org-part}$ and $N_{org-part}$ content was observed. This could be explained by an inhibitor effect of TM on microbial activity, as previously suggested for contaminated soils (Giller et al. 1998). Indeed, even low TM concentrations in soil have been shown to lead to the loss of functional diversity, reducing the diversity of met-abolic pathways and impacting mineralization and nutrient cycling (Kandeler et al. 1996). Although a decrease in abundance and bacterial diversity (Table 2) was observed for samples with high levels of TM_{part}, the correlation with TMs was not significant (Supplementary Fig. 1). Alterna-tively the high concentration of $C_{org-part}$, $N_{org-part}$ and TM, as well as the absence of a correlation of the later with bacterial diversity can be the result of processes of trans-port and speciation associated with the deposition and accumulation of sediments derived from the treated wastewater, which is enriched in organic C, N, and TM, and that were not considered in this study, but have been analyzed elsewhere (Gascon Diez et al. 2013).

However, the release of treated wastewater clearly influenced bacterial abundance, which was highest in the upper layer of sediments retrieved near the outlet pipe, similarly to findings from a previous report (Thevenon et al. 2011b). Along with the WWTP effluent, not only large quantities of nutrients but also bacteria are directly released and probably deposited in the nearby sediments. This hypothesis is supported by the results of pyrosequencing revealing that *A. defluvii* and *H. caeni*, both typically found in activated sludge (Chung et al. 2007; Schulze et al. 1999), are the most abundant species in these samples.

A clear correlation was observed between endospore-forming bacteria and high levels of total particulate TMs (Fig. 4). Indeed, endospore-forming bacteria related to *Clostridium* spp., belonging to the Firmicutes, were detected in large proportions (representing close to 50 % of the total bacterial community) in samples with high TM_{part} contents. The species identified as the closest relatives to the sequences from the sediments corresponded to bacterial species previously linked to wastewater contamination. For example, *C. lituseburense*, which represents up to 21 % of the communities and was the most dominant OTU in TM contaminated sediments, was previously found in human feces (Franks et al. 1998) or abundant in wastewater effluents from dairy farms (McGarvey et al. 2004). Other two dominant species, *C. bartlettii* and *C. disporicum*, were also detected in human (Song et al. 2004; Mangin et al. 2004) or animal feces (Kobayashi et al. 2011; Horn 1987). Similar to *H. caeni* and *A. defluvii*, these three *Clostridium* species were probably released with the WWTP effluent, but contrary to the first two species, they seemed to be favored in the highly TM contaminated sediments. This was clearly demonstrated when comparing the cores N1 and N2. Although bacterial communities were very similar in the upper layer of both cores, as revealed by the DGGE analysis (Fig. 2) and the CCA (Fig. 4), they were com-pletely different in the lower layer. For the low contaminated sample (N1lower) even though *C. litusebu-rense* was detected in small proportions, no bacteria seemed to dominate as indicated by a relatively high evenness (Table 2). In the sample with high TM_{part} levels (N2lower), *Clostridium* spp., and most particularly *C. lit-useburense* (Supplementary Fig. 3), clearly prevailed and the functional diversity was arguably lower. The domi-nance of bacteria known for a fermentative metabolism could represent a reduction of OM degradation and a blockage of mineralization at the level of fermentation products, which could explain the accumulation of $C_{org-part}$ and $N_{org-part}$ observed here. The same interpretation is valid for the other highly TM contaminated sediments (M1upper, M1lower and M2lower).

Even though the dominant Clostridia species originate at the WWTP, it is still unclear if the correlation between high TM_{part} concentrations and dominance of these endospore-forming Firmicutes simply represent variations in the functioning of the WWTP. For example, one explanation could be that these bacteria are dominant in the microbial community at the WWTP during phases in which TM values are higher. However, the dominance of Firmicutes in wastewater treatment plants has not been reported in recent analyses of microbial communities by high throughput sequencing (Hu et al. 2012; Xia et al. 2010). This will favor a second hypothesis, better supported by the results, involving a selection mechanism driven by high load of TMs and leading to the dominance of Clostridia.

However an open question still remaining is why do TMs accumulate in certain areas. As filaments of *Beggia-toa sp.* were observed during the sampling campaign at the surface of sediments corresponding to the core M1 (con-taining high TM_{part} levels), and as this genus is known for sulfur-oxidation (Hagen and Nelson 1997), a role of sulfur in the adsorption of TM onto the solid sediment in these underlying layers was suspected. Sulfur was quantified in order to assess the plausibility of this explanation. Even if a correlation was found between sediment total particulate iron and sulfur (Supplementary Fig. 1), possibly through the formation of pyrite (FeS_2) or mackinawite (FeS), no correlation was found between total particulate sulfur and TM, suggesting another mechanism for metal accumulation on the sediment.

In conclusion, sediments in the Vidy Bay were impacted by the WWTP effluent. The highest TM concentrations measured were 2 (Cu) to 18 (Hg) times higher than the probable effect levels (PELs) indicated by the Canadian Sediment Quality Guidelines (CCME 1999). These sediments are also significantly contaminated, accordingly to ICPR (2009) and therefore still represent a threat for aquatic organisms, as well as humans. The strong focus put on the methodology used to improve the recovery of DNA from sediments allowed a deeper characterization of the spatial distribution of the bacterial communities found in the surface sediments. In particular, while Proteobacteria and Bacter-oidetes have been reported in the past as the dominant bacterial groups found in the Vidy Bay (Haller et al. 2011), a large proportion of endospore-forming Firmicutes (*Clos-tridium* spp.) was found in the present study. *C. lituseburense, C. bartlettii* and *C. disporicum* are all typically found in human feces, suggesting that they are probably released by the WWTP sewage effluent together with the TMs. This is supported by the strong correlation observed between endospore-forming bacteria and high TM_{part} con-centrations (Fig. 4). However, the understanding of the underlying mechanisms explaining the relationship between endospore-forming bacteria and TM sorption on sediments requires further research. The fact that other bacterial groups associated to the WWTP effluent are not correlated with increased TM_{part} levels points out to a potentially active role of endospores in TM transport and deposition on sediments. Therefore endospore-forming bacteria can be a missing biological link to understand the role of sediment as sink or source of TM contamination. However, other alternatives including factors such as the spatiotemporal variation of the microbiological and chemical composition of the treated effluent, or processes controlling microbial communities such as predation, need to be explored in the future.

Acknowledgments This publication is part of the international, interdisciplinary research project ELEMO (http://www.elemo.ch) t o investigate the deep-waters of Lake Geneva using two Russian MIR submarines. Funding for this study was provided by the Fondation pour l'Etude des Eaux du Le´man (FEEL). Additional funding for the work described in this paper was provided by the Swiss National Science Foundation (Grant 31003A_132358). We are grateful for the support. We thank the Russian MIR crew members (http://www.elemo.ch/mir-team) for their excellent performance and the SA-GRAVE team who provided and operated the platform from which the dives were carried out. We also thank Ulrich Lemmin and Jean-Denis Bourquin for project coordination. The service of Mikhail Kranoperov (Russian Honorary Consulate) as liaison is greatly appreciated. Special thanks to Jean-Claude Lavanchy from the Centre of Mineral Analysis (CAM), University of Lausanne, for helping with elemental analysis, and Radu Slobodeanu, from the University of Neuchâtel, for helping on statistical analysis. The authors thank two anonymous reviewers for the valuable comments to improve this manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410

Anderson RT, Vrionis HA, Ortiz-Bernad I, Resch CT, Long PE, Dayvault R, Karp K, Marutzky S, Metzler DR, Peacock A, White DC, Lowe M, Lovley DR (2003) Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer. Appl Environ Microbiol 69(10):5884–5891

Bakke I, De Schryver P, Boon N, Vadstein O (2011) PCR-based community structure studies of Bacteria associated with eukaryotic organisms: a simple PCR strategy to avoid co-amplification of eukaryotic DNA. J Microbiol Methods 84(2):349–351

Batten KM, Scow KM (2003) Sediment microbial community composition and methylmercury pollution at four mercury mine-impacted sites. Microb Ecol 46(4):429–441

Benoit JM, Gilmour CC, Heyes A, Mason RP, Miller CL (2003) Geochemical and biological controls over methylmercury production and degradation in aquatic ecosystems. Biogeochem Environ Important Trace Elem 835:262–297

Berg J, Brandt KK, Al-Soud WA, Holm PE, Hansen LH, Sørensen SJ, Nybroe O (2012) Selection for Cu-tolerant bacterial communities with altered composition, but unaltered richness, via long-term Cu exposure. Appl Environ Microbiol 78(20):7438–7446

Blanck H (2002) A critical review of procedures and approaches used for assessing pollution-induced community tolerance (PICT) in biotic communities. Hum Ecol Risk Assess 8(5):1003–1034

Blanck H, Wangberg SA, Molander S (1988) Pollution-induced community tolerance—a new ecotoxicological tool. In: Cairns J, Pratt JR (eds) Functional testing of aquatic biota for estimating hazards of chemicals, ASTM STP 988. American Society for Testing and Materials, Philadelphia, pp 219–230

Bravo AG, Bouchet S, Amouroux D, Pote J, Dominik J (2011) Distribution of mercury and organic matter in particle-size classes in sediments contaminated by a waste water treatment plant: Vidy Bay, Lake Geneva, Switzerland. J Environ Monit 13(4):974–982

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336

CCME (1999) Canadian sediment quality guidelines for the protection of aquatic life. Reprinted in Canadian environmental quality guidelines, Chapter 6. Canadian Council of Ministers of the Environment, Winnipeg.

Chang YJ, Peacock AD, Long PE, Stephen JR, McKinley JP, Macnaughton SJ, Hussain AK, Saxton AM, White DC (2001) Diversity and characterization of sulfate-reducing bacteria in groundwater at a uranium mill tailings site. Appl Environ Microbiol 67(7):3149–3160

Chung BS, Ryu SH, Park M, Jeon Y, Chung YR, Jeon CO (2007) *Hydrogenophaga caeni* sp. nov., isolated from activated sludge. Int J Syst Evol Microbiol 57:1126–1130

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc Lond B Biol Sci 345(1311):101–118

Daims H, Lucker S, Wagner M (2006) daime, a novel image analysis program for microbial ecology and biofilm research. Environ Microbiol 8(2):200–213

Dineen SM, Aranda R, Anders DL, Robertson JM (2010) An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. J Appl Microbiol 109(6):1886–1896

Dorioz JM, Pelletier JP, Benoit P (1998) Physico-chemical properties and bioavailability of particulate phosphorus of various origin in a watershed of Lake Geneva (France). Water Res 32(2):275–286

Errington J (2003) Regulation of endospore formation in *Bacillus subtilis*. Nat Rev Microbiol 1(2):117–126

Ferris MJ, Muyzer G, Ward DM (1996) Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl Environ Microbiol 62(2):340–346

Förstner U, Wittmann GTW (1981) Metal pollution in the aquatic environment, 2nd rev. edn. Springer-Verlag, New York, Heidelberg, Berlin

Franks AH, Harmsen HJM, Raangs GC, Jansen GJ, Schut F, Welling GW (1998) Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. Appl Environ Microbiol 64(9):3336–3345

Gascon Diez E, Bravo AG, a Porta N, Masson M, Graham ND, Stoll S, Akhtman Y, Amouroux D, Loizeau JL (2013) Mercury content and speciation related to sediment surface patterns in contaminated Vidy Bay, Lake Geneva, Switzerland. Aquat Sci (this issue)

Giller KE, Witter E, McGrath SP (1998) Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. Soil Biol Biochem 30(10–11):1389–1414

Gough HL, Stahl DA (2011) Microbial community structures in anoxic freshwater lake sediment along a metal contamination gradient. ISME J 5(3):543–558

Hagen KD, Nelson DC (1997) Use of reduced sulfur compounds by *Beggiatoa* spp.: enzymology and physiology of marine and freshwater strains in homogeneous and gradient cultures. Appl Environ Microbiol 63(10):3957–3964

Haller L, Tonolla M, Zopfi J, Peduzzi R, Wildi W, Pote J (2011) Composition of bacterial and archaeal communities in freshwater sediments with different contamination levels (Lake Geneva, Switzerland). Water Res 45(3):1213–1228

Harris D, Horwath WR, van Kessel C (2001) Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. Soil Sci Soc Am J 65(6):1853–1856

Hines ME, Horvat M, Faganeli J, Bonzongo JC, Barkay T, Major EB, Scott KJ, Bailey EA, Warwick JJ, Lyons WB (2000) Mercury biogeochemistry in the Idrija river, Slovenia, from above the mine into the Gulf of Trieste. Environ Res 83(2):129–139

Horn N (1987) *Clostridium disporicum* sp-nov, a saccharolytic species able to form 2 spores per cell, isolated from a rat cecum. Int J Syst Bacteriol 37(4):398–401

Hu M, Wang X, Wen X, Xia Y (2012) Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. Bioresour Technol 117:72–79

Huerta-Diaz MA, Delgadillo-Hinojosa F, Siqueiros-Valencia A, Valdivieso-Ojeda J, Reimer JJ, Segovia-Zavala JA (2012) Millimeter-scale resolution of trace metal distributions in microbial mats from a hypersaline environment in Baja California, Mexico. Geobiology 10(6):531–547

ICPR (2009) Sediment management plan Rhine: summary. Report No. 175, International Commission for the Protection of the Rhine (ICPR), Koblenz

Kandeler E, Kampichler C, Horak O (1996) Influence of heavy metals on the functional diversity of soil microbial communities. Biol Fertil Soils 23(3):299–306

Kerin EJ, Gilmour CC, Roden E, Suzuki MT, Coates JD, Mason RP (2006) Mercury methylation by dissimilatory iron-reducing bacteria. Appl Environ Microbiol 72(12):7919–7921

Kobayashi Y, Itoh A, Miyawaki K, Koike S, Iwabuchi O, Iimura Y, Kobashi Y, Kawashima T, Wakamatsu J, Hattori A, Murakami H, Morimatsu F, Nakaebisu T, Hishinuma T (2011) Effect of liquid whey feeding on fecal microbiota of mature and growing pigs. Anim Sci J 82(4):607–615

Li H, Zhang Y, Li DS, Xu H, Chen GX, Zhang CG (2009) Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. Soil Biol Biochem 41(5):954–968

Loizeau J-L, Rozé S, Peytremann C, Monna F, Dominik J (2003) Mapping sediment accumulation rate by using volume magnetic susceptibility core correlation in a contaminated bay (Lake Geneva, Switzerland). Eclogae Geol Helv 96(Supplement 1):S73–S79

Loizeau J-L, Pardos M, Monna F, Peytremann C, Haller L, Dominik J (2004) The impact of a sewage treatment plant's effluent on sediment quality in a small bay in Lake Geneva (Switzerland–France). Part 2: temporal evolution of heavy metals. Lakes Reserv: Res Manag 9:53–63

Mangin I, Bonnet R, Seksik P, Rigottier-Gois L, Sutren M, Bouhnik Y, Neut C, Collins MD, Colombel JF, Marteau P, Dore J (2004) Molecular inventory of faecal microflora in patients with Crohn's disease. FEMS Microbiol Ecol 50(1):25–36

Masson M, Tercier-Waeber ML (2013) Trace metal speciation at the sediment-water interface of the Vidy Bay: influence of contrasting sediment characteristics. Aquat Sci (this issue)

McGarvey JA, Miller WG, Sanchez S, Stanker L (2004) Identification of bacterial populations in dairy wastewaters by use of 16S rRNA gene sequences and other genetic markers. Appl Environ Microbiol 70(7):4267–4275

Murphy J, Riley JP (1962) A modified single solution method for determination of phosphate in natural waters. Anal Chim Acta 27(1):31–36

Muyzer G, Dewaal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16 s rRNA. Appl Environ Microbiol 59(3):695–700

Nakagawa T, Hanada S, Maruyama A, Marumo K, Urabe T, Fukui M (2002) Distribution and diversity of thermophilic sulfate-reducing bacteria within a Cu-Pb-Zn mine (Toyoha, Japan). FEMS Microbiol Ecol 41(3):199–209

Nealson KH (1997) Sediment bacteria: who's there, what are they doing, and what's new? Annu Rev Earth Planet Sci 25:403–434

Nevin KP, Holmes DE, Woodard TL, Hinlein ES, Ostendorf DW, Lovley DR (2005) *Geobacter bemidjiensis* sp nov and *Geobacter psychrophilus* sp nov., two novel Fe(III)-reducing subsurface isolates. Int J Syst Evol Microbiol 55:1667–1674

Ogilvie LA, Grant A (2008) Linking pollution induced community tolerance (PICT) and microbial community structure in chronically metal polluted estuarine sediments. Mar Environ Res 65(2):187–198

Pardos M, Benninghoff C, Alencastro LP, Wildi W (2004) The impact of a sewage treatment plant's effluent on sediment quality in a small bay in Lake Geneva (Switzerland–France). Part 1:

spatial distribution of contaminants and the potential for biological impacts. Lakes Reserv: Res Manage 9:41–52

Poté J, Haller L, Loizeau J-L, Garcia Bravo A, Sastre V, Wildi W (2008) Effects of a sewage treatment plant outlet pipe extension on the distribution of contaminants in the sediments of the Bay of Vidy, Lake Geneva, Switzerland. Bioresour Technol 99(15):7122–7131

Ramette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 62(2):142–160

Sánchez-Andrea I, Knittel K, Amann R, Amils R, Sanz JL (2012) Quantification of Tinto river sediment microbial communities: importance of sulfate-reducing bacteria and their role in attenuating acid mine drainage. Appl Environ Microbiol 78(13):4638–4645

Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71(3):1501–1506

Schulze R, Spring S, Amann R, Huber I, Ludwig W, Schleifer KH, Kampfer P (1999) Genotypic diversity of *Acidovorax* strains isolated from activated sludge and description of *Acidovorax defluvii* sp nov. Syst Appl Microbiol 22(2):205–214

Schwarzenbach RP, Escher BI, Fenner K, Hofstetter TB, Johnson CA, von Gunten U, Wehrli B (2006) The challenge of micropollutants in aquatic systems. Science 313(5790):1072–1077

Selin NE (2009) Global biogeochemical cycling of mercury: a review. Annu Rev Environ Resour 34:43–63

SESA (2012) Bilans 2011 de l'épuration vaudoise. Service des eaux, sols et assainissement

Shipp WG, Zierenberg RA (2008) Pathways of acid mine drainage to Clear Lake: implications for mercury cycling. Ecol Appl 18(8 Suppl):A29–A54

Song YL, Liu CX, McTeague M, Summanen P, Finegold SM (2004) *Clostridium bartlettii* sp nov., isolated from human faeces. Anaerobe 10(3):179–184

Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Microbiol 44(4):846–849

Tercier-Waeber M-L, Stoll S, Slaveykova VI (2012) Trace metal behavior in surface waters: emphasis on dynamic speciation, sorption processes and bioavailability. Arch Sci 65:119–142

Thevenon F, Graham ND, Chiaradia M, Arpagaus P, Wildi W, Pote J (2011a) Local to regional scale industrial heavy metal pollution recorded in sediments of large freshwater lakes in central Europe (lakes Geneva and Lucerne) over the last centuries. Sci Total Environ 412:239–247

Thevenon F, Graham ND, Herbez A, Wildi W, Pote J (2011b) Spatio-temporal distribution of organic and inorganic pollutants from Lake Geneva (Switzerland) reveals strong interacting effects of sewage treatment plant and eutrophication on microbial abundance. Chemosphere 84(5):609–617

Weinbauer MG, Beckmann C, Hofle MG (1998) Utility of green fluorescent nucleic acid dyes and aluminum oxide membrane filters for rapid epifluorescence enumeration of soil and sediment bacteria. Appl Environ Microbiol 64(12):5000–5003

Wildi W, Dominik J, Loizeau JL, Thomas RL, Favarger P-Y, Haller L, Perroud A, Peytremann C (2004) River, reservoir and lake sediment contamination by heavy metals downstream from urban areas of Switzerland. Lakes Reserv: Res Manag 9:75–87

Xia S, Duan L, Song Y, Li J, Piceno YM, Andersen GL, Alvarez-Cohen L, Moreno-Andrade I, Huang CL, Hermanowicz SW (2010) Bacterial community structure in geographically distributed biological wastewater treatment reactors. Environ Sci Technol 44(19):7391–7396

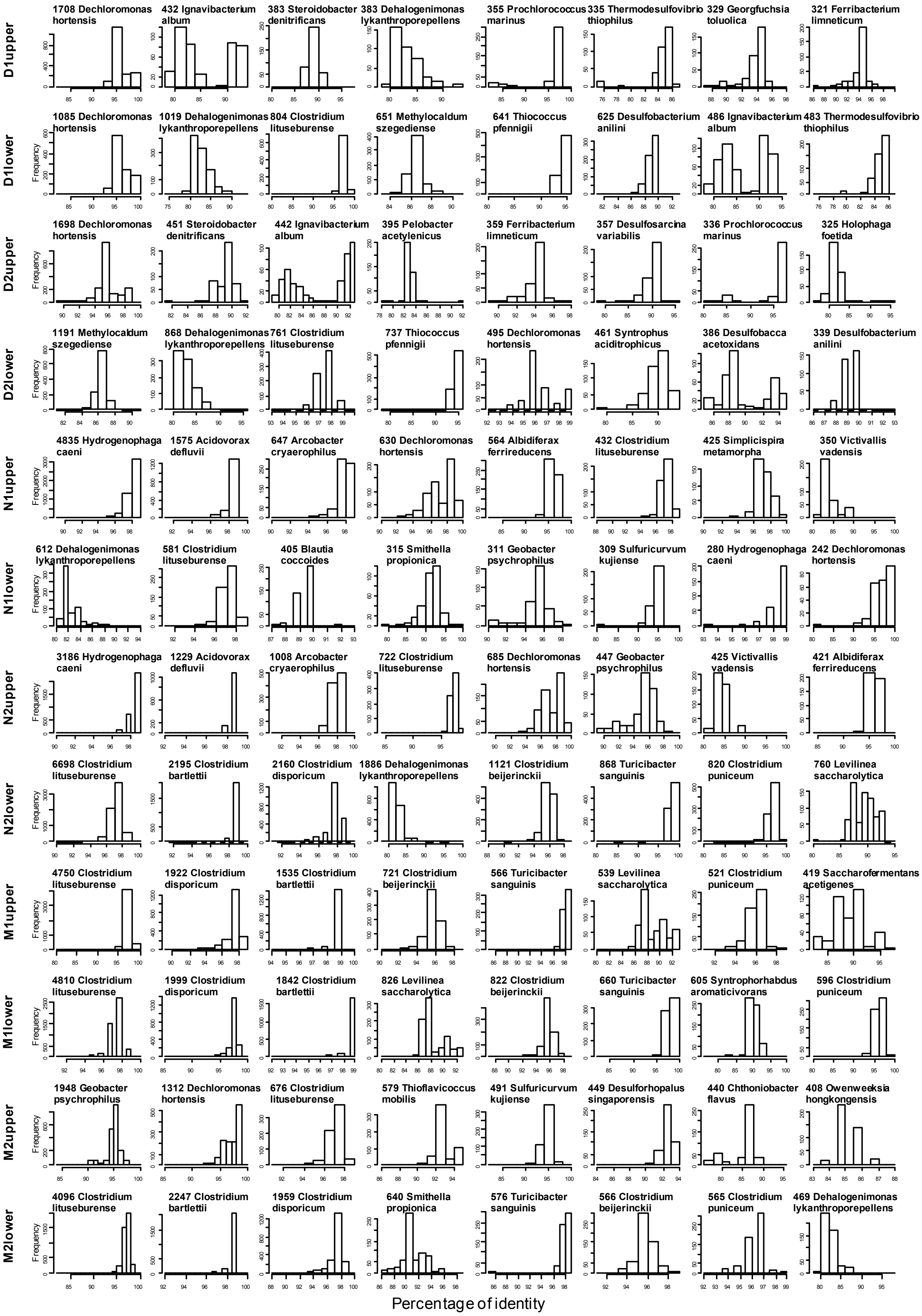
Yu R-Q, Flanders JR, Mack EE, Turner R, Mirza MB, Barkay T (2012) Contribution of coexisting sulfate and iron reducing bacteria to methylmercury production in freshwater river sediments. Environ Sci Technol 46(5):2684–2691

Zhang YJ, Bryan ND, Livens FR, Jones MN (1996) Complexing of metal ions by humic substances. Humic Fulvic Acids: Isol Struct Environ Role 651:194–206

Supplementary Informations

Supplementary Figure 1 Pearson correlation matrix between different chemical elements and/or bacterial abundance measured by fluorescence microscopy (abund1) or by real-time PCR (abund2). Significant correlations are indicated by “*” at *p*-values of 0.05 (*), 0.01 (**) or 0.001 (***).

Eight most abundant species in the samples



Supplementary Figure 2 Histograms representing the eight most abundant identified sequences per sample. Identification was carried out based on BLAST against a reference database containing 8911 sequences of type strains. For each identified sequence, the frequency of sequences from the sample (y-axis) and the identity level (x-axis) are shown. Depending the percentage of identity the sequences were assigned at species (>97%), genus (>95%) or phylum (>80%) levels (Schloss and Handelsman 2005; Stackebrandt and Goebel 1994).

Paper III

In which a qPCR method to quantify endospore-forming bacteria is developed...

Published in Applied and Environmental Microbiology 79, issue 17, 5302-5312, 2013
which should be used for any reference to this work

Quantification of Endospore-Forming *Firmicutes* by Quantitative PCR with the Functional Gene *spo0A*

Matthieu Bueche, Tina Wunderlin, Ludovic Roussel-Delif, Thomas Junier, Loic Sauvain, Nicole Jeanneret, Pilar Junier

Laboratory of Microbiology, Institute of Biology, University of Neuchatel, Neuchatel, Switzerland

Bacterial endospores are highly specialized cellular forms that allow endospore-forming *Firmicutes* (EFF) to tolerate harsh environmental conditions. EFF are considered ubiquitous in natural environments, in particular, those subjected to stress conditions. In addition to natural habitats, EFF are often the cause of contamination problems in anthropogenic environments, such as industrial production plants or hospitals. It is therefore desirable to assess their prevalence in environmental and industrial fields. To this end, a high-sensitivity detection method is still needed. The aim of this study was to develop and evaluate an approach based on quantitative PCR (qPCR). For this, the suitability of functional genes specific for and common to all EFF were evaluated. Seven genes were considered, but only *spo0A* was retained to identify conserved regions for qPCR primer design. An approach based on multivariate analysis was developed for primer design. Two primer sets were obtained and evaluated with 16 pure cultures, including representatives of the genera *Bacillus*, *Paenibacillus*, *Brevibacillus*, *Geobacillus*, *Alicyclobacillus*, *Sulfobacillus*, *Clostridium*, and *Desulfotomaculum*, as well as with environmental samples. The primer sets developed gave a reliable quantification when tested on laboratory strains, with the exception of *Sulfobacillus* and *Desulfotomaculum*. A test using sediment samples with a diverse EFF community also gave a reliable quantification compared to 16S rRNA gene pyrosequencing. A detection limit of about 10^4 cells (or spores) per gram of initial material was calculated, indicating this method has a promising potential for the detection of EFF over a wide range of applications.

Endospore formers are a paraphyletic group within the *Firmicutes* (1). This phylum comprises three classes: *Bacilli*, *Clostridia*, and *Erysipelotrichi* (2), but only the first two classes contain endospore-forming species. *Bacilli* principally include aerobic bacteria, whereas *Clostridia* encompass mainly anaerobic strains. Since the discovery of endospore-forming *Firmicutes* (EFF) in the late 19th century, most of the knowledge on their biology has been gained from laboratory studies on cultivable strains (3). In laboratory experiments, endosporulation is commonly triggered by starvation (4) or is coupled to other physiological changes, such as solventogenesis (5), as the ultimate mechanism when all other survival strategies have failed (6).

The survival advantage conferred by sporulation is considered one of the factors that has led to the ubiquitous distribution of EFF (7, 8). As spores remain viable for long periods of time, EFF can survive long distance transportation by climatic or biological agents, leading to a higher dispersion rate than non-EFF organisms (9, 10). In addition, the broad metabolic diversity found within this group is probably another factor that explains their ubiquity (11).

The presence of EFF is a positive factor in many domains related to human life, such as plant growth promotion, biocontrol (12), or the production of fermented food and probiotics by the health care industry (13). However, EFF are often the cause of contamination problems in the food (14) and medical (15) industries. Since endospores resist many disinfection treatments (chemical, heat, and UV sterilization) (7), it is difficult to remove them without a reduction in the quality of the end product. Also, the formation of biofilms is a major challenge encountered in production plants (16, 17). In the case of the food industry, in addition to the health concerns related to contamination by pathogenic species (e.g., *Bacillus cereus* or *Clostridium botulinum*), there is also the aspect of spoilage and short shelf life caused by nonpathogenic species. All these issues represent potentially significant economic losses (18).

Until now, the methods available to detect EFF were either laborious and time-consuming or not sensitive enough. For example, due to the heterogeneity under culture conditions, multiple assays have to be carried out to retrieve as many different EFF species as possible. Culture-independent approaches are mainly based on the quantification of dipicolinic acid (DPA), a unique biomarker for endospores (19). Different methods have been developed to extract and quantify DPA from endospores (10, 19–22), with detection limits ranging from 10^8 endospores/g of sediment (via high-performance liquid chromatography) (23), 10^5 endospores/g of soil (via complexation to terbium and time-resolved fluorometry) (20, 23), or 10^3 endospores/ml (via complexation to terbium and direct fluorescence microscopy) (24). Despite the relatively high sensitivity of DPA-based methods, a disadvantage of those assays is that they cannot take into account vegetative cells of EFF. Furthermore, terbium fluorescence is easily quenched with substances commonly found in environmental samples (e.g., humic acids [23] or organophosphates [19]), leading to a dramatic decrease in the detection limit.

To overcome these limitations, a molecular assay could provide a valid alternative. Although an assay for the detection of specific thermophilic *Bacilli* has been reported (25), a universal tool is still needed. The aim of this study was to develop and evaluate a molecular approach. Genes involved in the endosporulation process were considered for the identification of a functional molecular marker. Endosporulation is a complex mechanism that involves many regulatory genes (6). Several of them could be used to develop universal primers for endosporulation (T. Wunderlin, T. Junier, L. Roussel-Delif, N. Jeanneret, and P. Junier, unpublished data) and they were tested here. *In silico* tests led to the selection of *spo0A* as a promising marker gene. Quantitative PCR (qPCR) is an accurate method to quantify the frequency of a gene from a DNA extract (26), and therefore in this study qPCR primers targeting the gene *spo0A* were designed, tested, and validated in pure cultures and environmental samples.

MATERIALS AND METHODS

Evaluation of target genes involved in endosporulation. Candidate genes for the development of universal primers to target EFF by PCR assays were identified in a previous study of 27 EFF genomes (Wunderlin et al., unpublished) and included the following: *spo0A*, *gpr*, *spoIVB*, *spoVT*, *spoVAC*, and *spoVAD*. The gene coding for sigma-H factor (*sigH*) was added to this list because it has been shown to be a regulator of sporulation, as *spo0A* is (6). In order to estimate the prevalence of these seven genes among endospore-forming species, an initial list of 90 endospore-forming genera was established and confirmed manually by checking for sporulation as a feature in the *Bergey's Manual of Systematic Bacteriology* (2). This list was key to verify the prevalence of the selected sporulation genes retrieved from the automatic annotation of bacterial genomes that can contain annotation errors (27).

To check for the presence of the seven selected genes in EFF, TIGRFAM collections of proteins (28) generated by the TIGR annotation engine of the J. Craig Venter Institute (29) were downloaded from the Integrated Microbial Genome (IMG) system (30). The seven genes selected in this study corresponded to TIGR01441 (*gpr*), TIGR02875 (*spo0A*), TIGR02851 (*spoVT*), TIGR02860 (*spoIVB*), TIGR02845 (*spoVAD*), TIGR02838 (*spoVAC*), and TIGR02859 (*sigH*). The species in the TIGRFAM list were compared to the list of 90 endospore-forming genera described above. Only species corresponding to endospore formers were retained.

Refined database of EFF sequences. For the design of universal primers for EFF, only the *spo0A* gene was considered. A functional search was used to retrieve *spo0A* sequences in all microbial genomes published in the IMG-JGI database. A query with the keywords *spo0A* or "stage 0 sporulation" retrieved sequences from genes that were not linked to endospore formation. A manual selection was carried out to find the annotated functions that corresponded to the *spo0A* gene. Based on sequence size (250 to 290 amino acids), as well as name, the following keyword annotations seemed to correspond to *spo0A* and were retained for this study: spo0a, spo0a protein, sporulation initiation factor spo0a, sporulation transcription factor spo0a, sporulation transcriptional activator spo0a, and stage 0 sporulation protein A.

A database of 216 gene sequences corresponding to these annotations was constructed. The "seqret" command from the EMBOSS package (31) was used to check for the integrity of the generated fasta file. A multiple alignment was performed using the default settings in MAFFT (32). In order to obtain a database containing a minimum of homologous sequences, a Jukes-Cantor distance matrix was calculated with the command "distmat" from EMBOSS. Sorting of the distances and removal of the closest sequences was carried out with a customized R script (see the supplemental material for further details) (33). As the aim of this sorting step was to avoid redundancy and overrepresentation of specific groups of organisms in the database, an experimentally calibrated threshold value of 2 was set to accomplish this. Gaps were removed, and the 99 remaining sequences were realigned with MAFFT. The "consambig" command (EMBOSS) was used to produce a consensus sequence. To determine the most conserved regions in the alignment, the number of conserved bases for each position was calculated using an R script that ran a 20-bp sliding window with a skip of 1 base in the alignment. A nonmetric multidimensional scaling (NMDS) analysis was performed (vegan library [34]) to determine the homogeneity of the sequences in four conserved regions identified with the R script. This first analysis allowed the detection of atypical sequences from non-endospore-forming species (*Ethanoligenens* spp., *Acetivibrio* spp., and *Blautia* spp., among others) that strongly decreased the conservation signal. These atypical sequences were removed. Finally, a refined database containing 80 sequences was obtained and realigned using MAFFT.

Primer development. The 80 aligned sequences were used for a similar procedure to identify conserved regions. Taking into account the suitability of the amplicon for qPCR (50 to 250 bp) (35, 36), only a few sites were retained. Furthermore, in order to have amplicons of a constant length for all endospore-forming species, the alignment between the annealing sites for the forward and reverse primers could not contain gaps. Based on these parameters, only 11 sites were kept and evaluated by hierarchical clustering (Jaccard distance and average linkage) to identify the least variable sites (vegan library in R). A threshold of 0.4 was selected to retain only the most conserved regions. A Jaccard distance of 0.5 corresponds to an average change of 1 in 3 bases, which matches the variability of the genetic code for the majority of amino acids. Thus, for a 0.4 threshold, only five out of six codons contained in a 20-bp primer present one degeneracy. The 20-bp sequences from the most promising sites were analyzed manually to find degenerate primers that matched the majority of the endospore-forming species. The exact position of each primer in the alignment was then fine-tuned to include a perfectly conserved region for the 3′ end (the five last base pairs) to increase specificity and efficiency during PCR amplification (37).

Finally, an overview of the variability of the sequences inside the retained sites was obtained by a principal components analysis [PCA; rda() function of vegan library in R]. The heterogeneity for the three major sets of EFF sequences (from the genera *Bacillus*, *Clostridium*, and *Geobacillus*) was plotted as an ellipse corresponding to the standard deviation of the PCA coordinates for the members of these groups. All the combinations corresponding to the degenerate positions in the primers were also projected on the two axes of the ordination plot, and an ellipse was drawn around these points to estimate the specificity coverage of the primers. The script used for primer design is provided in the supplemental material.

In silico **specificity test using a BLAST search.** The Primer-BLAST online tool (38) provided by the NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to evaluate the specificity of the two different primer pairs that were designed. As this tool cannot take into account the degenerate positions, all the different combinations for the forward and reverse primers were inserted manually in the Web interface. The "primer pair specificity checking" parameter was set to use the nucleotide reference database and to restrict the query to only bacterial taxa (excluding uncultured sequences, which are not informative). All the other parameters were set to default values. For each primer set, a text file containing each individual query per degenerate combination was created. The redundancies were removed using low-level Unix commands (grep, awk, sort, and uniq) based on the gene identification number. Sequences were then grouped together by genus, and mismatches were quantified for both forward and reverse primers. This work was done for the entire annealing site (20 or 23 bp). The same analysis was carried out for only the last 6 bp on the 3′ end of the primers, to confirm the conservation and to more accurately evaluate the annealing specificity (data not shown).

Strains and culture conditions. In order to test the efficiencies of the newly developed primers, 16 strains of EFF and 2 non-EFF strains were cultivated for DNA extraction (Table 1). DNA extractions were performed with the innuPREP bacteria DNA kit (Analytik Jena AG, Jena, Germany) and following the manufacturer's instructions. The media

TABLE 1 Strains used for *in vitro* PCR and qPCR tests and culture conditions

Strain	Species	Medium	Temp (°C)	EFF
Neu 1062	*Bacillus licheniformis*	NB	37	Yes
Neu 1070	*Bacillus thuringiensis*	NB	30	Yes
Neu 1121	*Bacillus subtilis*	NB	30	Yes
Neu 1261	*Bacillus amyloliquefaciens*	NB	30	Yes
Neu 91	*Paenibacillus alvei*	NB	30	Yes
Neu1005	*Paenibacillus alvei*	NB	30	Yes
Neu 1003	*Lysinibacillus sphaericus*	NB	30	Yes
Neu 1014	*Brevibacillus brevis*	NB	30	Yes
Neu 1040	*Brevibacillus thermoruber*	DSM 572	45	Yes
Neu 1149	*Geobacillus thermoglucosidasius*	DSM 305	65	Yes
HFF 3	*Geobacillus thermoleovorans*	PTYG, pH 5.0	50	Yes
B 18	*Alicyclobacillus acidocaldarius*	PTYG, pH 2.9	50	Yes
JWO 13	*Sulfobacillus acidophilus*	FeSO	50	Yes
Neu 28	*Clostridium beijerinckii*	ST1 anaerobic	30	Yes
Neu 1155	*Clostridium pasteurianum*	DSM 54 anaerobic	37	Yes
MI-1	*Desulfotomaculum reducens*	Difco 2216*	37	Yes
Neu 1021	*Escherichia coli*	NB	37	No
Neu 1037	*Lactococcus lactis* subsp. *lactis*	DSM 92	37	No

used were standard nutrient broth (NB); DSM medium (http://www.dsmz.de/); standard I nutrient broth (ST1; Merck Millipore, Darmstadt, Germany); PTYG medium composed of 0.25 g/liter peptone, 0.25 g/liter tryptone, 0.5 g/liter yeast extract, 0.5 g/liter glucose, 0.6 g/liter $MgSO_4 \cdot 7H_2O$, 0.07 g/liter $CaCl_2 \cdot 2\ H_2O$ (pH was adjusted with NaOH and H_2SO_4); FeSO medium (described elsewhere [39]); Difco2216*, corresponding to a modified version of marine broth (also described elsewhere [40]) supplemented with 20 mM sodium pyruvate and 1 mg/liter of resazurin to verify anoxia. All cultures were in liquid medium under agitation, except for Neu 1155 (*Clostridium pasteurianum*) and Neu 28 (*Clostridium beijerinckii*), which were cultivated in petri dishes by spreading them on solid medium (15 g/liter agar). The incubation was at 37°C under anaerobic conditions attained by placing the plates in sealed plastic bags in an Anaerocult A mini apparatus and one strip of Anaerotest (Merck Millipore).

Experimental accuracy test. To assess the accuracy of the qPCR quantification, eight strains of EFF (*Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus thuringiensis*, *Brevibacillus brevis*, *Paenibacillus alvei*, *Bacillus amyloliquefaciens*, *Lysinibacillus sphaericus*, and *Clostridium beijerinckii*) and two strains of a non-EFF (*Escherichia coli* and *Herbaspirillum autotrophicum*) were cultivated and quantified by microscopic counting using a Neubauer improved chamber with a special depth of 10 μm (0642010; Paul Marienfeld, Lauda-Königshofen, Germany) prior to DNA extraction. *B. subtilis* was grown in 2× SG medium (41) and incubated at 30°C under agitation at 150 rpm. Vegetative cells were harvested from a fresh overnight culture, whereas endospores were obtained from an older culture in which phase-contrast microscopy revealed over 99% endospores. Cells from other strains were obtained by cultivation in NB and incubation at 37°C (*B. licheniformis* and *E. coli*), 24°C (*H. autotrophicum*), or 30°C (other strains) with agitation at 150 rpm. Cells were fixed in a solution with 4% Formalin and 1× phosphate-buffered saline until counting. Before counting, cells were dispersed in a solution of 1% sodium hexametaphosphate and vortexed for 30 s. Samples were then diluted in physiological solution and loaded by capillarity into the Neubauer improved counting chamber. Counting was carried out with a bright-field microscope with a magnification of ×400. Cells in several squares containing 2.5×10^{-5} μl of cell suspension were quantified, and the average number of cells per μl of fresh culture was calculated.

Finally, DNA extraction from cells and endospores was carried out with the FastDNA spin kit for soil (MP Biomedicals, Santa Ana, CA) following a modified procedure to extract DNA from resistant structures, including endospores (Wunderlin et al., unpublished). This kit was selected because it results in a better extraction yield than others when working with endospores of EFF (42).

PCR conditions. All PCR amplifications were done using *Taq* DNA polymerase with ThermoPol buffer from New-England Biolabs (Ipswich, MA). Reaction mixtures with a final volume of 20 μl contained 1× ThermoPol Buffer, either 0.75 μM (primer set 1, spo0A655f-spo0A923r [Table 2]) or 1 μM (primer set 2, spo0A655f-spo0A834r [Table 2]) of the forward primer, 0.45 μM reverse primer (both primer sets), 200 μM deoxynucleoside triphosphate mix (Promega Corporation, Madison, WI), 1.25 U *Taq* DNA polymerase, and 1 μl of diluted sample (1 to 2 ng of DNA). The volume was completed with PCR-grade water. PCRs were carried out with the Arktik thermal cycler from Thermo Scientific (VWR International, Radnor, PA). The PCR program consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles composed of a denaturation step at 95°C for 30 s, an annealing step at 52°C (primer set 1) or 54°C (primer set 2) for 30 s, and an elongation step at 68°C for 30 s. A final elongation step at 68°C for 10 min was performed. The PCR results were verified by electrophoresis in 2% agarose gels, loading 2 μl of the PCR product diluted in 10 μl of loading buffer (0.05% bromophenol blue, 8% sucrose, 1.6 mM Tris-HCl at pH 8, 0.16 mM EDTA at pH 8). DNA was revealed by staining for 30 min in a 3× GelRed bath (Biotium Inc., Hayward, CA). Gel imaging was carried out with a GenoPlex chamber (VWR).

qPCR conditions for the *spo0A* gene. qPCRs were carried out in a final reaction volume of 10 μl with 5 μl Rotor-Gene SYBR green PCR master mix (Qiagen GmbH, Hilden, Germany), 0.75 μM and 0.45 μM each primer (forward and reverse, respectively) for set 1 (spo0A655f-spo0A923r); and 1 μM and 0.45 μM each primer for set 2 (spo0A655f-spo0A834r). Two-microliter aliquots of diluted samples (approximately 5 ng) were added to the reaction tubes. The Corbett Rotor-Gene RG-3000A thermocycler (Qiagen) was used for the amplifications. An initial denaturation/polymerase activation step at 95°C for 10 min was first performed. For primer set 1, the 45 following cycles consisted of denaturation at 95°C for 30 s, annealing at 52°C for 30 s, and elongation at 72°C for 30 s. For primer set 2, the 45 following cycles consisted of denaturation 95°C for 15 s, annealing 54°C for 15 s, and elongation 68°C for 10 s. For both primers sets, a melting analysis was performed at the end of the reaction by

TABLE 2 Summary information on the three designed primers[a]

Site no.	Primer	Sense	Sequence (5′–3′)	Length (bp)	% GC content	Annealing temp (°C)	Hairpin formation	3 ′ complementarity
14	spo0a655f	Forward	5′-GGH GTD CCN GCN CAT ATH AA	20	40–65	47.7–57.9	None	None
15	spo0a923r	Reverse	5′-GCD ATG AAY TCD GAG TTN GTN GG	23	39–61	51.7–60.6	None	None
9	spo0a834r	Reverse	5′-CCA HGC NAC TTC WAT NGC RT	20	40–60	47.7–55.9	None	None

[a] GC content, annealing temperature, hairpin formation, and 3′ complementarity were obtained by using the Oligo Calc Web tool (60).

gradually increasing the temperature by steps of 1°C from the elongation temperature up to 95°C.

The standard curve for quantification was prepared from 10-fold dilutions (10^8 to 10^2 copies/μl) of a plasmid in which the *spo0A* gene of *B. subtilis* was inserted. The TOPO TA cloning kit (Invitrogen, Carlsbad, CA) was used to produce this plasmid in One Shot TOP10F′ chemically competent *E. coli* cells (Invitrogen), following the manufacturer's guidelines. Plasmid DNA was extracted with the Wizard Plus SV Miniprep DNA purification system (Promega) following the manufacturer's instructions. Finally, the number of gene copies was deduced from the DNA quantification carried out with a Qubit 2.0 fluorometer (Invitrogen).

qPCR conditions for the 16S rRNA gene. An additional qPCR quantification of the 16S rRNA gene was performed with DNA extracted from the pure cultures to verify the accuracy of the *spo0A* gene quantification. The qPCR was done on the hypervariable V3 region of the 16S rRNA gene. The primers used were 338f (5′-ACTCCTACGGGAGGCAGCAG-3′) and 520r (5′-ATTACCGCGGCTGCTGG-3′) (43, 44) at a concentration of 0.3 μM each. qPCR was carried out as described for the *spo0A* gene with a modified program consisting of an initial denaturation/activation step at 95°C for 15 min, followed by 40 cycles composed of denaturation at 95°C for 10 s, annealing at 55°C for 15 s, and elongation at 72°C for 20 s. The standard curve consisted of 10-fold dilutions (10^8 to 10^2 copies/μl) of a plasmid in which the V3 region of an environmental clone was inserted. The Zero Blunt TOPO PCR cloning kit for sequencing (Invitrogen) was used to produce the plasmid. Transformation, plasmid extraction, and quantification were performed as mentioned above for the *spo0A* gene.

The ratio between the qPCR quantification of both genes was calculated (qPCR 16S rRNA/qPCR *spo0A*) and divided by the number of copies of the 16S rRNA gene found in each individual species that was tested. The number of 16S rRNA gene copies was obtained from the rrnDB published by Michigan State University (45). A ratio of 1 would be expected in case of no bias, but to take into account different factors of variability such as, for example, differences of annealing specificity, pipetting errors, differences in amplification efficiency (for the *spo0A* sequences), and differences in the number of copies of the 16S rRNA gene between the tested strains and the rrnDB, an arbitrary value between 0.1 and 10 was deemed accurate.

Environmental sequences. Sediments from the Bay of Vidy (Lake Geneva, Switzerland) were used to test the efficiency of the qPCR EFF quantification method in environmental samples. In a first sampling campaign, six cores were retrieved from different areas (Table 3) and subsampled in two layers: depths of 0 to 3 cm and 3 to 9 cm. During a second sampling, four additional cores were retrieved and subsampled in three layers: 0 to 1.5 cm, 1.5 to 3 cm, and 3 to 9 cm. A modified protocol, including cell separation from the environmental matrix and sequential beadbeating steps, was used for DNA extraction (Wunderlin et al., unpublished). Inhibitor-free DNA was then retrieved by an extra ethanol purification step at the end of the extraction procedure.

Samples were sent to Eurofins MWG Operon (Ebersberg, Germany) for 454 pyrosequencing of a fragment of the 16S rRNA gene. Fragments of approximately 500 bp were retrieved using primers Eub8f (5′-AGAGTTTGATCCTGGCTCAG-3′) and Eub519r (5′-GTATTACCGCGGCTGCTGG-3′) as described elsewhere (46). Bioinformatic processing was performed with QIIME (47), following the default pipeline recommended in the online tutorial for 454 pyrosequences. The BIOM format table (48) was retrieved from the QIIME analyses and converted into a text file (convert_biom.py) suitable for importing data in R and to compute the ratio of *Firmicutes* sequences in each sediment sample. Because the reference database to assign taxonomy in QIIME does not allow in every case the identification of all the operational taxonomix units (OTUs) at the genus level, the ratio of *Firmicutes* was evaluated in place of the ratio of EFF. Nonetheless, a similar ratio has been obtained when considering only the *Bacilliales* and *Clostridiales* orders.

Bacterial abundance in environmental samples. In order to assess the abundance of total bacteria in each sample, a qPCR for the hypervariable V3 region of the 16S rRNA gene was carried out as described above for pure cultures. The ratio between the 16S rRNA gene and the *spo0A* gene was calculated.

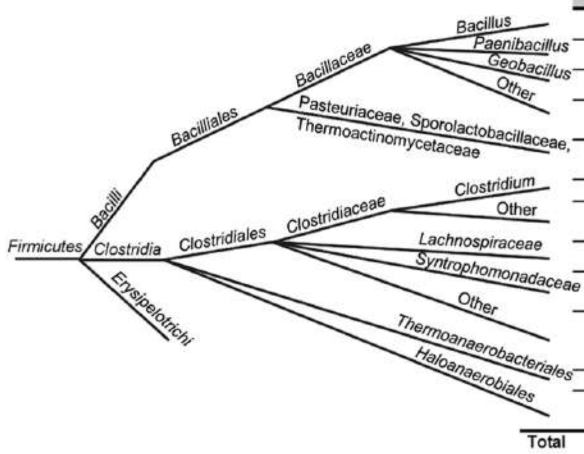
TABLE 3 Bacterial communities in environmental samples based on pyrosequencing data[a]

		Swiss coordinate			Result from bioinformatics analysis of pyrosequencing data		
Core	Sampling depth (m)	East	North	Depth of layer (cm)	No. of sequences	No. of OTUs	*Firmicutes*
C2.21	47	534509	151342	0–3	22,805	4,778	12,822
				3–9	22,573	3,794	13,398
C8.22	60	534368	151306	0–3	19,440	4,954	2,224
				3–9	20,393	4,157	11,898
C1.03	36	533735	151351	0–3	20,640	6,618	985
				3–9	21,977	6,745	2,424
C2.03	35	534063	151491	0–3	19,357	6,048	931
				3–9	17,830	5,624	2,351
C3.03	32	534676	151543	0–3	20,521	4,200	2,383
				3–9	11,995	4,346	2,385
C4.03	32	534676	151543	0–3	20,873	4,674	2,821
				3–9	31,024	5,415	17,973
C1b.24	32	534709	151526	0–1.5	31,533	5,729	7,169
				1.5–3	9,463	3,064	2,355
				3–9	9,409	3,409	2,767
C2b.24	31.5	534648	151567	0–1.5	10,812	3,221	1,337
				1.5–3	13,349	3,683	1,222
				3–9	5,848	2,217	1,370
C3b.24	31	534803	151513	0–1.5	10,199	2,885	1,147
				1.5–3	5980	1,852	567
				3–9	8,191	2,400	3,189
C4b.24	31	534961	151369	0–1.5	26,725	5,779	708
				1.5–3	7,203	2,570	509
				3–9	6,531	2,380	1,725

[a] All the environmental samples were retrieved from Lake Geneva, in the Bay of Vidy (Lausanne, Switzerland). Sampling locations are given in Swiss national coordinates (CH1903). Information on bacterial communities for layer subsamples were obtained by bioinformatics analysis on pyrosequencing data.

RESULTS

Selection of molecular markers for EFF. A search for EFF clearly demonstrated that the trait of endospore formation is not distributed uniformly in all bacteria from the phylum *Firmicutes* and can be found scattered in various genera of the classes *Clostridia* and *Bacilli*. A first set of genes annotated as part of the endosporulation process and common to EFF with different phylogenetic affiliations was previously identified in our laboratory (Wunderlin et al., unpublished). The comparison between TIGRFAM annotations and a list of 90 endospore-forming genera showed that six out of the seven chosen genes (*gpr*, *spo0A*, *spoVT*, *spoIVB*, *spoVAD*, and *spoVAC*) appeared to be very conserved and widespread in endospore formers (present in 277 to 282 out of 300 EFF species) (Fig. 1). For each of these genes, a large number of sequences from species belonging to the genera *Bacillus*, *Clostridium*, *Paenibacillus*, and *Geobacillus* were found. However, in the case of *Bacillus* and *Clostridium*, this was partially due to the fact that these genera were overrepresented in the data set relative to other EFF. Only the gene coding for the sigma-H factor (*sigH*) was clearly not as well represented as the other six (present in 256 out of 300 EFF species). Based on the results, all the genes but *sigH* could be used as molecular markers for endosporulation. Since the phylogeny of *spo0A* is similar to the phylogeny obtained with the 16S rRNA gene. As a consequence, the gene selected for the design of the qPCR primers was *spo0A*.



Reported sporulation gene in TIGR database

	gpr TIGR01441	***spo0A* TIGR02875**	*spoVT* TIGR02851	*spoIVB* TIGR02860	*spoVAD* TIGR02845	*spoVAC* TIGR02838	*sigH* TIGR02859	
Bacillus	123	**124**	124	123	126	126	117	
Paenibacillus	11	**12**	12	12	12	12	11	
Geobacillus	9	**9**	9	9	9	9	9	
Other	9	**9**	9	7	9	9	9	*Alicyclobacillus, Lysinibacillus, Brevibacillus, Others Bacillaceae (28 genera)*
Pasteuriaceae, Sporolactobacillaceae, Thermoactinomycetaceae	2	**2**	2	2	2	2	2	*Candidatus / Tuberibacillus(0) /Desmospora, 5 other Thermoactinomycetaceae(0)*
Clostridium	85	**86**	83	85	85	86	72	
Other	2	**2**	2	2	2	2	2	*Alkaliphilus, Caloramator(0), Caminicella(0), Metabacterium(0), Oxobacter(0), Tepidibacter(0)*
Lachnospiraceae	4	**4**	4	4	4	4	2	*Coprococcus, Sporobacterium(0)*
Syntrophomonadaceae	2	**2**	2	2	2	2	2	*Syntrophothermus, Syntrophomonas, Pelospora(0), Syntrophospora(0)*
Other	15	**15**	15	15	15	15	15	*Peptococcaceae / Acidaminococcaceae / Heliobacteriaceae / Ruminococcaceae(0) / Peptostreptococcaceae(0) / Incertae sedis*
Thermoanaerobacteriales	15	**15**	14	15	15	14	14	*Thermoanaerobacter and 11 other genera*
Haloanaerobiales	1	**1**	1	1	1	1	1	*Halonatronum(0), Natroniella(0), Orenia(0), Sporohalobacter(0) / Acetohalobium*
Total	278	**281**	277	277	282	282	256	**On 300 Endospore-forming species**

FIG 1 Phylogenetic overview of the major genera of endospore-forming *Firmicutes* (left) and frequency of sporulation related genes in each clade (table in the middle). *spo0A* (in bold) is the most prevalent gene in endospore-forming bacteria from different clades. For the branches "Other" the genera are indicated on the right. The number 0 indicates the genera that have no match for any of the queried genes in the IMG/JGI database.

Primer design. To overcome the variability of the sequences of a functional gene, an approach based on multivariate analysis was developed for the design of universal primers, a process that could possibly be automated in a primer design pipeline. Six highly conserved regions suitable for primer design were identified in a refined database of 80 *spo0A* sequences. In total, 15 starting positions for 20-bp primers containing at least 9 conserved bases (maximum 11 bases) could be defined in the alignment (see Fig. S1 in the supplemental material). However, due to the existence of several gaps in the alignment between the second and the third regions, only the last four (between positions 649 and 923 of the alignment, primer sites 5 to 15 [see Fig. S1]) were effectively suitable for designing qPCR primers, yielding amplicons of a constant length of approximately 150 bp.

From the 11 sites evaluated by clustering, 8 had sequences sufficiently conserved for grouping them together within a Jaccard distance below 0.4 (see Materials and Methods). These eight sites were grouped into three regions corresponding to positions 652 to 686, 815 to 834, and 901 to 923 in the alignment (see Fig. S1 in the supplemental material). To develop primers of 20 bp with this threshold, still an average of 5 degenerate bases were present within the region, and therefore the exact position and the length of the primers was fine-tuned manually to reduce degeneracies.

Initially, two sites were selected (site 14, positions 655 to 674, and site 15, 901 to 923) to develop a primer pair giving an amplicon of 268 bp. In the first site, the forward primer spo0A655f was defined (20 bp), with three degenerate bases and two inosines (Table 2). In the second site, a reverse primer (spo0A923r) of 23 bp, also with three degenerate bases and two inosines, was designed. A third site was selected to improve matching to *Clostridia* (see results of the specificity test below). In this third site, a new reverse primer (spo0A834r) of 20 bp with three degenerate bases and two inosines was designed. The combination of spo0A655f and spo0A834r resulted in a shorter amplicon of 180 bp, which was better suited for qPCR.

A PCA performed on the portions of the sequences corresponding to the primer sites (Fig. 2) showed clear differences in the specificities of the two reverse primers. The reverse primer spo0A923r did not have the same level of coverage as the other two primers. Indeed, in the PCA, the annealing sites for spo0A923r were clearly separated into two distinct groups that were not considered by the overall coverage of the degenerate primer that matched mainly *Bacillus*, and not *Clostridium* or *Geobacillus*. Concerning the two other sites, Fig. 2 illustrates that the sequence variations within the annealing sites for the primers spo0A655f and spo0A834r were much lower and that the degenerate primers covered almost all of the variation displayed in the two main principal components of the ordination plot.

Specificity test *in silico*. In the evaluation of the first primer set (spo0A655f -spo0A923r) with the NCBI Primer-BLAST online tool (38), 145 different species belonging to 26 genera were detected. For the second set (spo0A655f-spo0A834r), 150 species were distributed in 24 genera. Even though the major genera detected by Primer-BLAST were the same for the two different sets, there was some variation in the minor groups found.

The annealing specificity of both sets was evaluated by quantifying the mismatches reported between the primers used and the sequences detected in the NCBI database. These values were averaged by genera and are represented as a bar plot in Fig. 3. In agreement with the results of the PCA, the average number of mismatches for the reverse primer in the second set was much lower than that for the first set, and therefore the second set appears to be more universal.

Regarding specificity, only a few non-endospore-forming genera (*Eubacterium*, *Ruminococcus*, and *Faecalibacterium*) were detected with the two primer sets. These genera are members of the class of *Clostridia*. Furthermore, a closer look at the results in the TIGRfam *spo0A* collection showed that these genera possess a protein with a structure that is closely related to Spo0A. In addition, a recent genomic profiling for sporulation-related genes (49) showed that *Eubacterium* and *Ruminococcus* can be predicted as endospore-forming *Clostridia*, based on their genetic signature. In the case of *Faecalibacterium*, a search for endosporulation-related genes in the genome of *Faecalibacterium prausnitzii* A2-165, recently available from the Genome Institute at Washington University (50), using BioCyc revealed the presence of more than 30 additional genes related to endospore formation. Genera belong-

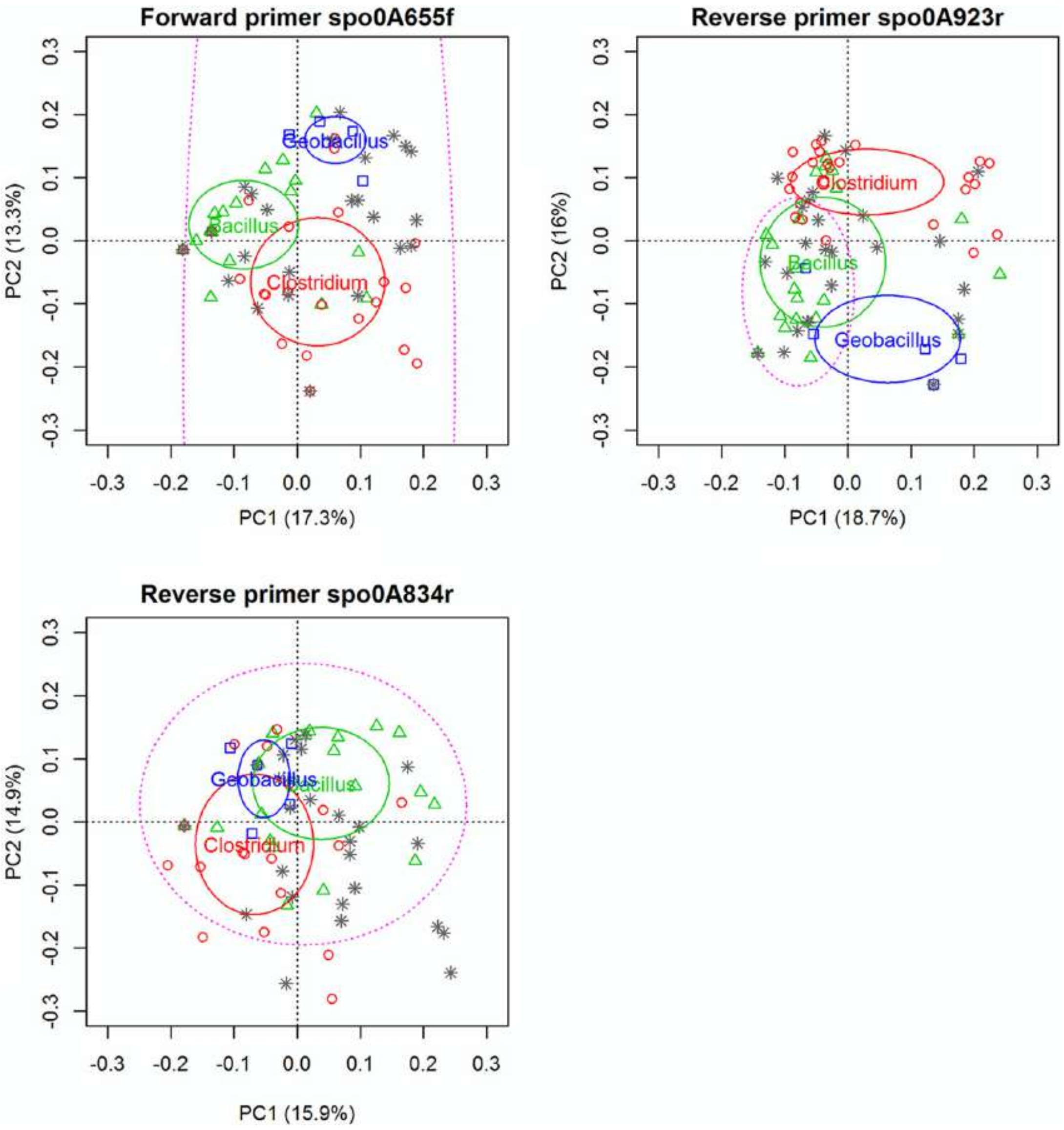


FIG 2 Principal component analysis of the three selected regions used for primer design. This representation gives an overview of the homogeneity of the sequences inside the annealing regions. The dimension of the ellipses (solid lines) gives an estimation of the sequence variability for the main genera found in endospore-forming *Firmicutes* (EFF). The individual points correspond to species used for the alignment. Δ, *Bacillus* spp.; □, *Geobacillus* spp.; O, *Clostridium* spp.; *, other EFF species. The ellipse in dashed lines gives an estimate of the coverage of the degenerate primers designed for each particular site.

ing to non-*Firmicutes* phyla were not reported from the Primer-BLAST analysis.

Laboratory experiments. The validation of both primer sets in laboratory experiments was first done by regular PCR with a set of seven species. The two non-endospore-forming strains, *Escherichia coli* and *Lactococcus lactis* subsp. *lactis*, gave no PCR amplification, whereas four out of five endospore formers gave positive results (*Bacillus subtilis*, *Clostridium pasteurianum*, *Alicyclobacillus acidocaldarius*, and *Geobacillus thermoleovorans*). *Sulfobacillus acidophilus*, which was the fifth endospore-forming strain tested, gave no amplification in the classical PCR. Based on these results, the first four endospore-forming strains were used to fine-tune the annealing temperature conditions of a gradient PCR assay between 44 and 56°C. For primer set 1, the best annealing temperature was found to be 52°C, whereas for the second set it was 54°C. These optimal annealing temperatures were then used to test additional endospore-forming strains with the two primer sets. As shown in Table 4, 12 out of the 13 (set 1) and 14 out of the 16 (set 2) endospore-forming strains assayed resulted in a positive PCR amplification. Only *S. acidophilus* did not give a PCR product with any of the two primer sets. For the second set (spo0A655f-spo0A834r), *Desulfotomaculum reducens*, a second strain of *Paenibacillus alvei*, and *Bacillus amyloliquefaciens* were also included in the tests. The two Bacilli gave positive amplification results, whereas *D. reducens* did not.

The first tests of qPCR were also done with the same strains mentioned above. The results were similar except with primer set 1, which gave no amplification for *Clostridia* or *Brevibacillus brevis*. Although we tested different parameters, such as increased $MgCl_2$ concentration, increased primer concentration, and lower annealing temperature, none was successful for amplification of these strains. The second set with the reverse primer (spo0A834r) gave better results for the two *Clostridium* strains tested, but the amplification efficiency was still low. The melting analysis, however, highlighted that the PCR products for the *Clostridium* strains have relatively low melting temperatures (close to the one used for elongation). As a consequence, the elongation was decreased from 72 to 68°C. With this modification, a drastic increase of the amplification efficiency for the *Clostridium* strains was obtained. Concerning *B. brevis*, positive amplification results were also observed with the second primer set.

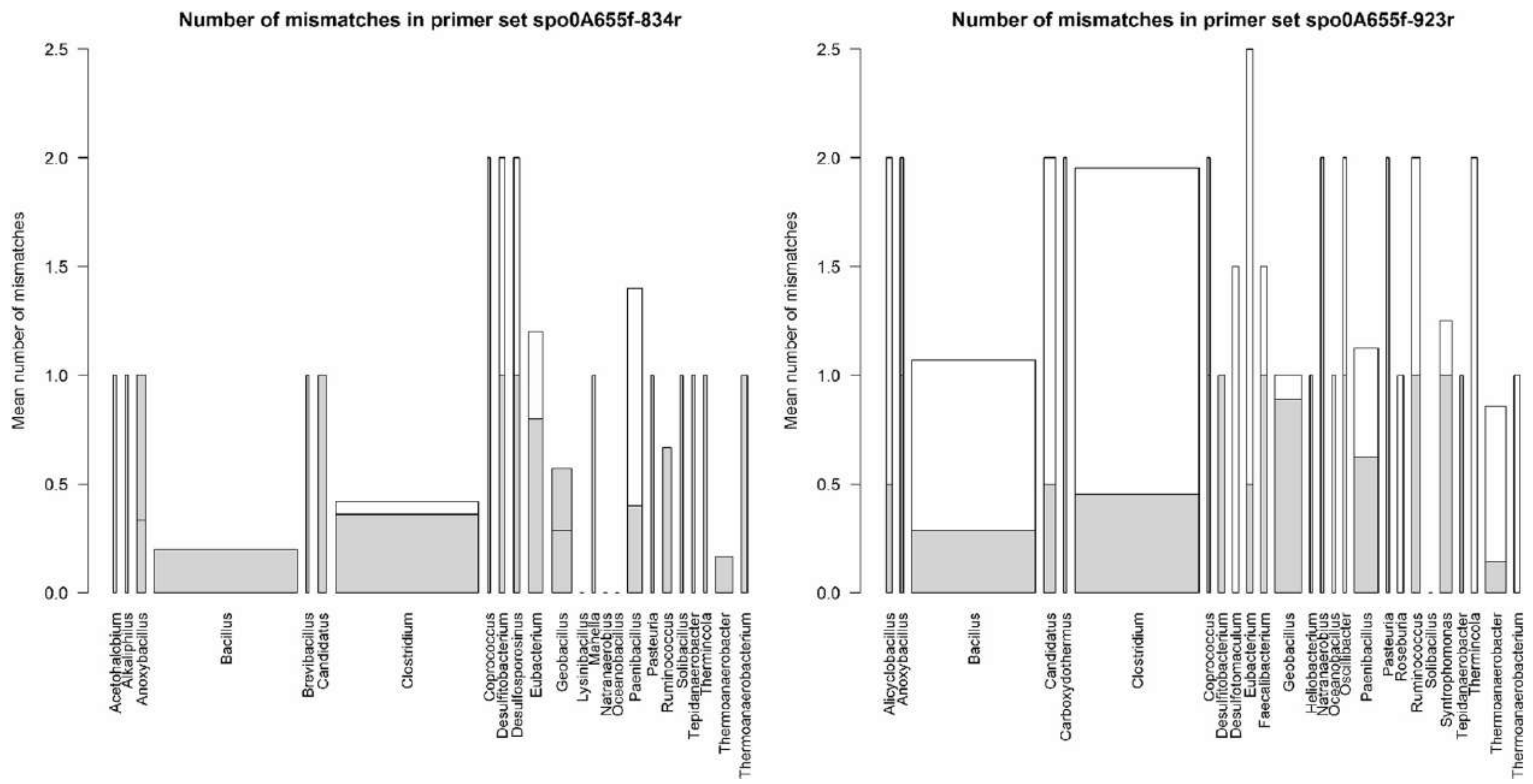


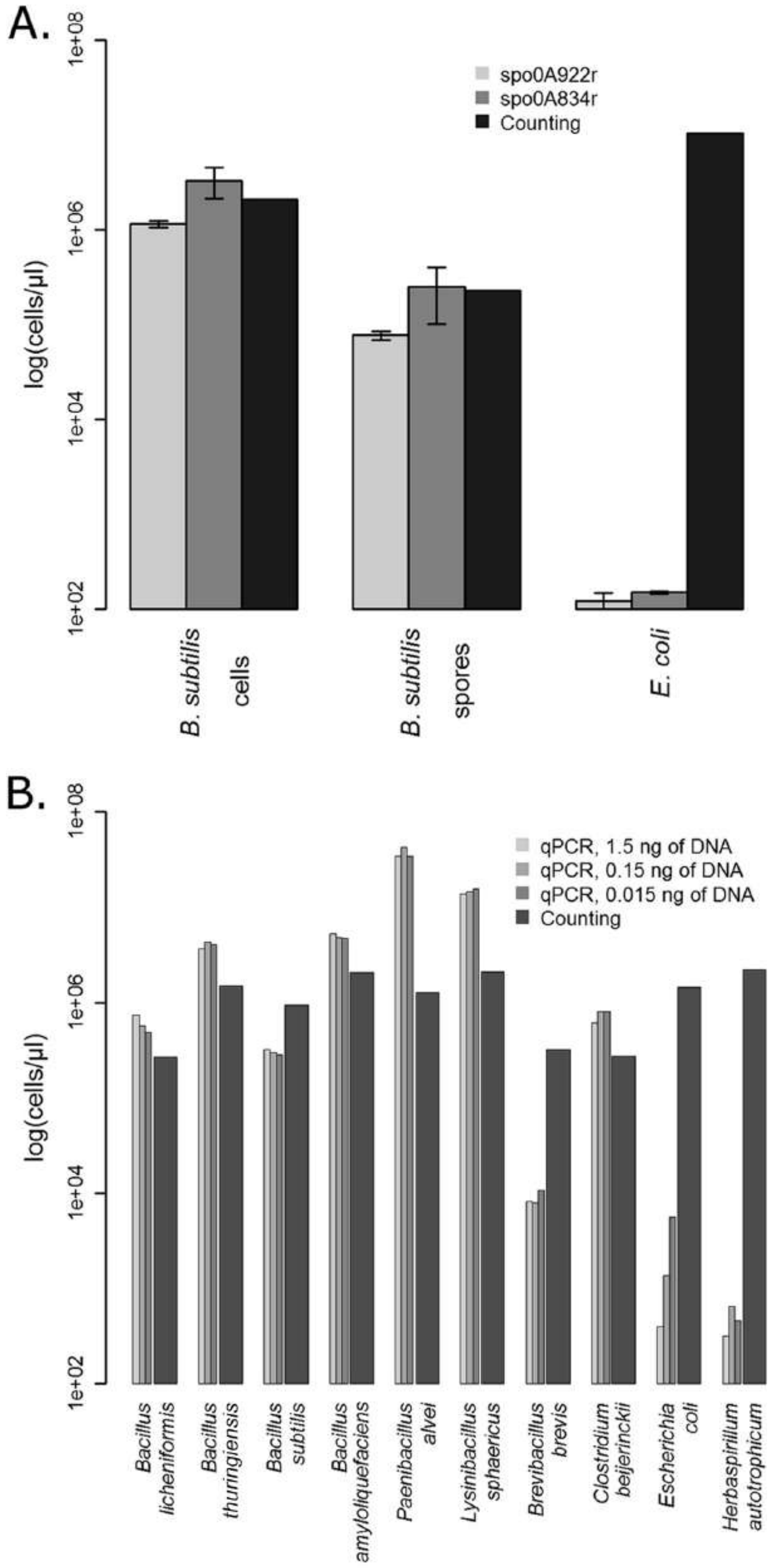
FIG 3 *In silico* test of the two primer sets assessed by NCBI primer blast. The height of the bars represents the number of mismatches detected in the alignment of the primers. The gray zone corresponds to the forward primer whereas the white zone to the reverse. The width of the bars represents the relative number of hits in the NCBI database for each genus.

Finally, the accuracy of the *spo0A* qPCR quantification (second primer set only) was evaluated by comparison with a qPCR for the 16S rRNA gene. For most of the strains that gave a positive amplification signal (above 10^2 copies of the *spo0A* gene per μl of DNA sample), quantification of the *spo0A* gene was consistent with the 16S rRNA gene results (taking into account the number of copies of the 16S rRNA gene in each strain). This was, however, not the case for *B. brevis*, *G. thermoleovorans*, or *A. acidocaldarius*, for

TABLE 4 PCR and qPCR amplification results for the two sets of primers for the *spo0A* gene[a]

			PCR amplification		qPCR amplification		*spo0A* vs 16S rRNA qPCR	
Strain	Species	EFF	Set 1	Set 2	Set 1	Set 2	Set 1	Set2
Neu1062	***Bacillus licheniformis***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↓ 1.38**
Neu1070	***Bacillus thuringiensis***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 6.41**
Neu1121	***Bacillus subtilis***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↓ 4.55**
Neu1261	***Bacillus amyloliquefaciens***	**Yes**	**ND**	**Yes**	**ND**	**Yes**	**ND**	**↑ 9.68**
Neu91	***Paenibacillus alvei***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 1.71**
Neu1005	***Paenibacillus alvei***	**Yes**	**ND**	**Yes**	**ND**	**Yes**	**ND**	**↑ 2.18**
Neu1003	***Lysinibacillus sphaericus***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 5.1**
Neu1014	*Brevibacillus brevis*	Yes	Yes	Yes	No	Yes	ND	↓ 9,906
Neu1040	***Brevibacillus thermoruber***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↓ 2.92**
Neu1149	***Geobacillus thermoglucosidasius***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 1.2**
HFF-3	*Geobacillus thermoleovorans*	Yes	Yes	Yes	Yes	Yes	ND	↓ 176,008
B 18	*Alicyclobacillus acidocaldarius*	Yes	Yes	Yes	Yes	Yes	ND	↓ 2,210
JWO13	*Sulfobacillus acidophilus*	Yes	No	No	No	No	ND	ND
Neu28	***Clostridium beijerinckii***	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**ND**	**↑ 7.52**
Neu1155	***Clostridium pasteurianum***	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**ND**	**↑ 3.14**
MI-1	*Desulfotomaculum reducens*	Yes	ND	No	ND	No	ND	ND
Neu1021	*Escherichia coli*	No	No	No	No	No	ND	ND
Neu1037	*Lactococcus lactis* subsp. *lactis*	No	No	No	No	No	ND	ND

[a] Set 1, spo0A655f-spo0A923r; set 2, spo0A655f-spo0A834r. Sixteen EFF strains and two non-endospore formers were tested. In the qPCR amplification test, results were considered positive when more than 10^2 copies/μl were detected, which corresponded to the lowest dilution used to establish the standard curve. Strains for which data are shown in boldface are those for which quantification for *spo0A* (using the second set of primers) was consistent with that obtained for the 16S rRNA gene. To assess the consistency of results, the ratio between the qPCR quantification of the 16S rRNA gene and the *spo0A* gene was computed and then divided by the number of copies of the 16S rRNA gene found in the strain (retrieved from the rrnDB database [45]). Finally, an accurate amplification was considered when this value was between 0.1 and 10 (see Material and Methods). ↑, overestimation of the number of *spo0A* gene copies; ↓, underestimation of the number of *spo0A* gene copies. ND, not determined.



C.

	Bacillus thuringiensis	Bacillus subtilis	Bacillus amyloliquefaciens	Paenibacillus alvei	Lysinibacillus sphaericus	Brevibacillus brevis	Clostridium beijerinckii	Escherichia coli *	Herbaspirillum autotrophicum *
Bacillus licheniformis	1.11	1.19	1.31	0.88	0.92	2.05	0.81	1.35	1.34
Bacillus thuringiensis		0.84	1.02	0.97	1.04	0.97	1.03	0.95	0.87
Bacillus subtilis			1.22	1.10	1.00	1.00	0.97	1.02	1.18
Bacillus amyloliquefaciens				1.02	1.04	1.11	1.11	1.35	1.39
Paenibacillus alvei					1.20	0.99	1.01	1.00	1.04
Lysinibacillus sphaericus						1.01	0.97	1.07	1.12
Brevibacillus brevis							0.85	0.95	1.37
Clostridium beijerinckii								0.79	0.78

which the results indicated either an over- or underestimation of the number of cells.

Another reliability test of the quantification of endospore formers with these primer sets was carried out by comparing qPCR results with the quantification obtained by microscopic cell counting (Fig. 4). For vegetative cells and endospores of *B. subtilis*, the quantification by qPCR with both primer sets gave values on the same order of magnitude as those obtained by counting. The standard deviations (Fig. 4A, error bars) were much lower for the first set of primers (spo0A655f-spo0A834r) than for the second set (spo0A655f-spo0A834r). The qPCR quantification of a non-EFF, *E. coli*, gave no significant qPCR amplification signal, as expected. Additional comparisons that considered different initial DNA concentrations for eight other EFF strains (Fig. 4B) showed that the quantification by qPCR was also on the same order of magnitude as the microscopic counting, even for the lowest DNA dilution (0.015 ng of DNA). Only *P. alvei* was surprisingly overestimated by qPCR or underestimated by counting, while *B. brevis* was clearly underestimated, which is consistent with the comparison between the qPCR results for *spo0A* and the 16S rRNA gene (Table 4). The two non-EFF strains tested in Fig. 4B (*E. coli* and *H. autotrophicum*) also gave no significant qPCR signal (below 10^2 copies µl of the *spo0A* gene), as expected. Finally, the potential interference that occurs when quantifying a mixture of EFF was evaluated by combining equal amounts of DNA from two EFF strains or one EFF and one non-EFF strain (Fig. 4C). A ratio was calculated between the theoretical expected value (the sum of the individual amplifications) and the obtained measurement. Except for the combination of DNA from *B. brevis* and *B. licheniformis*, the ratio was between 0.81 and 1.31. The same range was obtained when DNA from a non-EFF strain was added, for which only the EFF contributed to the final quantification, indicating that there is no or little interference between DNA from different origins when using this *spo0A* qPCR quantification assay.

Validation with environmental samples. A very good correlation (R^2, 0.926) was obtained when the qPCR quantification was compared to pyrosequencing results in determining the proportion of EFF in bacterial communities found in sediments of Lake Geneva (Fig. 5). The slope of the regression function (7.44) showed that the ratio calculated by qPCR was lower than the one calculated by pyrosequencing. Four samples that contained a very large proportion of EFF, close to 50% of the sequences by pyrosequencing, corresponded to the larger ratios between qPCR for the *spo0A* gene relative to the 16S rRNA gene (6.5%). In seven samples, the ratio of EFF measured by pyrosequencing varied between 19.88% and 38.93% and corresponded to intermediate values when considering qPCR ratios (2.22% to 3.93%). Finally, in the other 13 samples considered, the fraction of EFF according to the

FIG 4 Comparison between a classical quantification with a Neubauer improved counting chamber and the *spo0A* gene qPCR assay. A. Comparison of the quantification with the two sets of primers on vegetative cells and spores of *Bacillus subtilis*, as well as with cells of *Escherichia coli* as a negative non-endospore forming control. B. Comparison of quantification for eight other EFF strains and two non-EFF strains with the second set of primers (spo0A655f-spo0A834r) and different initial concentrations of DNA. C. Ratio calculated for the quantification with equal initial DNA concentration for mixtures of two-strains. The values represent the ratio between the theoretical value (sum of the individual quantifications) and the obtained measurement. *, only the EFF contributed to the final quantification.

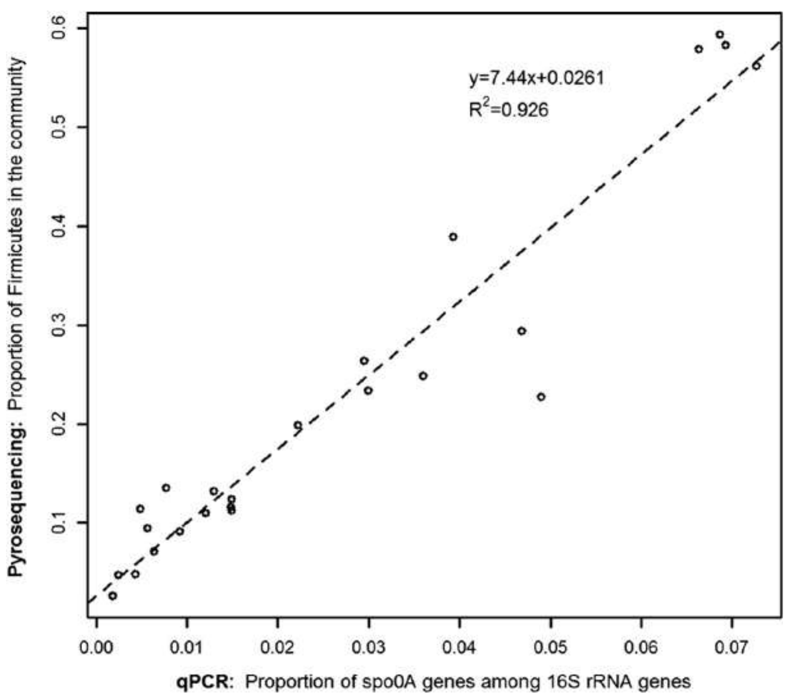


FIG 5 Correlation between two methods to quantify the ratio of endospore-formers in complex microbial communities. The proportion of endospore-forming bacteria in the community has been evaluated by two independent approaches. First (on the *y* axis): QIIME taxonomical identification of OTUs from pyrosequencing (V1-V3 region of the 16S rRNA gene) and quantification of *Firmicutes* in the overall sequences. Second (on the *x* axis): ratio of qPCR on the *spo0A* gene and qPCR on the 16S rRNA gene (V3 region). The slope of 7.44 represents approximately the bias induced by the number of 16S rRNA gene copies for both methods of quantification (see discussion section for details).

pyrosequencing data ranged from 2.65% to 13.52%. For those samples, the qPCR ratio varied from 0.18% to 0.76%.

DISCUSSION

The seven genes evaluated here are all involved in essential steps of the endosporulation pathway. The *spo0A* gene is the major regulator for the initiation of sporulation. The *gpr* gene encodes a protease active during germination (6). Stage IV sporulation protein B (*spoIVB*) is involved in the activation of the sigma-K factor (final transcription factor expressed during the sporulation cascade in the mother cell); stage V sporulation proteins A (spo *VAC* and *spoVAD*) have been suggested to be important transmembrane proteins that are probably involved in DPA transport into the endospore (51). The stage V sporulation protein T is a global regulator activated by sigma-G (1). Finally, the gene coding for the sigma-H factor is another positive regulator of sporulation in a way similar to *spo0A* (6). The prevalence rates of *spo0A*, *spoVT*, and *spoIVB* have already been observed in a genome-wide comparative study of the evolution of the endosporulation pathway (1). Here, we showed that the same is true for *spoVAD*, *spoVAC*, and *gpr*. However, in the case of the sigma-H factor, this is clearly not the case, as this factor directs the transcription of several genes not only during the initial steps of endospore formation but also during the transition from exponential growth to stationary phase and the entry into the state of genetic competence (52). As a consequence, homologs of this gene are also found in phyla like *Thermotogae* and *Fusobacteria* (data not shown), which are not related to EFF, making this gene unsuitable for the design of endosporulation-specific primers.

The two primer sets that were designed here are promising as universal primers for the detection of endospore-forming bacteria. For example, the results of the *in silico* tests gave a positive match not only for the major groups of endospore formers (*Bacillus*, *Clostridium*, and *Geobacillus*) but also for some less-studied groups, such as *Haloanaerobiales*, *Thermoanaerobacteriales*, *Lachnospiraceae*, *Syntrophomonadaceae*, *Peptococcaceae*, *Natranaerobiaceae*, and *Heliobacteriaceae*. However, a general concern for these *in silico* studies is the bias due to the number of DNA sequences that are deposited in published online databases. As an example, the higher representation rates of *Bacillus*, *Clostridium*, *Paenibacillus*, and *Geobacillus* found in this work are probably not due to better specificities of the primers for these genera but to the higher number of related sequences in genomic databases. As a consequence, the widths of the bars in Fig. 3 give only a vague indication of the stringency of the primers.

In laboratory experiments, the two primer sets designed (set 1, spo0A655f -spo0A923r, and set 2, spo0A655f-spo0A834r) gave good *spo0A* amplification results for the major groups of *Bacilliales*. However, as expected from the narrowed spectrum observed in the PCA and in the *in silico* analysis, the set spo0A655f-spo0A923r did not amplify *Clostridium* spp. efficiently, but this was improved during the design of primer set 2 (spo0A655f-spo0A834r). This was also the case for *B. brevis*, for which the lack of positive amplification with primer set 1 can be explained by the poor specificity of the reverse primer, as demonstrated in the PCA analysis (Fig. 2) and the *in silico* tests (Fig. 3). This bias was overcome by using primer set 2, which resulted in positive amplification of this strain in classical and quantitative PCRs.

No amplification was obtained for *S. acidophilus* or *D. reducens* with any of the primer sets. The full genomes of these two EFF species are available (*S. acidophilus* [53], *D. reducens* [54]), and a homolog protein structure to Spo0A was found in the TIGR02875 data set. When aligned with sequences from the refined reference database (80 *spo0A* sequences used for primer design), the sequence at the annealing site matched exactly the sequences of the primers that were designed. The PCR inhibitory effect of iron (55), which was present in the culture media of these two strains, probably explains the negative results that were obtained, and we have experimental evidence showing the inhibitory effect for these DNA extracts (data not shown).

The validation of the qPCR quantification in environmental samples showed that the ratio calculated by qPCR was lower than the one calculated by pyrosequencing. The first explanation for this is the difference in the number of copies of 16S rRNA genes compared to *spo0A*. As the ribosomal (rrn) operon is normally found in several copies, the quantification of this gene by qPCR gives values that are artificially greater than the real number of bacterial cells in the community, lowering the obtained qPCR ratio. Furthermore, the average number of rrn operon copies depends on the group of bacteria. An average value of 4.3 copies of 16S rRNA genes was found in the rrnDB in all bacterial phyla (45). This value is substantially greater (average, 7.01 copies) for the phylum *Firmicutes*. As the pyrosequencing was carried out for the 16S rRNA gene, the difference of average 16S rRNA copy numbers between bacteria and *Firmicutes* generated another bias by overestimating the ratio of EFF in the community by a factor of 7.01/4.3, or 1.63. Thus, these two bias sources combined (the overestimation of the 16S rRNA gene copy numbers an average of 4.3 times and the ratio calculated by pyrosequencing being 1.63 times greater) led to a correction factor of 7.01, which was very close to the slope of 7.44 obtained experimentally (Fig. 5).

The detection limit of this new method depends on different factors. The quantity of sample used for the DNA extraction as well as the extraction methodology are probably the more important ones. DNA extraction is very important, because cells from EFF are harder to lyse than cells from other bacterial groups. This is particularly true if most of the cells are present in the form of endospores. Furthermore, in order to have quantitatively comparable results, one must pay particular attention to the repeatability of the extraction methodology in terms of yield and quality of the final product. In our experimental evaluation, the quantification was linear down to 10^2 copies/μl (quantification curve) and we could detect down to 10^4 copies of *spo0A* gene/g of wet sediment from the environmental samples. Within the range covered by the standard curve, the qPCR quantification of the *spo0A* gene in EFF strains was independent of the amount of initial DNA used, down to 0.015 ng of DNA, and it was consistent with results of direct microscopic counting of cells prior DNA extraction (Fig. 4). No significant bias in the quantification were observed when mixing DNA from different strains, confirming the selectivity of this qPCR-based method for the quantification of a complex mixture of EFF strains, like those found in environmental samples. Thus, by concentrating DNA prior to qPCR quantification or by optimizing the DNA extraction protocol, this detection limit could probably be lowered one or more orders of magnitude. In comparison, the detection limit reported in the literature for the DPA assay is on the order of 10^5 endospores/g of soil sample. Furthermore, the DPA assay does not quantify vegetative cells, only endospores. As a consequence, these two techniques can be seen as complementary in order to obtain a full overview of EFF in the environment and their morphological state.

In conclusion, the experimental evaluation suggests that the quantification of endospore-forming bacteria by qPCR using the primers developed for the *spo0A* gene gives results that are in agreement with the pyrosequencing quantification of EFF in complex microbial communities. In addition, quantification of EFF by qPCR is fast, cheap, and very reliable compared to other modern techniques, such as next-generation sequencing. Although the information obtained is not identical (pyrosequencing has the advantage of species identification), in some cases a simple quantification can be sufficiently informative. Examples of this are deep oceanic drilling projects (56), exobiology research (8) paleontology (57), study of polluted environments (61), the packaging industry (58), and sterility assessments in the food and pharmaceutical industries (59). Also, although direct clinical or industrial examples could not be tested here, the level of sensitivity obtained with environmental samples, which were notable for their high level of PCR-inhibitory substances, makes likely an even better detection potential for this approach in much cleaner anthropogenic environments.

ACKNOWLEDGMENTS

We thank Ferring Pharmaceutical and the Consulat honoraire de la Fédération de Russie à Lausanne for funding this project through elemo grant 301, as well as the EPFL for coordinating the elemo project. Funding from the Swiss National Science Foundation (grant 31003A_132358) is also acknowledged.

We thank Patricia L. Siering from Department of Biological Sciences, Humboldt State University, CA, for the strains *Geobacillus thermoleovorans* B18, *Alicyclobacillus acidocaldarius* strain JWO13, and *Sulfobacillus acidophilus*. We thank Elena Dalla-Vecchia from the Ecole Polytechnique Federale de Lausanne for providing *Desulfotomaculum reducens*.

REFERENCES

1. **De Hoon MJL, Eichenberger P, Vitkup D.** 2010. Hierarchical evolution of the bacterial sporulation network. Curr. Biol. **20:**R735–R745.
2. **De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, Whitman WB.** 2009. Bergey's manual of systematic bacteriology, 2nd ed. Springer, Dordrecht, Netherlands.
3. **Nicholson W.** 2002. Roles of *Bacillus* endospores in the environment. Cell. Mol. Life Sci. **59:**410–416.
4. **Hageman J, Shankweiler G, Wall P, Franich K, Cowan MG, Cauble S, Grajeda J, Quinones C.** 1984. Single, chemically defined sporulation medium for *Bacillus subtilis* growth, sporulation, and extracellular protease production. J. Bacteriol. **160:**438–441.
5. **Tracy BP, Jones SW, Papoutsakis ET.** 2011. Inactivation of σ^E and σ^G in *Clostridium acetobutylicum* illuminates their roles in clostridial-cell-form biogenesis, granulose synthesis, solventogenesis, and spore morphogenesis. J. Bacteriol. **193:**1414–1426.
6. **Errington J.** 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. **1:**117–126.
7. **Nicholson W, Fajardo-Cavazos P, Rebeil R, Slieman T, Riesenman P, Law J, Xue Y.** 2002. Bacterial endospores and their significance in stress resistance. Antonie Van Leeuwenhoek **81:**27–32.
8. **Yung PT, Kempf MJ, Ponce A.** 2006. A rapid single spore enumeration assay, p 751–763. Aerospace Conference, 2006 IEEE. Institute of Electrical and Electronics Engineers, New York, NY. doi:10.1109/AERO.2006.1655788.
9. **Roberts MS, Cohan FM.** 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution **49:**1081–1094.
10. **Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, ReysenbachA-L, Smith VH, Staley JT.** 2006. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. **4:**102–112.
11. **Mandic-Mulec I, Prosser JI.** 2011. Diversity of endospore-forming bacteria in soil: characterization and driving mechanisms, p 31–59. *In* Logan NA, De Vos P (ed), Endospore-forming soil bacteria. Springer-Verlag, Berlin, Germany.
12. **McSpadden Gardener BB.** 2004. Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. Phytopathology **94:**1252–1258.
13. **Logan NA.** 2012. *Bacillus* and relatives in foodborne illness. J. Appl. Microbiol. **112:**417–429.
14. **Bibek R.** 2004. Fundamental food microbiology, 3rd ed. CRC Press LLC, Boca Raton, FL.
15. **De Clerck E, Vanhoutte T, Hebb T, Geerinck J, Devos J, De Vos P.** 2004. Isolation, characterization, and identification of bacterial contaminants in semifinal gelatin extracts. Appl. Environ. Microbiol. **70:**3664–3672.
16. **Heyndrickx M.** 2011. The importance of endospore-forming bacteria originating from soil for contamination of industrial food processing. Appl. Environ. Soil Sci. **2011:**11.
17. **Seale RB, Bremer PJ, Flint SH, McQuillan AJ.** 2010. Characterization of spore surfaces from a *Geobacillus* sp. isolate by pH dependence of surface charge and infrared spectra. J. Appl. Microbiol. **109:**1339–1348.
18. **Brown KL.** 2000. Control of bacterial spores. Br. Med. Bull. **56:**158–171.
19. **Fichtel J, Koster J, Scholz-Bottcher B, Sass H, Rullkotter J.** 2007. A highly sensitive HPLC method for determination of nanomolar concentrations of dipicolinic acid, a characteristic constituent of bacterial endospores. J. Microbiol. Methods **70:**319–327.
20. **Brandes Ammann A, Kolle L, Brandl H.** 2011. Detection of bacterial endospores in soil by terbium fluorescence. Int. J. Microbiol. **2011:**435281. doi:10.1155/2011/435281.
21. **Tabor M, Macgee J, Holland J.** 1976. Rapid determination of dipicolinic acid in spores of *Clostridium* species. Appl. Environ. Microbiol. **31:**25–28.
22. **Beverly MB, Basile F, Voorhees KJ, Hadfield TL.** 1996. A rapid approach for the detection of dipicolinic acid in bacterial spores using pyrolysis mass spectrometry. Rapid Commun. Mass Spectrom. **10:**455–458.
23. **Fichtel J, Koster J, Rullkotter J, Sass H.** 2007. Spore dipicolinic acid contents used for estimating the number of endospores in sediments. FEMS Microbiol. Ecol. **61:**522–532.
24. **Yang W-W, Ponce A.** 2011. Validation of a *Clostridium* endospore viability assay and analysis of Greenland ices and Atacama desert soils. Appl. Environ. Microbiol. **77:**2352–2358.

25. **Rueckert A, Ronimus RS, Morgan HW.** 2006. Development of a real-time PCR assay targeting the sporulation gene, spo0A, for the enumeration of thermophilic bacilli in milk powder. Food Microbiol. **23**:220–230.
26. **Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, Sindelka R, Sjoback R, Sjogreen B, Strombom L, Stahlberg A, Zoric N.** 2006. The real-time polymerase chain reaction. Mol. Aspects Med. **27**:95–125.
27. **Schnoes AM, Brown SD, Dodevski I, Babbitt PC.** 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput. Biol. **5**:e1000605. doi:10.1371/journal.pcbi.1000605.
28. **Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O.** 2007. TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res. **35**:D260–D264.
29. **Stothard P, Wishart DS.** 2006. Automated bacterial genome analysis and annotation. Curr. Opin. Microbiol. **9**:505–510.
30. **Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC.** 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. **40**:D115–D122.
31. **Rice P, Longden I, Bleasby A.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16**:276–277.
32. **Katoh K, Misawa K, Kuma K, Miyata T.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**:3059–3066.
33. **R Development Core Team.** 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
34. **Oksanen J.** 2011. Multivariate analysis of ecological communities in R: vegan tutorial. R Foundation for Statistical Computing, Vienna, Austria.
35. **Sharma S, Radl V, Hai B, Kloos K, Fuka MM, Engel M, Schauss K, Schloter M.** 2007. Quantification of functional genes from procaryotes in soil by PCR. J. Microbiol. Methods **68**:445–452.
36. **Smith CJ, Osborn AM.** 2009. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol. Ecol. **67**:6–20.
37. **Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M.** 2010. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. J. Mol. Diagn. **12**:109–117.
38. **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 Bioinformatics **13**:134.
39. **Johnson D.** 1995. Selective solid media for isolating and enumerating acidophilic bacteria. J. Microbiol. Methods **23**:205–218.
40. **Zobell CE.** 1941. Studies on marine bacteria: the cultural requirements of heterotrophic aerobes. J. Mar. Res. **4**:42–75.
41. **Leighton TJ, Doi RH.** 1971. The stability of messenger ribonucleic acid during sporulation in *Bacillus subtilis*. J. Biol. Chem. **246**:3189–3195.
42. **Dineen SM, Aranda R, Anders DL, Robertson JM.** 2010. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. J. Appl. Microbiol. **109**:1886–1896.
43. **Bakke I, De Schryver P, Boon N, Vadstein O.** 2011. PCR-based community structure studies of bacteria associated with eukaryotic organisms: a simple PCR strategy to avoid co-amplification of eukaryotic DNA. J. Microbiol. Methods **84**:349–351.
44. **Muyzer G, Dewaal E, Uitterlinden A.** 1993. Profiling of complex microbial populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16S ribosomal RNA. Appl. Environ. Microbiol. **59**:695–700.
45. **Lee ZM-P, Bussema C, Schmidt TM.** 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. **37**:D489–D493.
46. **Li H, Zhang Y, Li D, Xu H, Chen G, Zhang C.** 2009. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. Soil Biol. Biochem. **41**:954–968.
47. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7**:335–336.
48. **McDonald D, Clemente J, Kuczynski J, Rideout J, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso J.** 2012. The biological observation matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience **1**:7.
49. **Abecasis AB, Serrano M, Alves R, Quintais L, Pereira-Leal JB, Henriques AO.** 2013. A genomic signature and the identification of new sporulation genes. J. Bacteriol. **195**:2101–2115.
50. **Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD.** 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **38**:D473–D479.
51. **Vepachedu VR, Setlow P.** 2005. Localization of SpoVAD to the inner membrane of spores of *Bacillus subtilis*. J. Bacteriol. **187**:5677–5682.
52. **Britton RA, Eichenberger P, Gonzalez-Pastor JE, Fawcett P, Monson R, Losick R, Grossman AD.** 2002. Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. J. Bacteriol. **184**:4881–4890.
53. **Li B, Chen Y, Liu Q, Hu S, Chen X.** 2011. Complete genome analysis of *Sulfobacillus acidophilus* strain TPY, isolated from a hydrothermal vent in the Pacific Ocean. J. Bacteriol. **193**:5555–5556.
54. **Junier P, Junier T, Podell S, Sims DR, Detter JC, Lykidis A, Han CS, Wigginton NS, Gaasterland T, Bernier-Latmani R.** 2010. The genome of the Gram-positive metal- and sulfate-reducing bacterium *Desulfotomaculum reducens* strain MI-1. Environ. Microbiol. **12**:2738–2754.
55. **Teng F, Guan Y, Zhu W.** 2008. A simple and effective method to overcome the inhibition of Fe to PCR. J. Microbiol. Methods **75**:362–364.
56. **Lomstein BA, Langerhuus AT, D'Hondt S, Jorgensen BB, Spivack AJ.** 2012. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature **484**:101–104.
57. **Vreeland R, Rosenzweig W, Powers D.** 2000. Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. Nature **407**:897–900.
58. **Pirttijarvi TSM, Graeffe TH, Salkinoja-Salonen MS.** 1996. Bacterial contaminants in liquid packaging boards: assessment of potential for food spoilage. J. Appl. Bacteriol. **81**:445–458.
59. **Yung PT, Ponce A.** 2008. Fast sterility assessment by germinable-endospore biodosimetry. Appl. Environ. Microbiol. **74**:7669–7674.
60. **Kibbe WA.** 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. **35**:W43–W46.
61. **Sauvain L, Bueche M, Junier T, Masson M, Wunderlin T, Kohler R, Gascon Diez E, Loizeau J-L, Tercier-Waeber M-L, Junier P.** Bacterial communities in trace metals contaminated lake sediments are dominated by endospore forming bacteria. Aquat. Sci., in press.

1 gtgawgkbgddbrwdbbhdkbnnnvnrrbnndndddrnndnnnvngganrnnnnnnnnnnnnnnnnnnwnnvnnTNNBNNTNDBNGAYGAYMAYVDNVRDH

101 YNNBNNNNNYNHTNNNNVVNDWNNTNNNNNNNNNWNNVNRAHNTNNNNNYNNBNRVNVNNKSNNNNVRYGGNNNNVWNDBNNDNNVNNNNNTNNNNNVNnv

1 2 3

201 hvsmnnnNNNHYNGAYNTNNYNNTNHTNGAYDTNRTHATGCYNNDNHTNGAYGGNNTNRSNGYNHTNGRnnvnnknnnnnvnnNNNNnnntmaynnnnmv

4

301 nnvnnnNVBNNNNDYNNTHDTNNTNWSNGCNDYNRRNMRNGANNNNDTNRYNVNDVNNGCNNYNNNNHTNGGNGYNDVNTAYTWYRYNNTNAARCCNTTY

401 RRHNWNNHNNHNNTNDBNNVNMDNRTNHNNVDNNBnnnndnnvnndnndnnnndnnnnhnnnnnnnnnnnnnnnnnbnndhvvnnnnnnccatatcdydnh

501 dhndnhnhvdnhnvnnnnnvnvnnnnnrdnrnnnnddhwtcaayctaacnnnnnnnhnnnnnnnnnnnhnnnnnndnnnnhnnnnnnnnnnnnnnnnnnn

5 6 14 7 13

601 nnnnnnnnnnnnnnnnvnhTNGRNNNNNNNRTNACNNVNDTNHTNMAHSADVTNGGNRTNCCNSCNCAYRTNRMVGGNTAYHHNTAYHTDMGNGRNKCNR

8

701 THNNNHDNDYNDHNVRNRANNHNDVNNWNHTNDVNDSNRTNACNAARNHNYTNTAYCCNNNNRTHGCNVMVVRDYWYVMNACNRCNNSNHNNMRNGTNGA

9 11 10

801 RMGNGSNAYHMGNM**AYKCNATHGARNKNGCNTGG**DVNMGNRRNHRNNHNGANDBNNTNNVNNVNNWNTTYGGNYAYWCNRTNNVNNNNDNNMRDNV**NAAR**

15 12

901 **CCNACNAAYWSNGARTK**YRTHGCNHTDVTNDCNGAYMRNHTNMGNNTNDnnngcaNNVDNvhdNrmgtwsNNNNhwyttttcaacaaggtggttttggag

tnTRR

Figure S1. Consensus sequence corresponding to an alignment of 80 selected sequences of *spo0A* gene. The more conserved regions are highlighted in grey when they contain at least 9 conserved bases in regions defined by 20 consecutive base-pairs. Bases were in bold characters and underlined when they contain up to 11 conserved bases. Potential annealing sites for the development of primers are identified by the small numbers above the sequence.

```
# Sequences sorting:
#
# This script is making a sorting of sequences based on a distance matrix calculated elsewhere.
# Only the closest (more related sequences are keeped together and output in a new text file.
# This text file (list of sequences) can then be used to create a new reference fasta file with the seqret
# command from the EMBOSS package
#

distmat=read.table("spo0a_epali.distmat",header=F,dec=".",skip=8,fill=T)
noms=read.table("seqnames.txt",skip=1,header=F,fill=T,dec=".",sep="",flush=T)
numero=read.table("seqnum.txt",skip=1,header=F)

groupe=0

# Define the distance threshold value in the line below!
seuil=2
n.lignes=dim(distmat)[1]
num.select=numero[,1]

for (i in c(1:n.lignes))
{
 groupe[i]=paste(numero[(c(which(distmat[i,c(1:n.lignes)]<=seuil)[-1])+i-1),1],":",
                       noms[(c(which(distmat[i,c(1:n.lignes)]<=seuil)[-1])+i-1),1]," ",
                       noms[(c(which(distmat[i,c(1:n.lignes)]<=seuil)[-1])+i-1),2],
                       sep="",collapse=";")

 num=numero[(c(which(distmat[i,c(1:n.lignes)]<=seuil)[-1])+i-1),]
 if (length(num)>0)
 {
  for (j in c(1:length(num)))
  {
   num.select=num.select[c(which(num.select!=num[j]))]
  }
 }
}

write.table(paste("fasta::Spo0A_epali.fasta:",num.select,sep=""),file="selected_seq.txt",row.names=F,
              col.names=F)
```

```
# Search for best annealing sites
#
# This script is doing the following tasks:
#           1) For each position, it computes the number of conserved bases in the 20 following bases
#           2) Generate a CSV file to visualise the bases within selected annealing postitions
#           3) Do a cluster and NMDS analyses on most promising annealing sites
#           4) Do a principal component analyses on three selected sites and plot an estimate of the
#              extend of the more important genera of EFB as well as the coverage of the degenerated
#              primers that were designed.
#

library(bio3d)
library(vegan)
library(plotrix)
setwd("Z:/Matthieu/Primers design/New primers/All Spo0A in JGI")

consensus=read.fasta("consensus_sel2ali.fasta",to.upper=F)
DNA=read.fasta("Spo0A_sel2ali.fasta",to.upper=T)
variables=read.csv("spo0a_sel2ali.csv",header=T,sep=";",dec=".")

primer.length=20
more.conserved=site=0
nbr.sites=30


# Conservation signal for annealing sites of 20bp long
for (i in c(1:(length(consensus$ali)-primer.length)))
{
  more.conserved[i]=length(which(consensus$ali[c(i:(i+primer.length-1))]=="A"))+
                          length(which(consensus$ali[c(i:(i+primer.length-1))]=="C"))+
                          length(which(consensus$ali[c(i:(i+primer.length-1))]=="G"))+
                          length(which(consensus$ali[c(i:(i+primer.length-1))]=="T"))
}

for (i in c(1:nbr.sites))
{
  site[i]=which.max(more.conserved)
  more.conserved[site[i]]=-1
}

# To be inserted by hand!!!
# List of the more promising sites for primer design
site_prime=matrix(nrow=15,ncol=2)
site_prime[1,]=c(222,243)
site_prime[2,]=c(235,255)
site_prime[3,]=c(238,257)
site_prime[4,]=c(380,399)
site_prime[5,]=c(649,669)
site_prime[6,]=c(652,671)
site_prime[7,]=c(661,680)
site_prime[8,]=c(739,759)
site_prime[9,]=c(815,834)
site_prime[10,]=c(897,917)
site_prime[11,]=c(816,834)
site_prime[12,]=c(904,923)
site_prime[13,]=c(667,686)
site_prime[14,]=c(655,674)
site_prime[15,]=c(901,923)
```

```
#
# Compute Jaccard distance and do a cluster analysis on each site of "liste" variable
# Compute and plot an NMDS analysis to see the relationship between the sequences
# A pdf file is generated containing both representations
#

liste=c(15)
scale.factor.env <- 0.9
cex.pt=0.5
#par(mar=c(4,4,1,2),oma=c(0,2,1,0))

for (m in liste)
{
site_prim=c(site_prime[m,1]:site_prime[m,2])
file.name=paste("graph3NMDS",m,".pdf",sep="")
pdf(file.name)
j=1
seq=matrix(nrow=dim(DNA$ali)[1],ncol=length(site_prim)*4)
for (i in site_prim)
{
 for (k in c(1:(dim(DNA$ali)[1])))
 {
  if (DNA$ali[k,i]=="A") seq[k,c(j:(j+3))]=c(1,0,0,0)
  if (DNA$ali[k,i]=="T") seq[k,c(j:(j+3))]=c(0,1,0,0)
  if (DNA$ali[k,i]=="G") seq[k,c(j:(j+3))]=c(0,0,1,0)
  if (DNA$ali[k,i]=="C") seq[k,c(j:(j+3))]=c(0,0,0,1)
 }
 j=j+4
}
rownames(seq)=variables[,2]
seq2=seq
rownames(seq2)=paste(variables[,2],": ",variables[,13]," ",variables[,14],sep="")
varexpl=matrix(nrow=dim(DNA$ali)[1],ncol=23,0)
colnames(varexpl)=c(levels(variables[,13]),"Thermoph","Gram+")
for (i in c(1:dim(DNA$ali)[1]))
{
 varexpl[i,variables[i,13]]=1
 if ((variables[i,22] == "yes")||(variables[i,22] == "Extrem")) varexpl[i,22] = 1
 if (variables[i,23] == "plus") varexpl[i,23] = 1
}



data.clus.Q <- hclust(vegdist(seq2, method="jaccard"),"average")
data.NMDS <- metaMDS(seq, trace=F, zerodist="add", k=2, distance="jaccard")
envfit.NMDS <- envfit(data.NMDS, varexpl[,c(1:23)], permu = 1000)

titre=paste("Cluster analysis: position ",site_prime[m,1], " to ",site_prime[m,2],sep="")
plot(data.clus.Q,
    cex.axis=cex.pt,
    cex.lab= cex.pt,
    cex=cex.pt,
    main=titre,
    xlab="Q-mode analysis",
    sub="Distance Jaccard / Agrégation selon la moyenne")
rect.hclust(data.clus.Q, 4, border=c(3:6))
grp.clus.Q <- cutree(data.clus.Q, 4)
```

```
plot(       data.NMDS, type="p",
            choice=c(1,2),
            display=c("sites"),
            cex.axis = cex.pt,
            cex.lab = cex.pt,
            )
text(data.NMDS, choice=c(1,2),display = c("sites"), cex=0.6*cex.pt, col="black")
#text(data.NMDS, choice=c(1,2),display = c("species"), cex=cex.pt, col="red")

plot(   envfit.NMDS,
        col = "green",
        cex.axis = scale.factor.env*cex.pt,
        cex.lab = scale.factor.env*cex.pt,
        cex = scale.factor.env*cex.pt,
        p.max = 0.99
        )
abline(h=0,v=0)

}

dev.off()




#
# Generate a CSV file with the sequence of a selected annealing site for all sequences in the
# database. The position of the bases which are in the site are inserted manually.
#

for (i in data.clus.Q$order)
{
  Sequence=paste(DNA$ali[i,815:834],sep="",collapse=";")
  Ligne_seq=paste(variables[i,2],";",variables[i,13]," ",variables[i,14],";",Sequence,sep="")
  write.table(file="amorce9b.csv",Ligne_seq,append=T,col.names=F,row.names=F)
}
```

```
# Compute a principal componant analyses on the three more promising annealing sites that were
# retained.
# Plot the ordination and estimate the position of the more important genera of endospore-
# formers (Clostridium, Bacillus, Geobacillus), by ploting an ellipse which correspond to the
# standard deviation of the coordinates of the points corresponding to each taxonomic group.
# With the same methodology as above, estimate the level of coverage of the degenerated primers.
# The degenerated sequence of each primer has to be defined prior to this part of the analyse and
# stored into a CSV file!
#

file.name=paste("graph2ACP_comp4.png",sep="")
titre=c("Forward primer spo0A655f","Reverse primer spo0A834r","Reverse primer spo0A923r")


# Read the sequences of each degenerated primers with all the degeneration possibilities!
degen14=read.csv("amorce14alldegen.csv",sep=";",header=F,colClasses="character")
degen9=read.csv("amorce9alldegen.csv",sep=";",header=F,colClasses="character")
degen15=read.csv("amorce15alldegen.csv",sep=";",header=F,colClasses="character")


png(file.name, bg="transparent", width=2400, height= 2400, res=300)
par(mfcol=c(2,2),mar=c(4,4,3,2),oma=c(0,2,1,0))
for (m in c(14,9,15))
{
        degen=xdegen=ydegen=0
        if (m==14) degen=degen14
        if (m==9) degen=degen9
        if (m==15) degen=degen15
        site_prim=c(site_prime[m,1]:site_prime[m,2])
        j=1
        ADN=rbind(DNA$ali[,site_prim],degen)

        seq=matrix(nrow=dim(ADN)[1],ncol=length(site_prim)*4)
        for (i in c(1:length(site_prim)))
        {
                for (k in c(1:(dim(ADN)[1])))
                {
                        if (ADN[k,i]=="A") seq[k,c(j:(j+3))]=c(1,0,0,0)
                        if (ADN[k,i]=="T") seq[k,c(j:(j+3))]=c(0,1,0,0)
                        if (ADN[k,i]=="G") seq[k,c(j:(j+3))]=c(0,0,1,0)
                        if (ADN[k,i]=="C") seq[k,c(j:(j+3))]=c(0,0,0,1)
                }
                j=j+4
        }
        rownames(seq)=c(variables[,2],c(1:dim(degen)[1]))
        seq2=seq
        rownames(seq2)=c(as.character(variables[,13]),as.character(c(1:dim(degen)[1])))
        data.PCA <- rda(seq2[1:80,])
        Clostridium=which(attributes(data.PCA$CA$u[,"PC1"])$names=="Clostridium")
        Bacillus=which(attributes(data.PCA$CA$u[,"PC1"])$names=="Bacillus")
        Geobacillus=which(attributes(data.PCA$CA$u[,"PC1"])$names=="Geobacillus")
        n_Others=c(Clostridium,Bacillus,Geobacillus)

        a=1
        for (d in c(81:dim(seq)[1]))
        {
          xdegen[a]=sum(data.PCA$CA$v.eig[,"PC1"]*seq[d,])
          ydegen[a]=sum(data.PCA$CA$v.eig[,"PC2"]*seq[d,])
          a=a+1
        }
```

```
	plot(data.PCA,display="sites",type="none",scaling=0,xlim=c(-0.3,0.3),ylim=c(-0.3,0.3),
		xlab=paste("PC1 (",format(data.PCA$CA$eig["PC1"]*100/sum(data.PCA$CA$eig),
		digits=3),"%)",sep=""),
		ylab=paste("PC2 (",format(data.PCA$CA$eig["PC2"]*100/sum(data.PCA$CA$eig),
		digits=3),"%)",sep=""),main=titre[which(c(14,9,15)==m)])
	points(x=data.PCA$CA$u[Bacillus,"PC1"],data.PCA$CA$u[Bacillus,"PC2"],pch=2,col=3)
	points(x=data.PCA$CA$u[Clostridium,"PC1"],data.PCA$CA$u[Clostridium,"PC2"],pch=1,col=2)
	points(x=data.PCA$CA$u[Geobacillus,"PC1"],data.PCA$CA$u[Geobacillus,"PC2"],pch=0,col=4)
	points(x=data.PCA$CA$u[-n_Others,"PC1"],data.PCA$CA$u[-n_Others,"PC2"],pch=8,col=grey(0.4))


    if (m==9)
	{
	 xdegen=(xdegen+0.368)*0.19
	 ydegen=(ydegen-0.429)*0.2032
	}
	if (m==14)
	{
	 xdegen=(xdegen+0.44)*0.165
	 y=(ydegen+0.255)*0.214
	}

	if (m==15)
	{
	 xdegen=(xdegen+0.25)*0.119
	 ydegen=(ydegen-0.15)*0.138
	}
	#points(xdegen,ydegen,pch=10,col=6,cex=0.2)

	xclost=mean(data.PCA$CA$u[Clostridium,"PC1"])
	xclostsd=sd(data.PCA$CA$u[Clostridium,"PC1"])
	yclost=mean(data.PCA$CA$u[Clostridium,"PC2"])
	yclostsd=sd(data.PCA$CA$u[Clostridium,"PC2"])
	draw.ellipse(xclost,yclost,xclostsd,yclostsd,border=2)
	text(xclost,yclost,"Clostridium",col=2)

	xbac=mean(data.PCA$CA$u[Bacillus,"PC1"])
	xbacsd=sd(data.PCA$CA$u[Bacillus,"PC1"])
	ybac=mean(data.PCA$CA$u[Bacillus,"PC2"])
	ybacsd=sd(data.PCA$CA$u[Bacillus,"PC2"])
	draw.ellipse(xbac,ybac,xbacsd,ybacsd,border=3)
	text(xbac,ybac,"Bacillus",col=3)

	xgeobac=mean(data.PCA$CA$u[Geobacillus,"PC1"])
	xgeobacsd=sd(data.PCA$CA$u[Geobacillus,"PC1"])
	ygeobac=mean(data.PCA$CA$u[Geobacillus,"PC2"])
	ygeobacsd=sd(data.PCA$CA$u[Geobacillus,"PC2"])
	draw.ellipse(xgeobac,ygeobac,xgeobacsd,ygeobacsd,border=4)
	text(xgeobac,ygeobac,"Geobacillus",col=4)

	xext_prim=mean(xdegen)
	xext_primsd=sd(xdegen)
	yext_prim=mean(ydegen)
	yext_primsd=sd(ydegen)
	draw.ellipse(xext_prim,yext_prim,3*xext_primsd,3*yext_primsd,border=6,lty=3,lwd=1)
	#text(xext_prim,yext_prim,"Primer degen.",col=6)
}
dev.off()
```

Paper IV

In which the link between endospore-forming Firmicutes and trace metals is generalized to more samples which were analysed more in details...

Endosporulation as a resistance mechanism to metal-toxicity in sediments

For Envir science & tech(max 7000 words)

Authors: Matthieu Bueche, Loïc Sauvain, Pilar Junier

Affiliation: Laboratory of Microbiology, Institute of Biology, University of Neuchatel, CH-2000, Neuchatel, Switzerland

* Corresponding author. Mailing address: Laboratory of Microbiology, Institute of Biology, University of Neuchâtel, PO box 158, CH - 2000 Neuchâtel

Phone: +4132 7182244

Fax: +4132 7182231

E-mail: pilar.junier@unine.ch

E-mail co-authors: matthieu.bueche@unine.ch

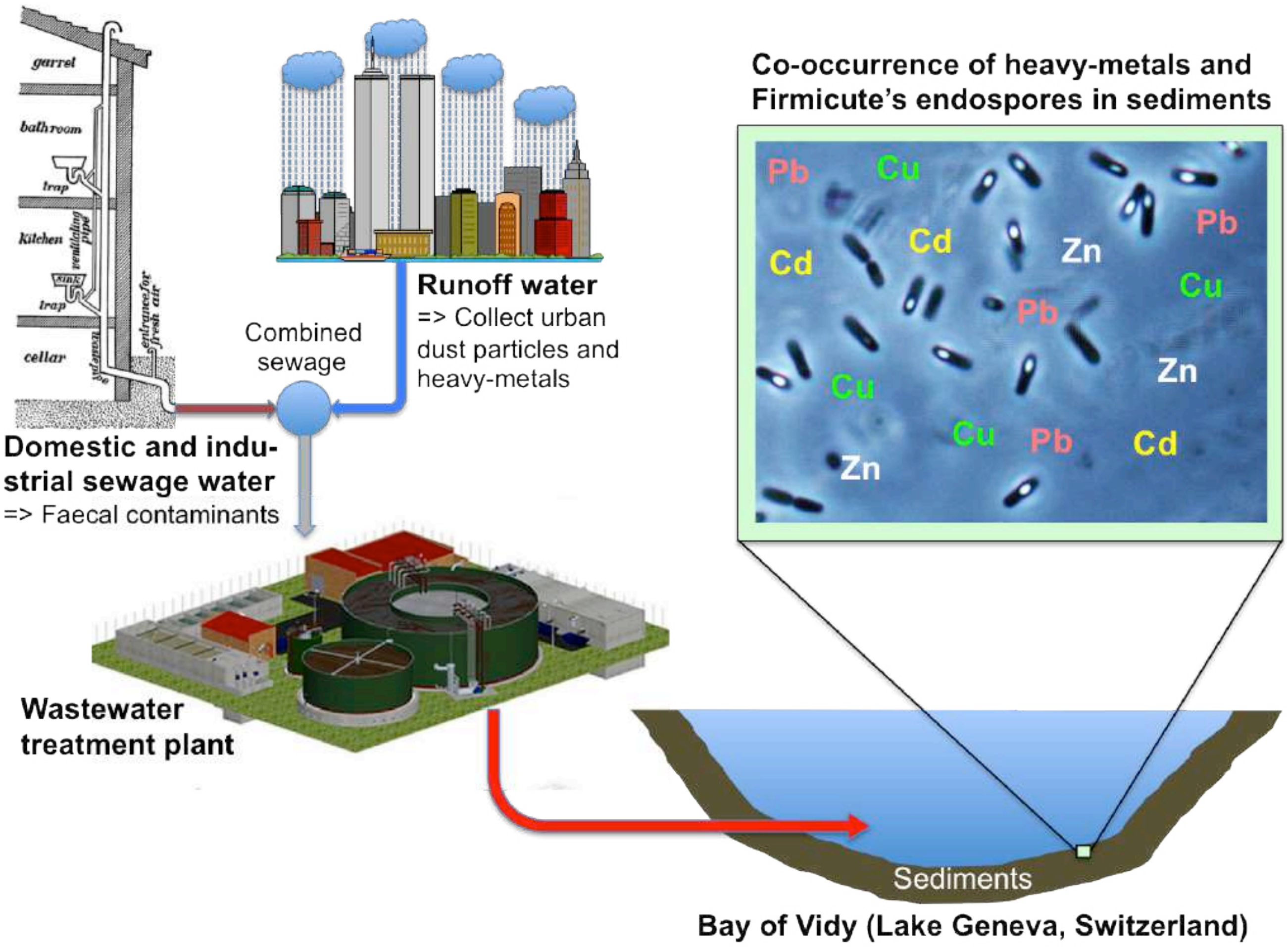
Loic.Sauvain@unil.ch

Running title: Metals and endospores in lake sediments

Key words: Clostridia, EPS, Heavy metals, Lake sediments, Wastewater Treatment Plant

Abstract (max 200 words; 165)

Lake sediments near wastewater treatment plants are considered as potential sinks of contaminants derived from water treatment. Metals as contaminants are of particular importance because of their resilience and the risk they pose to the health of the entire aquatic ecosystem. A link between metal contamination and the relative abundance of endospore-forming Firmicutes was clearly observed in sediments influenced by effluents from a wastewater treatment on the shore of a temperate lake. The specific analysis of the species found in the sediments allowed the identification of Clostridia species, previously known as part of the human microbiome. The absolute abundance of those Clostridia suggests that *in situ* growth is negligeable. The analysis of the relative concentration of dipicolinic acid in sediments showed a significant increase in the ratio of endospore-forming Firmicutes in the state of endospore in contaminated samples. The results indicate that endosporulation explains the persistence of Clostridia compared to all other microbes originated from the wastewater treatment or the native microbiota in contaminated sediments.



Introduction

Metals play an important role in many biochemical processes and are therefore necessary as micronutrients for all living organisms. However, at higher concentrations, they alter physiological functions leading to detrimental effects on growth and reproduction of living organisms[1]. This is also true for microorganisms[2] and in many cases, contamination of an ecosystem by toxic levels of metals leads to a dramatic drop in microbial diversity and a decrease in the metabolic functions provided by the microbial community[3].

Lake sediments are often a sink for metal contamination associated to urban runoff. In cities, the major input for heavy metals into runoff waters is the leaching of building walls and roofs, which are sources of Cu, Cd, Pb, and Zn. Roads also contribute to deposition of Cu and Zn in the environment as they collect all the particles emitted by cars, mainly brakes and tires' wear. Finally, atmospheric deposition contributes also to a large portion of Pb and Cd found in runoff[4]. During raining events, water flow on impermeable surfaces of an urban area, collect dust and waste deposited during the preceding dry period[5]. Once collected, runoff water can be either released directly into aquatic environments or directed into wastewater treatment plants (WWTP), leading to an accumulation of heavy metals in the activated sludge[6]. The latter is the case of many industrialized countries in which the separation of sewage and runoff water is not entirely achieved[7].

A previous study in Lake Geneva (Switzerland) has established that endospore-forming Firmicutes (EFF) are predominant members of the bacterial communities in sediments contaminated with metals[8]. The high levels of metals in those sediments have been associated to the release of treated waste water from a WWTP on the shore. In this previous study, Clostridia were identified as dominant members of the bacterial communities in the most contaminated samples[8]. The first explanation for this observation could be the natural tolerance of these bacteria to metals. However, this explanation is not entirely satisfactory and raised several questions. The main one concerns the origin of these Clostridia and their activity in sediments. According to their affiliation, the species identified are commonly found in human feces and are, most likely, released into the sediment together with the treated effluent. On spite of this allochthonous origin, they were still dominant members of the community in sediment layers with the highest concentrations of metals. By contrast, other species typical from activated sludge (*Hydrogenophaga caenii* or *Acidovorax defluvii*) were only dominant in sediments with lower levels of metals. Therefore, the present study was carried out to explore this intriguing correlation pattern (metals-Clostridia) in more detail. For this, four additional sediment cores sub-sampled at a finer scale (three layers) was collected and analyzed. Bacterial community composition was established using 454-pyrosequencing and the results were merged to our previous pyrosequencing effort. This resulted in an extended sequence repertory. A particular emphasis was placed on the bioinformatic treatment of 454–pyrosequencing data leading to a better comprehension of the presence of particular taxa in the different samples. Finally, the evaluation of the role of endospore-formation as a survival strategy of these Clostridia was assessed based on the estimation of the absolute abundance of endospore-formers, the concentration of endospores in sediments, and experiments on the effect of metals on sporulation in four culturable strains under laboratory conditions.

Material and methods

Sampling and storage

To confirm the link between EFF and metals in the contaminated sediments near the WWTP of Lausanne (bay of Vidy, Switzerland), four additional cores were retrieved in May 2012 with a gravity corer (UWITECH, Mondstein, AU) from the boat "La Licorne" of the Forel Institute (University of Geneva, Switzerland). The sampling locations (Figure 1 and Table 1) were selected on the basis of preliminary measurements of metal concentration[9] to recover sediments that were differently impacted. Two cores (N3 and N4) were retrieved near the outlet pipe of the WWTP. One (M3) was retrieved at approximately 130 m to the East of the pipe. The last one (M4) was retrieved at approximately 330 m to the South-East of the pipe. The process of coring was recorded by the use of a GoPro camera attached to the corer. Prior to further treatment, sediment cores were carefully transported into the lab in an ice box and stored in the dark at 4°C while bubbling air at the surface of the overlaying water to preserve the redox gradients at the sediment water interface. Subsampling was done to recover three different layers, 0 to 1.5 cm, 1.5 to 3 cm and 3 to 9 cm. These layers were referred as "Up", "Med" and "Low", respectively.

Chemical measurements on sediments

Air-dried and agate ground sediments were used for chemical measurements. To remove carbonates, a first acidic fumigation with HCl vapors (6 hours) was performed[10]. A CHN-analyzer (CHNEA1108-Elemental analyser, CE Instruments Ltd, Wigan, UK) was used to measure total organic carbon (C_{org}) and total nitrogen (N_{tot}). The concentration of total particulate trace metals was determined from sediments extracted by an aqua regia digestion method[11,12] and measured by ICP-MS (HP 4500, Agilent Technologies, Santa Clara, CA, USA), following the same method described elsewhere[12].

DNA extraction

An indirect DNA extraction method was selected to treat sediment samples. Briefly, three grams of wet sediment were dispersed in 15 ml of 1% sodium hexametaphosphate solution. Dispersion was performed using an Ultra-turrax® T18-basic (IKA, Stauffen, Germany) dispersing tool operated at 2x1min at 15'500rpm. Supernatant was recovered after 10 minutes of sedimentation at 1 x g. Another round of dispersion was carried out adding other 15 ml of Na-Hexametaphosphate to the remaining pellet, followed by Ultra-turrax 2 x 1min and sedimentation as described above. The two supernatants were pooled and a slow centrifugation (20 x g, 10 minutes) was performed to remove large sediment particles still remaining in the suspension. Cells were recovered by filtrating the supernatant through two 0.2µm nitro-cellulose filters (11407-47-CAN, Sartorius, Goettingen, Germany). One filter was frozen (-80°C) as backup, the second was cut in small fragments and one half of it was inserted into a lysing bead-beating tube of the FastDNA SPIN kit for soil (MP biomedicals, Santa Ana, USA) the second half was stored at -80°C in a 2ml eppendorf tube. DNA extraction was performed with a modified protocol including three sequential bead-beating steps[13]. After each bead-beating round, the supernatant was retrieved and extracted separately following the manufacturer's instruction. At the end of the procedure, the three DNA extracts were pooled together and an extra purification procedure by ethanol precipitation was performed. Total DNA was quantified with Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA).

Pyrosequencing and bioinformatics analysis

Purified DNA products were sent to Eurofins MWG Operon (Ebersberg, Germany) for amplicon-based 454-pyrosequencing (GS FLX+ technology, 454 Life Sciences, Branford, CT, USA). The hypervariable V1-V3 regions of the 16S rRNA gene were selected by the use of the primers Eub8f (5' -AGAGTTTGATCCTGG CTCAG-3') and Eub519r (5' -GTATTACCGCGGCTGCT GG-3') as suggested elsewhere[14]. The sequences obtained were merged with the sequences obtained from our previous study[8] (samples referred as D1, D2, M1, M2, N1, and N2). The merged dataset was analysed with Qiime[15] following the main pipeline for 454-pyrosequencing data. Briefly, sequences were first demultiplexed based on tag identifiers, then filtered to remove too long, too short, bad quality, as well as chimeric (with Usearch[16] method) sequences. Operational taxonomic units were defined using Uclust[16] based on 97% of identity. A representative set of sequences from defined OTUs were then aligned and taxonomically assigned thanks to a Greengenes[17] 16S rRNA gene database ending up with a BIOM[18] format community matrix. Rarefaction curves were computed with CHAO metrics and bootstrapped beta-diversity analyses were computed to compare sample communities.

Starting from the BIOM matrix, especially written R[19] scripts (supplementary material) were used to perform constrained correspondence analyses (CCA, vegan library[20]) and to visualize the link between microbial communities composition and chemical parameters measured in sediments. Specific scripts were also used to draw different plots for visualizing community composition.

To allow a better identification of the predominant organisms in the communities (OTUs representing more than 1% of relative abundance), the corresponding sequences were gathered from the main fasta file (resulting from pyrosequencing) and grouped together in a single fasta file per dominant OTU. This treatment was performed using the information from Qiime[15] output files, EMBOSS[21] tools and low-level UNIX commands. OTUs corresponding fasta files were then individually blasted[22] (`blastn` command) against a bacterial 16S rRNA reference database of 326'231 sequences (most complete database available, containing only good quality sequences of type and non-type strains) downloaded from the ribosomal database project web site (http://rdp.cme.msu.edu/). Finally, the more representative blast identifications were used to replace OTU identification number in the CCA ordination plots.

qPCR on the 16S rRNA and spoOA gene

To assess bacterial abundance used to calculate the absolute abundance of each taxa, a qPCR quantification of the 16S rRNA gene was performed. For this purpose, 0.3µM of primers 338f (5'-ACT-CCTACGGGAGGCAGCAG-3') and 520r (5'-ATTACCG CGGCTGCTGG-3')[23,24], 1 x Rotor-Gene SYBR Green PCR Master Mix (Qiagen), DNA template (between 0.5 and 20 ng) and PCR-grade water (up to final volume of 10 µl) were mixed into the reaction tubes. A Corbett Rotor-Gene 3000 (Qiagen, Hilden, Germany) was used to perform the amplification. Initial denaturation and polymerase activation was first performed (15 min at 95°C), followed by 40 cycles of denaturation 10 s at 95°C, annealing 15 s at 55°C and elongation 20s at 72°C. A standard curve was made of 10-fold dilutions of a known concentration of a plasmid obtained by cloning the 16S rRNA gene from an environmental strain.

A qPCR quantification of the *spo0A* gene[25] was also performed to assess the abundance of EFF. qPCRs were carried out in a final reaction volume of 10 µL with 5 µL Rotor-Gene SYBR green PCR master mix (Qiagen GmbH, Hilden, Germany), and 1 µM, respectively 0.45 µM of primers spo0A655f (5'-GGH GTD CCN GCN CAT ATH AA) and spo0A834r (5'-CCA HGC NAC TTC WAT NGC RT). Two-microliter aliquots of diluted samples (approximately 5 ng) were added to the reaction tubes. The Corbett Rotor-Gene RG-3000A thermocycler (Qiagen) was used for the amplifications. An initial denaturation/polymerase activation step at 95°C for 10 min was first performed, followed by 45 cycles consisted of denaturation 95°C for 15 s, annealing 54°C for 15 s, and elongation 68°C for 10 s. The standard curve was constituted by 10-fold dilutions of a known concentration of a plasmid obtained by cloning the *spo0A* gene from *Bacillus subtilis*.

DPA quantification

The quantification of dipicolinic acid (DPA) was performed by measuring the fluorescence emitted by a Terbium-DPA complex in a procedure developed for sediments and described elsewhere[26]. Briefly, 150 mg of freeze-dried sediments were placed into a screw cap glass tube, to which 4.3 mL of sodium acetate buffer (0.2 M, pH 5) and 480 µL of aluminium chloride ($AlCl_3$, 0.5 M) solution were added and mixed. Screw caps were tightly closed and tubes were autoclaved (121°C, 103 kPa) for 20 minutes. After 30 minutes of cooling, 1 mL of the clear supernatant was carefully recovered and mixed with 1 mL of terbium chloride solution ($TbCl_3$, 30 µM) directly into a quartz cuvette (Perkin-Elmer, Waltham, USA) to be used for fluorescence measurements. Fluorescence was then measured within a Perkin-Elmer LS50B fluorometer. The excitation wavelength was set at 272 nm with a slit width of 2.5 nm. Emission was measured at 545 nm (slit width 2.5 nm). The device was set in the phosphorescence mode (equivalent to time-resolved fluorescence). The delay between emission and measurement was set at 50 µs. Measurements were performed every 20 ms. The integration of signal was performed over a duration of 1.2 ms. Values recovered for each measurement corresponded to the mean of the relative fluorescence unit (RFU) values given by the instrument within the 30s following sample introduction in the device.

Finally, to transform RFU units into DPA concentrations, a 10-point standard curve was established using increasing concentrations of DPA from 0.1 μM up to 10 μM.

Results and discussion

Sediment characterization

Four sediment cores were collected near the outlet pipe releasing treated wastewater from the WWTP of the city of Lausanne (Lake Geneva, Switzerland). These additional cores completed a sampling effort encompassing a total of 17 sediment cores covering an area of 55 ha around the pipe. The cores were collected near the pipe (N cores) ar at 100 to 420 m (intermediate area; M cores) and 600 to 900 m (D cores) from it. During sampling with the MIR submersibles (our previous study[8]) and also when visualizing videos recorded with a GoPro camera attached to the sediment corer, escape of gas from the sediments was observed for the cores retrieved near the outlet pipe and in the intermediate area. In the intermediate area a patchy covering of the sediment by filamentous bacteria was also observed. A microscopic observation of these filaments (supplementary figure 1) showed that they corresponded to sulfur-oxidizing bacteria (presumably *Beggiatoa* spp).

<u>Table 1:</u> Summary of chemical measurements and core sampling locations. A complete table encompassing values from the first set of samples[8] is available as supplementary material.

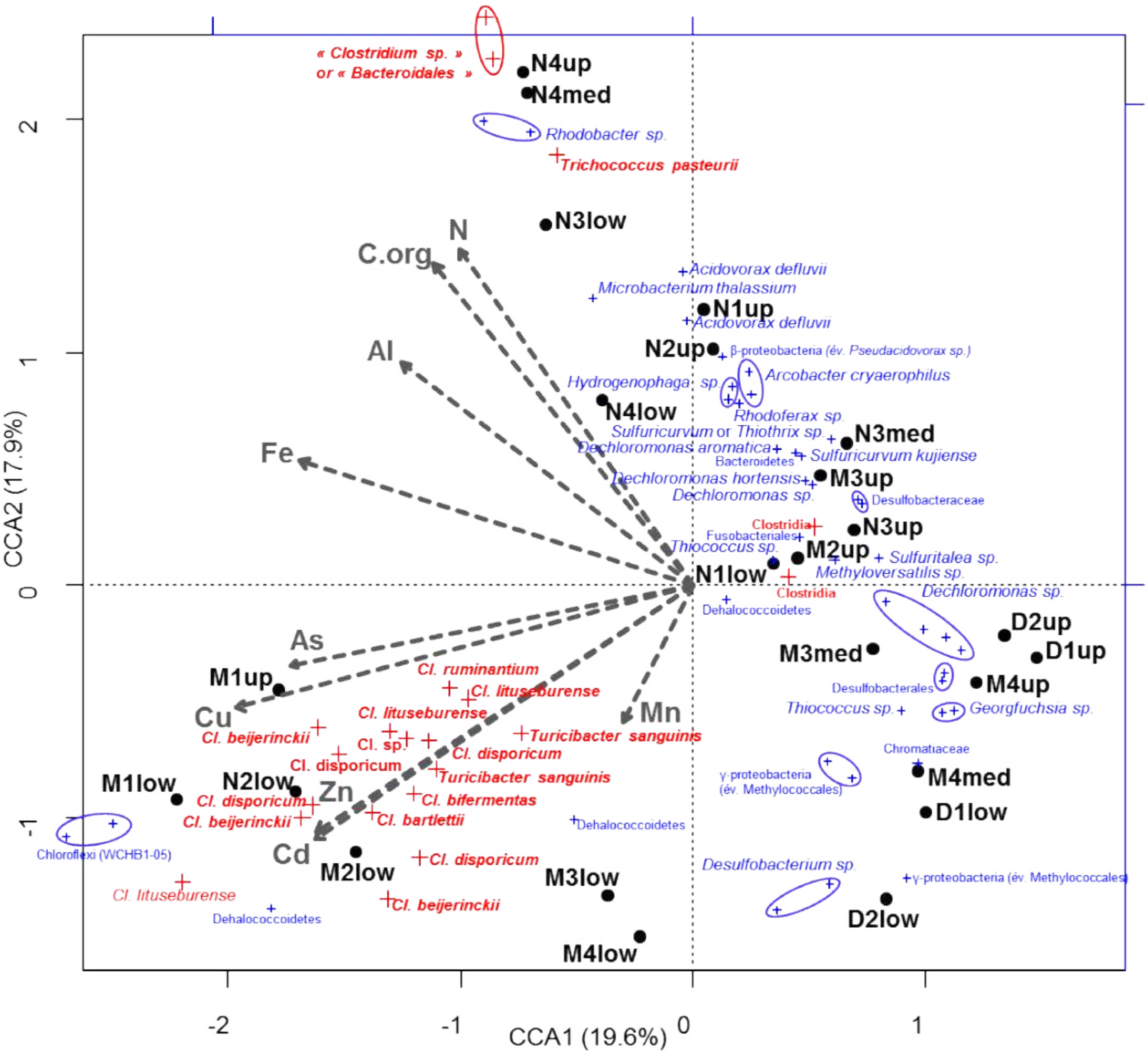
Samples, Layer		Swiss coord. East	Swiss coord. North	Depth (m)	Corg %	Ntot %	Cd mg/kg	Cu mg/kg	Fe g/kg	Mn mg/kg	Al g/ kg	Zn mg/kg	As mg/kg
	Up, 0-1.5cm				4.30	0.43	0.66	116	30.0	355	14.4	338	11.0
M3	Med, 1.5-3cm	534803	151513	31	4.40	0.41	0.77	111	31.6	345	13.2	341	11.7
	Low, 3-9cm				3.97	0.28	3.79	197	33.5	349	9.6	858	14.3
	Up, 0-1.5cm				3.96	0.31	1.00	109	22.1	327	9.7	364	11.3
M4	Med, 1.5-3cm	534961	151369	31	4.82	0.32	2.22	137	27.0	331	12.9	526	14.9
	Low, 3-9cm				3.97	0.27	5.03	266	39.4	443	16.6	1343	17.6
	Up, 0-1.5cm				4.70	0.49	0.18	111	28.7	376	14.4	268	12.2
N3	Med, 1.5-3cm	534648	151567	31	4.69	0.45	0.11	105	31.5	377	16.2	277	12.1
	Low, 3-9cm				4.33	0.35	1.39	171	34.0	457	26.3	493	13.3
	Up, 0-1.5cm				12.5	1.33	2.04	216	38.2	312	20.8	478	14.9
N4	Med, 1.5-3cm	534709	151526	32	13.2	0.99	0.42	168	35.7	337	21.4	381	13.4
	Low, 3-9cm				10.1	1.16	1.06	221	47.3	334	15.5	489	16.8

Chemical measurements carried out in the sediment cores obtained for this study are reported in Table 1. For comparison, chemical data from the entire dataset (including the values already reported in our previous study) are presented in Supplementary Table 1. Nutrients were estimated based on the content of C_{org} (ranging from 3.97 to 13.25 %) and N_{tot} (0.28 to 1.33 %). The highest values were found in the upper layers of a sediment core collected near the effluent pipe (N4), signalling the high load of nutrients released by the WWTP. Surprisingly, the next core showing the second highest values for nutrients is core M1, which was characterized in our previous study and was collected in the intermediate area. Overall, nutrients varied greatly on the area studied. A higher nutrient content in some cores collected near to the pipe and in the intermediate area could be explained by an intermittent release of incompletely-treated wastewater. Indeed, the treatment plant is episodically flooded by an overload of run-off water during rain events and yearly estimates suggest that almost 10% of the wastewater does not undergo a complete treatment and is released into the lake[27]. As a consequence, and depending on the local currents, a patchy distribution of nutrients in the sediments collected around the pipe could be expected, following the direction of the rejection's plume. For comparison, lower nutrient values were reported for distal samples (mean values of C_{org} and N_{tot} of 2.6% and 0.21%, respectively), which correspond to values reported for sediments in the oligotrophic Lake Michigan[28]. Based on the nutrient composition, the distal area could be considered as not (or less) influenced by the release of effluents from the treatment plant.

The levels of particulate metals measured in the sediments are consistent with previous studies in the same area[9]. Particulate metal levels ranged from 0.11 to 5.03 mg/kg Cd, 105 to 266 mg/kg Cu, 22.1 to 47.3 g/kg Fe, 312 to 457 mg/kg Mn, 9.6 to 26.3 g/kg Al, 268 to 1343 mg/kg Zn, and 11.0 to 16.8 mg/kg As (Table 1). For all the metals except Mn and Al, the highest levels were not measured close to the pipe (sample N4), but in the lowest layers of cores from the intermediate zone (M samples). In regard to the classification proposed for the management of Rhine sediments[29], which gives limit values for Cd, Cu and Zn, the more contaminated sediments in our study should be considered as posing a potential risk for the environment in the perspective of perturbation leading to metal re-mobilisation. On the contrary, the upper layers of sediments from the distal stations D1 and D2, presented Cd and Zn concentrations close to the indicative values (or natural background; Cd 0.8 mg/kg, Cu 40 mg/kg and Zn 150 mg/kg[30]) of the Swiss federal law for soil preservation (Osol), but exceed still the values expected for Cu (two-fold).

Bacterial community composition in sediments

The bacterial community composition in the 12 sediment samples was studied by pyrosequencing of the bacterial 16S rRNA gene using universal PCR primers. Sequencing resulted in 161'688 amplicons, which were merged to our previous dataset (268'498 amplicons[8]), leading to a total of 385'951 sequences after quality filtering and chimera removing. These new dataset was globally analysed to define the overall bacterial community composition in the area surrounding the WWTP outlet pipe, as well as the sediments in the intermediate and distal areas. With a level of identity of 97%, a total of 44'483 OTUs were defined. Figure 1 presents the link between physicochemical parameters measured in the sediment samples and the dominant bacterial species (representing at least 1% in one of the communities). Together with the second CCA plot for components 2 and 3, presented in supplementary figure 2, the ordination explain up to 50.5% of the correlation between the parameters analysed. Based on these two plots, four distinct groups of samples could be defined. The first group corresponded to samples highly impacted by the effluent from the WWTP (N4up, N4med, N3low, N1up, and N2up), which were all collected in the area near to the pipe at different depths and are characterized by high content of nutrients (N4up and N4med) or high level of Al (N3low). This first group was subdivided into two groups composed by N4up, N4med and N3low (upper part of Fig 1), and N1up, N2up samples, the latter separated by the second CCA ordination plot (Supplementary figure 2). The second group included sediments from the near and intermediate areas (N4low, N1low, M2up, N3up, M3up, and N3 med), which seemed to lack a differentiating chemical parameter. All the values for nutrients and metals were effectively close to the mean values, except for nutrient levels for sample N4low, which were higher. The third group corresponded to samples from the distal zone (D1up, D2up, D1low, and D2low) and three samples from the intermediate area (M3med, M4med, and M4up), which were characterized by low to intermediate values for nutrients and metals. The final group corresponded to samples mainly from the intermediate area (M2low, M1low, M1up, N2low, M3low, and M4low), which are strongly linked to Cu, Cd, Zn, Mn, and As as environmental variables, representing all the samples with high levels of particulate metals.



<u>Figure1:</u> CCA ordination plot computed with OTUs defined on merged pyrosequencing data of the two sets of samples collected in sediments from Lake Geneva, Switzerland. To refine the analysis, sequences from the most abundant OTUs were gathered from the dataset and blasted against a more complete 16S rRNA database. D samples were collected the more distantly from the rejection pipe of the WWTP, M in an intermediate zone and N close to the pipe. Suffixes in sample names correspond to subsampling depth in sediment cores ("up" for surface layer, "med" for intermediate layer, "low" for lower layer, as defined in table 1)

Figure 2 shows the results of the bioinformatic analysis of the pyrosequencing data to identify the composition of the bacterial communities of these four groups of samples. It is worth mentioning that even though the assignation of specific metabolic functions based on phylogenetic identity can be debatable, up to now, this is nonetheless the most accessible approach existing to understand the link between bacterial community composition and environmental factors. The bacterial community was separated into five major bacterial groups (Firmicutes; β-, δ-, and γ-proteobacteria; other taxa) taking into account the relative or absolute abundance of each identified OTU.

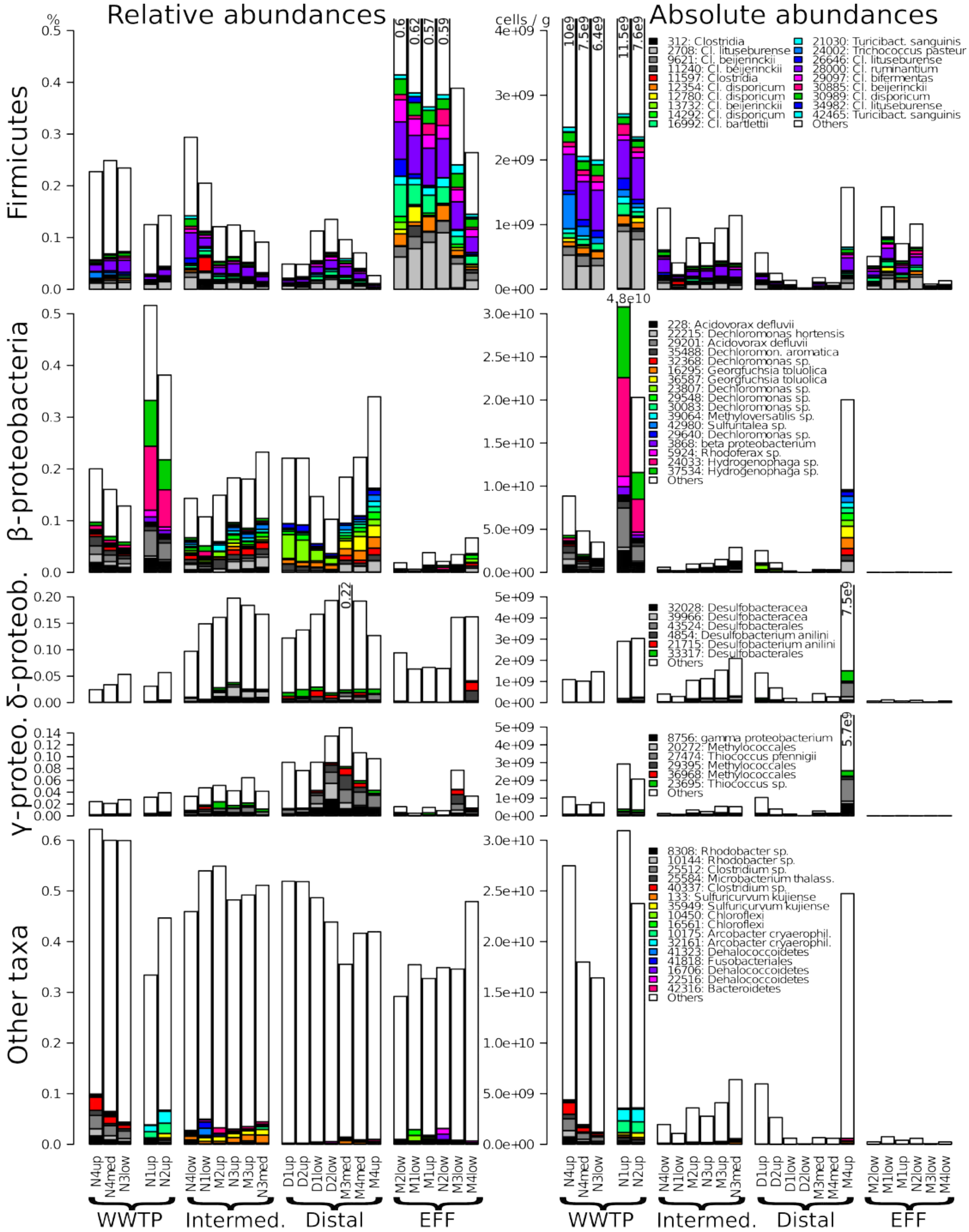


Figure 2: **Relative (%) and absolute (cells / gram of dry sediment) abundance of the bacterial communities found in sediment samples. Identity of each group is presented in the legend to the right. Most abundant OTUs (more than 1% of any community) are shown in color. Taxa below 1% are presented as "other" in white. To facilitate the analysis, the composition is presented for five major bacterial groups: Firmicutes; β-, δ-, and γ-proteobacteria; and other taxa. Samples were grouped based on their composition. WWTP correspond to samples clearly influenced by WWTP rejects, Interm correspond to samples that display an intermediate status, Distal correspond to samples with the lowest influence of WWTP, and EFF correspond to samples were endospore-forming Firmicutes are dominant.**

The relative abundance corresponded to the relative number of sequences in the dataset, while the second was calculated based on the total bacterial abundance estimated by quantification of the 16S rRNA gene by qPCR. Regarding the relative abundance, it is important to mention that in all the sediment samples, a large proportion of OTUs correspond to species representing less that 1% of the community (white bars in Figure 2), and that the sum of abundant taxa (OTUs representing more than 1% of the community in at least one sample; coloured bars in Figure 2), never represent more than 50% of the total community. The absolute abundances measured by qPCR (supplementary figure 3) were highly variable. Roughly, samples highly impacted by the effluent from the WWTP (labelled as WWTP in Figure 2), contained around 5e10 cells/g of dry sediment. Sediments from the near and intermediate areas (labelled as Intermed.) contained 6.6e9 cells/g, and those from the distal zone (labelled as Distal) 4.2e9 cells/g, except for the samples D2low and M4up (1.5e8 cells/g and 5.9e10 cells/g, respectively). Finally the samples associated to high levels of metals (indicated as EFF) contained 1.5e9 cells/g, except for the samples M3low and M4low (2.2e8 and 5e8 cells/g, respectively).

Bacterial community composition in the first subgroup of samples influenced by the WWTP was distinguished by the presence of two uncertain OTUs that were affiliated to the Bacteroidales with Qiime and to *Clostridium* by BLAST. Another typical OTU from these samples was affiliated to *Trichococcus pasteurii* (Firmicutes). The second closest BLAST affiliation for this OTU (with percentage of identity of 98%) is *Trichococcus floculiformis*, which is a typical species from activated sludge in WWTP[31], suggesting that the microbial composition of these samples was dominated by organisms from the treatment plant. Finally, two OTUs affiliated to *Rhodobacter* sp. (α-proteobacteria), were identified in the samples. This genus is known to be an anoxygenic phototroph from the group of the purple non-sulphur bacteria, but in the absence of light it can also grow by aerobic chemoheterotrophy[32], which would be a more likely metabolisms if those OTUs are hypothesized to originate from the WWTP. The second subgroup (N1up and N2up) was characterized by the predominance of β-proteobacteria, as shown in Figure 2. The identified OTUs that distinguish these communities are mainly related to *Hydrogenophaga* sp., *Acidovorax defluvii*, and *Rhodoferax* sp. for the β-proteobacteria, and *Arcobacter cryaerophilus* for ε-proteobacteria (other taxa). The genus *Hydrogenophaga* encompasses mainly chemolithoautotrophic hydrogen-oxidizers but two strains from this genus (*H. defluvii* and *H. atypical*) were isolated from activated sludge and are obligate aerobes, growing best as heterotrophs[33]. *A. defluvii* is also an aerobe from activated sludge[34]. *Rhodoferax* spp. encompasses mainly phototrophic bacteria, with one exception for the species *Rhodoferax ferrireducens,* which uses organic electron donors to reduce Fe(III) to Fe(II)[35]. As $FeCl_3$ is added as coagulation agent during the processing of wastewater[36], this latter metabolism seems plausible for this group. Furthermore this strain could also grow using O_2 as final electron acceptor[35]. Finally, *A. cryaerophilus* is a typical human or animal enteropathogen and could be used as an indicator of faecal contamination[37].

Bacterial communities for the second group of sediment samples (N4low, N1low, M2up, N3up, M3up and N3med) were almost equally composed by 10 to 15% of Firmicutes, 15 to 20% of β- and δ-proteobacteria, and 50% of other taxa. However, a slight predominance of Firmicutes was observed for samples N1low and N4low, which correspond to the lower layers of sediment cores collected close to the outlet pipe. This predominance is perhaps the result of strongly reductive conditions prevailing at this depth favouring fermentative metabolisms. This will be discussed later. This group of sediment samples represented an intermediate state between sediments influenced by the WWTP rejections and those that are not.

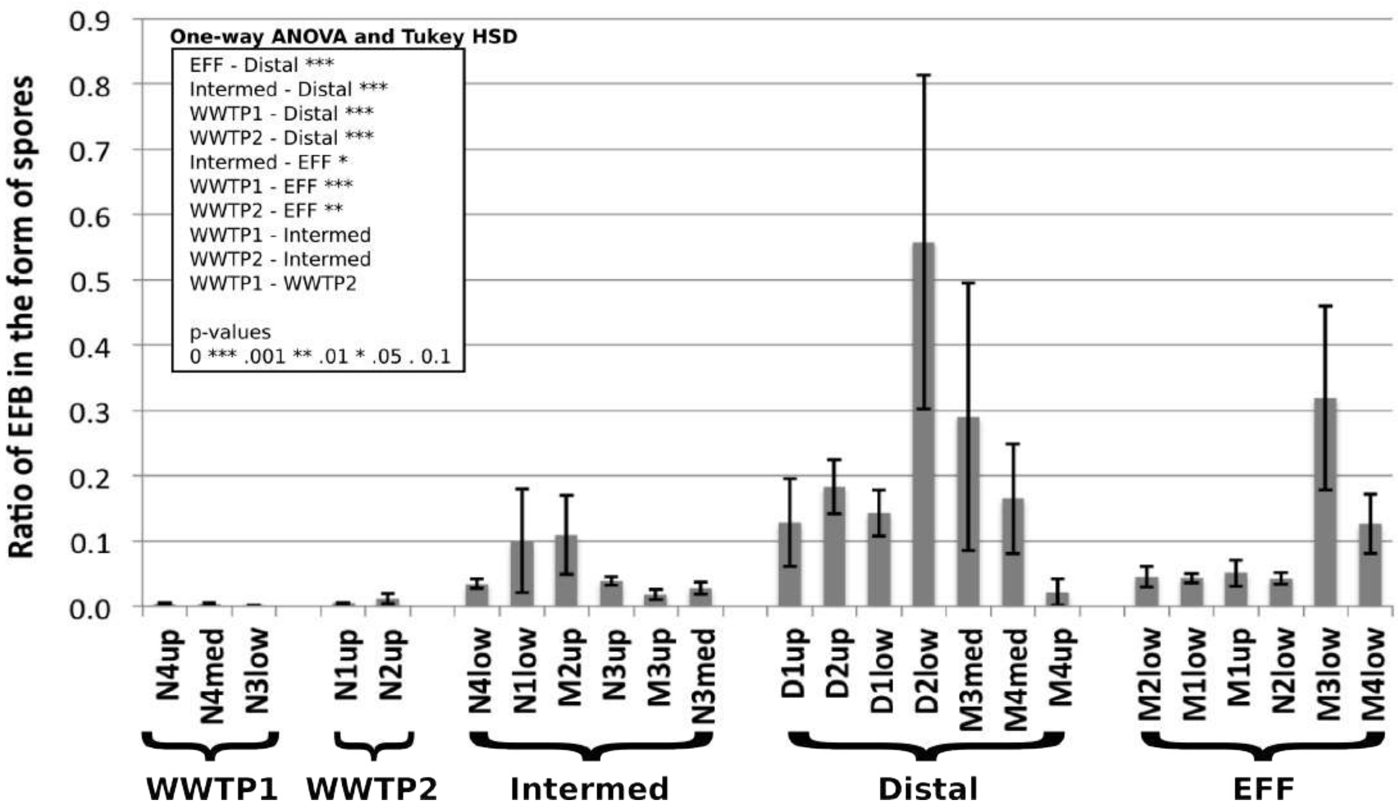
The bacteria composition in the third group, composed by samples collected away from the outlet pipe (D1up, D2up, D1low, D2low M3med, M4med and M4up) were characterized by a high diversity of β-, δ- and γ-proteobacteria, combined with the lowest relative abundance of Firmicutes. β-proteobacteria are mainly represented by *Dechloromonas* spp., *Georgfuchsia toluolica*, *Methyloversatilis* sp., and *Sulfuritalea hydrogenivorans*. δ-proteobacteria were represented by OTUs related to Desulfobacterales, like *Desulfobacterium* sp., showing the importance of sulphate reduction in anaerobic sediments even at relatively shallow depths. Finally, γ-proteobacteria are represented by some Methylococcales and *Thiococcus* sp. The former are known for producing energy by methane oxidation, the second is affiliated to Chromatiacea, obligate anaerobe phototrophic bacteria using sulphide or S° as electron donor for photosynthesis (purple-sulfur bacteria). These latter taxa are normal inhabitants of freshwater sediments containing hydrogen sulfide and exposed to light irradiation[38]. In Lake Geneva, transparency of around 10 m was reported for the beginning of October 2011 and the end of May 2012 based on Secchi disk measurements[39,40], leading to an euphotic zone down to 30m[41]. Purple sulphur bacteria are normally well established below this limit[42] and could proliferate down to the sediment surface at a depth of 35 m for stations D1 and D2 (Table 1), were sulphide is also present due to sulphate reduction in the underlying sediments.

Metals and endospore-forming Firmicutes, environmental evidence

The last group of sediment samples (M2low, M1low, M1up, N2low, M3low and M4low) was clearly the most interesting for the purpose of this study since it was characterized by samples with the highest concentrations of metals. Interestingly these samples originate from the deepest layers of the cores, except for M1up, or from the intermediate area, except for N2low. Bacterial communities were also particular because they were clearly dominated by Firmicutes (up to 62% of the community in sample M1low). In those samples Firmicutes sub-communities were mainly composed by fermentative Clostridia. The main identified strains were *Clostridium lituseburense*, *Clostridium beijerinckii*, *Clostridium disporicum*, *Clostridium bartletii*, *Clostridium ruminantium*, and *Clostridium bifermentas*. *Turicibacter sanguinis,* a normal inhabitant of human gut[43], was also observed in a small proportion. *T. sanguinis* is reported to be an anaerobic non-spore forming Firmicute with chemoorganotrophic or fermentative metabolism[44]. All the other Clostridia strains identified are obligate anaerobes with a fermentative metabolism. All strains have been isolated from faecal material. *Cl. lituseburense* has previously been also isolated from soil[44,45]. In summary, this last group of sediment samples contain mainly endospore-forming Firmicutes that originated from human gut or faeces, and therefore are present in the sediment most likely as a consequence of the WWTP rejections. Moreover, these strains are also present in samples from the first group described above (WWTP; Figure 2), and their prevalence could be explained by their survival to environmental conditions prevailing in the sediment, compared to other taxa also released from the WWTP. However, optimal growth temperature is reported to be at 30-37°C for all the strains except *Cl. disporicum* (25°C), suggesting that they cannot (or very slowly) develop under the conditions encountered in the sediments (temperature at the lake bottom close to 10°C). This is supported by the comparison of the relative and absolute abundance barplots in Figure 2, which did not show an increase in abundance in any of the samples that are not directly influenced by the WWTP. However, none of the former explain the co-occurrence of these EFF species and metals in sediments.

Selection of EFF by high metal levels

High concentrations of metal have often been related to a loss of functional diversity in microbial ecosystems[1,46]. However, in comparison to other microbial groups, EFF are well known for their capability to survive diverse environmental stressors such as nutrient deprivation, desiccation, UV-radiation, or hazardous chemicals. When environmental conditions become deleterious to vegetative cells, a specialized resistance mechanism, the formation of an endospore, is triggered[47,48]. With this particular resistance mechanism, one can imagine that EFF could be more tolerant to high concentrations of metals than other bacterial groups. This hypothesis was tested by evaluating sediment levels of dipicolinic acid (DPA), a biomarker only known to be present in endospores[49]. The DPA values were normalized to the quantification of endospore-formers carried out by qPCR of the *spo0A* gene[25], which is also specific to endospore-formers[48]. The ratio of EFF in the state of endospores is presented in Figure 3. The results showed very low values for the sediments more directly influenced by the wastewater treatment effluent (WWTP samples in Figure 2), while in the two other groups of samples (Distal and EFF), the values varied greatly but were overall significantly higher (p-value<0.01). More importantly, the ratio of endospore obtained in the samples highly contaminated with metals are significantly different (p-value<0.05) from those in the intermediate area, suggesting that fraction of EFF in the state of endospores in significantly higher. A selection process that leads to the co-occurrence of metals and endospores of EFF in sediments is, therefore, supported by the data. However, cultivation assays (Supplementary figure 4) performed in our laboratory on four Clostridia strains identified as dominant in the most contaminated sediments, showed that metals precipitate readily in the form of metal-sulphides under the anaerobic conditions of the culture medium (combination of low redox potential and a source of sulphur in the form of Na-thioglycolate). The free metal concentration measured in the culture vials coincide with the concentrations measured in a parallel study performed in pore water from sediment cores collected in the same area of the bay of Vidy[12]. The highest amount of dissolved Cu in interstitial water was 21 μg/L (approximately 10'000 times lower than the total amount of particulate Cu measured in sediment in the present study). The same is true for other metals like Cd and Pb. Furthermore, with such low concentrations of metals in solution, it is unlikely that metals alone trigger endosporulation. As an increase in the ratio of endospores with distance from the pipe can also be observed as a general trend (Figure 3), a complementary explanation is that endospore formation is also linked to sediment stabilisation overtime. When comparing community composition and abundance of bacteria (Figure 2), it appears clear that metals have a strong negative effect bacteria that are not capable to produce endospores as resistance structures, which indirectly favours EFF in the most contaminated sediments.



<u>Figure 3:</u> Ratio of endospore-formers found in the state of spores in different sediments samples.

However, the question that remains open is the origin of the high levels of metals rejected by the wastewater treatment plant. In principle, in this type of wastewater treatment plant, metals should be entrapped in activated sludge and recovered during the sedimentation of sludge at the end of the process[50]. The effluent should, as a consequence, not be a vector for carrying toxic concentration of metals into the aquatic environment. However, as mentioned above, evidence for species forming part of the activated sludge[51–54] is also found in the sediment (e.g. *T. pasteurii* or *T. floculiformis*, *Rhodobacter* sp., *Thricococcus* sp., *Acidovorax* sp., *Hydrogenophaga* sp., *Rhodoferax* sp., *Arcobacter* sp., and *Dechloromonas* sp.). showing that this plant is regularly subjected to overflow events leading to a patchy deposition of a mixture of sludge and untreated (or incompletely treated) wastewater on the sediment surface all around the rejection pipe. Once in sediments, the easily degradable portion of the released organic matter could be rapidly consumed by aerobic microbial activity (which is also predominant in the biological treatment line in this WWTP[36]) at the surface of sediments, leading to a rapid drop of dissolved oxygen levels (supplementary figure 5), and to the erosion of strictly aerobic bacterial communities. This leaves only those taxa resisting to anoxia, such as Clostridia, as the dominant bacterial members in the sediment. Furthermore, the release of metals complexed within sludge after mineralisation could perhaps lead to an increase in local metal concentrations, finally triggering endosporulation in EFF. To verify this hypothesis further experiments are needed, probably at a microcosm scale.

In conclusion, in this study, we demonstrated that endosporulation is a mean to survive to toxic concentrations of metals in sediments. However, the exact mechanism leading to this co-occurrence is still not completely understood and requires additional investigations. Even if the particular mechanisms have not yet been entirely elucidated, we are convinced that an in-deep comprehension of the association between EFF and metals could result in environmental-friendly bioremediation technologies in the future.

Acknowledgments

We thank Ferring Pharmaceutical and the Consulat honoraire de la Fédération de Russie à Lausanne for funding this project through elemo grant 301, as well as the EPFL for coordinating the elemo project. Funding from the Swiss National Science Foundation (grant 31003A_132358) is also acknowledged. A particular thanks will also be addressed to Marie-Louise Tercier-Waeber (University of Geneva, Switzerland) for her help in measuring particulate HTM in sediments by ICP-MS. HTM measurement in culture experiments were done by ICP-OES thanks to the help of Anna Simanova and Jasquelin Peña from the Institute of Earth Sciences at the University of Lausanne (Switzerland). We thank Anaele Simon from the Laboratory of Microbiology, University of Neuchâtel, for her help for the SEM-EDS analyses.

References

(1) Olaniran, A.; Balgobind, A.; Pillay, B. Bioavailability of heavy metals in soil: Impact on microbial biodegradation of organic compounds and possible improvement strategies. *Int. J. Mol. Sci.* **2013**, *14*, 10197–10228.

(2) Rial, D.; Vázquez, J.; Murado, M. Effects of three heavy metals on the bacteria growth kinetics: a bivariate model for toxicological assessment. *Appl. Microbiol. Biotechnol.* **2011**, *90*, 1095–1109.

(3) Hemme, C. L.; Deng, Y.; Gentry, T. J.; Fields, M. W.; Wu, L.; Barua, S.; Barry, K.; Tringe, S. G.; Watson, D. B.; He, Z.; et al. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J.* **2010**, *4*, 660–672.

(4) Davis, A. P.; Shokouhian, M.; Ni, S. Loading estimates of lead, copper, cadmium, and zinc in urban runoff from specific sources. *Chemosphere* **2001**, *44*, 997–1009.

(5) Marsalek, J.; Diamond, M.; Kok, S.; Watt, W. E. Urban runoff, 2008.

(6) Liu, Y.; Lam, M.; Fang, H. Adsorption of heavy metals by EPS of activated sludge. *Water Sci. Technol.* **2001**, *43*, 55–66.

(7) Krejci, V.; Frutiger, A.; Kreikenbaum, S.; Rossi, L. Impact des rejets pluviaux sur les milieux récepteurs, 2005.

(8) Sauvain, L.; Bueche, M.; Junier, T.; Masson, M.; Wunderlin, T.; Kohler-Milleret, R.; Gascon Diez, E.; Loizeau, J.-L.; Tercier-Waeber, M.-L.; Junier, P. Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria. *Aquat. Sci.* **2013**, 1–14.

(9) Pote, J.; Haller, L.; Loizeau, J. L.; Bravo, A. G.; Sastre, V.; Wildi, W. Effects of a sewage treatment plant outlet pipe extension on the distribution of contaminants in the sediments of the Bay of Vidy, Lake Geneva, Switzerland. *Bioresour. Technol.* **2008**, *99*, 7122–7131.

(10) Harris, D.; Horwáth, W. R.; van Kessel, C. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. *Soil Sci Soc Am J* **2001**, *65*, 1853–1856.

(11) Förstner, U.; Wittmann, G. *Metal pollution in the aquatic environment*; 2nd reviewed edition.; Springer-Verlag: New-York, Heidelberg, Berlin, 1981.

(12) Masson, M.; Tercier-Waeber, M.-L. Trace metal speciation at the sediment–water interface of Vidy Bay: influence of contrasting sediment characteristics. *Aquat. Sci.* **2013**, 1–12.

(13) Wunderlin, T.; Junier, T.; Roussel-Delif, L.; Jeanneret, N.; Junier, P. Stage 0 sporulation gene A as a molecular marker to study diversity of endospore-forming Firmicutes. *Environ. Microbiol. Rep.* **2013**, *5*, 911–924.

(14) Li, H.; Zhang, Y.; Li, D.; Xu, H.; Chen, G.; Zhang, C. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. *Soil Biol. Biochem.* **2009**, *41*, 954–968.

(15) Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336.

(16) Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461.

(17) DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072.

(18) McDonald, D.; Clemente, J.; Kuczynski, J.; Rideout, J.; Stombaugh, J.; Wendel, D.; Wilke, A.; Huse, S.; Hufnagle, J.; Meyer, F.; et al. The biological observation matrix (BIOM) format or: How I learned to stop worrying and love the ome-ome. *GigaScience* **2012**, *1*, 7.

(19) Team, R. D. C. *R: a language and environment for statistical computing*; Vienna, Austria, 2009.

(20) Oksanen, J. Multivariate analysis of ecological communities in R: vegan tutorial, 2011.

(21) Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **2000**, *16*, 276–277.

(22) Altschul, S.; Gish, W.; Miller, W.; Myers, E.; Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.

(23) Muyzer, G.; Dewaal, E.; Uitterlinden, A. Profiling of complex microbial populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16S ribosomal RNA. *Appl. Environ. Microbiol.* **1993**, *59*, 695–700.

(24) Bakke, I.; De Schryver, P.; Boon, N.; Vadstein, O. PCR-based community structure studies of bacteria associated with eukaryotic organisms: A simple PCR strategy to avoid co-amplification of eukaryotic DNA. *J. Microbiol. Methods* **2011**, *84*, 349–351.

(25) Bueche, M.; Wunderlin, T.; Roussel-Delif, L.; Junier, T.; Sauvain, L.; Jeanneret, N.; Junier, P. Quantification of endospore-forming Firmicutes by qPCR with the functional gene *spo0A*. *Appl. Environ. Microbiol.* **2013**.

(26) Brandes Ammann, A.; Kölle, L.; Brandl, H. Detection of bacterial endospores in soil by terbium fluorescence. *Int. J. Microbiol.* **2011**, *2011*.

(27) Jaton, J.-F. Bilan 2011 de l'épuration vaudoise, 2011.

(28) Meyers, P. A.; Ishiwatari, R. Lacustrine organic geochemistry—an overview of indicators of organic matter sources and diagenesis in lake sediments. *Org. Geochem.* **1993**, *20*, 867–900.

(29) *Sediment Management Plan Rhine : Summary*; International Commission for the Protection of the Rhine (ICPR): Koblenz, 2009.

(30) Osol - Ordonnance sur les atteintes portées aux sols, 1998.

(31) Scheff, G.; Salcher, O.; Lingens, F. Trichococcus flocculiformis gen. nov. sp. nov. A new gram-positive filamentous bacterium isolated from bulking sludge. *Appl. Microbiol. Biotechnol.* **1984**, *19*, 114–119.

(32) Mackenzie, C.; Eraso, J. M.; Choudhary, M.; Roh, J. H.; Zeng, X.; Bruscella, P.; Puskás, Á.; Kaplan, S. Postgenomic adventures with *Rhodobacter sphaeroides*. *Annu. Rev. Microbiol.* **2007**, *61*, 283–307.

(33) Kämpfer, P.; Schulze, R.; Jäckel, U.; Malik, K. A.; Amann, R.; Spring, S. *Hydrogenophaga defluvii* sp. nov. and *Hydrogenophaga atypica* sp. nov., isolated from activated sludge. *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 341–344.

(34) Schulze, R.; Spring, S.; Amann, R.; Huber, I.; Ludwig, W.; Schleifer, K.-H.; Kämpfer, P. Genotypic diversity of *Acidovorax* strains isolated from activated sludge and description of Acidovorax defluvii sp. nov. *Syst. Appl. Microbiol.* **1999**, *22*, 205–214.

(35) Finneran, K. T.; Johnsen, C. V.; Lovley, D. R. *Rhodoferax ferrireducens* sp. nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of Fe(III). *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 669–673.

(36) Margot, J.; Magnet, A.; Thonney, D.; Chèvre, N.; De Alencastro, F.; Rossi, L. Traitement des micropolluants dans les eaux usée: Rapport final sur les essais pilotes à la STEP de Vidy (Lausanne), 2011.

(37) Collado, L.; Figueras, M. J. Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clin. Microbiol. Rev.* **2011**, *24*, 174–192.

(38) Garrity, G.; Brenner, D. J.; Krieg, N. R.; Staley, J. R. *Bergey's Manual® of Systematic Bacteriology: Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria*; Springer, 2007.

(39) Lazzarotto, J.; Quetin, P.; Klein, A. *Evolution physico-chimique des eaux du Léman (éléments majeurs)*; Campagne 2012; Commission Internationale pour la protection des eaux du Léman, 2013; p. 25.

(40) Lazzarotto, J.; Klein, A. *Evolution physico-chimique des eaux du Léman (éléments majeurs)*; Campagne 2011; Commission Internationale pour la protection des eaux du Léman, 2012; p. 25.

(41) Koenings, J. P.; Edmundson, J. A. Secchi disk and photometer estimates of light regimes in Alaskan lakes: Effects of yellow color and turbidity. *Am. Soc. Limnol. Oceanogr.* **1991**, *36*, 91–105.

(42) Madigan, M. T.; Martinko, J. M.; Stahl, D.; Clark, D. P. *Brock biology of microorganisms*; Benjamin Cummings: San Francisco, 2012.

(43) Cuív, P. Ó.; Klaassens, E. S.; Durkin, A. S.; Harkins, D. M.; Foster, L.; McCorrison, J.; Torralba, M.; Nelson, K. E.; Morrison, M. Draft genome sequence of *Turicibacter sanguinis* PC909, isolated from human feces. *J. Bacteriol.* **2011**, *193*, 1288–1289.

(44) De Vos, P.; Garrity, G. M.; Jones, D.; Krieg, N. R.; Ludwig, W.; Rainey, F. A.; Schleifer, K.-H.; Whitman, W. B. *Bergey's manual of systematic bacteriology*; Second edition.; Springer: Dordrecht, Heidelberg, London, New-York, 2009; Vol. Volume 3: The Firmicutes.

(45) Maukonen, J.; Satokari, R.; Mättö, J.; Söderlund, H.; Mattila-Sandholm, T.; Saarela, M. Prevalence and temporal stability of selected Clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria. *J. Med. Microbiol.* **2006**, *55*, 625–633.

(46) Niklinska, M.; Chodak, M.; Laskowski, R. Characterization of the forest humus microbial community in a heavy metal polluted area. *Soil Biol Biochem* **2005**, *37*, 2185 – 2194.

(47) Nicholson, W.; Fajardo-Cavazos, P.; Rebeil, R.; Slieman, T.; Riesenman, P.; Law, J.; Xue, Y. Bacterial endospores and their significance in stress resistance. *Antonie Van Leeuwenhoek* **2002**, *81*, 27–32.

(48) Errington, J. Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Microbiol.* **2003**, *1*, 117–126.

(49) Fichtel, J.; Koster, J.; Scholz-Bottcher, B.; Sass, H.; Rullkotter, J. A highly sensitive HPLC method for determination of nanomolar concentrations of dipicolinic acid, a characteristic constituent of bacterial endospores. *J. Microbiol. Methods* **2007**, *70*, 319–327.

(50) Junier, P.; Frutschi, M.; Wigginton, N. S.; Schofield, E. J.; Bargar, J. R.; Bernier-Latmani, R. Metal reduction by spores of *Desulfotomaculum reducens*. *Env. Microbiol* **2009**, *11*, 3007–3017.

(51) Spiegel, S. J.; Farmer, J. K.; Garver, S. R. Heavy metal concentrations in municipal wastewater treatment plant sludge. *Bull. Environ. Contam. Toxicol.* **1985**, *35*, 38–43.

(52) Mielczarek, A. T.; Saunders, A. M.; Larsen, P.; Albertsen, M.; Stevenson, M.; Nielsen, J. L.; Nielsen, P. H. The microbial database for Danish wastewater treatment plants with nutrient removal (MiDas-DK) - a tool for understanding activated sludge population dynamics and community stability. *Water Sci. Technol.* **2013**, *67*, 2519–2526.

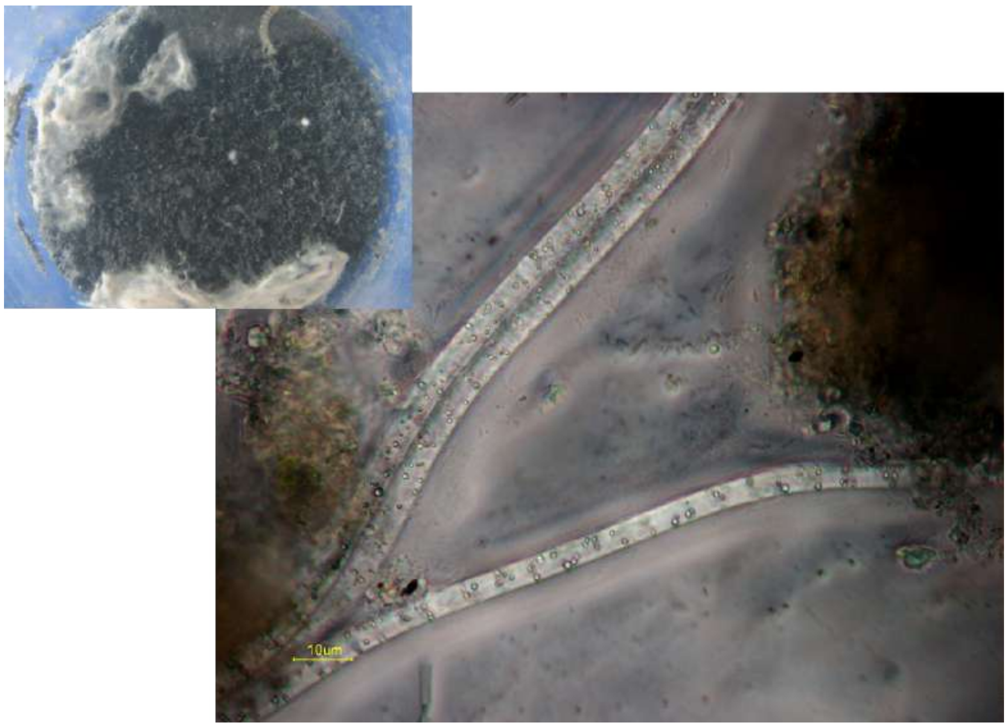
(53) Wagner, M.; Loy, A. Bacterial community composition and function in sewage treatment systems. *Curr. Opin. Biotechnol.* **2002**, *13*, 218–227.

(54) Bitton, G. *Wastewater microbiology*; third edition.; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2005.

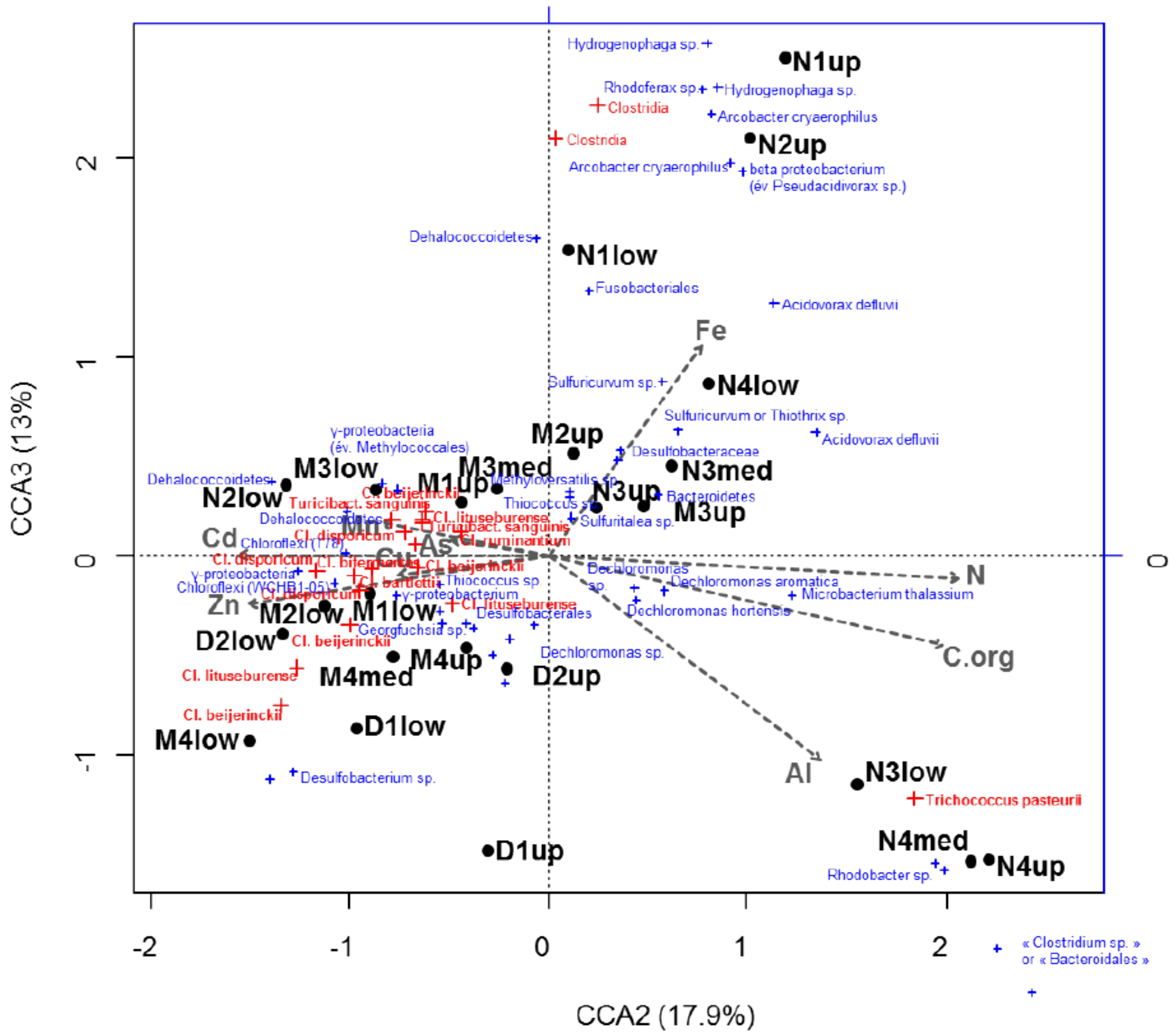
(55) Goel, R. K.; Sanhueza, P.; Noguera, D. R. Evidence of *Dechloromonas* sp. participating in enhanced biological phosphorus removal (EBPR) in a bench-scale aerated-anoxic reactor. *Proc. Water Environ. Fed.* **2005**, *2005*, 3864–3871.

(56) Chen, F.; Yang, Y.; Zhang, D.; Zhang, L. Heavy metals associated with reduced sulfur in sediments from different deposition environments in the Pearl River estuary, China. *Environ. Geochem. Health* **2006**, *28*, 265–272.

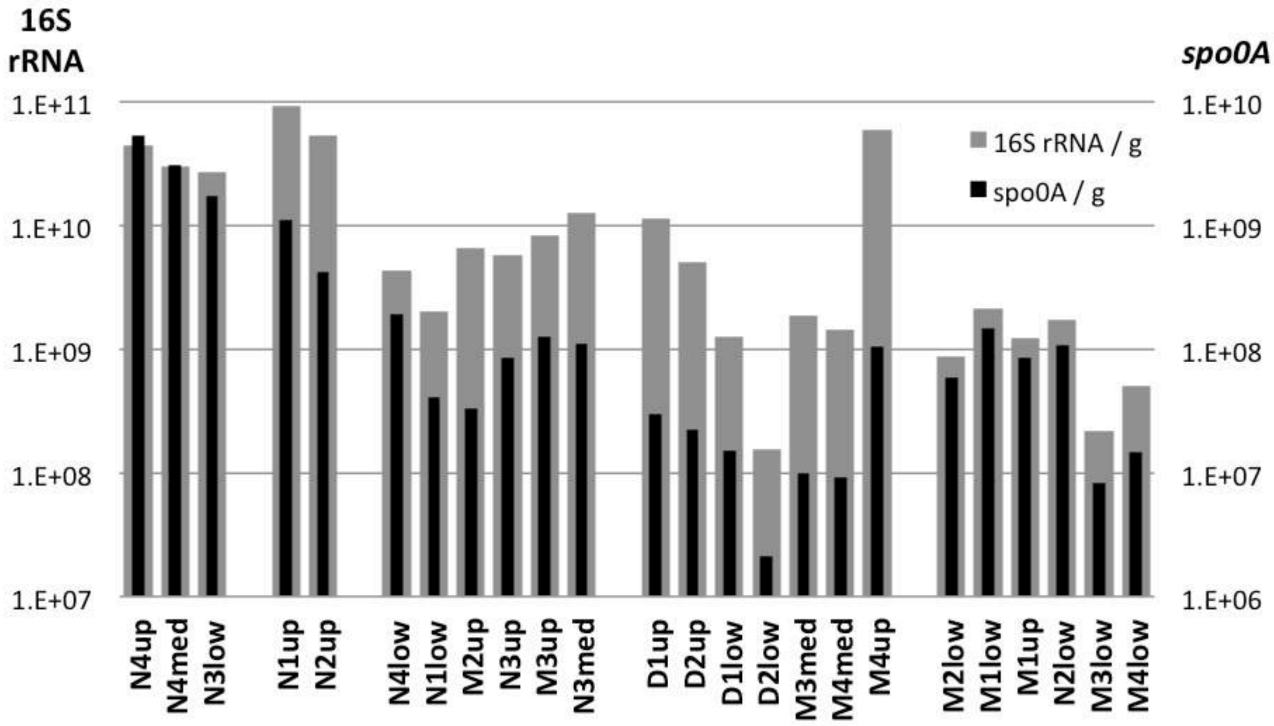
Supplementary Information



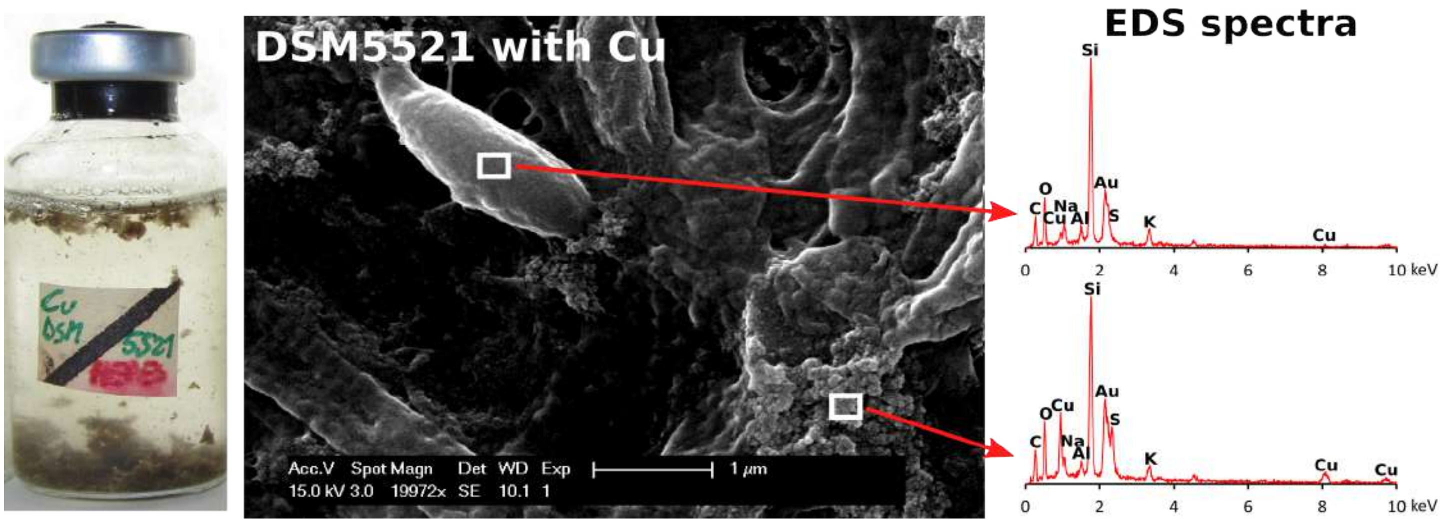
Supplementary figure 1. Aspect of the surface of a sediment core recovered around the rejection pipe. White filaments correspond to the sulphur-oxidizing *Beggiatoa* sp., as demonstrated by the light microscope observation (magnification 1000x). Typical sulphur granules are effectively clearly visible.



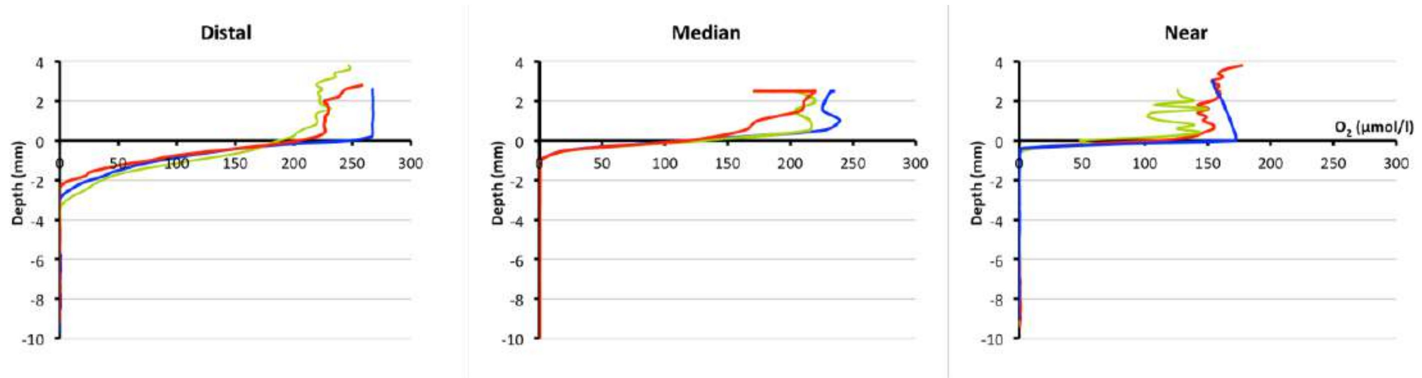
Supplementary figure 2. CCA ordination plot for components 2 and 3



Supplementary figure 3. qPCR quantification of 16S rRNA (grey bars, left axis) and *spo0A* (black bars, right axis) genes found per gram of dry sediment



Supplementary figure 4. Physico-chemical precipitation of copper-sulphide in culture vial (left part) when cultivating anaerobically *Clostridium disporicum* (DSM5521) in BDM medium amended with 50 mg / L of copper. Visualisation of the association between copper-sulphide precipitates and bacteria by scanning electron microscopy (middle part). EDS measurement of the enrichment in sulphur and copper in amorphous precipitates when comparing to a bacterial endospore (right part).



Supplementary figure 5. Profiles of dissolved oxygen in sediments measured with microelectrodes. The three stations represented correspond to the distance from the outlet pipe. In all the profiles oxigen disappear within the few first millimetres when going down into the sediment. However a little bit slower for distal samples.

Supplementary table 1. Summary of chemical measurements. Samples in bold correspond to cores described in this paper. Samples in italic with an asterix (*) correspond to the measurement published in another paper and used here for a deeper data mining of pyrosequencing data.

Samples,	Layer	Swiss coord. East	Swiss coord. North	Depth (m)	Corg %	Ntot %	Cd mg/kg	Cu mg/kg	Fe g/kg	Mn mg/kg	Al g/ kg	Zn mg/kg	As mg/kg	Shannon
D1*	Up, 0-3cm	533735	151351	36	2.67	0.23	0.74	75	18.2	408	16.3	170	7.8	10.25
	Low, 3-9cm				2.64	0.21	1.90	114	19.9	387	13.9	322	10.4	10.18
D2*	Up, 0-3cm	534063	151491	35	2.78	0.23	0.81	81	24.8	405	15.3	205	8.5	10.16
	Low, 3-9cm				2.36	0.17	2.76	137	22.5	396	9.2	542	9.1	10.02
M1*	Up, 0-3cm	534515	151335	47	7.00	0.53	6.27	292	43.5	390	18.6	1032	14.7	8.58
	Low, 3-9cm				7.72	0.65	10.13	362	47.9	413	20.8	1564	22.0	8.17
M2*	Up, 0-3cm	534379	151229	60	4.01	0.38	1.66	165	33.1	415	17.7	399	13.0	9.71
	Low, 3-9cm				5.08	0.40	4.38	293	34.3	400	16.3	974	15.6	8.23
M3	**Up, 0-1.5cm**	**534803**	**151513**	**31**	**4.30**	**0.43**	**0.66**	**116**	**30.0**	**355**	**14.4**	**338**	**11.0**	**9.72**
	Med, 1.5-3cm				**4.40**	**0.41**	**0.77**	**111**	**31.6**	**345**	**13.2**	**341**	**11.7**	**9.41**
	Low, 3-9cm				**3.97**	**0.28**	**3.79**	**197**	**33.5**	**349**	**9.6**	**858**	**14.3**	**9.27**
M4	**Up, 0-1.5cm**	**534961**	**151369**	**31**	**3.96**	**0.31**	**1.00**	**109**	**22.1**	**327**	**9.7**	**364**	**11.3**	**9.54**
	Med, 1.5-3cm				**4.82**	**0.32**	**2.22**	**137**	**27.0**	**331**	**12.9**	**526**	**14.9**	**9.85**
	Low, 3-9cm				**3.97**	**0.27**	**5.03**	**266**	**39.4**	**443**	**16.6**	**1343**	**17.6**	**9.85**
N1*	Up, 0-3cm	534676	151543	32	5.96	0.62	1.88	155	40.0	392	13.4	416	12.8	8.26
	Low, 3-9cm				3.79	0.31	1.76	123	33.9	443	16.2	349	11.4	10.06
N2*	Up, 0-3cm	534676	151543	32	4.69	0.46	1.59	129	35.8	369	13.0	346	11.2	8.95
	Low, 3-9cm				6.30	0.58	8.06	341	36.0	401	21.4	902	22.0	8.22
N3	**Up, 0-1.5cm**	**534648**	**151567**	**31**	**4.70**	**0.49**	**0.18**	**111**	**28.7**	**376**	**14.4**	**268**	**12.2**	**9.85**
	Med, 1.5-3cm				**4.69**	**0.45**	**0.11**	**105**	**31.5**	**377**	**16.2**	**277**	**12.1**	**9.83**
	Low, 3-9cm				**4.33**	**0.35**	**1.39**	**171**	**34.0**	**457**	**26.3**	**493**	**13.3**	**10.09**
N4	**Up, 0-1.5cm**	**534709**	**151526**	**32**	**12.5**	**1.33**	**2.04**	**216**	**38.2**	**312**	**20.8**	**478**	**14.9**	**9.55**
	Med, 1.5-3cm				**13.2**	**0.99**	**0.42**	**168**	**35.7**	**337**	**21.4**	**381**	**13.4**	**10.05**
	Low, 3-9cm				**10.1**	**1.16**	**1.06**	**221**	**47.3**	**334**	**15.5**	**489**	**16.8**	**10.05**
Mean					5.42	0.48	2.53	175	32.9	382	16.0	557	13.4	9.49
SD					2.89	0.29	2.57	82	7.9	39	4.1	364	3.6	0.70

4

A yet un- publishable experiments with heavy metals and Clostridia

This chapter introduces some not yet publishable results on experiments performed with the four Clostridia strains highlighted during this thesis. The particular link with heavy metals has been explored in cultivation experiments amended with different heavy metals.

Contents

The experiments presented in this chapter are the results of several months of work. As the results are not clear and sometimes contradictory, we decided last minute to remove this part from the third publication. In fact we have realized quite late that all the experiments done to a certain point give invalid results due to a physicochemical precipitation of heavy metals in the culture medium. The experiments as well as the results are, therefore, presented in a raw form with only few discussion elements.

4.1 Cultivation of Clostridia strains

Four different strains of Clostridia were used in this part of the study. *Clostridium disporicum* (DSM 5521), *Clostridium lituseburense* (DSM 797) and *Clostridium bartlettii* (DSM 16795) were purchased from the Leibniz institute DSMZ-German collection of microorganism and cell cultures. *Clostridium beijerinckii* (Neu 28) was retrieved from the culture collection of our laboratory and was initially isolated from lake mud.
The composition of basal defined medium (BDM) (Figure 4.1) used for their cultivation, is based on the defined medium first described by Haslam et al. (1986)[100] and then modified by Karasawa et al. (1995).[101] Phosphate buffer was replaced by HEPES in order to avoid heavy metals precipitation due to complexing with phosphate. To support growth of the bacteria, a small quantity of phosphorus is needed. It was, therefore, added in the form of K_2HPO_4. The mineral base of the BDM medium was prepared in 1 litter Erlenmeyer flask and supplemented with 1 mg/L of resazurin redox indicator. Mixture was heated and stirred until it boiled to remove as much as possible dissolved oxygen. Cooling down was operated under inert gas atmosphere (N_2 / CO_2 mixture) while stirring. Once at room temperature, 500 mg of sodium thioglycolate and 500 mg of ascorbic acid were added as reducing agents. 19 mL of medium was dispensed in 25 mL headspace crimp vials. Bottles were flushed 5 minutes with inert gas mixture (N_2 / CO_2), hermetically crimped and autoclaved at 121°C, 103 kPa during 20 minutes. Before each experiment, a twenty fold concentrated amino acid solution (Figure 4.1) was prepared. This mixture was supplemented by 80 g/L of glucose, 500x vitamin solution (40 μL per 20 mL bottle) and depending on the experiment, the needed concentration of heavy metal salt. At the end, mixture was flushed for 5 minutes with inert gas mixture (N_2 / CO_2) and 1 ml of solution is poured into crimp vials through a cellulose-acetate 0.2μm filter.

Inoculation was performed under a laminar flow hood by injecting 2.5% (V:V) of a previous culture, using a sterile syringe previously flushed with a sterilised inert gas mixture. Inoculated crimp vials were incubated in a shaking-incubator at 37°C and 150 rpm. In a first approach, growth was assessed by visual observation of the culture vials and confirmed by microscopic observation and counting (vegetative cells and spores) in a Neubauer improved counting chamber (0642010; Paul Marienfeld, Lauda-Königshofen, Germany). The Figure 4.2 shows the aspect of the four Clostridia strains after overnight cultivation in BDM medium. The four strains grew within less than 24 hours.

Mineral base (mg/l) in distilled water	
K_2HPO_4	400
HEPES	2380
NaCl	900
$CaCl_2 \cdot 2H_2O$	26
$MgCl_2 \cdot 6H_2O$	20
$MnCl_2 \cdot 4H_2O$	10
$(NH_4)_2SO_4$	40
$FeSO_4 \cdot 7H_2O$	4
$CoCl_2 \cdot 6H_2O$	1
$NaHCO_3$	5000

Autoclave sterilisation 121°C, 103kPa, 15min

500x Vitamine solution (mg/l) in distilled water	
Thiamin · HCl	500
Calcium- D – pantothenate	500
Nicotinic acid	500
Riboflavin	500
Pyridoxine	500
p-Aminobenzoic acid	25
Folic acid	6.25
Biotin	6.25
B_{12} vitamin	2.5

Sterilisation by 0.2µm filtration

Carbon source (mg/l) in distilled water	
Glucose	4000

Sterilisation by 0.2µm filtration

20x Amino acids solution (mg/l) in distilled water	
L - Histidine	100
L – Tryptophane	100
Glycine	100
L - Tyrosine	100
L – Arginine	200
L – Phenylalanine	200
L – Methionine	200
L – Threonine	200
L – Alanine	200
L – Lysine	300
L – Serine	300
L – Valine	300
L – Isoleucine	300
L – Aspartic acid	300
L – Leucine	400
L – Cysteine	500
L – Proline	600
L – Glutamic acid	900

Sterilisation by 0.2µm filtration

Figure 4.1: Composition of the BDM medium used for the cultivation of the four Clostridia strains

4.2 Amendment with heavy metals

The initial hypothesis evaluated was that heavy metals could trigger sporulation in Clostridia. A first set of experiments was done to verify this by cultivating the four Clostridia strains with different amounts of copper and cadmium. As these treatments seemed to make no difference on sporulation rate, the effect of heavy metals on growth dynamic was then evaluated.

4.2.1 Effect on sporulation

A preliminary experiment was done with the strain Neu28 (*Clostridium beijerinckii*), cultivated in anaerobically prepared crimp vials filled with Merck Standard 1 nutrient broth

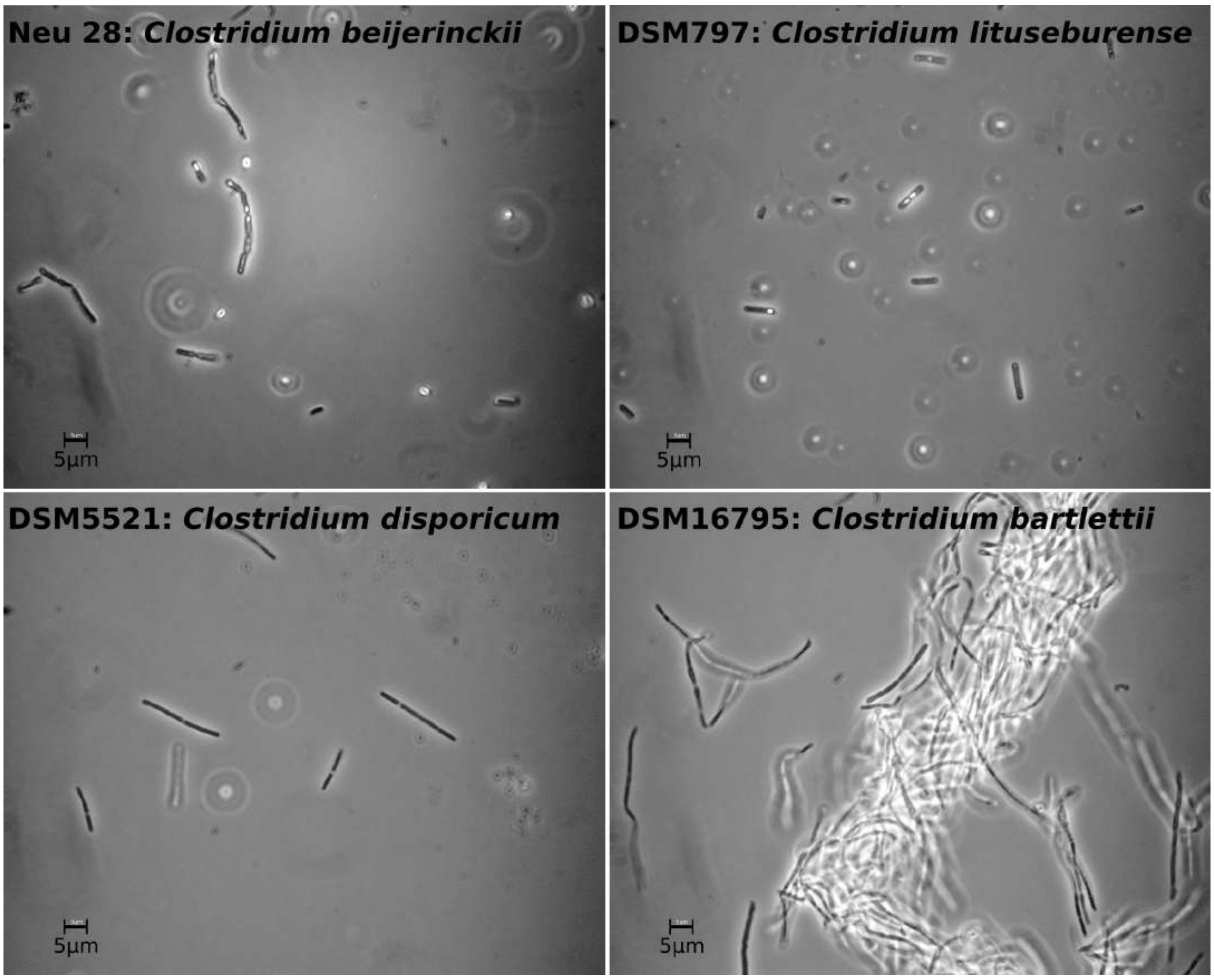
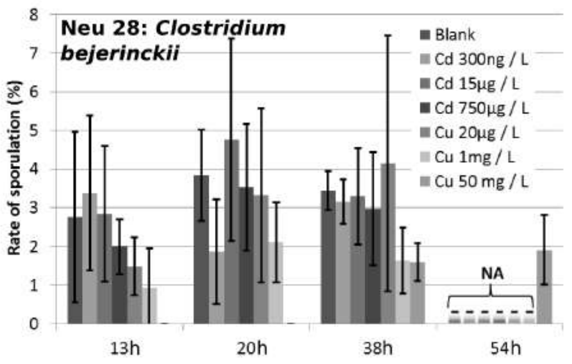


Figure 4.2: Aspect of the four Clostridia strains when growing in BDM medium

(Merck KGaA, Darmstadt, Germany) and amended with three different concentrations of copper (0.02, 1 and 50 mg/L) and cadmium (0.3, 15 and 750 $\mu g/L$) sulphate salts. Blanks were also prepared by amending crimp vials with an equivalent volume of sterile deionised water. As mentioned above, the sporulation rate was assessed by counting cells and spores within a Neubauer improved counting chamber. Experiment were performed in triplicates and counting at three different time points (13, 20, and 38 hours of incubation) for all the treatments, except the amendment with 50 mg/L of copper for which growth was delayed (or slowed down) and required an extra counting after 54 hours of incubation.
Results shown by Figure 4.3 (left part) clearly demonstrate that these two heavy metals have no effect on sporulation rate. For *Clostridium beijerinckii* (Neu28), sporulation rate does not increase above 5 % even after 1 day and a half of incubation, whereas microscopic observations show senescent cells. The only effect observed in this first experiment is a longer lag phase for the treatment with 50 mg / L of copper. This point wil be discussed below.

In a second experiment, the three other Clostridia strains (DSM797, DSM5521, and DSM16795) were cultivated in BDM medium amended with deionized (DI) water as blanks and two different concentrations of copper (Cu^{2+} 1 and 150 mg/L) and cadmium (Cd^{2+} 0.75 and 5 mg/L) sulphate salts. The sporulation rate was also assessed by counting (as above), but only after 20 hours and 100 hours of incubation. Figure 4.3 (right part) shows the sporultation rates for *Clostridium lituseburense* (DSM797), *Clostridium disporicum* (DSM5521), and *Clostridium bartlettii* (DSM16795). In this experiment, every strain grew well in all the treatments. After 20 hours of incubation, DSM797 (*Clostridium lituseburense*) has sporulated while DSM5521 (*Clostridium disporicum*) and DSM16795 (*Clostridium bartlettii*) did not sporulated, without any link to the concentration and type of heavy metal. After 100h of incubation, an observation of DSM797 cultures under light microscope revealed only spores and fragments of lysed cells. Vials with DSM5521 contained no spores, but still relatively well formed cells. Finally DSM16795 vials harboured no spores but a lot of senescent cells.



	After 20 h of incubation					100h of incub.
	H_2O	Cu 1mg/l	Cu 150mg/l	Cd 750µg/l	Cd 5mg/l	All traitements
DSM797: *Cl. lituseburense*	~1%	33% ±17	7% ±3	4%±5	8% ±11	100% ±0
DSM5521: *Cl. disporicum*	0% ±0	0% ±0	3% ±2	0% ±0	0% ±0	0% ±1
DSM16795: *Cl. bartlettii*	0% ±0	0% ±0	0% ±0	0% ±0	0% ±0	0% ±0

Figure 4.3: Rate of sporulation of the four Clostrida strains when exposed to different heavy metals. Left part of the figure: Cultivation of *Clostridium beijerinckii* in Merck ST1 medium. Right part: Cultivation of strains from the DSMZ in BDM medium.

Discussion on sporulation rate. As demonstrated by the results, the rate of sporulation did not seem to be influenced by the level of heavy metals in the vials, even when we are close to an inhibitory concentration as demonstrated by the delayed growth of *Clostridium beijerinckii* when cultivated with 50 mg/L of copper in the ST1 medium. The longer lag phase observed in this treatment could be due to the necessity of this strain to synthesize new proteins to resist high levels of copper. *Cop* and *pco* genes (which encode for multi-copper oxidase) are well known for conferring copper resistance to different Gram-negative bacteria.[102] Concerning Gram-positives, several studies have demonstrated the antimicrobial effect of copper surfaces on *Clostridium difficile* strains in clinical environments.[103,104] On the other hand, copper remain an essential micronutrient involved in many cuproenzymes. The development of a copper homoeostasis mechanism is, therefore, necessary to maintain the "right" copper concentration inside the cell.[105] When concentrations become to high copper have, however, deleterious effect on the bacterial communities. Experiments performed in a waste water treatment plant showed 50% of inhibition for different key microbial communities when amending the plant with around 25 mg/L of copper(II).[28] As *Clostridium beijerinckii* (Neu28) is able to growth with higher concentrations it is perhaps particularly tolerant to heavy metals. This will be tested in the other experiments!
The second part of the experiment, with DSM strains, allowed us to distinguish between the phenomenon of sporulation itself, that results from a specific response of individual strains against non-optimal environmental factors,[48] from the senescence that leads to the death

of vegetative cells when nutrients are consumed. If sporulation occurs, both phenomena increase the ratio of spores observed in the culture vials. The differences of sporulation rate observed for *Clostridium lituseburense* (DSM797) after 20 hours of incubation (right part of Figure 4.3) are probably not due to the presence of heavy metals, because the highest rate was measured for the lowest concentration of copper (1 mg / L) rather than for the highest (150 mg / L). Furthermore, after 100 hours of incubation, all the treatments achieved a sporulation rate of 100% because all vegetative cells died. For *Clostridium disporicum* (DSM5521) and *Clostridium bartlettii* (DSM16795), no sporulation or low sporulation was observed even after 100 hours of incubation, also leading to the conclusion that heavy metals is not the triggering factor for sporulation in the conditions of these experiments. In conclusion, and as mentioned by other authors,[53] sporulation does not seem to be triggered in response to a single unfavourable environmental factor, like the presence of heavy metals, but to a multitude of them. Furthermore, as described in the Bergey's manual of systematic bacteriology,[106] readily sporulation is not a constitutive characteristic of the four strains tested. *Clostridium beijerinckii* (Neu28) and *Clostridium lituseburense* (DSM 797) are described as sporulating readily, whereas *Clostridium disporicum* (DSM 5521) and *Clostridium bartletii* (DSM16795) are not.

4.2.2 Effect on growth dynamic

This section describe a set of experiments that were performed to evaluate the effect of heavy metals on the growth dynamic of the four strains (Neu28, DSM797, DSM5521, and DSM16795). The strains were cultivated in crimp vials filled with resazurin-free BDM medium, supplemented by DI water as blank or 6 different sources of heavy metals (100 $\mu g/L$ of Cd^{2+}, 5 mg/L of Cu^{2+}, Ni^{2+} or Zn^{2+}, 10 mg/L of Pb^{2+} or Cr^{3+}). Sources of heavy metals were provided in the form of sulphate salts, except Pb and Cr which were provided as chloride salts. Growth curves were established by measuring optical density with a spectrophotometer (Genesys 10S UV-VIS; Thermo Scientific, Waltham, Massachusetts, USA) at the wavelength of 600 nm. Measurements were done every hour, except during the exponential phase where they were done every 30 minutes. Sampling in the crimp vials was done with sterile syringes previously flushed with a sterile inert gas mixture (N_2 / CO_2). To prevent any entrance of oxygen from external air, a positive pressure was maintained by injecting an equivalent volume of sterile gas (N_2 / CO_2) in the headspace of crimp vials before each sampling. When values of optical densities raised above 0.3, dilutions were performed with the sterile mineral base of the BDM medium. Finally, cells were counted in a Neubauer improved counting chamber at the end of the culture experiments to establish the relation between optical density and cell abundance for each strain.
Figure 4.4 shows the growth curves for every treatment and every strain tested. The three DSM strains grew well with all treatments except the one with 5 mg/L of Ni, which seemed to totally inhibit their growth. *Clostridium Beijerinckii* (Neu28) grew also well with every treatment, except the one with Ni in which the growth was much slower. For the four

strains, apart from Ni, every treatment behave in the same way, resulting in approximately the same growth dynamic, depending on the strain tested. Cr treatment seemed, however, to delay a little bit the growth of DSM5521 (*Clostridium disporicum*), as well as the growth of Neu28 (*Clostridium beijerinckii*). The same could also be observed for the Pb treatment on this last strain. Finally, the lack of smoothness observed in some of the growth curves is an effect of cell aggregation in exopolymeric substances (EPS) matrix, leading to a lack of homogeneity during the optical density measurements.

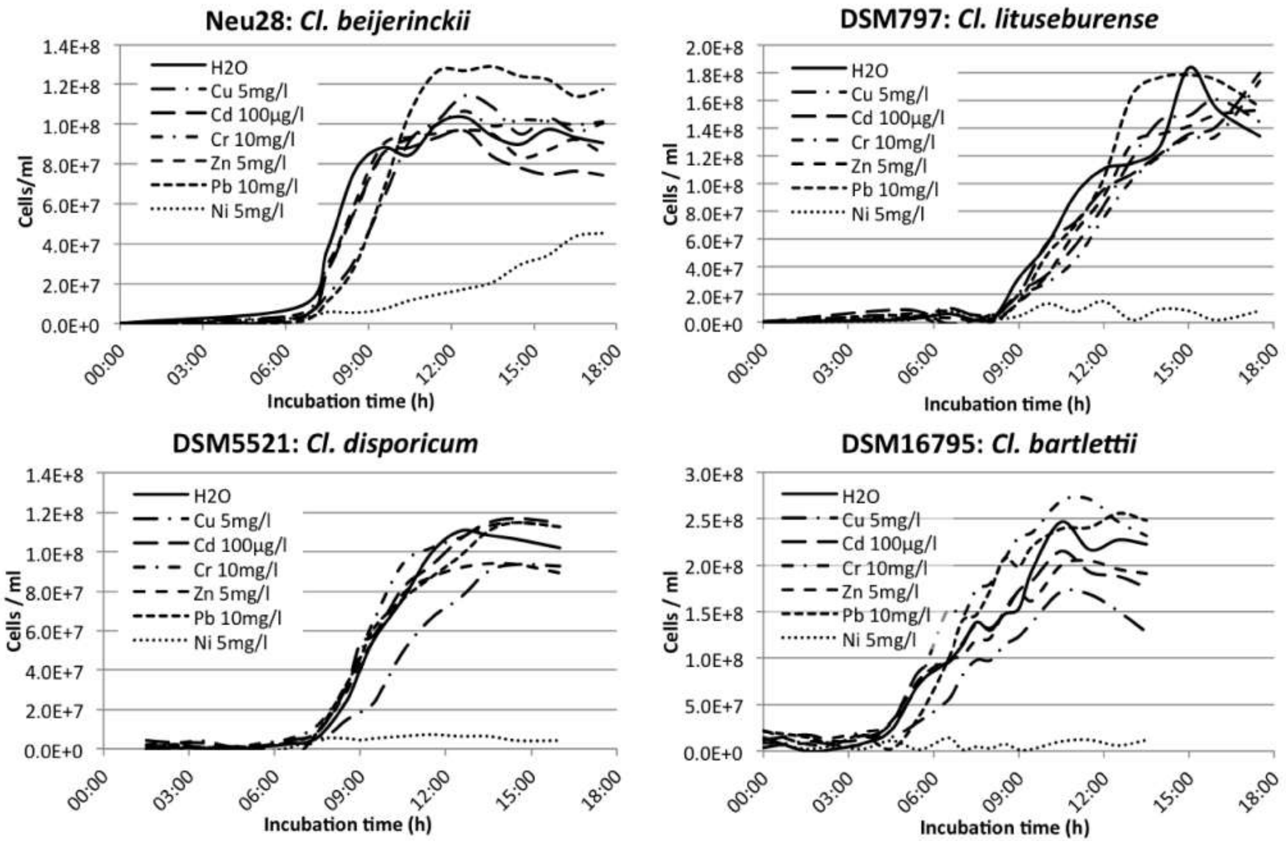


Figure 4.4: Growth curves for the four Clostridia strains cultivated anaerobically in BDM medium amended with 6 different heavy metals

96-well microplates were also used to assess the effect of three different concentrations of copper (1.5, 15 and 150 mg/L) and cadmium (50, 500 and 5000 $\mu g/L$) on the growth dynamic of the four Clostridia strains. Microplates are a very valuable and easy way for testing, in a single experiment, different strains confronted to different heavy metals concentrations including multiple replicates. 200 μL of inoculated culture media were retrieved from an inoculated crimp vial and deposited in each well in an anaerobic atmosphere (Glove-box, Plas Labs inc., flushed with N_2 / CO_2 mixture). Microplate cover was then carefully sealed with hot melt glue and plate was incubated in a microplate reader (Biotek, Synergy HT) heated at 37°C. Growth was evaluated by measuring optical density every twenty minutes in each well at 595 nm and at 350 nm. Growth curves were then treated with R software to fit a logistic growth function on them and to assess the maximum growth rate (μ) for each treatment.
Figure 4.5 is showing that only two strains DSM797 (*Clostridium lituseburense*) and DSM5521 (*Clostridium disporicum*) are able to grow in the particular conditions offered by the microplates. For every treatment, except the ones with 15 and 150 mg/L of copper, the calculated growth rates for DSM797 is around $\mu = 0.03$. Growth is tree times slower with 15 mg/L of copper and completely inhibited with 150 mg/L. For DSM5521, the

growth rate is around $\mu = 0.012$ for all the treatments, except the one with 150 mg/L of copper which completely inhibit the growth. Finally, Neu28 *Clostridium beijerinckii* and DSM 16795 *Clostridium bartlettii* were not able to grow in microplates.

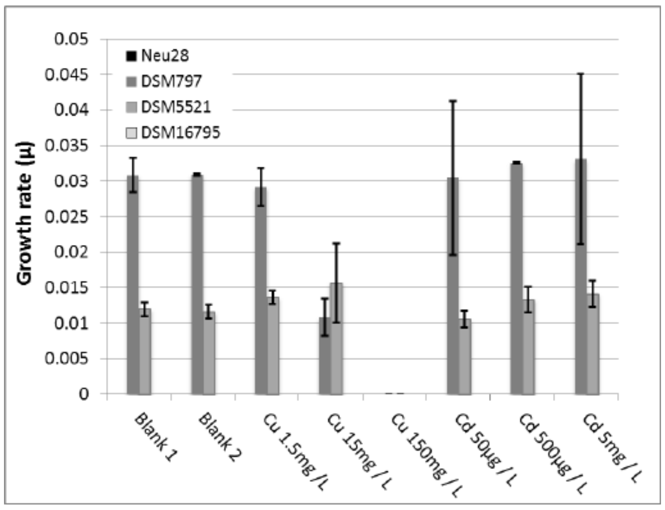


Figure 4.5: Growth rate determined by fitting a logistic curve on growth curves measured in microplaques

Discussion of growth dynamic's experiments. The first lesson that can be drawn from the growth experiments in microplates is related to the sensitivity of the four Clostridia strains to oxidative stress. During the filling of the wells of the microplate in the "anaerobic" atmosphere of the glove box, a very slight pink colouration of the medium was observed every time the experiment was attempted (3 different trials). This colouration was due to the oxydation of the resazurin dye, expressing a rise of the redox potential in reaction to residual traces of oxygen in the glove box. In fact, this colouration was observed even after 10 or 12 flushing of the working space of the glove box with an inert gas mixture (N_2 / CO_2). It is very difficult to remove any traces of oxygen. The fact that *Clostridium beijerinckii* and *Clostridium bartletti* could not grow in microplates might be the sign of their extremely low tolerance to oxidative stress.

In microplates, an inhibitory effect started to be visible for *Clostridium lituseburense* (DSM 797) when cultivated with 15 mg/L of copper, expressing a lower tolerance than *Clostridium disporicum* (DSM 5521). Such a concentration of copper is four times below the minimal inhibitory concentration (MIC) assessed for *Escherichia coli*,[63] but 700 times higher than the highest copper concentration measured in the sediment interstitial water from the most impacted area of the Bay of Vidy (work performed by another group of researcher from the Elemo project).[107] With cadmium, no inhibitory effect was measured with concentration up to 5 mg/L, which is 11 times below the MIC for *E. coli*, but 14'000 times above the maximum concentration found in sediment interstitial water from the Bay of Vidy. Finally, the total inhibition of the growth for both strains (DSM 797 and DSM 5521) cultivated in microplates with 150 mg/L of Cu is a little bit strange because the same strains grew well within crimp vials amended with an equal concentration of Cu during previous experiments (see above).

Growth curves established for Pb treatment, showed that no inhibitory effect could be measured with a concentration 10 times below the MIC for *E. coli*, but 10'000 times above values measured in sediment interstitial water. For Zn and Cr no clear inhibition was observed for the concentrations that were tested, which correspond to values 12 and 26 times below the MIC (no values are available for sediments interstitial water). Surprisingly, the growth of the four strains was completely inhibited with concentration of Ni 12 times below the MIC for *E. coli*. In summary, one can conclude from this part of the study that the levels of heavy metals found in solution in the interstitial water from anaerobic sediments of the Bay of Vidy are not high enough to inhibit the growth of the four Clostridia strains tested in this study.

Questioning of the work!
For different reasons, the exact determination of the concentrations at which heavy metals inhibit the growth of the four Clostridia strains is out of the scope of this work, which is not aiming at a physiological characterization of these four Clostridia strains. Furthermore, with the methodology used in this study it is impossible to give a reliable value for the concentrations of heavy metals at which the growth of the different strains is effectively inhibited. A close look at the Eh-pH diagram established for these heavy metals (Appendix A.7) demonstrated effectively that they almost all precipitate in a low soluble form at low redox potential (that are encountered in anoxic culture vials) and reduced sulphur molecules. At neutral pH, when Eh is changing from +820mV to -100mV (imposed by the couple Na-thioglycolate / ascorbic acid in the medium[108]), Cu^{2+} ions will precipitate as elemental copper or a copper sulphide (chalcocite or covellite) form by interacting with the thiol group from Na-thioglycolate (something like $TGA - S^{-}\ {}^{+}Cu^{+}\ {}^{-}S - TGA$). In oxygenated water lead can exist as Pb^{2+} cations which can precipitate in the form of lead carbonates or galena (PbS) when reduced sulphur is present in the medium.[109] Similarly, Zn^{2+} cations can precipitate as zinc sulphide (ZnS = sphalerite), and Cd^{2+} as cadmium sulphide (CdS = greenockite) - not reported in the Eh-pH diagram of the appendix. For nickel Ni^{2+} and chromium Cr^{3+} the speciation with sulphur ion is, however, not clear to me. But anyway, with such precipitates the exact concentrations of dissolved heavy metals that can interact with bacteria is only accessible if we measure it specifically.
The second problem with such precipitates rely on the fact that a reliable growth estimation by optical density measurement becomes impossible when metal concentrations are high, since metal precipitates interfere with optical density measurement. As growth of some of these strains (DSM 5521 for instance) produces a lot of gas an estimation of the growth by measuring the pressure developed inside the vials could be a good alternative to optical density measurement. Another way could be the determination of the growth dynamic via an isothermal microcalorimetric assay.[110]

Real concentrations of dissolved heavy metals The real concentrations of dissolved heavy metals in another growth experiment (Cu^{2+} 50 mg/L, Cd^{2+} 1mg/L, Cr^{3+} 100mg/L, Zn^{2+} 50mg/L, Ni^{2+} 50mg/L, Pb^{2+} 100 mg/L) were measured by inductively coupled plasma optical emission spectrometry (ICP-OES). At the end of the experiment, the

cultures were centrifuged (3000 x g during 5 minutes). Supernatant were carefully recovered, filtrated through sterile 0.45 μm acetate-cellulose filters and acidified at pH 1 with HNO_3. If necessary, samples were diluted with milliQ water to be in the right measurement range (highest standard concentration was 30 mg/L for every element measured). Measurements were performed with a Perkin-Elmer Optima 8300 device (Waltham, USA). For each element, emissions were measured at three different wavelengths allowing a better resolution in case of matrix interference. Intensities at each wavelengths were corrected against interferences from the matrix (interfering compounds from the culture media, like sulphates and phosphates) thanks to a solution of scandium used as internal standard.[111] Finally, concentration of each elements were recovered thanks to a 12 points calibration curve composed by 2% HNO_3 dilutions of the Perkin-Elmer Pure Plus Instrument Calibration Standard II.
As shown in Figure 4.6, there is no remaining copper, cadmium and lead in solution, whereas chromium and nickel seem only to slightly precipitate under the physico-chemical conditions of the culture vials. Finally, the results for zinc are not so clear. Either it does not precipitate readily in the physico-chemical conditions found in the culture vial and dissolved zinc ions are differentially adsorbed by the EPS matrix produced by the growth of the strains, either the redox potential was not low enough in all the vials to precipitate ZnS during the preparation of the media. In the light of these measurements, the results of sporulation and growth dynamic experiments should be interpreted in a different way, taking into account that all the Cu, Cd, Pb, and eventually the Zn precipitated by physico-chemical interaction with sulphides. These experiments should in fact be re-conducted without any sulphidic compound in the medium.

Condition tested	Cu	Cd	Cr	Zn	Ni	Pb
	mg / L					
Initial concentration	**50**	**1**	**100**	**50**	**50**	**100**
Without strain	0.0	0.0	59.7	35.9	42.2	0.0
Neu 28: *Cl. beijerinckii*	0.0	0.0	70.7	1.7	25.1	0.0
DSM797: *Cl. Lituseburense*	0.0	0.0	64.6	0.0	42.2	0.0
DSM5521: *Cl. disporicum*	0.0	0.0	71.5	15.5	38.2	0.0
DSM16795: *Cl. Bartlettii*	0.0	0.0	81.0	0.4	35.1	0.0

Figure 4.6: ICP-OES measurement of the dissolved heavy metals found in the culture vials during anaerobic cultivation experiments

4.3 Heavy metals immobilisation by Clostridia

As the four strains of Clostridia were producing a large amount of exopolymeric substances (EPS) and as these substances are well known for their adsorbing properties on metals, it seemed interesting to us to perform some experiments in this direction. These experiments were, however, also performed before we understood the metal precipitation effect of the Na-thioglycolate from the culture medium.

4.3.1 Immobilisation in anaerobic conditions

Two other sets of crimp vials were filled with BDM medium and amended with 2 different heavy metals (5 mg/L of Cd^{2+}, 150 mg/L of Cu^{2+}) for a first experiment, then 6 different heavy metals (1 mg/L of Cd^{2+}, 50 mg/L of Cu^{2+}, Ni^{2+} or Zn^{2+}, 100 mg/L of Pb^{2+} or Cr^{3+}) for a second. The two sets of vials were prepared to allow triplicates with the four Clostridia strains. Pictures of the vials were taken regularly until the end of the growth to allow comparison between treatments. At the end of the experiment, the cultures were centrifuged (3000 x g during 5 minutes) and pellets were recovered for microscopic observations.
When growing with high amounts of heavy metals the EPS matrix produced by the four Clostridia strains become clearly visible macroscopically since it is encrusted by coloured heavy metal precipitates. Figure 4.7 shows this evidence for the four Clostridia strains and the 2 heavy metals tested. Precipitates of heavy metals seemed to be tightly linked to EPS matrix, even in the case of strain Neu28 (*Clostridium beijerinckii*) that produces a thin and compact EPS matrix stuck at the bottom of the vials. In comparison, the three DSM strains produce much more EPS, but the appearance of each of them was different. DSM797 (*Clostridium lituseburense*) produce also an EPS structure that is stuck at the bottom of the vials, but it seems to be more filamentous and less compact than the one from Neu28. DSM5521 (*Clostridium disporicum*) is the strain that reaches the highest level of turbidity when growing. At the end of its growth, visual observation of the vials reveals a very slight matrix of mucus like EPS. In this case, heavy metals precipitates were also stuck at the bottom of the vials, but a lot of EPS were still in suspension in the overall spent medium. Finally, DSM16795 (*Clostridium bartlettii*) produces a gelatinous EPS matrix encrusted by heavy metals precipitates. This structure was also stuck at the bottom of the vials. A close look at the bottom of the blank vials (not inoculated BDM media amended with heavy metals) reveal that precipitation of heavy metals also occur without the help of any of the Clostridia strain. Without sticky EPS matrix, heavy metals precipitates are simply more easily dispersed when shaking the vials.
The pictures of the vials amended with 6 different heavy metals are presented in appendix A.8. The visual aspect of heavy metals precipitates varied from one to another. Copper and nickel were dark brown, cadmium was greenish, chromium was pinkish, zinc was whitish and lead was black.

Epifluorescence microscopy To visualize more precisely the link between EPS and heavy metals, a mixture of calcofluor white (Fluorescent brightener 28) and Nile red was used to stain EPS and more particularly β-polysaccharides and hydrophobic sites of lipids, respectively.[112,113] SYBR Green® I was also added to the staining procedure to highlight living bacterial cells.[114] The procedure was adapted from a published work[113] and consisted in the following steps: Bacterial and EPS pellets were first recovered and diluted in 5 mL physiological solution for conservation. For staining, some pieces of pellets were taken with a Pasteur pipette and resuspended in 1 mL sterile milliQ water supplemented with 100 μL

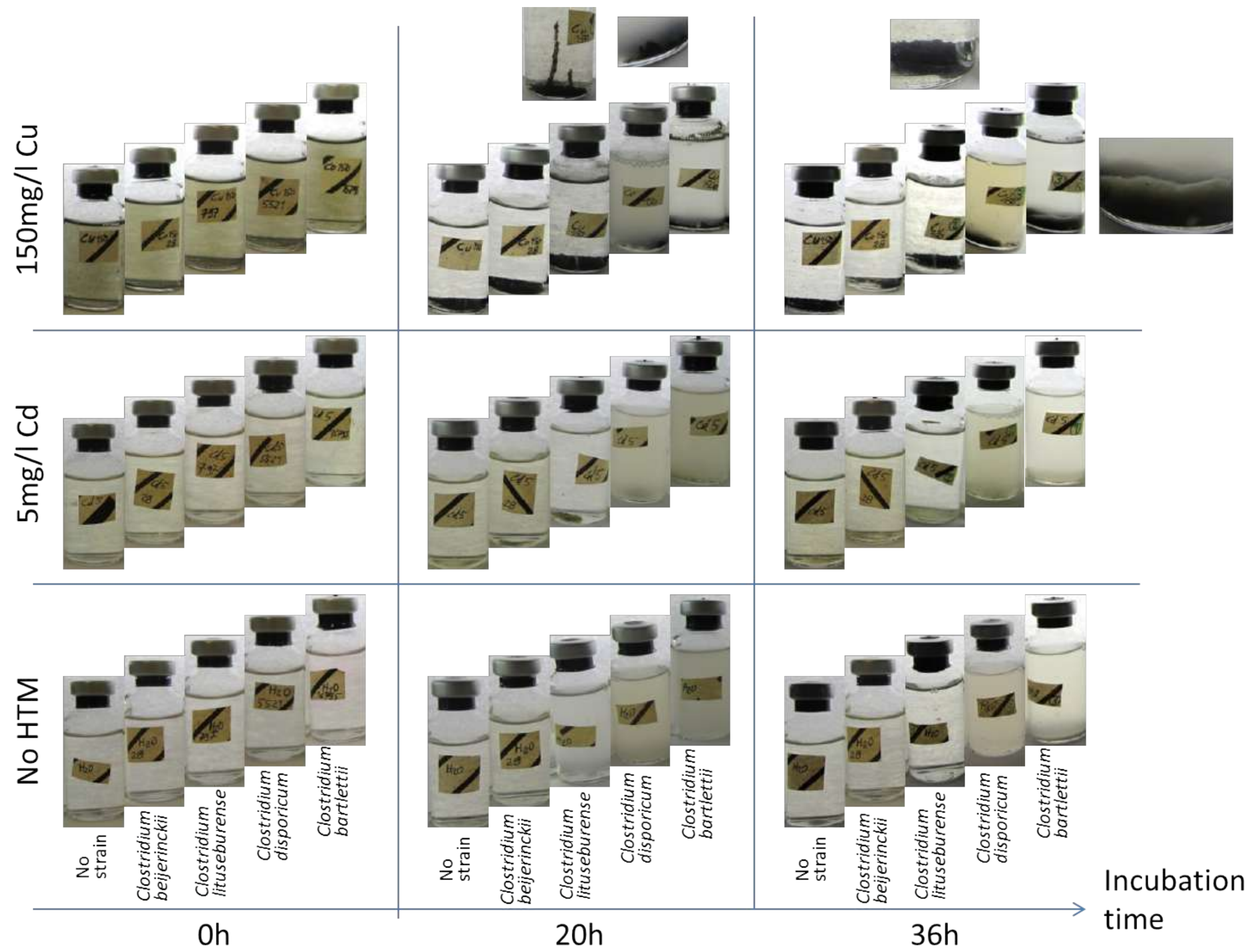


Figure 4.7: Precipitation of heavy metals in culture vials and agglomeration of precipitates in exopolymeric substances produced by the four Clostridia strains

of bicarbonate buffer solution (pH 10.8, 0.1M). Fluorescent brightener 28 (Sigma-Aldrich, St-Louis, MO, USA) was added at a final concentration of 100 mg/L and mixture was incubated at room temperature in the dark (RT-D) during 30 minutes. Nile red (Sigma-Aldrich, St-Louis, MO, USA) dissolved in DMSO was added at the final concentration of 600 $\mu g/L$ and mixture was incubated at RT-D during 10 minutes. Finally, SYBR green® I (Sigma-Aldrich, St-Louis, MO, USA) was added at a final concentration of 5x and mixture was incubated at RT-D for 15 minutes. A little piece of pellet was sampled in the tube and deposited on a glass slide and covered with a coverslip. Microscopic observation was performed with a Leica DMR light microscope (Wetzlar, Germany) equipped with a 100 W Hg ultra-high pressure lamp for epifluorescence and filter cubes A (BP340-380, LP425) for calcofluor fluorescence, Y3 (BP545/30, BP610/30) for Nile-red fluorescence and L5 (BP480/40, BP527/30) for SYBR green® fluorescence. (Leica, Wetzlar, Germany). Pictures of the more interesting zones were taken with each filter cube as well as with phase contrast and assembled with an imaging software.

Figure 4.8 shows the pictures obtained by the combinations of the 6 heavy metals and the 4 strains tested. For heavy metals like Cu, Zn, and Pb, dense aggregates of dark material are observed within either a red (hydrophobic sites of lipids) or a blue (β-polysaccharides) halo. Some heavy metals, like copper and lead, seems to inhibit the fluorescence of calcofluor, but not of Nile red nor SYBR green® I. Interestingly, endospores from DSM797 (*Clostridium* *lituseburense*) and DSM5521 (*Clostridium disporicum*) are highlighted by calcofluor staining. Finally, except this funny observation on the highlighting of endospores by calcofluor, this EPS staining experiment did not give so much interesting informations fur the purpose of this work.

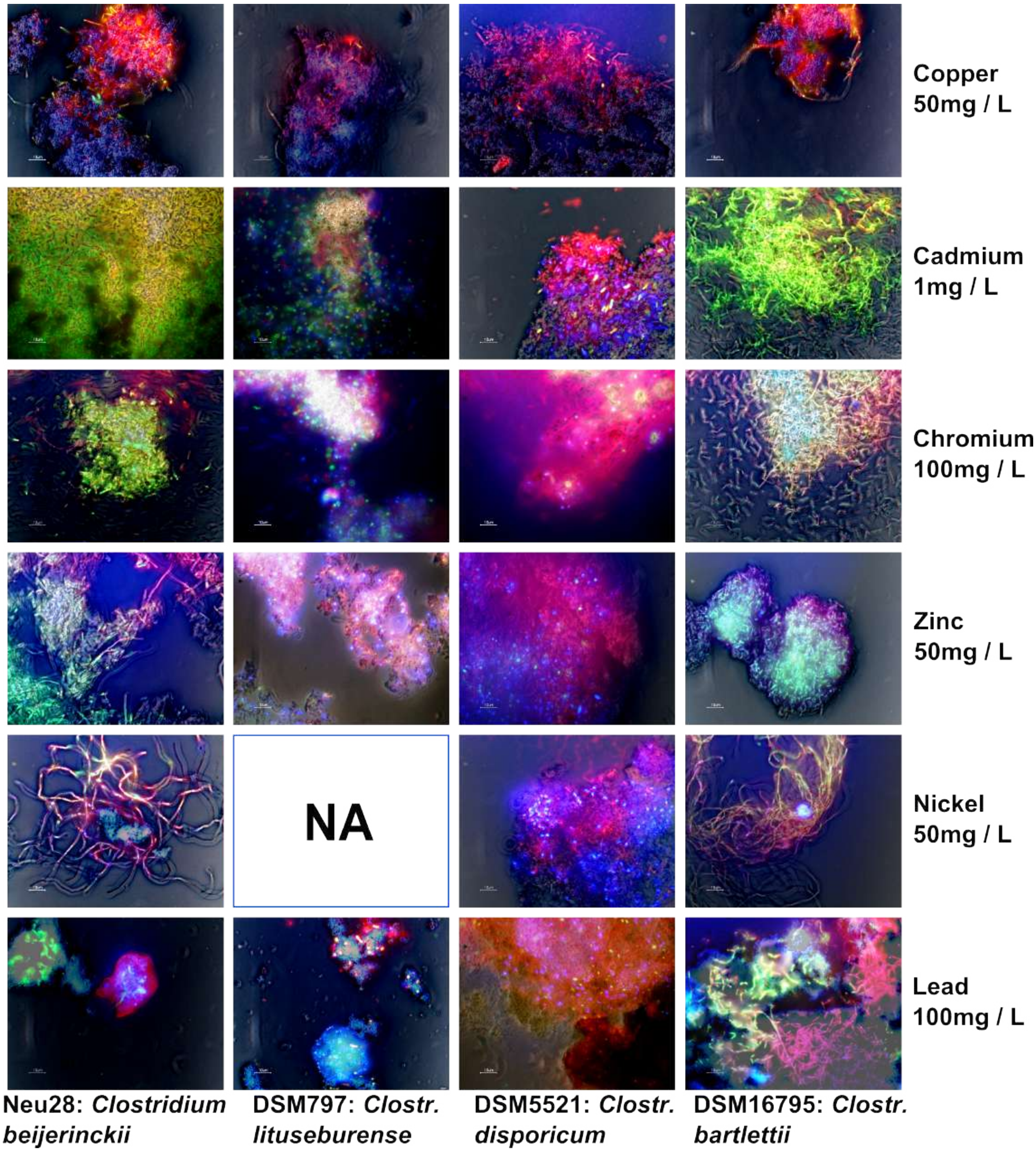


Figure 4.8: Epifluorescence microscopy for the visualisation of exopolimeric substances matrix produced by the four clostridia strains cultivated in BDM medium amended with 6 different heavy metals. Staining procedure was performed using Calcofluor, Nile Red, and SYBR green® I fluorescent dyes.

Scanning electron microscopy Pellets from the 4 clostridia strains and 1 blank (without bacteria) amended either with 50 mg/L of copper or 1 mg/L of cadmium were recovered and treated for SEM observation. Samples were fixed in 2.5% of glutaraldehyde and

dehydrated by an increasing ethanol series before sputter coating with gold.[115] Specimen observation was performed on a high vacuum Scanning Electron Microscope (SEM) using a Phillips ESEM-FEG XL30 Field Emission Gun Scanning Electron Microscope (FEG-SEM), coupled to an EDAX Energy Dispersive Spectrometer (EDS).
Figure 4.9 shows a selection of two pictures showing two strains (DSM797 and DSM5521) amended one with cadmium and the second with copper, respectively. In both cases, precipitates of minerals didn't form any crystal like structures, but amorphous aggregates. When looking at the composition of mineral precipitates, the EDS data of both pictures show an enrichment of cadmium and sulphur, or copper and sulphur, when compared to the surface of bacterial endospores. These elements are also highlighted when EDS measurement are performed on mineral precipitates found in blank samples (without bacterial growth). These results corroborate the hypothesis that copper and cadmium effectively precipitated in the form of metal-sulphur minerals by a purely physico-chemical phenomenon occurring in the culture vials.

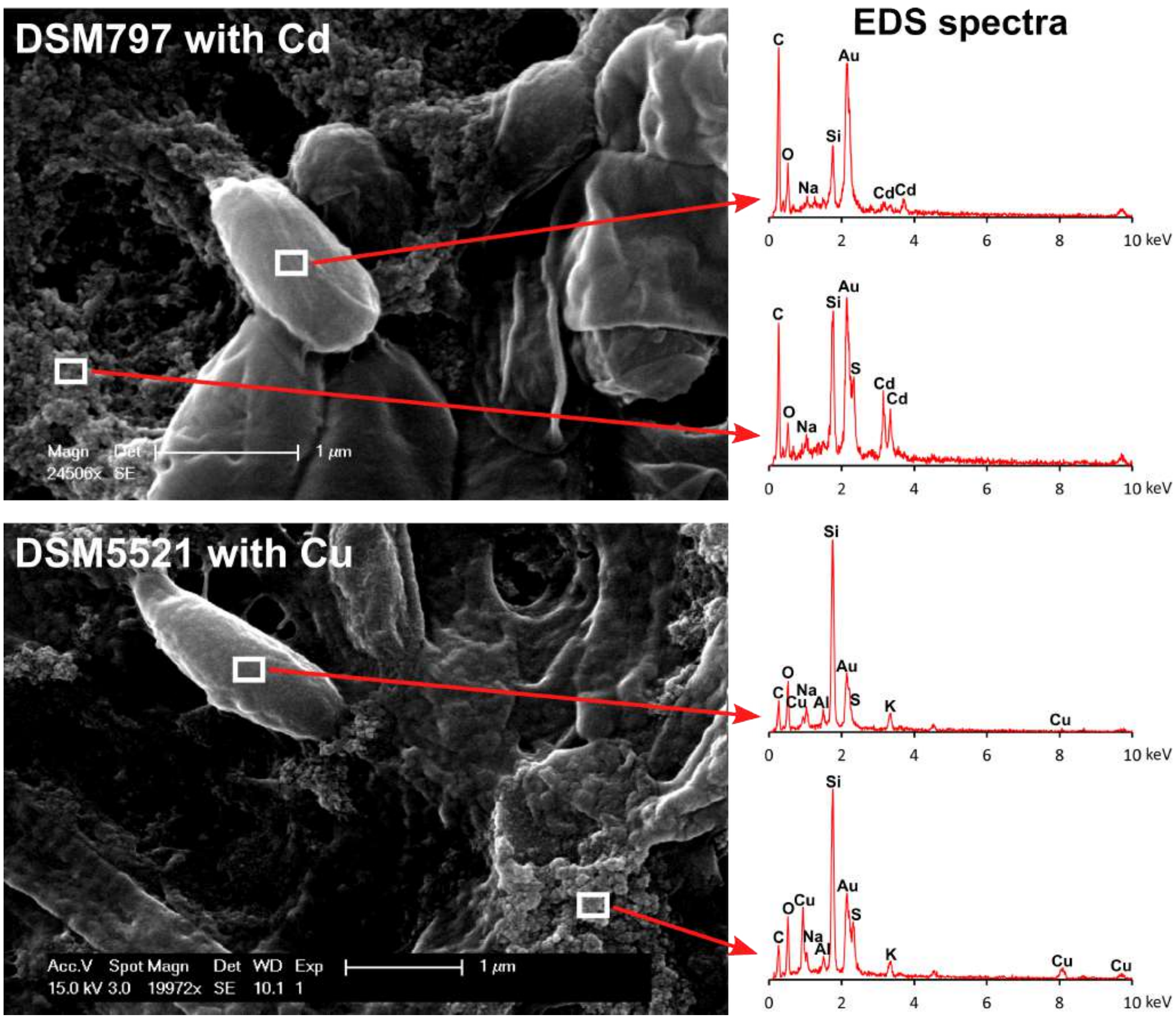


Figure 4.9: Scanning-electron microscopy on culture pellets of two Clostridia strains amended with heavy metals. Assessment of the sample composition by EDS probe measurement.

4.3.2 Complexation in aerobic conditions

Bacterial and EPS pellets from metals-free experiments were recovered and re-suspended in 20 mL physiological water amended with 6 mg/L of copper (from a solution of $CuSO_4$). 1.5 mL of this mixture was distributed in 12 eppendorf tubes of 2 mL and incubated for different durations at room temperature while shaking at 5 Hz in a bead beater. Four blank tubes without culture pellets were also prepared. Before measurement of dissolved copper, tubes were centrifuged during 2 minutes at 3000 rpm in a minispin to pellet cells and EPS-metals complex. Finally, bicinchoninic-acid method was used to measure the remaining dissolved copper in the supernatant.
Three reagents were prepared: reagent A (30 g trichloroacetic acid in 100 ml of DI water), reagent B (35.2 mg ascorbic acid in 100 mL DI water) and reagent C (6 mg bicinchoninic acid, 3.6 g $NaOH$, 15.6 g HEPES in 90 mL DI water). For measurement, 250 μL of reagent A was added to 750 μL of sample. Mixture was centrifuged in a minispin at maximium speed (13'000 rpm) for 5 minutes. 500 μL of supernatant was recovered and mixed with 100 μL of reagent B in a new tube, then 400 μL of reagent C was added. Mixture was incubated at room temperature for 2 minutes and optical density measurement was performed with a spectrophotomer at the wavelength of 354.5 nm. A standard curve was assessed with copper concentrations up to 5 mg/L to convert optical density units to dissolved copper concentrations.
Figure 4.10 shows a rapid diminution of dissolved copper from 4.6 mg / L to a level below 2 mg /L within only a few minutes of incubation. After 1 hour and a half of incubation, levels of residual copper reached values between 0.9 and 1.2 mg / L. This last experiment clearly demonstrated that EPS from the four Clostridia strains are able to adsorb copper in aerobic conditions. This property could be interesting for removing heavy metals during bioremediation process of contaminated environments. More experiments are, however, still required to extended these findings to other metals.

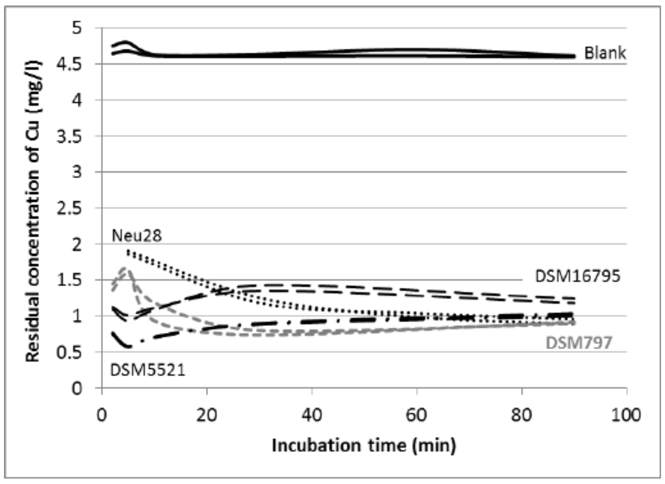


Figure 4.10: Adsorption of copper by EPS and cells pellets from the four Clostridia strains in aerobic conditions. Within few seconds, the Clostridia's pellets reduced readily the dissolved copper concentration.

4.3.3 Heavy metals remobilisation

Once entrapped in sediments, what is the fate of heavy metals? Do they remain there *ad eternam*? I guess not! The main idea behind this experiment was to test if the EPS matrix secreted by the Clostridia strains could have a stabilisation effect on heavy metals in the environment when they are released by sediments. To simulate this situation, heavy metals amended cultures were exposed to oxygen prior measurement of dissolved heavy metals by ICP-OES. The four Clostridia strains were cultivated in heavy metals amended BDM media as described above. After 16h (Neu28, DSM5521 and DSM16795) or 24h (DSM797) of incubation, the pressure inside the culture vials were first carried back to atmospheric pressure by inserting a needle through the caps. After pressure equilibration, 5 mL of pure oxygen was inserted into the vials through a 0.2 μm cellulose-acetate filter. The vials were finally slightly shaken and stored at $4°C$ for 72 hours. Finally, cultures were centrifuged (3000 x g, 10 minutes), supernatants were recovered and filtrated through 0.45 μm cellulose acetate filters. Samples were then acidified at pH 1 with HNO_3 and stored at $4°C$ for their preservation until ICP-OES measurement.
The results shown in Figure 4.11 demonstrate that once precipitated with sulphur anions, heavy metals remain quite stable even after an exposed to oxidative conditions. Effectively, Cd, Cr, Zn, and Pb did not resolubilize as demonstrated by their absence in the ICP measurements. Values for dissolved Cr are, however, contradictory when compared to previous anaerobic values measured by ICP-OES to assess dissolved heavy metals (Figure 4.6). This could perhaps be explained by small methodological variations during the preparation of the culture vials, leading to small differences in the redox potential from one series to another. This should be mastered more properly for further experiments. Concerning Ni, it remains in solution in all the vials as already shown in Figure 4.6. Finally, as demonstrated by the level measured in the blank, Cu-sulphur precipitates dissolved again when exposed to oxygen. However, it is not clear if precipitates have also been dissolved in cultures with the four Clostridia strains and then back adsorbed by EPS matrix for Neu28, DSM 5521, and DSM 16795, or if these results are also artefacts due to methodological variations. More work need to be done in this direction.

72h exposure to oxygen

Condition tested	Cu	Cd	Cr	Zn	Ni	Pb
	mg / L					
Initial concentration	**50**	**1**	**100**	**50**	**50**	**100**
Without strain	17.2	0.1	0.0	4.6	43.6	0.1
Neu 28: *Cl. beijerinckii*	0.5	0.1	0.0	4.0	44.3	0.3
DSM797: *Cl. Lituseburense*	14.5	0.1	0.0	1.4	24.2	0.1
DSM5521: *Cl. disporicum*	2.4	0.0	0.0	5.1	33.2	0.0
DSM16795: *Cl. Bartlettii*	1.4	0.0	0.0	7.9	30.0	0.2

Figure 4.11: ICP-OES measurements of the quantity of heavy metals released from heavy metal precipitates and/or EPS complex after an exposition to oxygen for 72 hours.

4.4 Chapter conclusions

As the results of the experiments presented in this chapter are not so easily valuable, it is difficult to draw conclusions on this chapter. However, the lack of success encountered in many experiments performed in this chapter constitute a bases for avoiding the same errors in the design of future experiments with anaerobic cultures and heavy metals. Furthermore, the different questions highlighted during these experiments were also helpful for imagining the mechanisms that could occurs in heavy metals polluted sediments. Effectively, as long as heavy metals are entrapped in anaerobic sediments, either in the form of metal-sulphur precipitates or in the form of complexes with EPS (or other complexation sites of the organic matter), one can imagine that heavy metals can last for a quite long time in this compartment of the ecosystem. However, at one time or another, sediments could experience some changes leading to the release of heavy metals in the aquatic environment. Some processes that could occur are for instance, their remobilization and exposition to oxygen, the biomineralization of organic matter, and the bio-oxidation of sulphur compounds. In the first case, if heavy metals are precipitated as metal-sulphur minerals, they probably will not be dissolved so readily as demonstrated above. In contrast, if heavy metals are adsorbed to EPS, they could probably be leached more rapidly due to more intensive degradation of the EPS matrix in aerobic conditions. In any case, if EPS matrix from Clostridia strains is still present in the sediments it could contribute to avoid the release of heavy metals in the environment. Microcosm experiments with sediments could be an interesting way to verify different hypothesis expressed in this chapter.

5

Conclusions and Outlook

This chapter concludes the thesis by first providing the main outline of the work that was done, then a critical review of some key points of this thesis, and finally the potential outlook and the future development that could be made.

Contents

5.1 Summary

Sediments of the Bay of Vidy (Lake Geneva, Switzerland) are polluted by heavy metals due to the release into the bay of the effluents of the waste water treatment plant of Lausanne. Due to their capacities to escape harsh environmental conditions through endospore formation, it was first hypothesised that endospore-forming Firmicutes will be favoured in lake sediments harbouring high levels of heavy metals pollution. To verify this hypothesis it has been chosen, on the one hand, to perform molecular analysis of the *in situ* bacterial community, on the other hand, to perform a chemical characterization of sediments. As the second was performed thanks to the help of different specialist, the first, the molecular part, was done entirely by ourselves and necessitated the development of especially tailored tools to analyse the endospore-forming Firmicute's portion of the overall bacterial community. It has effectively been demonstrated in our laboratory that endospore-forming Firmicutes are often absent (or are underrepresented) in molecular inventories from complex environments where they should normally be present. As it was hypothesised that the bottleneck was at the level of the DNA extracted, a new method was developed in our laboratory.[116] The first part of this work was devoted to the evaluation of this new extraction methodology to see if it meets the need of this particular study. As it seems to be the case, we decided to continue with this method.
Once extracted, DNA served different purposes. The quantification of endospore-forming Firmicutes in bacterial communities was performed thanks to quantitative PCR (qPCR) on the gene *spo0A*, which is reported to be specific to this taxa. As "universal" primers for targeting this gene did not exist, it was, therefore, necessary to develop and to validate them. This constitute the second part of this thesis.
Finally, 454-pyrosequencing was performed on 24 DNA samples extracted from different layers of 10 sediment cores selected for their representativeness of contrasted pollution concentrations in the Bay of Vidy. The confrontation of both, chemical and pyrosequencing data, highlighted that endospore-forming Firmicutes were predominant members of bacterial communities (constituting up to 60% of the 16S rRNA gene copies in the community) corresponding to the most polluted sediments. This was also supported by the results of qPCR quantification of the *spo0A* gene *versus* 16S rRNA gene (for the overall community). Four Clostridia strains (*Clostridium beijerinckii*, *Clostridium lituseburense*, *Clostridium disporicum*, and *Clostridium bartlettii*) were identified as dominant in these samples.

To try to understand why this particular link could be observed in the environment, different cultivation experiments amended with heavy metals were then performed with these four Clostridia strains. In the conditions found in sediments, concentration of dissolved heavy metals were, however, too low to trigger endosporulation or to affect growth dynamic of these strains. As a counterpart, large production of exopolymeric substances (EPS) was reported during the growth of these strains. Either under aerobic or anaerobic conditions,

this matrix of sticky material seems to entrap heavy metal ions, by sticking metal-sulphur precipitates in anaerobic conditions, or by chelating metallic cations in aerobic conditions. This part of the work is not yet finished and gives some contradictory results, maybe due to not yet entirely mastered methodological aspects of the anaerobical growth with heavy metals amendment. This part of the work constitutes, however, the first steps for an eventual post-doctoral research on this topic.
But what about the presence of these Clostridia strains in sediments? A search in the literature showed that these strains are typical inhabitant of human faeces, growing best at around 30 to $37°C$, therefore, excluding the hypothesis of a selective growth of this sub-population in sediments at a temperature around $10°C$. A more careful analyses of pyrosequencing data, taking also total bacterial abundances into account, reveal that the absolute abundances of Firmicutes taxa found in the most contaminated sediments were lower than those found in sediments collected close to the releasing pipe, leading to the hypothesis that these Firmicutes constitute a remainder of bacterial communities released by the waste waster treatment plant. Logically, the next question was: Why does a waste water treatment plant, based on an aerobic treatment, release so much anaerobic bacteria in the environment? Here again, the answer comes from a literature review, showing that Clostridia should normally not be found (or in very small quantities) in effluents from a waste water treatment plant.[117] Figure 5.1 shown below, intends to represent the structure of bacterial communities in the different steps of the fate of waste water, from its source (mainly, human faeces) to its release into the lake, after the treatment in the waste water treatment plant. This figure is based on the results obtained during this thesis, but also in two publications: first a study of microbial communities found at different stages in a waste water treatment plant of Hong-Kong[117] and second a study on the core microbial community of human faeces.[118] Of course, the identification at the order level is not accurate enough for doing precise comparisons on bacterial community structure, but the published data does not go deep enough into the taxonomical identification. It, however, permits to see in which steps the dominant groups are the same, particularly in the case of Firmicutes which are clearly indicating a faecal signature.

This figure slightly support the hypothesis of the regular rejection of activated sludge and incompletely treated waste water into the aquatic environment. A close look at the "Bilan de l'épuration Vaudoise 2011", published by Canton Vaud, shows that the waste water treatment plant of Lausanne experienced some flooding events leading to the rejection of more that 10% of untreated (or un-properly treated) waste water into the lake. During these flooding events one can easily imagine that some flocs and aggregates from activated sludge are also carried away from the plant and finish on the floor of the lake, around the rejection pipe. When talking about this, it seems effectively relatively straightforward that the material deposited around the pipe should be composed of heavy aggregated particles rather than liquid treated waste water that can be mixed with lake water and carried away from the Bay of Vidy.
As the activated sludge is well known for its adsorptive capacity for heavy metals, its deposition around the rejection pipe would also explain why so high levels of heavy metals are found in this area.

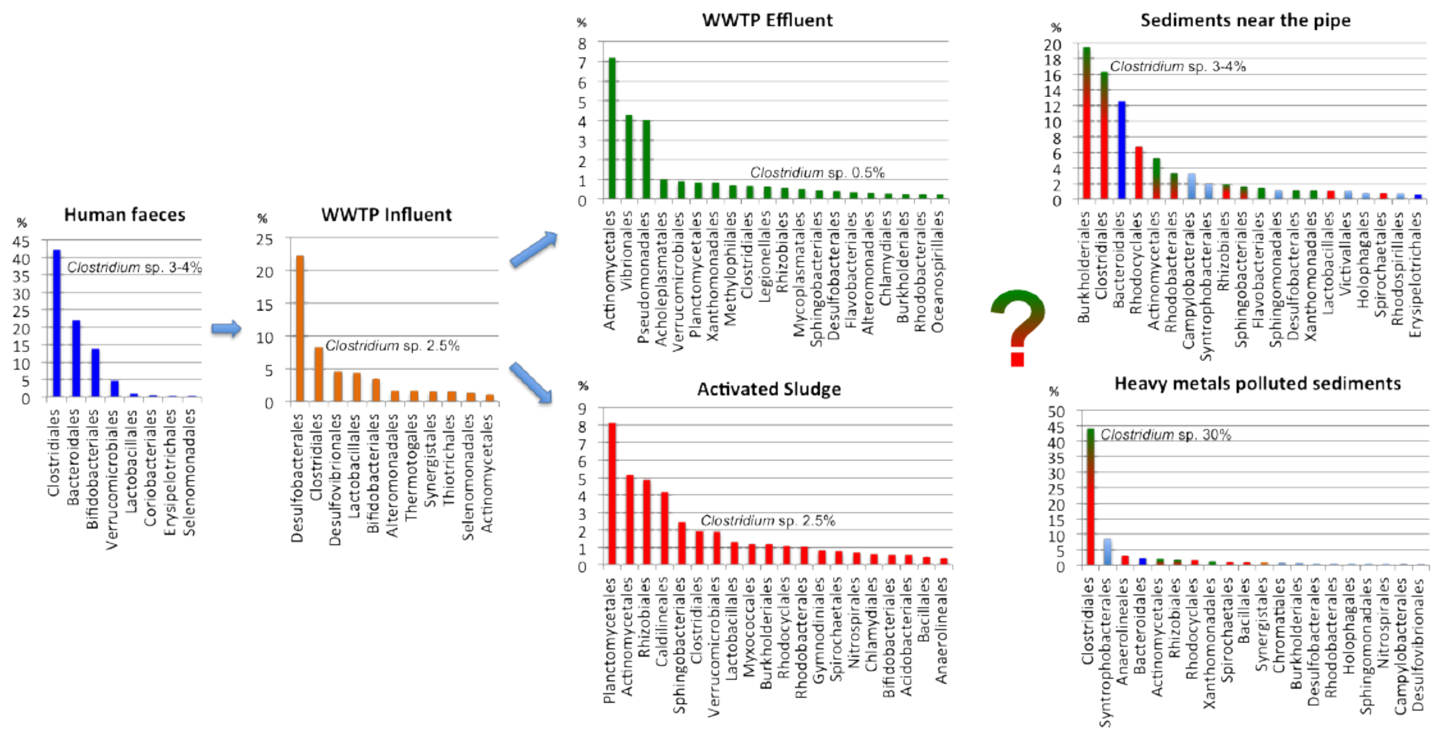


Figure 5.1: Relative abundances of bacterial orders in different steps of waste water treatment, from its source (faecal core microbiota[118]) to its sink in the sediments of Lake Geneva(this study), via its processing in waste water treatment plant (WWTP influent, effluent, and activated sludge[117]) The colors in the last bar plots (to the right) tries to track (based on presence-absence) the origin of the taxa found in sediments(blue: human faeces, orange: WWTP influent, green: WWTP effluent, and red: activated sludge). The results are not so clear, but seem, however, to indicate that the structure of microbial communities found in sediments are more similar than those found in activated sludge, WWTP influent, and human faeces.

As is often the case in science, a research gives unexpected results that open new questions, but also answer unexpected aspects of the research. Here, the main aim was to demonstrate the incredible skills of endospore-forming Firmicutes to resist to heavy metals pollution, but rather than this we demonstrate for the first time in this context that the pollution observed in the Bay of Vidy is a consequence of a dysfunction of the waste water treatment plant, which is flooded by the rejection of an excess of runoff water. This is, however, not the only plant that behave like that. Many other plants in Switzerland and probably in other countries experienced the same problem. That is why the political authorities in developed countries invest so much in the recent years, to separate runoff water from waste water. But anyway, if the effect of heavy metals on endosporulation had not been clearly demonstrated during this thesis, endospore-formation is, however, the resistance mechanism that permits Clostridia to survive in sediments. Why do they are dominant in highly contaminated sediments is another question that requires more investigations.

5.2 Some unanswered questions

At the end of this work there are principally two main questions that remain unanswered. The first one is: Why such high concentrations of heavy metals are found in sediments? Of course, the hypothesis of a regular flooding of the plant, rejecting sludge and flocs into the lake, could give an answer. But in this case, why concentrations are so high in sludge? From where do these heavy metals come? An accumulation mechanism by the gradual sequestration of heavy metals from runoff water could be a scenario. But is it a plausible explanation? A close look at the data published in 2011 for the functioning of the waste water treatment plant of Lausanne[76] shows that the quantity of runoff water flowing into the plant is comprised between 7000 and 23'000 m^3/day (depending on the values used) in an annual mean. The total amount of dehydrated sludge produced is around 20 t/day, containing 259, 0.8, 24, and 520 ppm of Cu, Cd, Pb, and Zn, respectively. Every kilogram of dehydrated sludge has, therefore, concentrated the content of heavy metals contained in around 1000 L of runoff water. In runoff water, concentration of the same metals are typically in the range of 25 $\mu g/L$ of Cu, 0.2 $\mu g/L$ of Cd, 15 $\mu g/L$ of Pb, and 200 $\mu g/L$ for Zn.[13,119] The concentration factor when compared to values found in dehydrated sludge stand, therefore, between 1000 and 10'000 times, depending on the metal. The scenario of an accumulation from runoff water is, therefore, quite plausible as values stand in the same order of magnitude. Taking into account the meteorological variations in terms of raining events in this particular climate, such differences in the contents of heavy metals does not seem impossible. Of course this rough evaluation does not constitutes an ultimate evidence. It, however, demonstrates that this concentration phenomena could happen. But what about the concentrations of heavy metals found in sediments? If they are in the same range as the concentrations found in dehydrated sludge for Cu, Zn, and Hg, they are around 10 times higher in sediments for Cd and Pb. Either the evaluations for concentrations in runoff are too low, or sequestration by activated sludge is not so much efficient for these two metals as it is for the others. In both cases, one can hypothesized a concentration phenomena that takes place in the sediments. This could happen most probably through a loss of mass due to the mineralisation of organic matter, or by precipitation of metal-sulphur minerals in the sulphur producing zone down into the sediments. Both phenomena probably occur in parallel. With the data collected during this thesis, it is, however, difficult to give a definitive answer to this question!

If we accept the hypothesis of the regular waste water treatment plant flooding, the second unanswered question is: Why Clostridia are found as dominant in the most polluted sediments whereas they are not either in microbial inventories of human faeces, in the influent or activated sludge from waste water treatment plant (see figure 5.1)? To answer this, the more plausible explanation is that the four Clostridia strains identified could perhaps be more resistant to different kind of stress factors experienced in the pipes driving sewage into the treatment plant, during the aeration treatment in the basins, and during their release into the lake and burying into the sediments. When studying dominant Firmicutes taxa enumerated in the publication on the core microbiota of human faeces, it seems that the only endospore-formers belong to *Clostridium*. Event if this taxa is not so much represented in the beginning (around 3-4% of the bacterial community of human faeces), through endospore-formation it can easily become dominant as the other taxa are eliminated by the selective pressure of the different environments encountered during waste water treatment and their dispersal to sediment. When considering all of this, one can draw the hypothesis that the absolute abundance of the endospore portion of the communities remain almost constant during all the process, but that the other taxa fluctuate depending on the environmental factors. This hypothesis needs, however, to be verified and different experiments should be done to understand in a more detailed way this particular system, assessing for example the quantity of *Clostridium*'s endospores in human faeces or at different stages of the treatment process at the waste water treatment plant.

5.3 Critical review of this work

As mentioned in the preface, the thin red line followed all along this thesis is centred around bias problematic. Of course, much of these biases are already known since long date and their enumeration does not constitute a big advance for science. However, this slow awareness of the limits of science in this particular topic constituted a big advance for my personal understanding of the different methods used during this thesis. Here are the particular bias I had to deal with during the last three years.

1. **Coring process bias**
   During coring process, especially with a gravity corer, sediments experience a compression.[120] The importance of this compression mainly depends on the design of the corer and on the coring tube dimensions. Depending on the intensity of this compression the dating of the different layers of the core could be seriously altered. In the case of this thesis project this bias should, however, not induce so much consequences on the interpretation of our results, since no absolute dating was performed. However, as two different kinds of corer were used in this project (corer of the MIR submersible and gravity corer from "La Licorne"'s boat) some small differences in the real thickness of the layers that were subsequently subsampled could arise. A more important bias could occur when gravity corer reaches the bottom of the lake. It is effectively, relatively frequent that all the corer tilts when the tube is planted into the sediment, leading to a redistribution and a resuspension of the enclosed material.[120] The visualisation of the coring process by the GoPro camera attached to the corer during the second sampling campaign from the boat shows, however, that this did not occur with this particular type of corer. The third coring bias that could occur is related to the contamination of the *in situ* bacterial community of the sediment by the coring tube. Effectively, this material is not sterile and could contaminate a little bit the peripheral layer of the core. That is the principal reason explaining why we choose to take samples in the middle of the core for microbial community assessment. Even if a small contamination can arise by this way, the contribution of these few *ex situ* bacteria is probably insignificant when compared to the many times higher abundances of bacterial taxa from the *in situ* community.

2. **Core storage bias**
   As discussed in the methodological chapter of this document, the microbial community

could evolve a lot during a long and improper storage of the sediment core. During this step of the work we had done a maximum to minimize this bias as much as possible. Except contradictory literature references as well as our common sense, we have, however, no mean to clearly verify that any significant changes in composition of the bacterial community has occurred during this storage period. The CCA ordination plot shown below (Figure 5.2), however, demonstrate that this factor has only few incidences on the composition of the bacterial community during a storage period ranging from one to seven days (see the short length of the storage arrow in the CCA below)

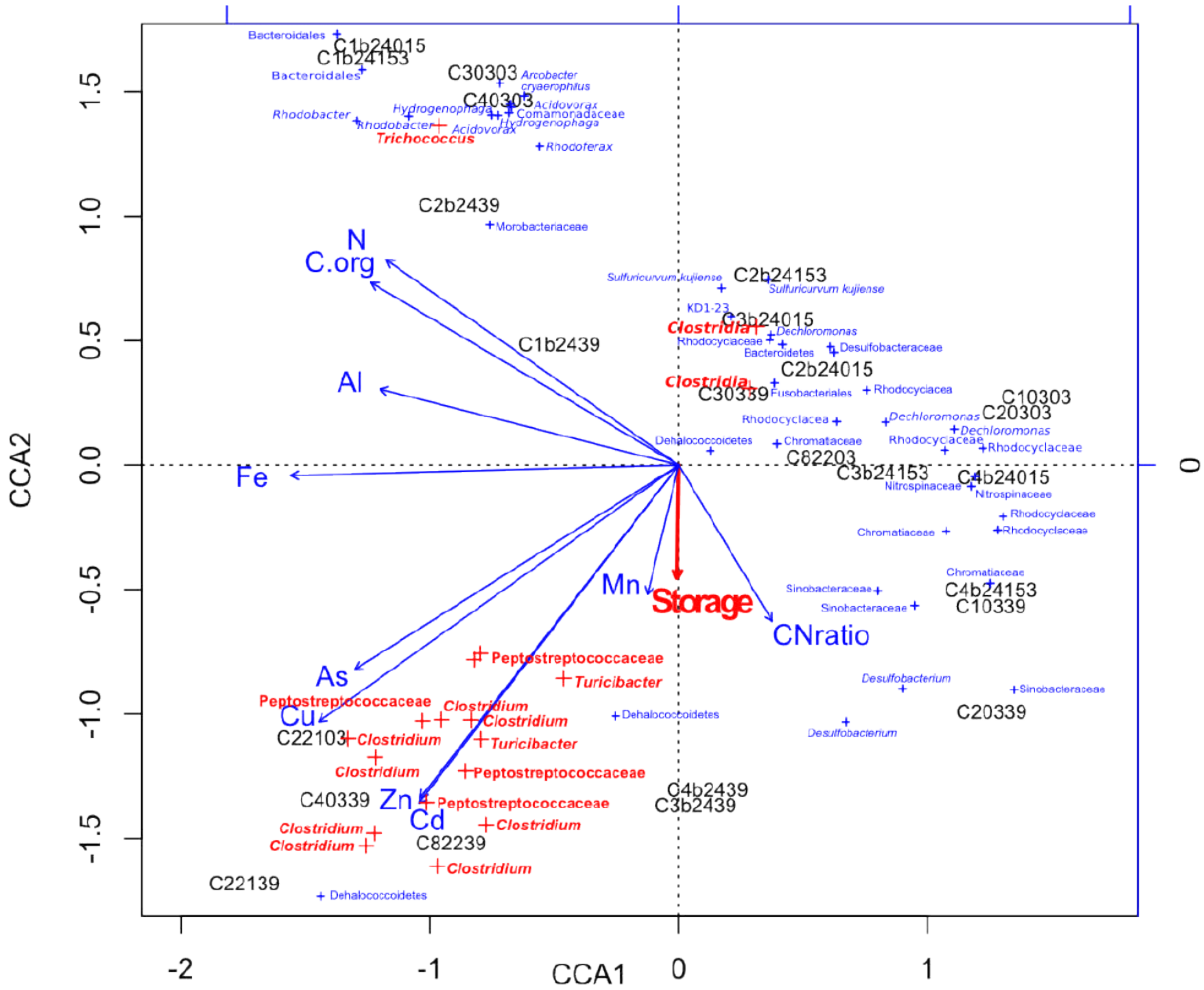


Figure 5.2: CCA ordination plot representing the relations between sediment samples based on microbial communities composition, as well as environmental parameters. In this figure the effect of the storage period of the cores prior to their treatment is represented by the red "Storage" arrow. When compared to other factors its effect is not significant

3. **Cell-separation methodological bias**
   Since the beginning of this work I was quite convinced that the separation protocol with the ultra-turrax was inducing an important bias favouring the more resistant structures, like Gram-positives and endospores. Even if such a bias was occurring this was, however, not really a problem for this work since it aimed at studying these taxa. An overestimation of this portion of the community should not have any important incidences on the conclusions that can be drawn. Ultra-turrax is composed by a rotor which run at high speed (14'000 rpm) in the sample, forcing

the biggest particles to pass through a small aperture in the stator. A search in the literature shows, surprisingly (but hopefully for this work!), that this treatment does not induce an important change in the communities when compared to softer method like stomacher homogenization.[121] Figure 5.3 presented below shows the results of such an evaluation.

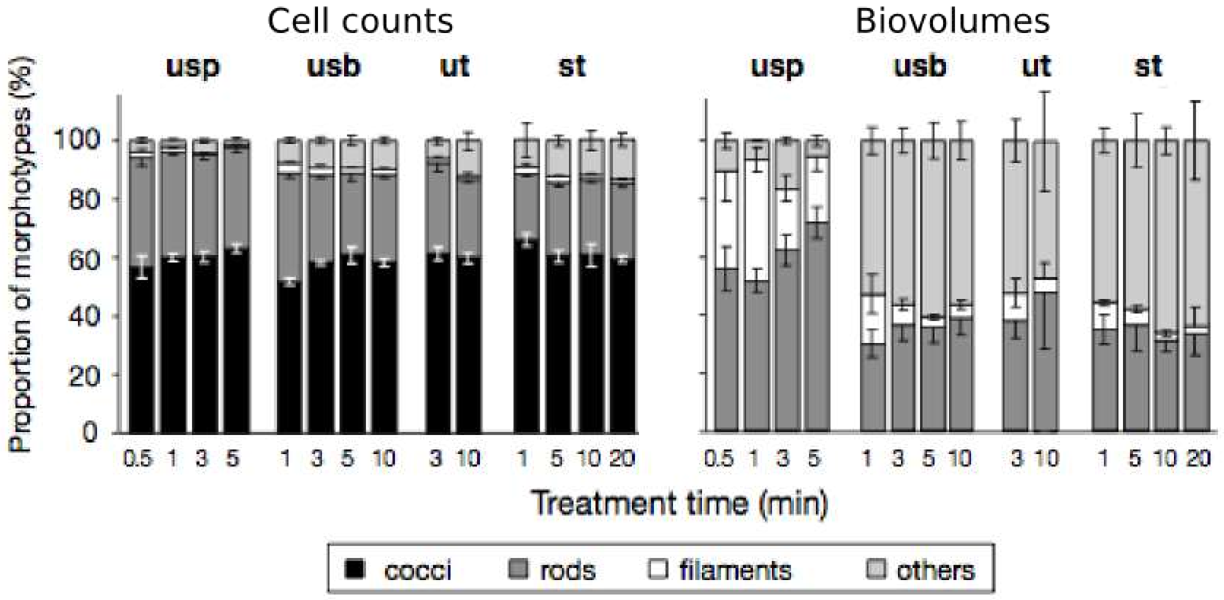


Figure 5.3: Effect of different homogenization method on the repartition of bacterial morphotypes from a sediment sample: ultasonic probe (usp), ultrasonic bath (usb), ultra-turrax (ut), and stomacher (st). No significant differences is reported by the use of one or another method.[121]

4. **DNA extraction bias**
   As mentioned in previous chapter, DNA extraction methodology, and more particularly the lysis protocol used, could induce an important bias favouring particular groups of bacteria. In this case, the extraction kit used was chosen on the bases of its good performance to extract DNA from endospores. The lysis was performed thanks to a bead beating step with a mix of beads of different sizes (provided by the manufacturer of the kit). To minimize the harshness of the treatment, several rounds of short shaking were performed with a TissueLyser LT (Qiagen). For disrupting bacterial sample, the manufacturer normally recommends to shake the lysing tubes with glass beads for 5 minutes at 50Hz. Here we chose to perform 4 sequential rounds at 50Hz during 5 minute each, collecting supernatant at the end of every steps. By doing this, the strength of the treatment increased with the bead beating rounds, leading to the collection of DNA from every kind of bacterial cells, avoiding in particular, to destroy DNA from softer cells which is readily released during the treatment and could be sheared by too long bead beating procedure.

5. **PCR amplification bias**
   This bias is well known by every scientists doing molecular biology. PCR primers design is always the result of a trade-off between their annealing specificity to a target portion of a particular gene and their universality to amplify DNA from a given group of taxa. As a consequence, even for very conserved genes, like the one coding for the small ribosomal subunit (16S rRNA) of bacteria, it is impossible to design universal primers that have the same annealing specificity for every bacteria,

resulting to a bias favouring a particular group of taxa rather than another. This problem has been extensively discussed in the published article on qPCR primer design for the *spo0A* gene, as well as in the chapter describing the choice of primers for the amplicon-based pyrosequencing. In both cases, if one would like to do a better job, the main solution is to perform the same work (qPCR or pyrosequencing) with different pairs of primers, specially designed for different groups of bacteria, and to mix the results together at the end. This approach is of course much more expensive and time consuming, but could result in less biased view of the composition of the *in situ* bacterial communities.[94] In this work, we, however, did not have the time and the financial resources to do it in that way. Furthermore, as the aim of this work was to compare bacterial communities from different area of the Bay of Vidy, an absolute determination of the composition of the communities was not necessary. As every sample experienced the same biases, the comparison between them is, therefore, still valuable.

6. **Sequencing bias**
Sequencing technologies are not bias-less. The weakness of 454-pyrosequencing when compared to Illumina, stands in an higher error rate when sequencing homopolymeric regions of the DNA molecule, and more particularly when a long succession of A's or T's occurs in the sequences. With such a bias, the error rate is higher for genomes with a low G+C% content, like Firmicutes, leading to their rejection during the quality control check at the beginning of the bioinformatic analyses pipeline. An error rate of around 20% more important is reported with these kind of sequences when comparing with Illumina which is less sensible to this particular bias.[122] As a counterpart, 454-pyrosequencing data gives longer sequences which are more easily and firmly identified in the databases. As discussed previously, sequences of too bad quality (therefore containing too much errors) or resulting from PCR chimeras are removed during the first steps of bioinformatic treatments.

7. **OTU picking bias**
By default, Qiime pipeline use "uclust" tool to group together sequences with 97% of identity into OTUs. But what is the reality of the relation between this 97% of identity and the real species composing the community? This question is strongly debated by the scientist since the advent of next generation sequencing technologies which raise the need to rapidly and efficiently attribute sequences to taxa. This threshold of 97% is in fact a consequence of the DNA-DNA hybridization experiments started in the 70s, which define a species as a group of organisms, which have at least 70% of their DNA that bind spontaneously in standard conditions. When transposed to 16S rRNA gene, which is much more conserved than the entire genome, this threshold of 70% becomes 97%. If it is true that empirically no body have ever found members of a species sharing less than 97% of identity with the type strain, the contrary is not always true. Effectively, it has been demonstrated that a species is sometimes nested into another one, as shown by Figure 5.4, sharing more 97% of identity with the neighbour species.

Another example is given by the *Bacillus cereus* group within which an insect

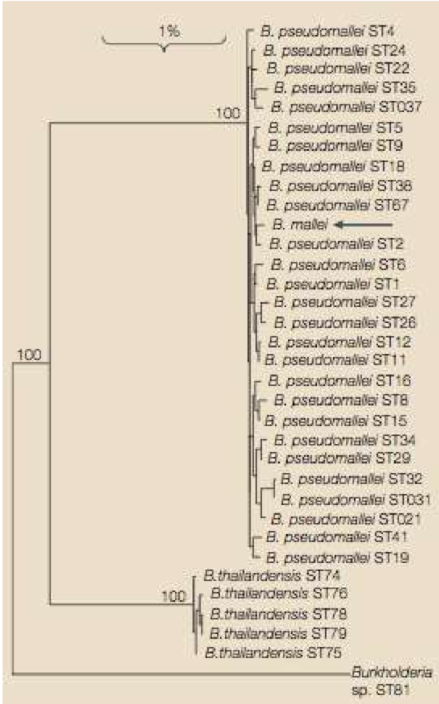


Figure 5.4: Example of a *Burkholderia* strain enclosed inside a cluster constituted by an other strain based on the 97% of identity criterion with the 16S rRNA gene[121]

pathogenic strain *B. thuringiensis* is defined by different lineages into the phylogeny of *B. cereus*, a typical soil bacteria. Similarly, *B. anthracis* is a kind specialized clone of *B. cereus* which has the particularity to cause anthrax.[123]
If this threshold results in a failure at identifying correctly some particular taxa when working with the entire 16S rRNA gene, what happens when only one third of it is used, as in this work? This is difficult to know and nobody has a clear answer to this. However, taxonomic assignment based on the 16S rRNA gene is, for the moment, still the best option we have, mainly because of the size of the available databases on this gene. Being more precise in the assignation will probably become possible in the future by using several conserved genes rather than a unique one as it is the case actually.

8. **Taxonomical assignment bias**
The taxonomical assignment of unknown sequences is not a straightforward task. Some algorithms are clearly influenced by the existing data in the taxonomic table, others less. In the case of the standard Qiime workflow used during this work, the uclust consensus taxonomy classifier has been used. It based its assignment on a Greengenes reference database of 16S rRNA gene, as well as a corresponding taxonomic table defined on the 97% of identity criterion. The main bias with such an approach is that it assumes that the sequence to be assigned is necessarily in the reference table. If not it makes a consensus between the closest relatives for determining its taxonomical position in the tree of life. When working with environmental samples most of the bacterial sequences remain unknown in the reference database. As the homogeneity of the sequences is not necessarily the same from one taxonomical group

to another, it is, therefore, difficult to infer the exact taxonomical position of an unknown sequence with an approach only based on the taxonomical distance. Some algorithms, like naive-bayesian classifier, seems to perform better. In this work, to verify the assignation for the dominants OTU, we decided to retrieve their sequences and to blast them against another 16S rRNA gene database. This "old school" method was of course possible because the number of sequences was not too high. Such an approach could not probably be applied for an Illumina dataset with much more sequences.

9. **16S rRNA gene copy number bias**
The number of copies from the 16S rRNA gene ivaries greatly (1 to 15 copies) from one strain to another, leading to an overestimation of the real abundance of the strains having the highest number of 16S rRNA gene copies. To correct this, some on-line databases, like rrnDB,[124] publishes the number of 16S rRNA gene copies found in an ever increasing number of bacterial strains. However, with the 1322 strains contained in this database it is likely not to find all the bacterial taxa that are present in an environmental sample. An attempt at extrapolating the missing values throughout a bacterial taxonomic table for correcting the absolute abundances values in bioinformatic treatment of the sequences, was, however, not very convincing. Effectively, with such an approach it is difficult to know if we are really correcting a bias rather than adding a new one. The extrapolation approach is effectively not valid because the variations of the number of 16S rRNA gene copies are not uniformly distributed and does not depend on the taxonomic level. High variations are effectively possible between two close related species.[125] Furthermore, the number of copies published in the rrnDB database comes mainly from laboratory strains which are not facing the same kind of constraints than bacteria from the environment. Some evidence effectively demonstrate that environmental strains isolated from stressful environments tends to harbour more copies of the 16S rRNA gene.[126] Finally, in some strains, and particularly in those harbouring the highest number of copies of the 16S rRNA gene (typically the Gamma-proteobacteria and the Firmicutes), all the sequences are not identical, but can harbour dissimilarity levels up to 1%.[125] If this dissimilarity does not induced a mis-identification of most of the strains when identity level reported by the classifier is high, it must, however, be taken into account when identity level drops below 98% when identifying Firmicutes or Gamma-proteobacteria at the species level. For this study, the overestimation of the Firmicutes in the community could be roughly corrected by diminishing their abundances by a factor of 1.6 (mean copy number in Firmicutes: 7.01 / mean copy number in bacteria: 4.3). This overestimation of Firmicutes is partially counterbalanced by the pyrosequencing bias on this taxa. But anyway, as we are comparing communities between them, this correction would not have any effect on the conclusions we drew during this thesis. Figure 5.5 shows the composition of the community corrected with the number of 16S rRNA gene copies extrapolated from the data available in the rrnDB database.

10. **Seasonal changes bias**
All around the year, the lake ecosystem experience cyclic changes in term of energy

input, biomass production, nutrient flow, aeration of deep layers, temperature. In standard temperate monomictic lake sediments, these variations could have a deep impact on the dynamic of the bacterial community, particularly in the upper layer of the sediments, which depends on nutrients and oxygen inflow from the upper water column. With the effluent from the waste water treatment plant, things are a little bit different, as nutrient inflow is almost constantly provided all around the year. With such a big input of organic matter, dissolved oxygen is readily consumed leading to a very fast transition toward anoxia in sediments. In such situation, the dissolved oxygen provided by the upper layer of water should not have a big impact on the functioning of the *in situ* bacterial communities. The remaining factors that could affect this dynamic is the temperature drop, mainly occurring during water turnover in winter, but also seasonal differences in the composition of the releases from the waste water treatment plant. During this project, sediment cores were recovered during three different sampling campaigns that took place in Summer, in Autumn, and in Spring. As shown by the well mixed repartition of the sample points in the different CCA ordination plots presented in this work, this seasonal factor does not seems to have a big influence on the bacerial composition of the sediments that were studied.

11. **Macroscopic scale homogenisation bias**
    When studying a microbial ecosystem there is inevitably a scale effect. The main question here is how to understand mechanisms that occur at the microscale (few μm) when working with grams of initial sample harbouring billions of bacteria? The problem is quite of the same complexity as observing the space from earth with a telescope for trying to find traces of life in the universe. What is the heterogeneity of the environment perceived by a bacteria at the micrometer scale? What are the local variation of the dissolved concentration of a particular heavy metal ion when moving in and out from EPS matrix, or aggregates of sediments? What is the local effect of the decomposition of a nematode on the surrounding bacterial community? It is difficult to answer to such questions. With the tools we dispose today in microbial ecology, its only possible to draw big generalities on a highly diverse population of living organisms. Of course, tools and analytical methods become more and more sensible, but there is still a hard job to do to understand the role of every bacterial cell in a complex environment.

With regards to these biases, we often view publications that hesitate to be too much precise in the identification of bacteria during bioinformatic treatment of next generation sequencing data. Most of them limit their description of the dataset at the order or family level, sometimes genus level. Of course, the different bias exposed above could lead to a miss-assignation of a sequence to a particular species. But in my opinion the potentiality of making an error is better than saying something like there is bacteria in a bacterial community sample! Effectively, with the high metabolic diversity found in bacteria, sometimes within members of a single genus, giving percentage composition of a bacterial community at the order level or higher always seemed to me a little bit information-less. But in this case what is the level of information carried by the species level? In my understanding of the species definition in modern microbiology I'm tempted to say that members of a single species have at least 7 chances on 10 to harbour the same genes, and to be able to perform the same metabolic or physiological functions. It is better than nothing! At the genus level this probability fall to half-half! Of course, this is a conceptual view, there is big variations in the level of conservation of the different genes and there is also non-coding portion of the DNA molecule that is more prone to variations. But anyway, in my opinion, with the tools we have to our disposition today, one must attempt to be as precise as possible in the identification of the DNA sequences as the reference databases permit to do. It makes also no sense to say that a particular species is present in the sample if it is identified by the sequence classifier with an identity level below 97%. In this case, the identification at the level of the genus or even higher is better.

In the same order of idea's, there is also the question of the correspondence between the real abundance of a particular taxa in the community and the measured relative abundance of the corresponding OTU detected by the sequencing tools. As seen in the little review of the different bias exposed above, different factors like PCR primers specificity or taxonomic affiliation can have an incidence on the number of the 16S rRNA gene copies that are gathered for a single taxon. Most of the next-generation sequencing analyses tools, like Qiime, provide the possibility to calculate distances between samples based on the presence and absence of OTUs rather than their relative abundances. The first option, the unweighted distance matrix calculation, involved a threshold for determining the condition of the presence of an OTU in the sample. Here, the main question is to find the "right" threshold that best discriminates between samples... I found this Idea a little bit strange and I prefer to use weighted distance matrix calculation based on biased relative abundances. Even if the relative abundance for a particular OTU is two times overestimated, it seems effectively more correct to me to say that it represents more than half of the community (for instance) rather than saying that it is present at the same level than another OTU which have an abundance just a little bit above the threshold defined for unweighted distance calculation. This last topic is effectively subject to intense debate in the scientific community, and there is probably not a single right way to do things. In my opinion, as its often the case in science, the right way to perform distance calculation depends on the initial hypothesis driving the analyses. Here, we tried to demonstrate the dominance of a particular group of organisms related to an environmental factor. An information based on its presence or absence in the community is, therefore, not sufficiently discriminating. A weighted approach was, therefore, much better.

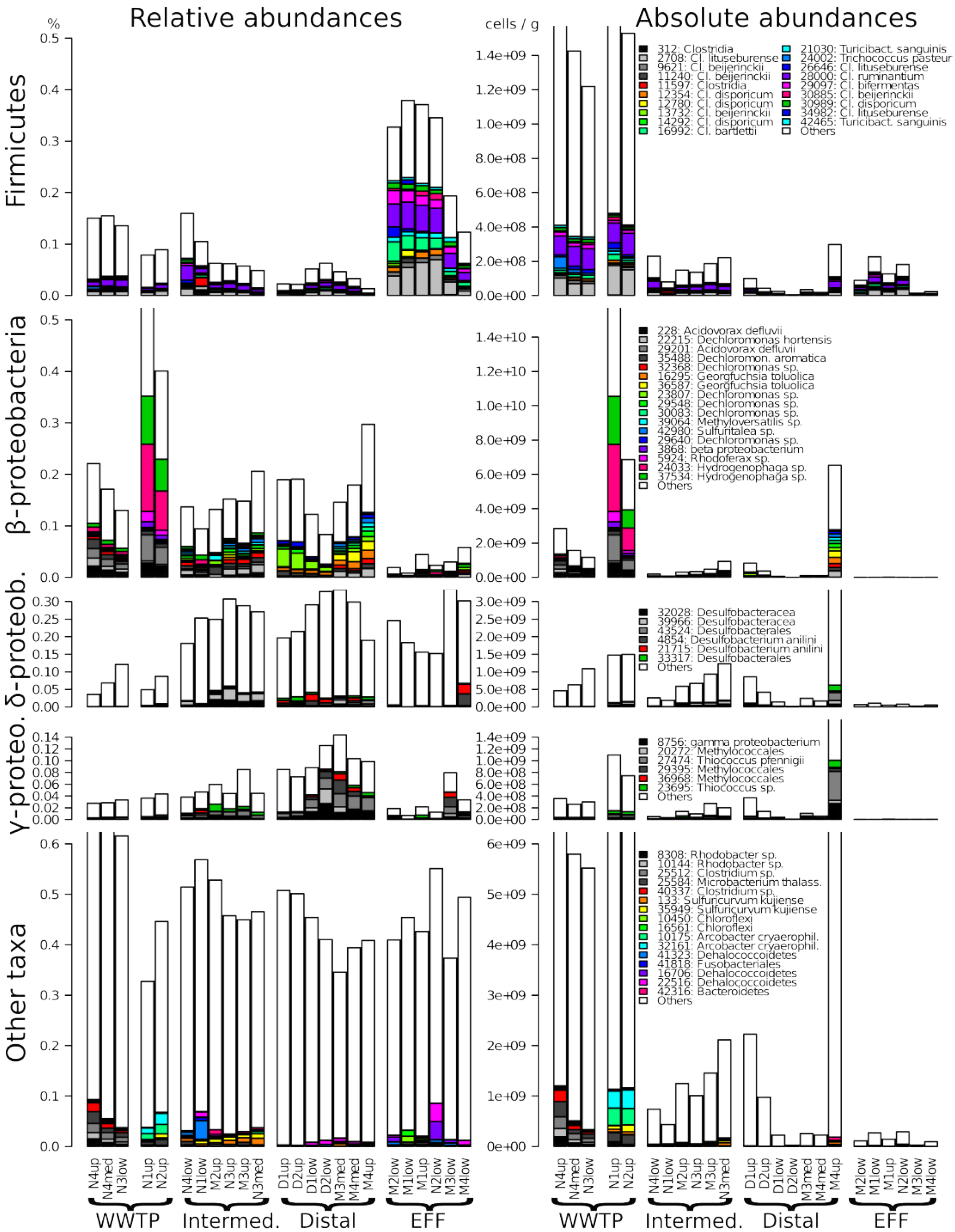


Figure 5.5: Microbial community composition corrected with the extrapolate number of copies of the 16S rRNA gene

5.4 Outlook

What happen to Clostridia spores when buried in sediments? Do they germinate and grow slowly? What is the effect of the exopolymeric substances excreted by the four Clostridia strains on the stability of heavy metals in sediments? What is the residence time of heavy metals in sediments?

At the end of this thesis a lot of questions remain open and necessitate more investigation. Depending on the results obtained by such experiments, the four Clostridia strains highlighted during this thesis could perhaps be interesting in the context of the bioremediation of anaerobic sediments polluted by heavy metals.

As seen in many places in this document, urban runoff water is the main transport vector for heavy metals down to the aquatic environment. As the trend of municipalities today, is to try to separate as much as possible waste water from runoff water, the problem of the contamination of the aquatic environment by heavy metals will become an ever bigger environmental issue in the coming years. Finding an environmental friendly way to remove these pollutants from runoff before they reach the aquatic environment is, therefore, necessary. Here again, a better comprehension of the fate in the environment of the four Clostridia strains identified during this thesis could perhaps lead to biotechnological applications.

To achieve such bioremediation objectives, the main points that need to be explored are the following:

1. How and at which rate heavy metals entrapped in sediments are released into the environment?
2. Does it exist a mean to accelerate this release process?
3. What about the exopolymeric substances excreted by Clostridia strains? Are they an efficient way to trap heavy metals?
4. What are their sorptive capacities and efficiencies when compared to exopolymeric substances from other species?
5. What is their stability in the environment?
6. What is the best mean to recovered exopolymeric substances matrix and heavy metals after the treatment?
7. Do bacterial strains exist with the same properties as Clostridia, but more adapted to a temperate aquatic environment?
8. Do the flocs from activated sludge of a waste water treatment plant contain such an organism?
9. If yes, how could their growth be stimulated when deposited in an oligotrophic environment, such as runoff water?

6

Acknowledgements

It is a pleasure for me to express my gratitude to the many people who have helped me or have been an important part of my life the last three years.

First of all, I am grateful to my supervisor in Neuchâtel, Pilar Junier, for supporting me during this three years of hard work. Pilar has always found the right way to force me to surpass myself. And God knows that I do not like when someone forces me to do things! Sometimes it has probably been very difficult to orient me in the right way. But now I know that it was necessary! Thanks Pilar for not having given up during these hard times! Thanks also for all the hours you spent for correcting my awful English. And thanks also for having always encouraged me.

I also would like to thank my colleagues of the laboratory of microbiology at the University of Neuchâtel. First of all, I'd like to thank Nicole, the "Mum" of the lab, for the always good technical advice she provides for doing the good choices in the lab. It is also mainly thanks to her and to her incredible vigilance for calling back to order careless people, that the work conditions in the lab are so nice. Thanks Nicole! I also would like to say a little word on other colleagues with whom I spent some good time in Neuchâtel and elsewhere. Thanks to Daniel and Sevi for the unforgettable times we spent during the TP, but also during the field trip in Graubunden and the sampling campaign in Greece (for Sevi). Thanks to Thomas for having directed me during my beginnings in bioinformatics, but also for the funny discussions we had sometimes between "two older men". Thanks to Ludo for his always good humor. Thanks to Tina for letting me tame her a little bit. Thanks also to Anaële for her help with electron microscopy. Among all, I would also like to thanks Loïc, for the beautiful and high quality work we have done together at the beginning of my thesis.

I would also like to thank the colleagues of the University of Geneva. My gratitude goes to Marylou, for the time we spent during the sampling in MIR, but also on-board of "La Licorne". It was also very nice to have the opportunity to do some chemical measurements in your lab. Thanks also to Jean-Luc for having giving us the opportunity to complete our sediment core collection by using the boat of the Forel-institute. Thanks also to Roxanne for the help with the chemical measurements at the university of Neuchâtel. Thanks to Jasqueline and Anna from the University of Lausanne for the support with the ICP-OES. It was very nice to come to your lab.

Finally, I'd like to thank Anita and Zacharie for having given me the necessary time to do this thesis. Your support was effectively very helpful during these three years of hard work. First for having helped me to focus my mind towards the essential things of life, second for having done my part of housework when I was doing the "copyist monk" behind my computer... I hope the time I spent in this work did not require too many sacrifices! I've done my best to find the right equilibrium!

Matthieu Bueche

Neuchâtel, Switzerland, March 2014

Bibliography

[1] L. Margulis and D. Sagan. *What is Life?* University of California Press, 1995.

[2] Claude Faurie, Christiane Ferra, Paul Médori, Jean Dévaux, and Jean-Louis Hemptinne. *Ecologie: Approche scientifique et pratique 5ème édition*, volume 1. Lavoisier, Londres, Paris, New-York, TEC & DOC edition, 2006.

[3] Laura Sigg, Werner Stumm, and Philippe Behra. *Chimie des milieux aquatiques*, volume 1. Masson, Paris, Milan, Barcelone, Bonn, 1992.

[4] Moritz Bollmann, Thomas Bosch, Franciscus Colijn, Ralf Ebinghaus, Rainer Froese, Kerstin Güssow, Setareh Khalilian, Sebastian Krastel, Arne Körtzinger, Martina Langenbuch, Mojib Latif, and Birte Matthiessen. world ocean review 2010 - living with the oceans, 2010.

[5] Cristina Del Bianco and Sheref S. Mansy. Nonreplicating protocells. *Accounts of Chemical Research*, 45(12):2125–2130, December 2012. WOS:000312825200012.

[6] Nicholas Cox, Dimitrios A. Pantazis, Frank Neese, and Wolfgang Lubitz. Biological water oxidation. *Accounts of Chemical Research*, 46(7):1588–1596, July 2013. WOS:000322086600022.

[7] J. M. Wilkinson. Silage and animal health. *Natural Toxins*, 7(6):221–232, 1999.

[8] C. M. Cook, E. Vardaka, and T. Lanaras. Toxic cyanobacteria in greek freshwaters, 1987-2000: Occurrence, toxicity, and impacts in the mediterranean region. *Acta Hydrochimica Et Hydrobiologica*, 32(2):107–124, June 2004. WOS:000222717400001.

[9] N. S. Shifrin. Pollution management in the twentieth century. *Journal of Environmental Engineering-Asce*, 131(5):676–691, May 2005. WOS:000228743100002.

[10] Neil Campbell, Brian D'Arcy, Alan Frost, Vladimir Novotny, and Anne Sansom. *Diffuse pollution: An introduction to the problems and solutions*, volume 1. IWA Publishing, London, IWA publishing edition, 2004.

[11] LEaux - loi fédérale sur la protection des eaux, 1991.

[12] J. Marsalek, M. Diamond, S. Kok, and W.E. Watt. Urban runoff, 2008.

[13] Allen P Davis, Mohammad Shokouhian, and Shubei Ni. Loading estimates of lead, copper, cadmium, and zinc in urban runoff from specific sources. *Chemosphere*, 44(5):997–1009, August 2001.

[14] Jean-Michel Gobat, Michel Aragno, and Willy Matthey. *Le sol vivant*, volume 1. Presses polytechniques et universitaires romandes, Lausanne, 3 edition, 2010.

[15] Martin Komarek and Josef Zeman. Dynamics of cu, zn, cd, and hg release from sediments at surface conditions. *Bulletin of Geosciences*, 79(2):99–106, 2004.

[16] S. C. Deb and T. Fukushima. Metals in aquatic ecosystems: mechanisms of uptake, accumulation and release - ecotoxicological perspectives. *International Journal of Environmental Studies*, 56(3):385–417, April 1999.

[17] Jian-feng Peng, Yong-hui Song, Peng Yuan, Xiao-yu Cui, and Guang-lei Qiu. The remediation of heavy metals contaminated sediment. *Journal of Hazardous Materials*, 161(2–3):633–640, January 2009.

[18] Walter Wildi, Janusz Dominik, Jean-Luc Loizeau, Richard Linn Thomas, Pierre-Yves Favarger, Laurence Haller, Antoine Perroud, and Christophe Peytremann. River, reservoir and lake sediment contamination by heavy metals downstream from urban areas of switzerland. *Lakes and Reservoirs: Research and Management*, 9:75–87, 2004.

[19] Sediment management plan rhine : Summary, 2009.

[20] Bigdeli Mohsen and Seilsepour Mohsen. Investigation of metals accumulation in some vegetables irrigated with waste water in shahre rey-iran and toxicological implications. *American-Eurasian Journal of Agricultural & Environmental Sciences*, 4(1):86–92, 2008.

[21] Monu Arora, Bala Kiran, Shweta Rani, Anchal Rani, Barinder Kaur, and Neeraj Mittal. Heavy metal accumulation in vegetables irrigated with water from different sources. *Food Chemistry*, 111(4):811–815, December 2008.

[22] Radu Lacatusu and Anca-Rovena Lacatusu. Vegetable and fruits quality within heavy metals polluted areas in *Romania. Carpathian Journal of Earth and Environmental Sciences*, 3(2):115–129, 2008.

[23] Wikipedia. Pollution, January 2014.

[24] Ademola Olaniran, Adhika Balgobind, and Balakrishna Pillay. Bioavailability of heavy metals in soil: Impact on microbial biodegradation of organic compounds and possible improvement strategies. *International Journal of Molecular Sciences*, 14(5):10197–10228, 2013.

[25] Lars Jarup. Hazards of heavy metal contamination. *British Medical Bulletin*, 68(1):167–182, 2003.

[26] Diego Rial, JoséAntonio Vázquez, and MiguelAnxo Murado. Effects of three heavy metals on the bacteria growth kinetics: a bivariate model for toxicological assessment. *Applied Microbiology and Biotechnology*, 90(3):1095–1109, 2011.

[27] Christopher L Hemme, Ye Deng, Terry J Gentry, Matthew W Fields, Liyou Wu, Soumitra Barua, Kerrie Barry, Susannah G Tringe, David B Watson, Zhili He, Terry C Hazen, James M Tiedje, Edward M Rubin, and Jizhong Zhou. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *The ISME Journal*, 4(5):660–672, 2010.

[28] Valeria Ochoa-Herrera, Glendy Leon, Qais Banihani, Jim A. Field, and Reyes Sierra-Alvarez. Toxicity of copper(II) ions to microorganisms in biological wastewater treatment systems. *Science of the Total Environment*, 412:380–385, December 2011. WOS:000298534300044.

[29] Sabine Brantl. Bacterial gene regulation: metal ion sensing by proteins or RNA. *Trends in Biotechnology*, 24(9):383–386, September 2006.

[30] Makino Tomoyuki. Heavy metal pollution of soil and a new approach to its remediation: Research experiences in japan, 2009.

[31] Gang Wu, Hubiao Kang, Xiaoyang Zhang, Hongbo Shao, Liye Chu, and Chengjiang Ruan. A critical review on the bio-removal of hazardous heavy metals from contaminated soils: Issues, progress, eco-environmental concerns and opportunities. *Journal of Hazardous Materials*, 174(1-3):1–8, February 2010. WOS:000273984000001.

[32] Wei-yu Shi, Hong-bo Shao, Hua Li, Ming-an Shao, and Sheng Du. Progress in the remediation of hazardous heavy metal-polluted soils by natural zeolite. *Journal of Hazardous Materials*, 170(1):1–6, October 2009.

[33] G. Porebska and A. Ostrowska. Heavy metal accumulation in wild plants: Implications for phytoremediation. *Polish Journal of Environmental Studies*, 8(6):433–442, 1999.

[34] Fenglian Fu and Qi Wang. Removal of heavy metal ions from wastewaters: A review. *Journal of Environmental Management*, 92(3):407–418, March 2011.

[35] Cristina Quintelas, Zélia Rocha, Bruna Silva, Bruna Fonseca, Hugo Figueiredo, and Teresa Tavares. Biosorptive performance of an *Escherichia coli* biofilm supported on zeolite NaY for the removal of cr(VI), cd(II), fe(III) and ni(II). *Chemical Engineering Journal*, 152(1):110–115, October 2009.

[36] R.S Prakasham, J.Sheno Merrie, R Sheela, N Saswathi, and S.V Ramakrishna. Biosorption of chromium VI by free and immobilized rhizopus arrhizus. *Environmental Pollution*, 104(3):421–427, March 1999.

[37] Michael T. Madigan, John M. Martinko, David Stahl, and David P. Clark. *Brock biology of microorganisms*. Benjamin Cummings, San Francisco, 2012.

[38] Beate Köpke, Reinhard Wilms, Bert Engelen, Heribert Cypionka, and Henrik Sass. Microbial diversity in coastal subsurface sediments: a cultivation approach using various electron acceptors and substrate gradients. *Applied and Environmental Microbiology*, 71(12):7819–7830, 2005.

[39] Reinhard Wilms, Henrik Sass, Beate Köpke, Jürgen Köster, Heribert Cypionka, and Bert Engelen. Specific bacterial, archaeal, and eukaryotic communities in tidal-flat sediments along a vertical profile of several meters. *Applied and Environmental Microbiology*, 72(4):2756–2764, 2006.

[40] IsabelaC. Torres, KanikaS. Inglett, and K.R. Reddy. Heterotrophic microbial activity in lake sediments: effects of organic electron donors. *Biogeochemistry*, 104(1-3):165–181, July 2011.

[41] Y. Koizumi, H. Kojima, and M. Fukui. Characterization of depth-related microbial community structure in lake sediment by denaturing gradient gel electrophoresis of amplified 16S rDNA and reversely transcribed 16S rRNA fragments. *Fems Microbiology Ecology*, 46(2):147–157, November 2003. WOS:000186251200003.

[42] Laurence Haller, Mauro Tonolla, Jakob Zopfi, Raffaele Peduzzi, Walter Wildi, and John Pote. Composition of bacterial and archaeal communities in freshwater sediments with different contamination levels (lake geneva, switzerland). *Water Research*, 45(3):1213–1228, January 2011. WOS:000287054500024.

[43] S. Spring, R. Schulze, J. Overmann, and K. H. Schleifer. Identification and characterization of ecologically significant prokaryotes in the sediment of freshwater lakes: molecular and cultivation studies. *Fems Microbiology Reviews*, 24(5):573–590, December 2000. WOS:000165362500005.

[44] M. Wagner and A. Loy. Bacterial community composition and function in sewage treatment systems. *Current Opinion in Biotechnology*, 13(3):218–227, June 2002. WOS:000175903900005.

[45] A. T. Mielczarek, A. M. Saunders, P. Larsen, M. Albertsen, M. Stevenson, J. L. Nielsen, and P. H. Nielsen. The microbial database for danish wastewater treatment plants with nutrient removal (MiDas-DK) - a tool for understanding activated sludge population dynamics and community stability. *Water Science and Technology*, 67(11):2519–2526, 2013. WOS:000320248500018.

[46] Michiel J. L. de Hoon, Patrick Eichenberger, and Dennis Vitkup. Hierarchical evolution of the bacterial sporulation network. *Current Biology*, 20(17):R735–R745, September 2010. WOS:000281941100019.

[47] P. Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. H. Schleifer, and W. B. Whitman. *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes.* Springer, 2011.

[48] WL Nicholson. Roles of *Bacillus* endospores in the environment. *Cellular and Molecular Life Sciences*, 59(3):410–416, March 2002. WOS:000174840500004.

[49] Michael S. Roberts and Frederick M. Cohan. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution*, 49(6):1081–1094, 1995. ArticleType: research-article / Full publication date: Dec., 1995 / Copyright © 1995 Society for the Study of Evolution.

[50] Jennifer B. Hughes Martiny, Brendan J.M. Bohannan, James H. Brown, Robert K. Colwell, Jed A. Fuhrman, Jessica L. Green, M. Claire Horner-Devine, Matthew Kane, Jennifer Adams Krumins, Cheryl R. Kuske, Peter J. Morin, Shahid Naeem, Lise Ovreas, Anna-Louise Reysenbach, Val H. Smith, and James T. Staley. Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2):102–112, 2006.

[51] Ines Mandic-Mulec and James I. Prosser. Diversity of endospore-forming bacteria in soil: Characterization and driving mechanisms. In *Endospore-forming Soil Bacteria*, number 27 in Soil Biology, pages 31–59. N.A. Logan and P. De Vos, Berlin Heidelberg, springer-verlag edition, 2011.

[52] JH Hageman, GW Shankweiler, PR Wall, K. Franich, GW MCCowan, SM Cauble, J. Grajeda, and C. Quinones. Single, chemically defined sporulation medium for *Bacillus subtilis* growth, sporulation, and extracellular protease production. *Journal of Bacteriology*, 160(1):438–441, 1984. WOS:A1984TM02800068.

[53] J. Errington. Regulation of endospore formation in *Bacillus subtilis*. *Nature Reviews Microbiology*, 1(2):117–126, November 2003.

[54] Wayne Nicholson, Patricia Fajardo-Cavazos, Roberto Rebeil, Tony Slieman, Paul Riesenman, Jocelyn Law, and Yaming Xue. Bacterial endospores and their significance in stress resistance. *Antonie van Leeuwenhoek*, 81(1):27–32, 2002.

[55] Pun To Yung, Michael J. Kempf, and Adrian Ponce. A rapid single spore enumeration assay. In *2006 IEEE Aerospace Conference, Vols 1-9*, pages 751–763. 2006 IEEE Aerospace Conference, Vols 1-9, 2006 IEEE aerospace conference, vols 1-9 edition, 2006. WOS:000241583101016.

[56] Ray Bibek. *Fundamental food microbiology.* CRC Press LLC, Boca Raton, Londo, New York, Whashington D. C., 3 edition, 2004.

[57] E. De Clerck, T. Vanhoutte, T. Hebb, J. Geerinck, J. Devos, and P. De Vos. Isolation, characterization, and identification of bacterial contaminants in semifinal gelatin extracts. *Applied and Environmental Microbiology*, 70(6):3664–3672, June 2004. WOS:000221981100061.

[58] Marc Heyndrickx. The importance of endospore-forming bacteria originating from soil for contamination of industrial food processing. *Applied and Environmental Soil Science*, 2011:11, 2011.

[59] R. B. Seale, P. J. Bremer, S. H. Flint, and A. J. McQuillan. Characterization of spore surfaces from a *Geobacillus* sp. isolate by pH dependence of surface charge and infrared spectra. *Journal of Applied Microbiology*, 109(4):1339–1348, October 2010. WOS:000281895200024.

[60] K. L. Brown. Control of bacterial spores. *British Medical Bulletin*, 56(1):158–171, 2000. WOS:000087549800014.

[61] Brian B. McSpadden Gardener. Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. *Phytopathology*, 94(11):1252–8, November 2004. MEDLINE:18944463.

[62] N. A. Logan. *Bacillus* and relatives in foodborne illness. *Journal of Applied Microbiology*, 112(3):417–429, March 2012. WOS:000300044300001.

[63] D. H. Nies. Microbial heavy-metal resistance. *Applied Microbiology and Biotechnology*, 51(6):730–750, June 1999. WOS:000081207300003.

[64] Massimiliano Marvasi, Pieter T. Visscher, and Lilliam Casillas Martinez. Exopolymeric substances (EPS) from *Bacillus subtilis*: polymers and genes encoding their synthesis. *FEMS Microbiology Letters*, 313(1):1–9, 2010.

[65] Franziska L. Lederer, Ulrike Weinert, Tobias J. Günther, Johannes Raff, Stephan Weiß, and Katrin Pollmann. Identification of multiple putative s-layer genes partly expressed by *Lysinibacillus sphaericus* JG-B53. *Microbiology*, 159(Pt 6):1097–1108, 2013.

[66] Mohamed L. Merroun, Johannes Raff, André Rossberg, Christoph Hennig, Tobias Reich, and Sonja Selenska-Pobell. Complexation of uranium by cells and s-layer sheets of *Bacillus sphaericus* JG-A12. *Applied and Environmental Microbiology*, 71(9):5532–5543, 2005.

[67] O. Gonet. *Etude gravimétrique du Lac Léman à bord du mésoscaphe "Auguste Piccard"*. Kümmerly u. Frey, Geograph. Verlag, 1969.

[68] Stéphanie Girardclos. *Sismostratigraphie et structure sédimentaire en 3D d'un bassin lacustre, du retrait glaciaire à nos jours (Lac Léman, Suisse)*. PhD thesis, Université de Genève, Faculté des sciences, Institut F.-A. Forel, 2001.

[69] P. Blanc. Evaluation de l'étendue de la zone désoxygènée dans la plaine centrale du grand lac, en automne 1997. *Rapport du CIPEL*, 1998.

[70] C. Giroud. *Evaluation du risque pour l'environnement des métaux lourds issus du milieu urbain*. PhD thesis, EPFL - Section Sciences et Ingénierie de l'Environnement, 2007.

[71] M. Pardos, C. Benninghoff, L. F. de Alencastro, and W. Wildi. The impact of a sewage treatment plant's effluent on sediment quality in a small bay in lake geneva (switzerland-france). part 1: Spatial distribution of contaminants and the potential for biological impacts. *Lakes & Reservoirs: Research and Management*, 9:41–52, 2004.

[72] J.-L. Loizeau, M. Pardos, F. Monna, C. Peytremann, L. Haller, and J. Dominik. The impact of a sewage treatment plant's effluent on sediment quality in a small bay in lake geneva (switzerland-france). part 2: Temporal evolution of heavy metals. *Lakes & Reservoirs: Research and Management*, 9:53–63, 2004.

[73] J. Pote, L. Haller, J. L. Loizeau, A. G. Bravo, V. Sastre, and W. Wildi. Effects of a sewage treatment plant outlet pipe extension on the distribution of contaminants in the sediments of the bay of vidy, lake geneva, switzerland. *Bioresource Technology*, 99(15):7122–7131, October 2008.

[74] Jonas Margot, Anoÿs Magnet, Denis Thonney, Nathalie Chèvre, Felippe De Alencastro, and Luca Rossi. Traitement des micropolluants dans les eaux usée: Rapport final sur les essais pilotes à la STEP de vidy (lausanne), 2011.

[75] JL Loizeau, S Roze, C Peytremann, F Monna, and J Dominik. Mapping sediment accumulation rate by using volume magnetic susceptibility core correlation in a contaminated bay (lake geneva, switzerland). *Eclogae Geologicae Helvetiae*, 96:S73–S79, 2003. WOS:000185970000009.

[76] J.-F. Jaton. Bilan 2011 de l'épuration vaudoise, 2011.

[77] Andre Preisler, Dirk de Beer, Anna Lichtschlag, Gaute Lavik, Antje Boetius, and Bo Barker Jorgensen. Biological and chemical sulfide oxidation in a *Beggiatoa* inhabited marine sediment. *ISME J*, 1(4):341–353, June 2007.

[78] Camino García-Balboa, Dina Cautivo, M.L. Blázquez, F. González, J.A. Muñoz, and A. Ballester. Successive ferric and sulphate reduction using dissimilatory bacterial cultures. *Water, Air, and Soil Pollution*, 207(1-4):213–226, March 2010.

[79] Dominica Nichols. Cultivation gives context to the microbial ecologist. *Fems Microbiology Ecology*, 60(3):351–357, June 2007. WOS:000246708800001.

[80] Patrick Robe, Renaud Nalin, Carmela Capellano, Timothy M. Vogel, and Pascal Simonet. Extraction of DNA from soil. *European Journal of Soil Biology*, 39(4):183–190, October 2003.

[81] Matthieu Bueche, Tina Wunderlin, Ludovic Roussel-Delif, Thomas Junier, Loic Sauvain, Nicole Jeanneret, and Pilar Junier. Quantification of endospore-forming firmicutes by qPCR with the functional gene *spo0A*. *Applied and Environmental Microbiology*, 2013.

[82] Loïc Sauvain, Matthieu Bueche, Thomas Junier, Matthieu Masson, Tina Wunderlin, Roxane Kohler-Milleret, Elena Gascon Diez, Jean-Luc Loizeau, Mary-Lou Tercier-Waeber, and Pilar Junier. Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria. *Aquatic Sciences*, pages 1–14, October 2013.

[83] Christian L. Lauber, Nicholas Zhou, Jeffrey I. Gordon, Rob Knight, and Noah Fierer. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters*, 307(1):80–86, 2010.

[84] Benjamin E. R. Rubin, Sean M. Gibbons, Suzanne Kennedy, Jarrad Hampton-Marcell, Sarah Owens, and Jack A. Gilbert. Investigating the impact of storage conditions on microbial community composition in soil samples. *PLoS ONE*, 8(7):e70460, July 2013.

[85] S. Günther, T. Hübschmann, M. Rudolf, M. Eschenhagen, I. Röske, H. Harms, and S. Müller. Fixation procedures for flow cytometric analysis of environmental bacteria. *Journal of Microbiological Methods*, 75(1):127–134, September 2008.

[86] Åsa Frostegård, Sophie Courtois, Vincent Ramisse, Sylvie Clerc, Dominique Bernillon, Francoise Le Gall, Pascale Jeannin, Xavier Nesme, and Pascal Simonet. Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology*, 65(12):5409–5420, 1999.

[87] Tom O. Delmont, Patrick Robe, Ian Clark, Pascal Simonet, and Timothy M. Vogel. Metagenomic comparison of direct and indirect soil DNA extraction approaches. *Journal of Microbiological Methods*, 86(3):397–400, September 2011.

[88] Tom O. Delmont, Patrick Robe, Sébastien Cecillon, Ian M. Clark, Florentin Constancias, Pascal Simonet, Penny R. Hirsch, and Timothy M. Vogel. Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology*, 77(4):1315–1324, 2011.

[89] J. Kallmeyer, D. C. Smith, A. J. Spivack, and S. D'Hondt. New cell extraction procedure applied to deep subsurface sediments. *Limnology and Oceanography-methods*, 6:236–245, June 2008.

[90] Nadia Khammar, Luc Malhautier, Valérie Degrange, Robert Lensi, and Jean-Louis Fanlo. Evaluation of dispersion methods for enumeration of microorganisms from peat and activated carbon biofilters treating volatile organic compounds. *Chemosphere*, 54(3):243–254, January 2004.

[91] S Courtois, A Frostegard, P Goransson, G Depret, P Jeannin, and P Simonet. Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environmental Microbiology*, 3(7):431–439, July 2001. WOS:000171022900002.

[92] S. M. Dineen, R. Aranda, D. L. Anders, and J. M. Robertson. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *Journal of Applied Microbiology*, 109(6):1886–1896, December 2010.

[93] Soumitesh Chakravorty, Danica Helb, Michele Burday, Nancy Connell, and David Alland. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, 69(2):330–339, May 2007.

[94] Hui Li, Ying Zhang, Dong-sheng Li, Hui Xu, Guan-xiong Chen, and Cheng-gang Zhang. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. *Soil Biology and Biochemistry*, 41(5):954–968, 2009.

[95] Jian Ye, George Coulouris, Irena Zaretskaya, Ioana Cutcutache, Steve Rozen, and Thomas Madden. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1):134, 2012.

[96] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990.

[97] Patrick D. Schloss, Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, Brian B. Oakley, Donovan H. Parks, Courtney J. Robinson, Jason W. Sahl, Blaz Stres, Gerhard G. Thallinger, David J. Van Horn, and Carolyn F. Weber. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23):7537–7541, 2009.

[98] J. Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, Antonio Gonzalez Pena, Julia K. Goodrich, Jeffrey I. Gordon, Gavin A. Huttley, Scott T. Kelley, Dan Knights, Jeremy E. Koenig, Ruth E. Ley, Catherine A. Lozupone, Daniel McDonald, Brian D. Muegge, Meg Pirrung, Jens Reeder, Joel R. Sevinsky, Peter J. Tumbaugh, William A. Walters, Jeremy Widmann, Tanya Yatsunenko, Jesse Zaneveld, and Rob Knight. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5):335–336, May 2010. WOS:000277175100003.

[99] Patrick D. Schloss and Jo Handelsman. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*, 71(3):1501 –1506, March 2005.

[100] S. C. Haslam, J. M. Ketley, T. J. Mitchell, J. Stephen, D. W. Burdon, and D. C. A. Candy. Growth of *Clostridium difficile* and production of toxins a and b in complex and defined media. *Journal of Medical Microbiology*, 21(4):293–297, June 1986. WOS:A1986C748400003.

[101] T. Karasawa, S. Ikoma, K. Yamakawa, and S. Nakamura. A defined growth medium for *Clostridium difficile*. *Microbiology-Uk*, 141:371–375, February 1995. WOS:A1995QH15900018.

[102] Fabiola Altimira, Carolina Yanez, Guillermo Bravo, Myriam Gonzalez, Luis Rojas, and Michael Seeger. Characterization of copper-resistant bacteria and bacterial communities from copper-polluted agricultural soils of central chile. *BMC Microbiology*, 12(1):193, 2012.

[103] Christophe Espírito Santo, Paula Vasconcelos Morais, and Gregor Grass. Isolation and characterization of bacteria resistant to metallic copper surfaces. *Applied and Environmental Microbiology*, 76(5):1341–1348, March 2010.

[104] L. J. Wheeldon, T. Worthington, P. A. Lambert, A. C. Hilton, C. J. Lowden, and T. S. J. Elliott. Antimicrobial efficacy of copper surfaces against spores and vegetative cells of *Clostridium difficile*: the germination theory. *Journal of Antimicrobial Chemotherapy*, 62(3):522–525, September 2008.

[105] Marc Solioz, HelgeK. Abicht, Mélanie Mermod, and Stefano Mancini. Response of gram-positive bacteria to copper stress. *JBIC Journal of Biological Inorganic Chemistry*, 15(1):3–14, January 2010.

[106] Paul De Vos, George M. Garrity, Dorothy Jones, Noel R. Krieg, Wolfgang Ludwig, Fred A. Rainey, Karl-Heinz Schleifer, and William B. Whitman. *Bergey's manual of systematic bacteriology*, volume Volume 3: The Firmicutes. Springer, Dordrecht, Heidelberg, London, New-York, second edition edition, 2009.

[107] Matthieu Masson and Mary-Lou Tercier-Waeber. Trace metal speciation at the sediment–water interface of vidy bay: influence of contrasting sediment characteristics. *Aquatic Sciences*, pages 1–12, November 2013.

[108] M. V. Smith and M. D. Pierson. Effect of reducing agents on oxidation-reduction potential and the outgrowth of *Clostridium botulinum* type-e spores. *Applied and Environmental Microbiology*, 37(5):978–984, 1979. WOS:A1979GV99600033.

[109] Richard J. Reeder, Martin A. A. Schoonen, and Antonio Lanzirotti. Metal speciation and its role in bioaccessibility and bioavailability. *Reviews in Mineralogy and Geochemistry*, 64(1):59–113, January 2006.

[110] Daniel Bravo, Olivier Braissant, Anna Solokhina, Martin Clerc, Alma U. Daniels, Eric Verrecchia, and Pilar Junier. Use of an isothermal microcalorimetry assay to characterize microbial oxalotrophic activity. *FEMS Microbiology Ecology*, 78(2):266–274, 2011.

[111] Xiandeng Hou and Bradley T. Jones. Inductively coupled plasma-optical emission spectrometry. In *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Ltd, 2006.

[112] P. C. Shih and C. T. Huang. Effects of quorum-sensing deficiency on *Pseudomonas aeruginosa* biofilm formation and antibiotic resistance. *Journal of Antimicrobial Chemotherapy*, 49(2):309–314, February 2002. WOS:000173960000012.

[113] Ming-Yuan Chen, Duu-Jong Lee, Joo-Hwa Tay, and Kuan-Yeow Show. Staining of extracellular polymeric substances and cells in bioaggregates. *Applied Microbiology and Biotechnology*, 75(2):467–474, May 2007. WOS:000246842500025.

[114] RT Noble and JA Fuhrman. Use of SYBR green i for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology*, 14(2):113–118, February 1998. WOS:000072138900001.

[115] Elizabeth R. Fischer, Bryan T. Hansen, Vinod Nair, Forrest H. Hoyt, and David W. Dorward. Scanning electron microscopy. In *Current Protocols in Microbiology*. John Wiley & Sons, Inc., 2005.

[116] Tina Wunderlin, Thomas Junier, Ludovic Roussel-Delif, Nicole Jeanneret, and Pilar Junier. Stage 0 sporulation gene a as a molecular marker to study diversity of endospore-forming firmicutes. *Environmental Microbiology Reports*, 5(6):911–924, 2013.

[117] Lin Ye and Tong Zhang. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Applied Microbiology and Biotechnology*, 97(6):2681–2690, March 2013.

[118] Inés Martínez, Catherine E. Muller, and Jens Walter. Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS ONE*, 8(7):e69621, July 2013.

[119] Muthukrishnan Swarna. Treatment of heavy metals in stormwater runoff using wet pond and wetland mesocosms. In *Proceedings of the Annual International Conference on Soils, Sediments, Water and Energy*, volume 11, Amherst (USA), 2006. Paul T. Kostecki, Edward J. Calabrese, James Dragun.

[120] Sven Blomqvist. Quantitative sampling of soft-bottom sediments: problems and solutions. *Marine ecology progress series*, 72:295–304, 1991.

[121] Nanna Buesing and Mark O. Gessner. Comparison of detachment procedures for direct counts of bacteria associated with sediment particles, plant litter and epiphytic biofilms. *Aquatic Microbial Ecology*, 27:29–36, 2002.

[122] Chengwei Luo, Despina Tsementzi, Nikos Kyrpides, Timothy Read, and Konstantinos T. Konstantinidis. Direct comparisons of illumina vs. roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE*, 7(2):e30087, February 2012.

[123] Dirk Gevers, Frederick M. Cohan, Jeffrey G. Lawrence, Brian G. Spratt, Tom Coenye, Edward J. Feil, Erko Stackebrandt, Yves Van de Peer, Peter Vandamme, Fabiano L. Thompson, and Jean Swings. Re-evaluating prokaryotic species. *Nat Rev Micro*, 3(9):733–739, September 2005.

[124] Zarraz May-Ping Lee, Carl Bussema, and Thomas M. Schmidt. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Research*, 37:D489–D493, January 2009. WOS:000261906200088.

[125] Tomáš Větrovský and Petr Baldrian. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*, 8(2):e57923, February 2013.

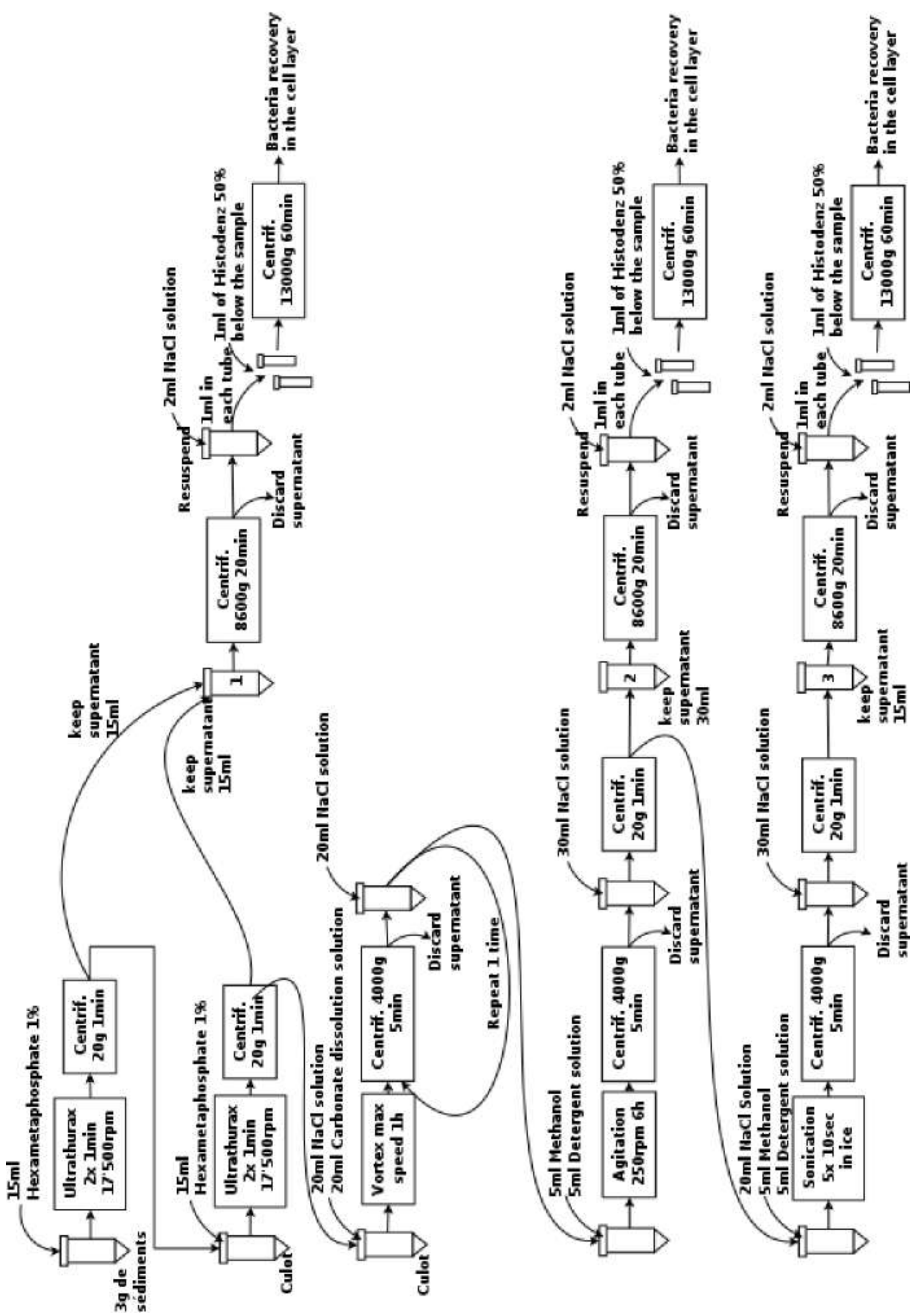
[126] Bradley S. Stevenson and Thomas M. Schmidt. Life history implications of rRNA gene copy number in *Escherichia coli*. *Applied and Environmental Microbiology*, 70(11):6670–6677, 2004.

A

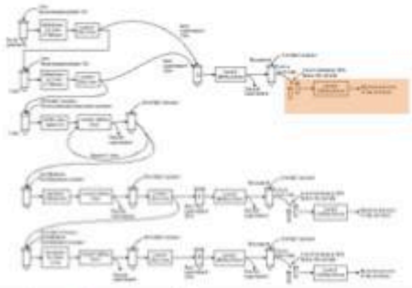
Appendix

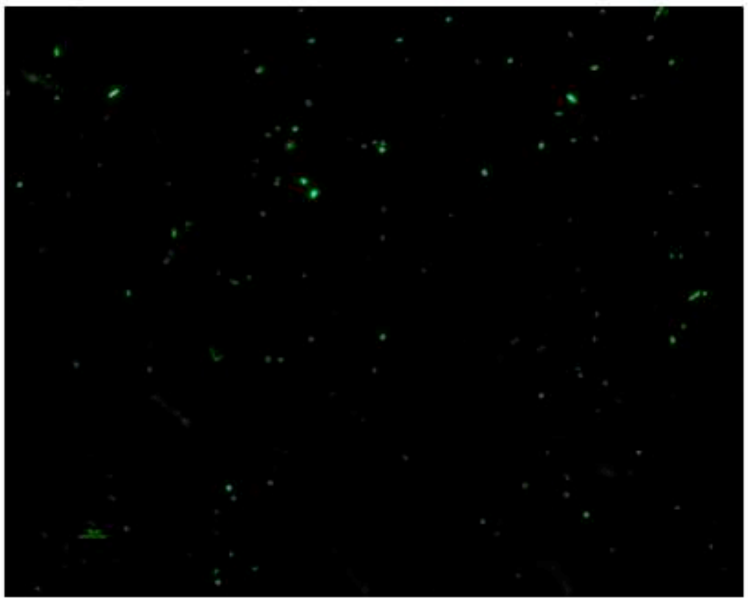
Contents:

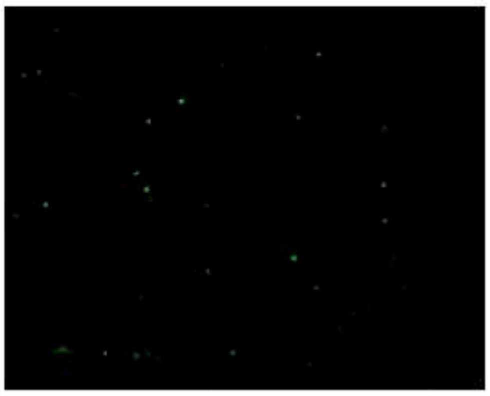
First separation method tested (based on Kallmeyer et al)



Results of step 1
after ultraturrax
homogeneisation



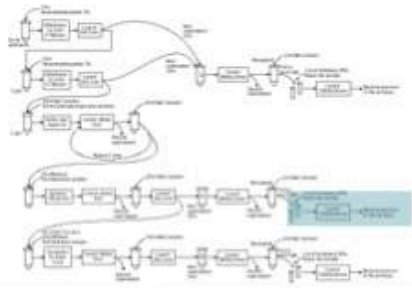


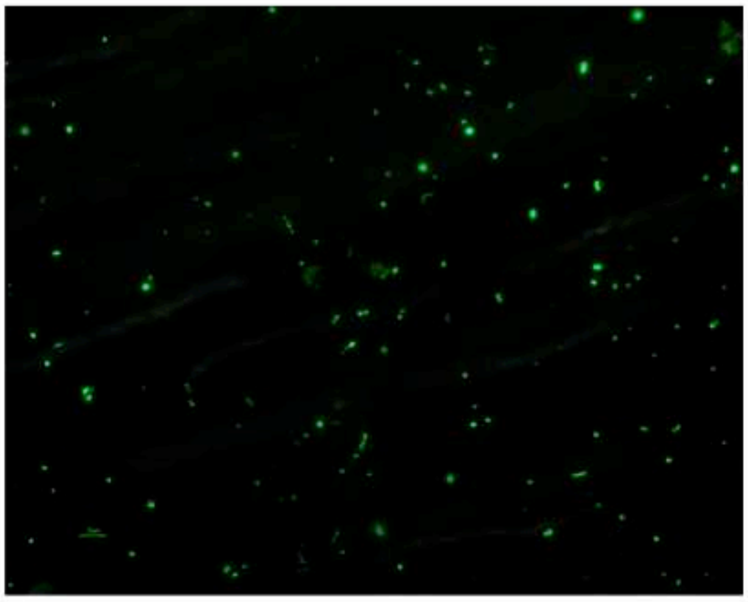


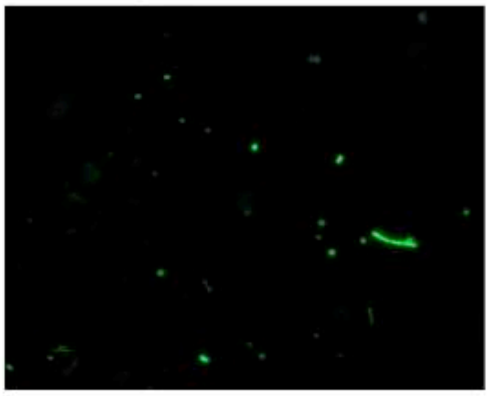
Supernatant: 30cells x 10’300 x 100
=> 30’900’000 cells

Cell layer: 180 cells x 10’300 x 100 => 185’400’000 cells

Results of step 2
after carbonates
and EPS dissolution



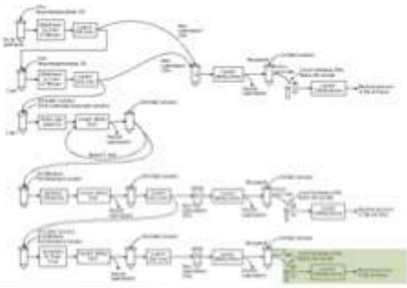


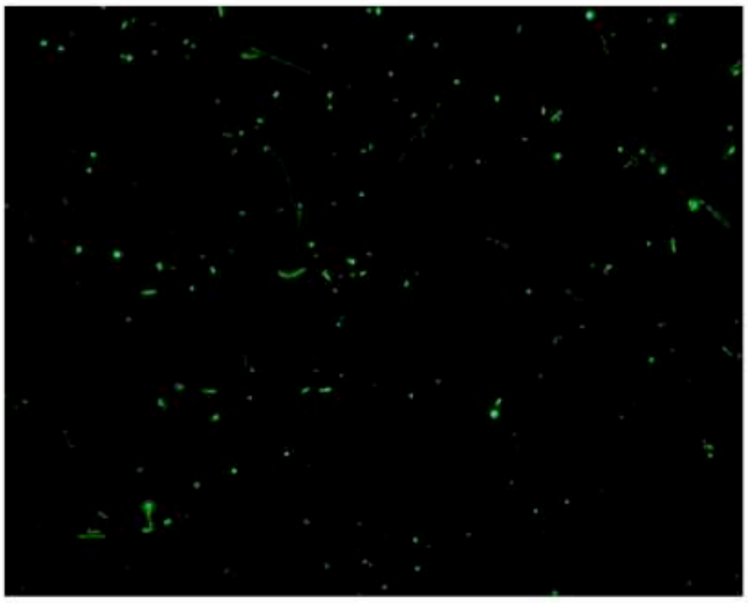


Supernatant: 70cells x 10’300 x 100
=> 72’100’000 cells

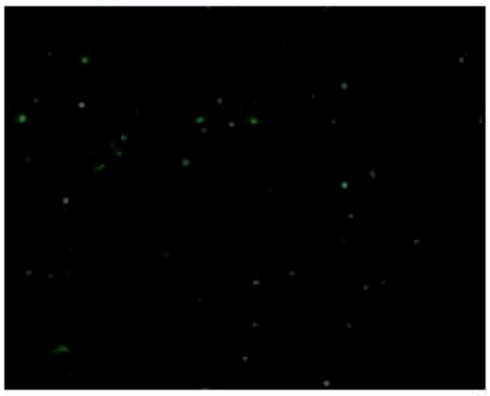
Cell layer: 250 cells x 10’300 x 100 => 257’000’000 cells

Results of step 3 after ultrasonic treatment





Cell layer: 270cells x 10’300 x 100 => 278’100’000 cells



Supernatant: 50cells x 10’300 x 100 => 51’500’000 cells

Summary

Step	Method	Histodenz 50%	Cell Layer	Supernatant	Dilution Factor	Nbr pictures per filter	Cell recovery	Supernatant
1	Ultrathurax 2x et centri 20g	without	300		30	10'300	9.27E+07	
	Ultrathurax 2x and centri 20g	with	180	30	100	10'300	1.85E+08	3.09E+07
2	Acid dissolution and detergent	with	250	70	100	10'300	2.58E+08	7.21E+07
3	Sonication	with	270	50	100	10'300	2.78E+08	5.15E+07

Next steps

- Improve the separation of the cells from the supernatant
  - Longer centrifugation
  - Lowering the density of NaCl solution by addition of ethanol before centrifugation
  - Filtration of all supernatants for quantification and DNA extraction

- Improve the quantification method
  - Increase the number of picture that we use for counting
  - Counting the cells in the different focal layer of the microscope
  - Take care of the different cell morphologies

DNA quantification during the second series of separation tests

DNA Extraction

B: Direct DNA extraction from sediment P4: Final sediment pellet after separation

Extraction of DNA from sediment

- Method mainly developed by Tina Wunderlin
- Test and adaptation was performed by Loïc Sauvain and Matthieu Bueche

Extraction of biomass from sediment:

1. Work under sterile laminar flow hood or around a Bunsen burner flame to create a sterile environment
2. Add 3 g of sediment sample to a sterile 50 ml tube
3. Add 15 ml of extraction solution (1% Na-Hexametaphosphate)
   (always keep ratio sediment-solution of 1/5)
4. Homogenize with Ultra-Thurrax (autoclaved rotor or cleaned in 70% ethanol) for 2 x 1 min at 15’500 rpm (speed 4)
5. Let the sample stand for 10 min to sediment the heaviest particles
6. Transfer supernatant to a new 50 ml tube
7. Add again 15 ml of extraction solution (1/5 vol) to the sediment pellet
8. Homogenize with Ultra-Thurrax (autoclaved rotor) for 2 x 1 min at 15’500 rpm (speed 4)
9. Let the sample stand for 10 min to sediment the heaviest particles
10. Transfer this supernatant to the previous supernatant
11. Weigh the pellet staying after removing the two supernatants
12. centrifuge at 20 x g for 1 min to remove some heavy particles
13. transfer the supernatant (contains the biomass) into a new 50 ml tube
14. Weigh the new pellet
15. Weigh the tube with the supernatant and write down the exact volume
16. Filter half of the volume through a sterile 0.2 µm nitrocellulose membrane (Millipore) using vacuum filtration (Filter apparatus has to be sterile)
    A) Retain the remaining supernatant at 4°C for further tests (spore treatment) or
    B) Filter the other half of supernatant through a second sterile 0.2 µm nitrocellulose membrane and keep it at -80°C
17. Remove the filter with sterile forceps and cut in 1/2 with sterile (flamed) scissors
18. add ½ of a filter to the lysing matrix tube of MP FAST DNA extraction kit (cut into several little pieces)
19. Add the other half of the filter to an 1.5 ml centrifuge tube and freeze at -80°C (BACKUP)

Protocol DNA extraction from sediment January 2012

DNA extraction from filters:

Preparation:

1. Clean the laminar flow hood and all material thoroughly with alcohol
2. Prepare pipette tips, pipettes, 1.5 ml Eppendorf tubes, waste bucket, tube racks and marker under the flow hood
3. Sterilize with UV light for 15 min

DNA extraction with kit from MP Biomedicals (FastDNA spin kit for soil) / Modified manufacturer procedure

1. Add **978 µl of Sodium Phosphate Buffer** to Lysing Matrix E tubes containing filters
2. Add **122 µl MT Buffer**
3. Bead beat in Tissue Lyzer for **10 min** at 50 Hz (**2 x 5 min and put on ice in between**)
4. Centrifuge at 14,000 x g for 10 minutes to pellet debris
5. Transfer supernatant (ca. 1 ml) to a clean 2.0 ml microcentrifuge tube -> **(S1)**
6. Add again **978 µl of Sodium Phosphate Buffer** to the same Lysing Matrix tube
7. Add **122 µl MT Buffer**
8. Bead beat in Tissue Lyzer for another 5 min at 50 Hz
9. Centrifuge at 14,000 x g for 10 minutes to pellet debris
10. Transfer supernatant to a clean 2.0 ml microcentrifuge tube -> **(S2)**
11. Add again **978 µl of Sodium Phosphate Buffer** to the same Lysing Matrix tube
12. Add **122 µl MT Buffer**
13. Bead beat in Tissue Lyzer for another 5 min at 50 Hz
14. Centrifuge at 14,000 x g for 10 minutes to pellet debris
15. Transfer supernatant to a clean 2.0 ml microcentrifuge tube -> **(S3)**
16. continue the protocol with three separate samples (S1 to S3)
17. Add **250 µl PPS (Protein Precipitation Solution)** and mix by shaking the tube by hand 10 times
18. Put at 4°C for 10 min to help precipitate the proteins
19. Centrifuge at 14,000 x g for 5 minutes to pellet the precipitate
20. Carefully transfer the supernatant (ca 1.5 ml) to a clean 15 ml tube (it is important to avoid taking any protein with (white precipitate))
21. Resuspend the **Binding Matrix suspension**

22. Add **1.0 ml Binding Matrix suspension** to the supernatant in the 15 ml tube
23. Place on rotator or invert by hand for 2 minutes to allow binding of DNA onto the Matrix
24. Place tube in a rack for 3 minutes to allow settling of Binding Matrix
25. Remove and discard 500 µl of supernatant being careful to avoid aspiring the settled Binding Matrix
26. Resuspend the Binding Matrix
27. Transfer approximately **600 µl of the mixture** to a SPIN™ Filter and centrifuge at 14,000 x g for 1 minute. Empty the catch tube and add the remaining mixture to the SPIN™ Filter and centrifuge as before. Empty the catch tube again. Repeat those manipulations until the all mixture is filtered
28. Add **500 µl prepared SEWS-M** and gently resuspend the pellet using the force of the liquid from the pipet tip (by aspiring and rejecting the liquid)
29. Centrifuge at 14,000 x g for 1 minute. Empty the catch tube and replace
30. Centrifuge a second time - without any addition of liquid- at 14,000 x g for 2 minutes to “dry” the matrix of residual wash solution. Discard the catch tube and replace with a new, clean catch tube.
31. Turn on heat block at 55°C
32. Air dry the SPIN™ Filter for 30 minutes at room temperature
33. Gently resuspend the Binding Matrix with **100 µl of water (DES)**
34. Incubate at 55°C in a heat block for 5 min
35. Centrifuge at 14,000 x g for 1 minute to bring eluted DNA into the clean catch tube Discard the SPIN filter. DNA is now ready for PCR and other downstream applications
36. Store at -20°C for extended periods or 4°C until use

Protocol DNA extraction from sediment January 2012

Precipitation of DNA

1. Pool the three sequential extractions together (S1, S2 and S3)
2. Measure the DNA concentration with the QuBit and determine the volume of the sample
3. Add 1/10 final volume of non-buffered sodium acetate 3M (final concentration 0.3M)
4. Mix the solution well
5. Add exactly 2 volumes of ice-cold absolute ethanol (-80°C) and again mix the solution well. Store the solution on ice for 1h or overnight at -20°C to allow precipitation of DNA
6. Centrifuge at maximum speed for 1h at 4°C
7. Carefully remove the supernatant
8. Fill the tube half way with 70% ethanol, **do not shake** and centrifuge at maximum speed for 30 minutes at 4°C
9. Carefully remove the supernatant
10. Store the open tube under the laminar flow hood at room temperature until the traces of alcohol have evaporated or use the speedvac for 2 to 5min
11. Dissolve the DNA pellet (which is often invisible) in the desired volume of nanopure water (50µl)
12. Measure the DNA concentration with the QuBit and calculate the recovery ratio of the precipitation step

Merging the two FASTA files together

Linux command in a terminal window in Ubuntu.

Modifying one of the two FASTA file

```
$ sed -f sedScript Pool_AllReads_clip.fasta > Pool_AllReads_clip2.fasta
```

sedScript for adding TAT at the beginning of each barcode sequences

```
s/^TAGTGTAGA/TATTAGTGTAGA/g
s/^ACACGTGAT/TATACACGTGAT/g
s/^TCGCACT/TATTCGCACT/g
s/^GTACTACTC/TATGTACTACTC/g
s/^AGCGTCGTCT/TATAGCGTCGTCT/g
s/^CACGCTACGT/TATCACGCTACGT/g
s/^TACACACACT/TATTACACACACT/g
s/^CTCGCGTG/TATCTCGCGTG/g
s/^GTCTAGTA/TATGTCTAGTA/g
s/^ACAGCGT/TATACAGCGT/g
s/^AGTACGCTAT/TATAGTACGCTAT/g
s/^AGAGATAC/TATAGAGATAC/g
```

Modifying one of the two QUAL file:

```
$ awk 'RS=">" {print ">"$1, $2; printf "40 40 40 "; for (i=3; i<NF; i++)
  printf $i" " ;print $NF}' Pool_AllReads_clip.qual >
  Pool_AllReads_clip2.qual
```

⇨ Manual correction of the beginning of the file generated by awk

Merging the two files:

```
$ cat C_pool_AllReads_clip.fasta Pool_AllReads_clip2.fasta >
  Pool_AllReads_clipm.fasta

$ cat C_pool_AllReads_clip.qual Pool_AllReads_clip2.qual >
  Pool_AllReads_clipm.qual
```

Qiime worflow used to treat 454-pyrosequencing data

```
#!/bin/bash -e
# interactive commands are commented out
#print_qiime_config.py

# Pre-processing
echo "Check mapping file"
rm -rf mapping_output ; check_id_map.py -B -m C_pool_info.txt -o mapping_output -v

echo "Demultiplexing"
rm -rf split_library_output ; split_libraries.py -b variable_length -z truncate_only -m
C_pool_info.txt -f Pool_AllReads_clipm.fasta -q Pool_AllReads_clipm.qual -o
split_library_output
```

Continuation:

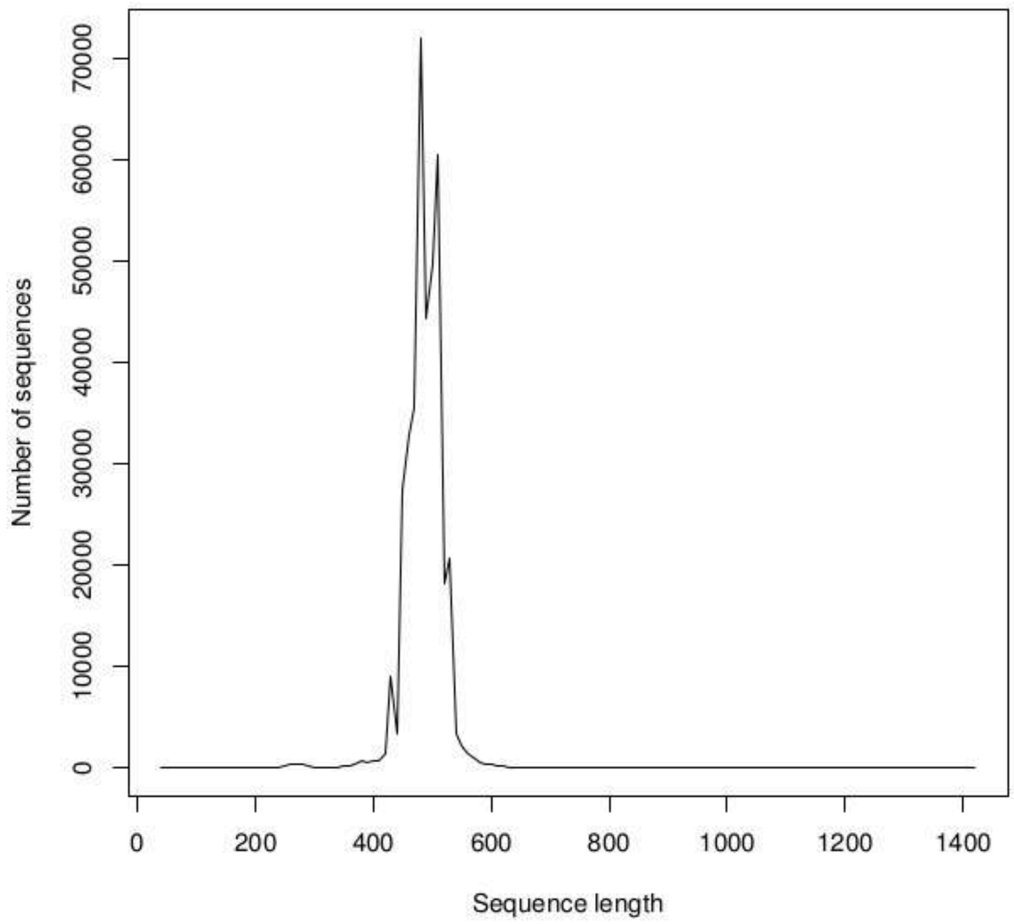
```
# Identify chimeric sequences and OUT picking without chimeric sequences
echo “Identify Chimeric sequences”
cd split_library_output; identify_chimeric_seqs.py –i seqs.fna –m usearch61 –o
usearch_checked_chimeras/ –r $HOME/qiime_software/gg_otus-12_10-release/rep_set/97_otus.fasta

filter_fasta.py –f seqs.fna –o seqs_chimeras_filtered.fna –s
usearch_checked_chimeras/chimeras.txt –n

cd ..; pick_otus.py –m usearch61 –i ./split_library_output/seqs_chimeras_filtered.fna –o
otus_chimeras_filtered/usearch61_picked_otus/

cd otus_chimeras_filtered; pick_rep_set.py –i usearch61_picked_otus/seqs_chimeras_filtered.txt –
f ../split_library_output/seqs_chimeras_filtered.fna –o rep_set

align_seqs.py –i rep_set/seqs_rep_set.fasta

assign_taxonomy.py –i pynast_aligned/seqs_rep_set_aligned.fasta

filter_alignment.py –i pynast_aligned/seqs_rep_set_aligned.fasta

make_phylogeny.py –i pynast_aligned/seqs_rep_set_aligned_pfiltered.fasta –o rep_set.tre

make_otu_table.py –i usearch61_picked_otus/seqs_chimeras_filtered_otus.txt –t
rdp_assigned_taxonomy/seqs_rep_set_aligned_tax_assignments.txt –o otu_table.biom

# otus
echo "Pick OTUs through OTU table"
rm -rf otus ; pick_de_novo_otus.py -i split_library_output/seqs.fna -o otus

#biom summarize-table –i otus/otu_table.biom –o otus/otu_table_summary.txt

echo "Convert BIOM OUT table in text file with taxonomy data"
convert_biom.py –I otus/otu_table.biom –o otus/table.from_biom_w_taxonomy.txt –b  --
header_key=taxonomy; cd ..


#OTU Heatmap
echo "OTU Heatmap"
make_otu_heatmap_html.py -i otus/otu_table.biom -o otus/OTU_Heatmap/

#OTU Network
echo "OTU Network"
make_otu_network.py -m C_pool_info.txt -i otus/otu_table.biom -o otus/OTU_Network

#Make Taxa Summary Charts
echo "Summarize taxa"
rm -rf wf_taxa_summary ; summarize_taxa_through_plots.py -i otus/otu_table.biom -o
wf_taxa_summary -m C_pool_info.txt

echo "Alpha rarefaction"
#alpha_diversity.py -h
echo "alpha_diversity:metrics shannon,PD_whole_tree,chao1,observed_species" > alpha_params.txt

rm -rf wf_arare ; alpha_rarefaction.py -i otus/otu_table.biom -m C_pool_info.txt -o wf_arare/ -p
alpha_params.txt -t otus/rep_set.tre

echo "Beta diversity and plots"
rm -rf wf_bdiv_even146 ; beta_diversity_through_plots.py -i otus/otu_table.biom -m
C_pool_info.txt -o wf_bdiv_even146/ -t otus/rep_set.tre -e 146

echo "Jackknifed beta diversity"
rm -rf wf_jack ; jackknifed_beta_diversity.py -i otus/otu_table.biom -t otus/rep_set.tre -m
C_pool_info.txt -o wf_jack -e 110

echo "Make Bootstrapped Tree"
make_bootstrapped_tree.py -m wf_jack/unweighted_unifrac/upgma_cmp/master_tree.tre -s
wf_jack/unweighted_unifrac/upgma_cmp/jackknife_support.txt -o
wf_jack/unweighted_unifrac/upgma_cmp/jackknife_named_nodes.pdf

echo "Make Bi-Plots"
rm -rf 3d_biplot ; make_emperor.py -i wf_bdiv_even146/unweighted_unifrac_pc.txt -m
C_pool_info.txt -t wf_taxa_summary/otu_table_L3.txt --n_taxa_to_keep 5 -o 3d_biplot
```

Input files for Qiime workflow :

1. **Merged FASTA file containing the sequences from the two rounds of pyrosequencing (2 times 12 samples)**
2. **Merged QUAL file containing sequences quality informations for all the sequences of the two pyrosequencing rounds**
3. **Mapping file with onformations on sequencing barcodes and primers that were used**

SampleID	BarcodeSequence	LinkerPrimerSequence	Treatment	ReversePrimer	Description
C1b24015	TAGTGTAGA	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	015cm	GTATTACCGCGGCTGCTGG	C1b24015core
C1b24153	ACACGTGAT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	153cm	GTATTACCGCGGCTGCTGG	C1b24153core
C1b2439	TCGCACT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	39cm	GTATTACCGCGGCTGCTGG	C1b2439core
C2b24015	GTACTACTC	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	015cm	GTATTACCGCGGCTGCTGG	C2b24015core
C2b24153	AGCGTCGTCT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	153cm	GTATTACCGCGGCTGCTGG	C2b24153core
C2b2439	CACGCTACGT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	39cm	GTATTACCGCGGCTGCTGG	C2b2439core
C3b24015	TACACACACT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	015cm	GTATTACCGCGGCTGCTGG	C3b24015core
C3b24153	CTCGCGTG	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	153cm	GTATTACCGCGGCTGCTGG	C3b24153core
C3b2439	GTCTAGTA	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	39cm	GTATTACCGCGGCTGCTGG	C3b2439core
C4b24015	ACAGCGT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	015cm	GTATTACCGCGGCTGCTGG	C4b24015core
C4b24153	AGTACGCTAT	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	153cm	GTATTACCGCGGCTGCTGG	C4b24153core
C4b2439	AGAGATAC	AGAGTTTGATCCTGGCTCAG,GTATTACCGCGGCTGCTGG	39cm	GTATTACCGCGGCTGCTGG	C4b2439core

Output from Demultiplexing operation



```
Number raw input seqs	430186

Length outside bounds of 200 and 1000	1597
Num ambiguous bases exceeds limit of 6	146
Missing Qual Score	0
Mean qual score below minimum of 25	2588
Max homopolymer run exceeds limit of 6	18942
Num mismatches in primer exceeds limit of 0: 12442

Number of sequences with identifiable barcode but without identifiable reverse primer: 255559

-z truncate_only option enabled; sequences without a discernible reverse primer as well as sequences with a valid barcode not
found in the mapping file may still be written.

Sequence length details for all sequences passing quality filters:
Raw len min/max/avg	200.0/781.0/529.9
Wrote len min/max/avg	170.0/732.0/490.9

Barcodes corrected/not	0/0
Uncorrected barcodes will not be written to the output fasta file.
Corrected barcodes will be written with the appropriate barcode category.
Corrected but unassigned sequences will not be written unless --retain_unassigned_reads is enabled.

Total valid barcodes that are not in mapping file	0
Sequences associated with valid barcodes that are not in the mapping file will not be written.

Barcodes in mapping file
Num Samples	24
Sample ct min/max/mean: 5849 / 31542 / 16222.08
Sample	Sequence Count	Barcode
C1b24015	31542	TATTAGTGTAGA
C40339	30468	GTCTAGTA
C4b24015	26736	TATACAGCGT
C22139	22187	AGAGATAC
C22103	21981	CTCGCGTG
C10339	21560	TCGCACT
C40303	20424	TACACACACT
C10303	20244	TAGTGTAGA
C30303	20076	CACGCTACGT
C82239	19911	AGTACGCTAT
C82203	19092	ACAGCGT
C20303	18951	ACACGTGAT
C20339	17423	GTACTACTC
C2b24153	13353	TATAGCGTCGTCT
C30339	11734	AGCGTCGTCT
C2b24015	10812	TATGTACTACTC
C3b24015	10200	TATTACACACACT
C1b24153	9464	TATACACGTGAT
C1b2439	9410	TATTCGCACT
C3b2439	8193	TATGTCTAGTA
C4b24153	7203	TATAGTACGCTAT
C4b2439	6531	TATAGAGATAC
C3b24153	5986	TATCTCGCGTG
C2b2439	5849	TATCACGCTACGT

Total number seqs written	389330
```

Output from OTU picking steps

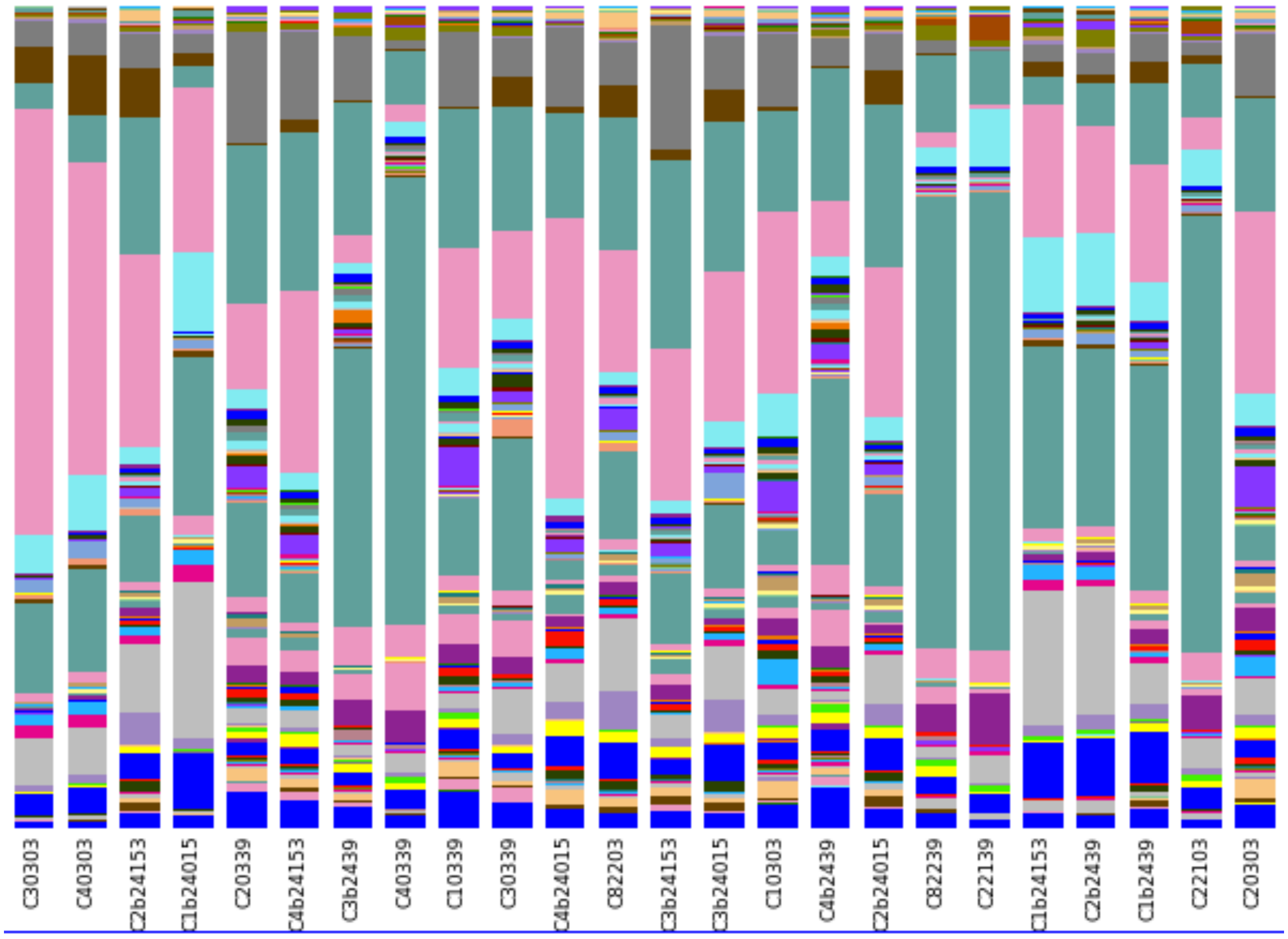
```
Num samples: 24
Num observations: 44483
Total count: 389330.0
Table density (fraction of non-zero values): 0.0919
Table md5 (unzipped): 038631cc8af925d1a77f0b0f44b6d4c6

Counts/sample summary:
 Min: 5849.0
 Max: 31542.0
 Median: 18187.0
 Mean: 16222.0833333
 Std. dev.: 7535.84775654
 Sample Metadata Categories: None provided
 Observation Metadata Categories: taxonomy

Counts/sample detail:
 C2b2439: 5849.0
 C3b24153: 5986.0
 C4b2439: 6531.0
 C4b24153: 7203.0
 C3b2439: 8193.0
 C1b2439: 9410.0
 C1b24153: 9464.0
 C3b24015: 10200.0
 C2b24015: 10812.0
 C30339: 11734.0
 C2b24153: 13353.0
 C20339: 17423.0
 C20303: 18951.0
 C82203: 19092.0
 C82239: 19911.0
 C30303: 20076.0
 C10303: 20244.0
 C40303: 20424.0
 C10339: 21560.0
 C22103: 21981.0
 C22139: 22187.0
 C4b24015: 26736.0
 C40339: 30468.0
 C1b24015: 31542.0
```

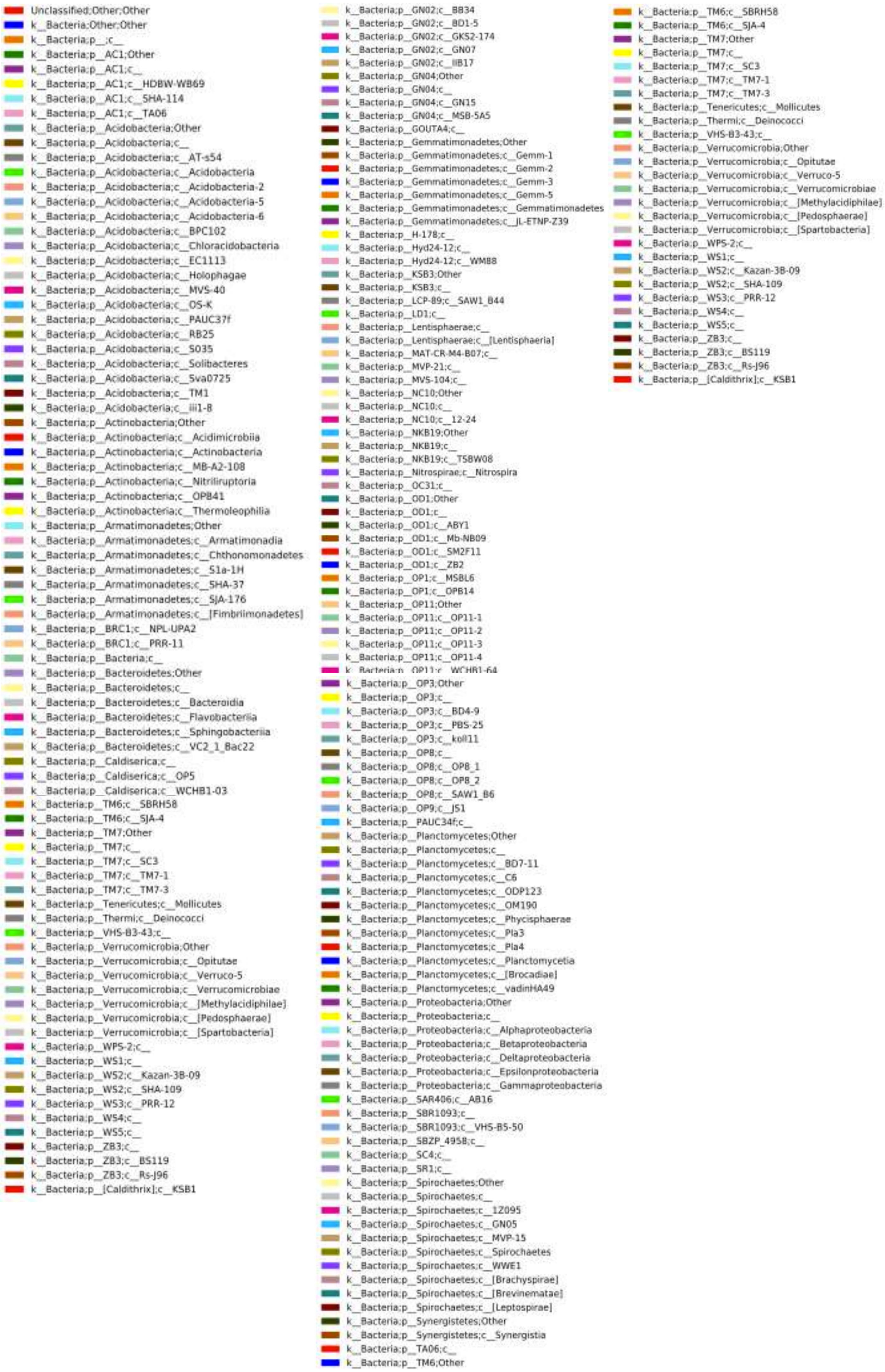
Output of OTU Heatmap and Network steps

Not so informative in this context !

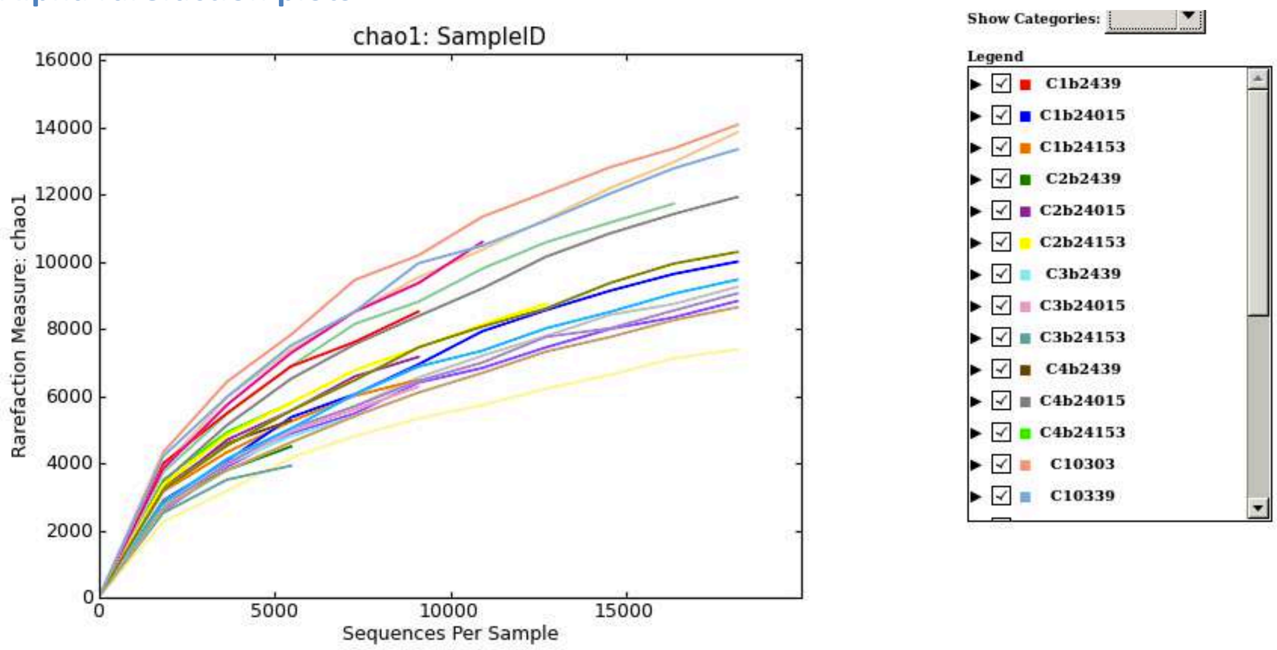
Summarize taxa plots

-Taxa plot at the Classes level :



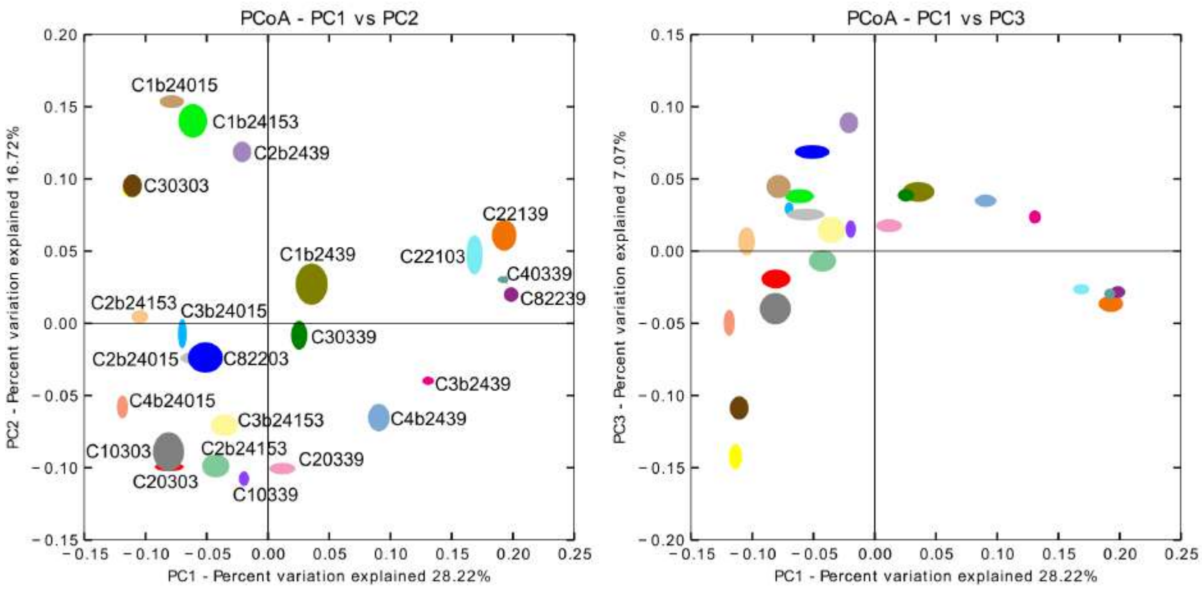

Unclassified;Other;Other
k__Bacteria;Other;Other
k__Bacteria;p__;c__
k__Bacteria;p__AC1;Other
k__Bacteria;p__AC1;c__
k__Bacteria;p__AC1;c__HDBW-WB69
k__Bacteria;p__AC1;c__SHA-114
k__Bacteria;p__AC1;c__TA06
k__Bacteria;p__Acidobacteria;Other
k__Bacteria;p__Acidobacteria;c__
k__Bacteria;p__Acidobacteria;c__AT-s54
k__Bacteria;p__Acidobacteria;c__Acidobacteria
k__Bacteria;p__Acidobacteria;c__Acidobacteria-2
k__Bacteria;p__Acidobacteria;c__Acidobacteria-5
k__Bacteria;p__Acidobacteria;c__Acidobacteria-6
k__Bacteria;p__Acidobacteria;c__BPC102
k__Bacteria;p__Acidobacteria;c__Chloracidobacteria
k__Bacteria;p__Acidobacteria;c__EC1113
k__Bacteria;p__Acidobacteria;c__Holophagae
k__Bacteria;p__Acidobacteria;c__MVS-40
k__Bacteria;p__Acidobacteria;c__OS-K
k__Bacteria;p__Acidobacteria;c__PAUC37f
k__Bacteria;p__Acidobacteria;c__RB25
k__Bacteria;p__Acidobacteria;c__S035
k__Bacteria;p__Acidobacteria;c__Solibacteres
k__Bacteria;p__Acidobacteria;c__Sva0725
k__Bacteria;p__Acidobacteria;c__TM1
k__Bacteria;p__Acidobacteria;c__iii1-8
k__Bacteria;p__Actinobacteria;Other
k__Bacteria;p__Actinobacteria;c__Acidimicrobiia
k__Bacteria;p__Actinobacteria;c__Actinobacteria
k__Bacteria;p__Actinobacteria;c__MB-A2-108
k__Bacteria;p__Actinobacteria;c__Nitriliruptoria
k__Bacteria;p__Actinobacteria;c__OPB41
k__Bacteria;p__Actinobacteria;c__Thermoleophilia
k__Bacteria;p__Armatimonadetes;Other
k__Bacteria;p__Armatimonadetes;c__Armatimonadia
k__Bacteria;p__Armatimonadetes;c__Chthonomonadetes
k__Bacteria;p__Armatimonadetes;c__S1a-1H
k__Bacteria;p__Armatimonadetes;c__SHA-37
k__Bacteria;p__Armatimonadetes;c__SJA-176
k__Bacteria;p__Armatimonadetes;c__[Fimbriimonadetes]
k__Bacteria;p__BRC1;c__NPL-UPA2
k__Bacteria;p__BRC1;c__PRR-11
k__Bacteria;p__Bacteria;c__
k__Bacteria;p__Bacteroidetes;Other
k__Bacteria;p__Bacteroidetes;c__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia
k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia
k__Bacteria;p__Bacteroidetes;c__VC2_1_Bac22
k__Bacteria;p__Caldiserica;c__
k__Bacteria;p__Caldiserica;c__OP5
k__Bacteria;p__Caldiserica;c__WCHB1-03
k__Bacteria;p__TM6;c__SBRH58
k__Bacteria;p__TM6;c__SJA-4
k__Bacteria;p__TM7;Other
k__Bacteria;p__TM7;c__
k__Bacteria;p__TM7;c__SC3
k__Bacteria;p__TM7;c__TM7-1
k__Bacteria;p__TM7;c__TM7-3
k__Bacteria;p__Tenericutes;c__Mollicutes
k__Bacteria;p__Thermi;c__Deinococci
k__Bacteria;p__VHS-B3-43;c__
k__Bacteria;p__Verrucomicrobia;Other
k__Bacteria;p__Verrucomicrobia;c__Opitutae
k__Bacteria;p__Verrucomicrobia;c__Verruco-5
k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae
k__Bacteria;p__Verrucomicrobia;c__[Methylacidiphilae]
k__Bacteria;p__Verrucomicrobia;c__[Pedosphaerae]
k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria]
k__Bacteria;p__WPS-2;c__
k__Bacteria;p__WS1;c__
k__Bacteria;p__WS2;c__Kazan-3B-09
k__Bacteria;p__WS2;c__SHA-109
k__Bacteria;p__WS3;c__PRR-12
k__Bacteria;p__WS4;c__
k__Bacteria;p__WS5;c__
k__Bacteria;p__ZB3;c__
k__Bacteria;p__ZB3;c__BS119
k__Bacteria;p__ZB3;c__Rs-J96
k__Bacteria;p__[Caldithrix];c__KSB1
k__Bacteria;p__GN02;c__BB34
k__Bacteria;p__GN02;c__BD1-5
k__Bacteria;p__GN02;c__GKS2-174
k__Bacteria;p__GN02;c__GN07
k__Bacteria;p__GN02;c__IIB17
k__Bacteria;p__GN04;Other
k__Bacteria;p__GN04;c__
k__Bacteria;p__GN04;c__GN15
k__Bacteria;p__GN04;c__MSB-5A5
k__Bacteria;p__GOUTA4;c__
k__Bacteria;p__Gemmatimonadetes;Other
k__Bacteria;p__Gemmatimonadetes;c__Gemm-1
k__Bacteria;p__Gemmatimonadetes;c__Gemm-2
k__Bacteria;p__Gemmatimonadetes;c__Gemm-3
k__Bacteria;p__Gemmatimonadetes;c__Gemm-5
k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes
k__Bacteria;p__Gemmatimonadetes;c__JL-ETNP-Z39
k__Bacteria;p__H-178;c__
k__Bacteria;p__Hyd24-12;c__
k__Bacteria;p__Hyd24-12;c__WM88
k__Bacteria;p__KSB3;Other
k__Bacteria;p__KSB3;c__
k__Bacteria;p__LCP-89;c__SAW1_B44
k__Bacteria;p__LD1;c__
k__Bacteria;p__Lentisphaerae;c__
k__Bacteria;p__Lentisphaerae;c__[Lentisphaeria]
k__Bacteria;p__MAT-CR-M4-B07;c__
k__Bacteria;p__MVP-21;c__
k__Bacteria;p__MVS-104;c__
k__Bacteria;p__NC10;Other
k__Bacteria;p__NC10;c__
k__Bacteria;p__NC10;c__12-24
k__Bacteria;p__NKB19;Other
k__Bacteria;p__NKB19;c__
k__Bacteria;p__NKB19;c__TSBW08
k__Bacteria;p__Nitrospirae;c__Nitrospira
k__Bacteria;p__OC31;c__
k__Bacteria;p__OD1;Other
k__Bacteria;p__OD1;c__
k__Bacteria;p__OD1;c__ABY1
k__Bacteria;p__OD1;c__Mb-NB09
k__Bacteria;p__OD1;c__SM2F11
k__Bacteria;p__OD1;c__ZB2
k__Bacteria;p__OP1;c__MSBL6
k__Bacteria;p__OP1;c__OPB14
k__Bacteria;p__OP11;Other
k__Bacteria;p__OP11;c__OP11-1
k__Bacteria;p__OP11;c__OP11-2
k__Bacteria;p__OP11;c__OP11-3
k__Bacteria;p__OP11;c__OP11-4
k__Bacteria;p__OP11;c__WCHB1-64
k__Bacteria;p__OP3;Other
k__Bacteria;p__OP3;c__
k__Bacteria;p__OP3;c__BD4-9
k__Bacteria;p__OP3;c__PBS-25
k__Bacteria;p__OP3;c__koll11
k__Bacteria;p__OP8;c__
k__Bacteria;p__OP8;c__OP8_1
k__Bacteria;p__OP8;c__OP8_2
k__Bacteria;p__OP8;c__SAW1_B6
k__Bacteria;p__OP9;c__JS1
k__Bacteria;p__PAUC34f;c__
k__Bacteria;p__Planctomycetes;Other
k__Bacteria;p__Planctomycetes;c__
k__Bacteria;p__Planctomycetes;c__BD7-11
k__Bacteria;p__Planctomycetes;c__C6
k__Bacteria;p__Planctomycetes;c__ODP123
k__Bacteria;p__Planctomycetes;c__OM190
k__Bacteria;p__Planctomycetes;c__Phycisphaerae
k__Bacteria;p__Planctomycetes;c__Pla3
k__Bacteria;p__Planctomycetes;c__Pla4
k__Bacteria;p__Planctomycetes;c__Planctomycetia
k__Bacteria;p__Planctomycetes;c__[Brocadiae]
k__Bacteria;p__Planctomycetes;c__vadinHA49
k__Bacteria;p__Proteobacteria;Other
k__Bacteria;p__Proteobacteria;c__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria
k__Bacteria;p__SAR406;c__AB16
k__Bacteria;p__SBR1093;c__
k__Bacteria;p__SBR1093;c__VHS-B5-50
k__Bacteria;p__SBZP_4958;c__
k__Bacteria;p__SC4;c__
k__Bacteria;p__SR1;c__
k__Bacteria;p__Spirochaetes;Other
k__Bacteria;p__Spirochaetes;c__
k__Bacteria;p__Spirochaetes;c__1Z095
k__Bacteria;p__Spirochaetes;c__GN05
k__Bacteria;p__Spirochaetes;c__MVP-15
k__Bacteria;p__Spirochaetes;c__Spirochaetes
k__Bacteria;p__Spirochaetes;c__WWE1
k__Bacteria;p__Spirochaetes;c__[Brachyspirae]
k__Bacteria;p__Spirochaetes;c__[Brevinematae]
k__Bacteria;p__Spirochaetes;c__[Leptospirae]
k__Bacteria;p__Synergistetes;Other
k__Bacteria;p__Synergistetes;c__Synergistia
k__Bacteria;p__TA06;c__
k__Bacteria;p__TM6;Other
k__Bacteria;p__TM6;c__SBRH58
k__Bacteria;p__TM6;c__SJA-4
k__Bacteria;p__TM7;Other
k__Bacteria;p__TM7;c__
k__Bacteria;p__TM7;c__SC3
k__Bacteria;p__TM7;c__TM7-1
k__Bacteria;p__TM7;c__TM7-3
k__Bacteria;p__Tenericutes;c__Mollicutes
k__Bacteria;p__Thermi;c__Deinococci
k__Bacteria;p__VHS-B3-43;c__
k__Bacteria;p__Verrucomicrobia;Other
k__Bacteria;p__Verrucomicrobia;c__Opitutae
k__Bacteria;p__Verrucomicrobia;c__Verruco-5
k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae
k__Bacteria;p__Verrucomicrobia;c__[Methylacidiphilae]
k__Bacteria;p__Verrucomicrobia;c__[Pedosphaerae]
k__Bacteria;p__Verrucomicrobia;c__[Spartobacteria]
k__Bacteria;p__WPS-2;c__
k__Bacteria;p__WS1;c__
k__Bacteria;p__WS2;c__Kazan-3B-09
k__Bacteria;p__WS2;c__SHA-109
k__Bacteria;p__WS3;c__PRR-12
k__Bacteria;p__WS4;c__
k__Bacteria;p__WS5;c__
k__Bacteria;p__ZB3;c__
k__Bacteria;p__ZB3;c__BS119
k__Bacteria;p__ZB3;c__Rs-J96
k__Bacteria;p__[Caldithrix];c__KSB1

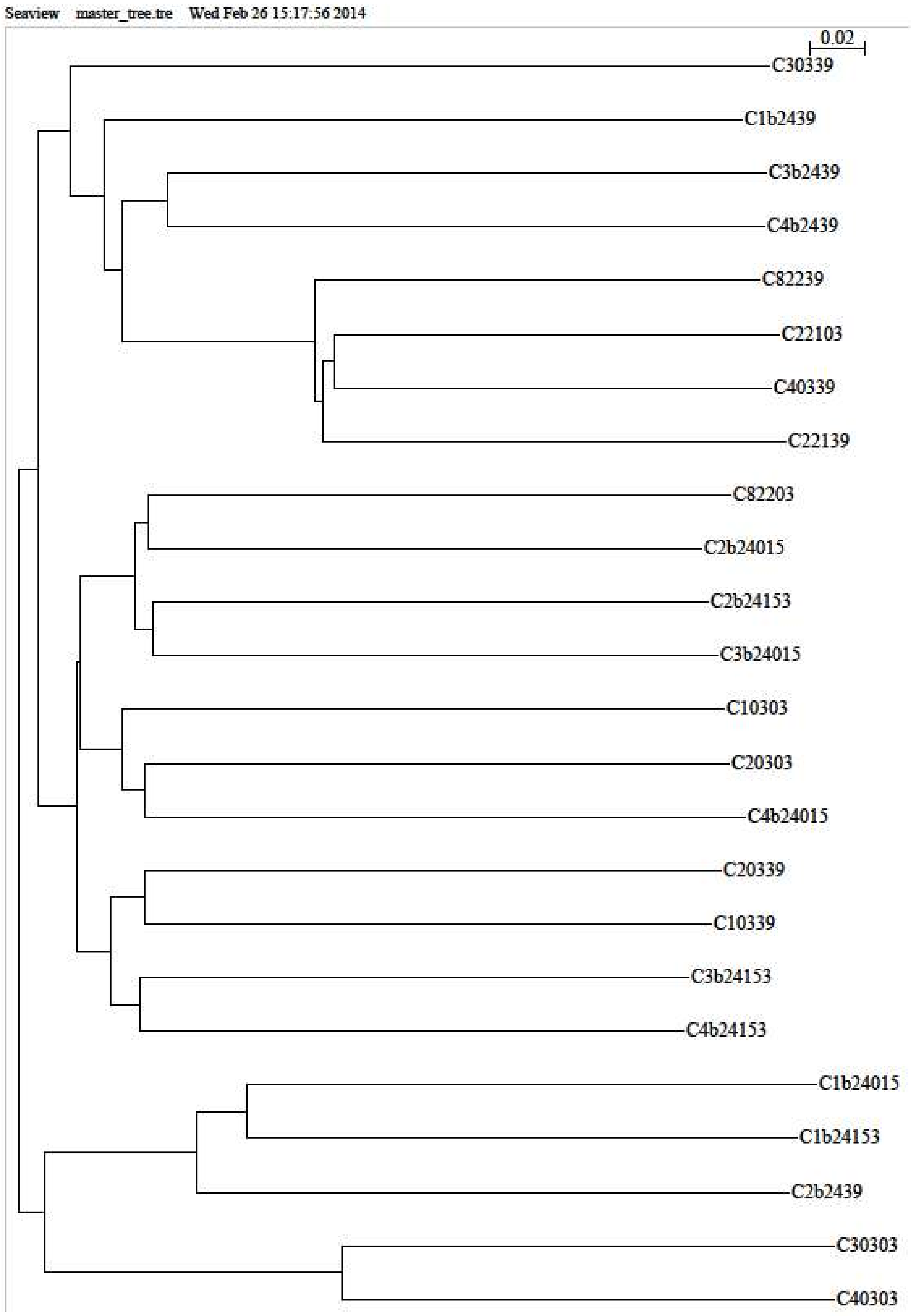
Alpha rarefaction plots



Jackknifed beta-diversity PCoA plots

⇨ Weighted unifrac distance




Seaview master_tree.tre Wed Feb 26 15:17:56 2014
0.02
C30339
C1b2439
C3b2439
C4b2439
C82239
C22103
C40339
C22139
C82203
C2b24015
C2b24153
C3b24015
C10303
C20303
C4b24015
C20339
C10339
C3b24153
C4b24153
C1b24015
C1b24153
C2b2439
C30303
C40303

R script for performing CCA analyses on pyrosequencing data

```
library(vegan)
setwd("/Users/matthieubueche/Documents/Unine/PhD
thesis/Matthieu/Pyrosequencing/MultiVarAnal_afterQiime")
#setwd("D:/Matthieu/Pyrosequencing/MultiVarAnal_afterQiime")
varenv=read.table("Summary_chimique_MergedSets.csv",dec=".",header=T,as.is=T,sep=";",row.names=1)

#setwd("Z:/Matthieu/Primers design/Article qPCR spo0A/Figure 1")
EFBlist=read.table("EFB_Genera_list.csv",dec=".",header=F,as.is=T,sep=";")

#OTUblasted=read.table("Blast assignedOTU 2Set.csv",dec=".",header=T,as.is=T,sep=";")

#setwd("D:/Bioinfo analysis/Pyr_data_taxonomy/FASTA/wf_taxa_summary/")
data=read.table("table.from_biom_w_taxonomy_MergedSets.txt",header=F,fill=T,blank.lines.skip=T)
colnames(data)=c("OTU ID","C30303","C40303","C2b24153","C1b24015","C20339","C4b24153",
                      "C3b2439","C40339","C10339","C30339","C4b24015","C82203","C3b24153",
                      "C3b24015","C10303","C4b2439","C2b24015","C82239","C22139","C1b24153",
                      "C2b2439","C1b2439","C22103","C20303",
                      "kingdom","phylum","class","order","family","genus","species")
rm_list=1

#
varenv[,14]=varenv$C.org/varenv$N
colnames(varenv)[14]="CNratio"

#for (i in c(1:length(data[,1])))
#{
#       if (is.na(data[i,c(2:13)])) data[i,c(2:13)]=0
#       #if ((sum(data[i,c(2:13)]))<=(15-(length(which(data[i,2:13]==0))/2))) rm_list=c(rm_list,i)
#       if (((sum(data[i,c(2:13)]))<=4) || ((max(data[i,c(2:13)]))<=2)) rm_list=c(rm_list,i)
#}
#data=data[-rm_list[-1],]

qPCR=read.csv("qPCR_MergedSets.csv",sep=";",header=T)
#taxonomy=read.csv("taxonomy.csv",header=T,sep=";")
#rrnDB=read.csv("rrnDB_search_results.csv",header=T,sep=";")
#copies16S=matrix(nrow=dim(taxonomy)[1],ncol=dim(taxonomy)[2])

#for (i in c(1:(dim(rrnDB)[1]-1)))
#{
#  liste=which(taxonomy[,8]==as.character(rrnDB[i,1]))
#  copies16S[liste,8]=mean(rrnDB[(which(rrnDB[,1]==as.character(rrnDB[i,1]))),4])
#}

#for (i in c(1:dim(taxonomy)[1]))
#{
#  if (!is.na(copies16S[i,8]))
#  {
#     j=7
#     while(taxonomy[i,j]=="") j=j-1
#     liste=which(taxonomy[,j]==as.character(taxonomy[i,j]))
#     copies16S[liste,j]=mean(copies16S[liste,8],na.rm=T)
#  }
#}

#for (k in c(7:2))
#{
#  for (Taxa in unique(taxonomy[,k]))
#  {
#    if (Taxa!="")
#    {
#       j=k-1
#       i=which(taxonomy[,k]==as.character(Taxa))[1]
```

```
#       while(taxonomy[i,j]=="") j=j-1
#       liste=which(taxonomy[,j]==as.character(taxonomy[i,j]))
#       if(is.na(copies16S[liste[1],j])) copies16S[liste,j]=mean(copies16S[liste,8],na.rm=T)
#    }
#  }
#}

#for (k in c(3:8))
#{
#  for (i in c(1:dim(taxonomy)[1]))
#  {
#    if (is.na(copies16S[i,k]) && (taxonomy[i,k]!=""))
#    {
#      j=k-1
#      while(taxonomy[i,j]=="") j=j-1
#      copies16S[i,k]=copies16S[i,j]
#    }
#  }
#}


data[,1]=c(0:(length(data[,1])-1))
data.transp=t(data[,1:25])
lst=1
tax_sum1=c("k__","p__","c__","o__","f__","g__","s__")
tax_sum2=c("k__;","p__;","c__;","o__;","f__;","g__;","s__;")
for (i in c(1:length(data[,1])))
{
  for (j in c(26:32))
  {
       if ((data[i,j]!="") && (data[i,j]!=tax_sum1[which(c(26:32)==j)]) &&
(data[i,j]!=tax_sum2[which(c(26:32)==j)])) lst[i]=as.character(data[i,j])
  }
}
colnames(data.transp)=lst
commun_table_pc=data.transp[c(1,5,21,23,18,4,22,15,14,8,12,7,17,16,10,25,6,24,20,2,11,3,9,13,19),
]
nbre_sequences_ech=0
for (i in c(2:25))
{
  nbre_sequences_ech[i]=sum(commun_table_pc[i,])
  commun_table_pc[i,]=commun_table_pc[i,]/nbre_sequences_ech[i]
}
commun_table=commun_table_pc[-1,]
EFB=paste("g__",EFBlist[,1],sep="")


#for (i in c(1:length(qPCR$X)))
#{
#  ligne=which(rownames(commun_table_pc)==qPCR$X[i])
#  print(ligne)
#  commun_table[ligne,]=commun_table_pc[ligne,]*qPCR$copies.g[i]
#}

#nbr16S=0
#for (k in c(1:dim(commun_table)[2]))
#{
#  i=colnames(commun_table)[k]
#  nbr16S[k]=copies16S[1,1]
#  if (length(grep("k__", i))>0)
#  {
#    taxon=gsub("k__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,1]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,1])
#  }
#  if (length(grep("p__", i))>0)
#  {
```

```
#     taxon=gsub("p__","",i)
#     taxon=gsub("Candidatus","",taxon)
#     taxon=gsub(";","",taxon)
#     taxon=gsub("\\[","",taxon)
#     taxon=gsub("\\]","",taxon)
#     liste=which(taxonomy[,2]==taxon)
#     if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,2])
#   }
#   if (length(grep("c__", i))>0)
#   {
#     taxon=gsub("c__","",i)
#     taxon=gsub("Candidatus","",taxon)
#     taxon=gsub(";","",taxon)
#     taxon=gsub("\\[","",taxon)
#     taxon=gsub("\\]","",taxon)
#     liste=which(taxonomy[,3]==taxon)
#     if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,3])
#   }
#   if (length(grep("o__", i))>0)
#   {
#     taxon=gsub("o__","",i)
#     taxon=gsub("Candidatus","",taxon)
#     taxon=gsub(";","",taxon)
#     taxon=gsub("\\[","",taxon)
#     taxon=gsub("\\]","",taxon)
#     liste=which(taxonomy[,5]==taxon)
#     if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,5])
#   }
#   if (length(grep("f__", i))>0)
#   {
#     taxon=gsub("f__","",i)
#     taxon=gsub("Candidatus","",taxon)
#     taxon=gsub(";","",taxon)
#     taxon=gsub("\\[","",taxon)
#     taxon=gsub("\\]","",taxon)
#     liste=which(taxonomy[,7]==taxon)
#     if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,7])
#   }
#   if (length(grep("g__", i))>0)
#   {
#     taxon=gsub("g__","",i)
#     taxon=gsub("Candidatus","",taxon)
#     taxon=gsub(";","",taxon)
#     taxon=gsub("\\[","",taxon)
#     taxon=gsub("\\]","",taxon)
#     liste=which(taxonomy[,8]==taxon)
#     if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,8])
#   }
#   if (is.na(nbr16S[k])) nbr16S[k]=copies16S[1,1]
#   commun_table[,k]=as.numeric(commun_table[,k])/nbr16S[k]
#}
#paste(colnames(commun_table),nbr16S,sep=": ")

orders=attributes(data[,29])$levels
tableOrders=matrix(nrow=length(orders),ncol=24)
for (i in c(1:length(orders))) {tableOrders[i,]=commun_table[,which(data[,29]==orders[i])[1]]; if
(length(which(data[,29]==orders[i]))>1)
tableOrders[i,]=rowSums(commun_table[,(which(data[,29]==orders[i]))])}


#Perform multivariate analyses
hc=hclust(dist(commun_table,"euclidean"),"average")
data.NMDS <- metaMDS(commun_table, trace=2, zerodist="add", k=2,
distance="bray",trymax=20,autotransform=FALSE)
envfit.NMDS <- envfit(data.NMDS, varenv[,c(4:12,14)], permu = 1000)
data.RDA <- rda(commun_table~ N + Fe + C.org + Mn + Cu + Cd + Al + Zn + As,data=varenv,scale=F)
data.CCA=cca(commun_table~ N + Fe + C.org + + CNratio + Mn + Al + Zn + As + Cu + Cd
,data=varenv,scale=T)
```

```
#Plot Clustering and NMDS data
scale.factor.env=0.8
cex.pt=0.7
pdf("pyroseq_BIOM_MergetSets.pdf")
plot(hc)

plot(
       data.NMDS, type="n",
       choice=c(1,2),
       cex.axis = cex.pt,
       cex.lab = cex.pt
       )
#points(data.NMDS, choice=c(1,2),display = c("sites"), cex=0.6*cex.pt, col="black")
text(data.NMDS, choice=c(1,2),display = c("sites"), cex=0.6*cex.pt, col="black")
arrows(x0=rep(0,10),y0=rep(0,10),x1=data.NMDS$species[1:10,"MDS1"],y1=data.NMDS$species[1:10,"MDS
2"],length=0.1,col="red")
text(x=data.NMDS$species[1:10,"MDS1"],y=data.NMDS$species[1:10,"MDS2"],row.names(data.NMDS$specie
s[1:10,0]),col="red",cex=0.4)

plot(   envfit.NMDS,
col = "green",
cex.axis = scale.factor.env*cex.pt,
cex.lab = scale.factor.env*cex.pt,
cex = scale.factor.env*cex.pt,
p.max = 0.99
)
abline(h=0,v=0)

dev.off()



#Sort data to plot as a function of a given threshold
above_threshold=above_threshold2=0
threshold=0.01
for (i in c(2:25))
{
       if (i == 1) above_threshold=which(commun_table_pc[i,]>threshold)
       if(i > 1) above_threshold=append(above_threshold,which(commun_table_pc[i,]>threshold))
}
above_threshold=unique(above_threshold)

threshold2=0.01
for (i in c(2:25))
{
       if (i == 1) above_threshold2=which(commun_table_pc[i,]>threshold2)
       if(i > 1) above_threshold2=append(above_threshold2,which(commun_table_pc[i,]>threshold2))
}
above_threshold2=unique(above_threshold2)




#Plot RDA ordination
win.metafile("pyroseq RDA_BIOM_MergedSets.wmf")
#pdf("pyroseq RDA_qPCR_16S.pdf")
plot(data.RDA,display = c("sites","species","cn"),type="p",cex=cex.pt,scaling=0)
text(labels=row.names(data.RDA$CCA$u[,c("RDA1","RDA2")]),x=data.RDA$CCA$u[,"RDA1"],y=data.RDA$CCA
$u[,"RDA2"],cex=0.7)
maj_sp=which(sqrt(rowSums(data.RDA$CCA$v[,c("RDA1","RDA2")]^2))>0.08)
text(labels=row.names(data.RDA$CCA$v[maj_sp,c("RDA1","RDA2")]),x=data.RDA$CCA$v[maj_sp,"RDA1"],y=
data.RDA$CCA$v[maj_sp,"RDA2"],cex=0.7,col="red")
dev.off()

scale.factor.env=0.8
cex.pt=0.9
```

```
#Plot CCA ordination
win.metafile("pyroseq CCA_BIOM_MergedSets.wmf")
#pdf("pyroseq CCA_qPCR_16S.pdf")
plot(data.CCA,display = c("sites","cn"),type="p",cex=cex.pt,scaling=0)
text(labels=row.names(data.CCA$CCA$u[,c("CCA1","CCA2")]),x=data.CCA$CCA$u[,"CCA1"],y=data.CCA$CCA
$u[,"CCA2"],cex=0.7)
#points(data.CCA$CCA$v[,"CCA1"],data.CCA$CCA$v[,"CCA2"],col="grey",pch=3,cex=0.2)
sp_HM=data.CCA$CCA$v[which(data.CCA$CCA$v[,"CCA1"]>0.2),c("CCA1","CCA2")]
sp_HM=sp_HM[which(sp_HM[,"CCA1"]<1),]
sp_HM=sp_HM[which(sp_HM[,"CCA2"]<(-0.9)),]
sp_HM=sp_HM[which(sp_HM[,"CCA2"]>(-1.7)),]

#points(sp_HM[,"CCA1"],sp_HM[,"CCA2"],col="red",pch=3,cex=0.8)
#text(x=sp_HM[,"CCA1"],y=sp_HM[,"CCA2"],rownames(sp_HM),cex=0.5,col="red")

Firmicutes=c(which(data[,27]=="p__Firmicutes"),which(data[,27]=="p__Firmicutes;"))
Clostridia=c(which(data[,28]=="c__Clostridia"),which(data[,28]=="c__Clostridia;"))
Clostridiales=c(which(data[,29]=="o__Clostridiales"),which(data[,29]=="o__Clostridiales;"))
Clostridiaceae=c(which(data[,30]=="f__Clostridiaceae"),which(data[,30]=="f__Clostridiaceae;"))
Bacilli=c(which(data[,28]=="c__Bacilli"),which(data[,28]=="c__Bacilli;"))
Bacillales=c(which(data[,29]=="o__Bacillales"),which(data[,29]=="o__Bacillales;"))
Bacillaceae=c(which(data[,30]=="f__Bacillaceae"),which(data[,30]=="f__Bacillaceae;"))
#Firmicutes=uniq(c(Firmicutes,Clostridia,Clostridiales,Clostridiaceae,Bacilli,Bacillales,Bacillac
eae))
for (i in c(1:length(EFB)))
{
  if (i==1) Endospore=which(data[,19]==EFB[i])
  if (i > 1) Endospore=append(Endospore,which(data[,19]==EFB[i]))
}

#names=colnames(commun_table_pc)
names=as.character(commun_table_pc[1,])


#Firmicutes in red
nivtaxon=Firmicutes; couleur="red"
nivtax=above_threshold[na.omit(match(nivtaxon,above_threshold))]
points(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],col=couleur,pch=3,cex=0.8)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],names[nivtax],cex=0.3,col=c
ouleur)
text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],rownames(data.CCA$CCA$v[nivt
ax,]),cex=0.5,col=couleur)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],paste(OTUblasted[match(name
s[nivtax],OTUblasted[,1]),2],OTUblasted[match(names[nivtax],OTUblasted[,1]),3]),cex=0.3,col=coule
ur)

plotted=nivtax

#nivtaxon=Bacillales; couleur="green"
#nivtax=above_threshold[na.omit(match(nivtaxon,above_threshold))]
#points(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],col=couleur,pch=3,cex=0.4
)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],rownames(data.CCA$CCA$v[niv
tax,]),cex=0.3,col=couleur)
#plotted=append(plotted,nivtax)

#nivtaxon=Endospore; couleur="red"
#nivtax=above_threshold[na.omit(match(nivtaxon,above_threshold))]
#points(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],col=couleur,pch=3,cex=0.8
)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],rownames(data.CCA$CCA$v[niv
tax,]),cex=0.5,col=couleur)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],nivtax,cex=0.5,col=couleur,
pos=2)
#plotted=append(plotted,nivtax)
```

```
#Other taxa in blue
nivtaxon=above_threshold2; couleur="blue"
nivtax=nivtaxon[which(is.na(match(nivtaxon,plotted)))]
points(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],col=couleur,pch=3,cex=0.4)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],names[nivtax],cex=0.3,col=c
ouleur)
text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],rownames(data.CCA$CCA$v[nivt
ax,]),cex=0.3,col=couleur)
#text(x=data.CCA$CCA$v[nivtax,"CCA1"],y=data.CCA$CCA$v[nivtax,"CCA2"],paste(OTUblasted[match(name
s[nivtax],OTUblasted[,1]),2],OTUblasted[match(names[nivtax],OTUblasted[,1]),3]),cex=0.3,col=coule
ur)


dev.off()



#Plot NMDS ordination with the same colors as above
win.metafile("pyroseq NMDS_BIOM_MergedSets.wmf")
plot(
        data.NMDS, type="n",
        choice=c(1,2),
        cex.axis = cex.pt,
        cex.lab = cex.pt
        )
points(data.NMDS, choice=c(1,2),display = c("sites"), cex=0.6*cex.pt, col="black")
text(data.NMDS, choice=c(1,2),display = c("sites"), cex=0.6*cex.pt, col="black")

nivtaxon=Firmicutes; couleur="red"
nivtax=above_threshold[na.omit(match(nivtaxon,above_threshold))]
points(x=data.NMDS$species[nivtax,"MDS1"],y=data.NMDS$species[nivtax,"MDS2"],col=couleur,pch=3,ce
x=0.8)
#text(x=data.NMDS$species[nivtax,"MDS1"],y=data.NMDS$species[nivtax,"MDS2"],names[nivtax],col=cou
leur,cex=0.5)
text(x=data.NMDS$species[nivtax,"MDS1"],y=data.NMDS$species[nivtax,"MDS2"],row.names(data.NMDS$sp
ecies[nivtax,0]),col=couleur,cex=0.5)

plotted=nivtax

nivtaxon=above_threshold2; couleur="blue"
nivtax=nivtaxon[which(is.na(match(nivtaxon,plotted)))]
points(x=data.NMDS$species[nivtax,"MDS1"],y=data.NMDS$species[nivtax,"MDS2"],col=couleur,pch=3,ce
x=0.4)
#text(x=data.NMDS$species[nivtax,"MDS1"],y=data.NMDS$species[nivtax,"MDS2"],names[nivtax],col=cou
leur,cex=0.5)
text(x=data.NMDS$species[nivtax,"MDS1"],y=data.NMDS$species[nivtax,"MDS2"],row.names(data.NMDS$sp
ecies[nivtax,0]),col=couleur,cex=0.3)


plot(   envfit.NMDS,
col = "green",
cex.axis = scale.factor.env*cex.pt,
cex.lab = scale.factor.env*cex.pt,
cex = scale.factor.env*cex.pt,
p.max = 0.99
)
abline(h=0,v=0)

dev.off()



#Export data on the OTUs with abundances above defined threshold
Important_OTU=matrix(ncol=3,nrow=length(c(nivtax,plotted))-1)
Important_OTU[,1]=c(names[nivtax],names[plotted])
Important_OTU[,2]=colnames(commun_table_pc)[c(nivtax,plotted)]
Important_OTU[,3]="EFB"
Important_OTU[c(1:length(nivtax)),3]="Non-EFB"
write.table(Important_OTU,file="Important_OTU_MergedSets.csv",sep=";",col.names=c("OTU
id","Taxon.","EFB"),row.names=F)
```

```
pc_EFB=rowSums(commun_table[1:12,unique(c(Bacillaceae,Clostridiaceae,Endospore))])/rowSums(commun
_table)
write.table(pc_EFB,file="pc_EFB_BIOMatrix_MergedSets.csv",sep=";")
pc_Firmicutes=rowSums(commun_table[1:12,unique(Firmicutes)])/rowSums(commun_table)
write.table(pc_Firmicutes,file="pc_Firmicutes_BIOMatrix_MergetSets.csv",sep=";")
```

R script for ploting relatives and absolute abundances of OTUs as barplots

```
library(vegan)
setwd("/Users/matthieubueche/Documents/Unine/PhD
thesis/Matthieu/Pyrosequencing/MultiVarAnal_afterQiime")
setwd("D:/Matthieu/Pyrosequencing/MultiVarAnal_afterQiime")


#setwd("Z:/Matthieu/Primers design/Article qPCR spo0A/Figure 1")
EFBlist=read.table("EFB_Genera_list.csv",dec=".",header=F,as.is=T,sep=";")

#OTUblasted=read.table("Blast assignedOTU 2Set.csv",dec=".",header=T,as.is=T,sep=";")

#setwd("D:/Bioinfo analysis/Pyr_data_taxonomy/FASTA/wf_taxa_summary/")
data=read.table("table.from_biom_w_taxonomy_MergedSets.txt",header=F,fill=T,blank.lines.skip=T)
colnames(data)=c("OTU ID","C30303","C40303","C2b24153","C1b24015","C20339","C4b24153",
                      "C3b2439","C40339","C10339","C30339","C4b24015","C82203","C3b24153",
                      "C3b24015","C10303","C4b2439","C2b24015","C82239","C22139","C1b24153",
                      "C2b2439","C1b2439","C22103","C20303",
                      "kingdom","phylum","class","order","family","genus","species")
blastNames=read.table("Blast assignedOTU MergedSets.csv",header=T,fill=T,sep=";")
rm_list=1

#for (i in c(1:length(data[,1])))
#{
#      if (is.na(data[i,c(2:13)])) data[i,c(2:13)]=0
#      #if ((sum(data[i,c(2:13)]))<=(15-(length(which(data[i,2:13]==0))/2))) rm_list=c(rm_list,i)
#      if (((sum(data[i,c(2:13)]))<=4) || ((max(data[i,c(2:13)]))<=2)) rm_list=c(rm_list,i)
#}
#data=data[-rm_list[-1],]

qPCR=read.csv("qPCR_MergedSets.csv",sep=";",header=T)
#taxonomy=read.csv("taxonomy.csv",header=T,sep=";")
#rrnDB=read.csv("rrnDB_search_results.csv",header=T,sep=";")
#copies16S=matrix(nrow=dim(taxonomy)[1],ncol=dim(taxonomy)[2])

#for (i in c(1:(dim(rrnDB)[1]-1)))
#{
#  liste=which(taxonomy[,8]==as.character(rrnDB[i,1]))
#  copies16S[liste,8]=mean(rrnDB[(which(rrnDB[,1]==as.character(rrnDB[i,1]))),4])
#}

#for (i in c(1:dim(taxonomy)[1]))
#{
#  if (!is.na(copies16S[i,8]))
#  {
#     j=7
#     while(taxonomy[i,j]=="") j=j-1
#     liste=which(taxonomy[,j]==as.character(taxonomy[i,j]))
#     copies16S[liste,j]=mean(copies16S[liste,8],na.rm=T)
#  }
#}

#for (k in c(7:2))
#{
#  for (Taxa in unique(taxonomy[,k]))
#  {
#    if (Taxa!="")
#    {
#       j=k-1
#       i=which(taxonomy[,k]==as.character(Taxa))[1]
#       while(taxonomy[i,j]=="") j=j-1
#       liste=which(taxonomy[,j]==as.character(taxonomy[i,j]))
#       if(is.na(copies16S[liste[1],j])) copies16S[liste,j]=mean(copies16S[liste,8],na.rm=T)
```

```
#    }
#  }
#}

#for (k in c(3:8))
#{
#  for (i in c(1:dim(taxonomy)[1]))
#  {
#    if (is.na(copies16S[i,k]) && (taxonomy[i,k]!=""))
#    {
#      j=k-1
#      while(taxonomy[i,j]=="") j=j-1
#      copies16S[i,k]=copies16S[i,j]
#    }
#  }
#}



#Format data for further treatment
data[,1]=c(0:(length(data[,1])-1))
data.transp=t(data[,1:25])
lst=1
tax_sum1=c("k__","p__","c__","o__","f__","g__","s__")
tax_sum2=c("k__;","p__;","c__;","o__;","f__;","g__;","s__;")
for (i in c(1:length(data[,1])))
{
  for (j in c(26:32))
  {
       if ((data[i,j]!="") && (data[i,j]!=tax_sum1[which(c(26:32)==j)]) &&
(data[i,j]!=tax_sum2[which(c(26:32)==j)])) lst[i]=as.character(data[i,j])
  }
}
colnames(data.transp)=lst
commun_table_pc=data.transp[c(1,5,21,23,18,4,22,15,14,8,12,7,17,16,10,25,6,24,20,2,11,3,9,13,19),
]
nbre_sequences_ech=0
for (i in c(2:25))
{
  nbre_sequences_ech[i]=sum(commun_table_pc[i,])
  commun_table_pc[i,]=commun_table_pc[i,]/nbre_sequences_ech[i]
}
commun_table=commun_table_pc[-1,]
EFB=paste("g__",EFBlist[,1],sep="")


for (i in c(1:length(qPCR$X)))
{
  ligne=which(rownames(commun_table_pc)==qPCR$X[i])
  print(ligne)
  commun_table[(ligne-1),]=commun_table_pc[ligne,]*qPCR$copies.g[i]
}


#The lines commented below are useful for extrapolating the number of 16S rRNA copies from rrnDB
to current dataset
#nbr16S=0
#for (k in c(1:dim(commun_table)[2]))
#{
#  i=colnames(commun_table)[k]
#  nbr16S[k]=copies16S[1,1]
#  if (length(grep("k__", i))>0)
#  {
#    taxon=gsub("k__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,1]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,1])
#  }
```

```
#  if (length(grep("p__", i))>0)
#  {
#    taxon=gsub("p__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,2]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,2])
#  }
#  if (length(grep("c__", i))>0)
#  {
#    taxon=gsub("c__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,3]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,3])
#  }
#  if (length(grep("o__", i))>0)
#  {
#    taxon=gsub("o__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,5]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,5])
#  }
#  if (length(grep("f__", i))>0)
#  {
#    taxon=gsub("f__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,7]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,7])
#  }
#  if (length(grep("g__", i))>0)
#  {
#    taxon=gsub("g__","",i)
#    taxon=gsub("Candidatus","",taxon)
#    taxon=gsub(";","",taxon)
#    taxon=gsub("\\[","",taxon)
#    taxon=gsub("\\]","",taxon)
#    liste=which(taxonomy[,8]==taxon)
#    if (length(liste)>0) nbr16S[k]=mean(copies16S[liste,8])
#  }
#  if (is.na(nbr16S[k])) nbr16S[k]=copies16S[1,1]
#  commun_table[,k]=as.numeric(commun_table[,k])/nbr16S[k]
#}
#paste(colnames(commun_table),nbr16S,sep=": ")



#Sort OTUs with abundances above threshold
above_threshold=above_threshold2=0
threshold=0.01
for (i in c(2:25))
{
        if (i == 1) above_threshold=which(commun_table_pc[i,]>threshold)
        if(i > 1) above_threshold=append(above_threshold,which(commun_table_pc[i,]>threshold))
}
above_threshold=unique(above_threshold)
above_threshold=above_threshold[-1]

threshold2=0.01
for (i in c(2:25))
```

```
{
	if (i == 1) above_threshold2=which(commun_table_pc[i,]>threshold2)
	if(i > 1) above_threshold2=append(above_threshold2,which(commun_table_pc[i,]>threshold2))
}
above_threshold2=unique(above_threshold2)
above_threshold2=above_threshold2[-1]



#Define some parameters for barplots
ToPlot=blastNames$OTU.id+1
barSpaces=c(.1,.1,.1,1,.1,1,.1,.1,.1,.1,.1,1,.1,.1,.1,.1,.1,.1,1,.1,.1,.1,.1,.1)

LegText=paste(blastNames$OTU.id,": ",blastNames$Genus," ",blastNames$Species,sep="")
OrdrePlot=c(1,2,6,19,21,3,20,23,4,7,5,13,15,14,16,8,11,10,24,18,17,22,9,12)
LabelTxt=c("N4up","N4med","N3low","N1up","N2up","N4low","N1low","M2up","N3up","M3up","N3med","D1u
p","D2up","D1low","D2low","M3med","M4med","M4up","M2low","M1low","M1up","N2low","M3low","M4low")


#Trace barplots by taxonomic groups
pdf("barGraphsMerged2.pdf",width=7,height=9,pointsize=11)
#win.metafile("barGraphsMerged.wmf",width=7,height=9.167,res=300,pointsize=11)
layout(matrix(c(1:10), 5, 2, byrow = TRUE),width=c(1,1),
heights=c(0.5,0.5,0.23,0.2,0.58),respect=F)
par(cex=0.5,mar=c(1,3,1,1),oma=c(4,3,1,2))

#pdf("Firmicutes bar_plotMerged.pdf",width=12, height=8)
#  par(mfcol=c(1,2),cex=0.75,oma=c(3,1,1,2))

  couleurs=c(1,grey(3:1/4),rainbow(12,alpha = 1),3,4,5,"white")

  Firmicutes=c(which(data[,27]=="p__Firmicutes"),which(data[,27]=="p__Firmicutes;"))
  FirmToPlot=Firmicutes[na.omit(match(ToPlot,Firmicutes))]
  OtherFirmicuteAbovT2=above_threshold2[na.omit(match(Firmicutes,above_threshold2))]
  OtherFirmicuteAbovT2=OtherFirmicuteAbovT2[-na.omit(match(ToPlot,OtherFirmicuteAbovT2))]
  OtherFirmicutes=Firmicutes[-na.omit(match(c(ToPlot,OtherFirmicuteAbovT2),Firmicutes))]

  LegendeTxt=c(LegText[match(FirmToPlot-1,blastNames$OTU.id)],"Others")

  dataToPlot=t(commun_table_pc[OrdrePlot+1,c(FirmToPlot,OtherFirmicuteAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table_pc[OrdrePlot+1,OtherFirmicutes]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
	names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F, ylim=c(0,0.5),
	space=barSpaces)

  dataToPlot=t(commun_table[OrdrePlot,c(FirmToPlot,OtherFirmicuteAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table[OrdrePlot,OtherFirmicutes]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
	names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F,
	space=barSpaces,ylim=c(0,4e9))
  legend(x=6,y=4e9, LegendeTxt,fill=couleurs,bty="n",y.intersp=0.75,ncol=2,text.width=9)
#dev.off()



#pdf("Betaproteo bar_plotMerged.pdf",width=12, height=8)
#  par(mfcol=c(1,2),cex=0.75,oma=c(3,1,1,2))

  couleurs=c(1,grey(3:1/4),rainbow(12,alpha = 1),3,"white")


Betaproteo=c(which(data[,28]=="c__Betaproteobacteria"),which(data[,28]=="c__Betaproteobacteria;")
)
  BetaPToPlot=Betaproteo[na.omit(match(ToPlot,Betaproteo))]
  OtherBetaPAbovT2=above_threshold2[na.omit(match(Betaproteo,above_threshold2))]
  OtherBetaPAbovT2=OtherBetaPAbovT2[-na.omit(match(ToPlot,OtherBetaPAbovT2))]
  OtherBetaP=Betaproteo[-na.omit(match(c(ToPlot,OtherBetaPAbovT2),Betaproteo))]

  LegendeTxt=c(LegText[match(BetaPToPlot-1,blastNames$OTU.id)],"Others")

  dataToPlot=t(commun_table_pc[OrdrePlot+1,c(BetaPToPlot,OtherBetaPAbovT2)])
```

```
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table_pc[OrdrePlot+1,OtherBetaP]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F, ylim=c(0,0.5),
          space=barSpaces)

  dataToPlot=t(commun_table[OrdrePlot,c(BetaPToPlot,OtherBetaPAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table[OrdrePlot,OtherBetaP]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F,
          space=barSpaces,ylim=c(0,3e10))
  legend(x=6,y=3e10, LegendeTxt,fill=couleurs,bty="n",y.intersp=0.75)
#dev.off()


#pdf("Deltaproteo bar_plotMerged.pdf",width=12, height=8)
#  par(mfcol=c(1,2),cex=0.75,oma=c(3,1,1,2))

  couleurs=c(1,grey(3:1/4),2,3,"white")


Deltaproteo=c(which(data[,28]=="c__Deltaproteobacteria"),which(data[,28]=="c__Deltaproteobacteria
;"))
  DeltaPToPlot=Deltaproteo[na.omit(match(ToPlot,Deltaproteo))]
  OtherDeltaPAbovT2=above_threshold2[na.omit(match(Deltaproteo,above_threshold2))]
  OtherDeltaPAbovT2=OtherDeltaPAbovT2[-na.omit(match(ToPlot,OtherDeltaPAbovT2))]
  OtherDeltaP=Deltaproteo[-na.omit(match(c(ToPlot,OtherDeltaPAbovT2),Deltaproteo))]

  LegendeTxt=c(LegText[match(DeltaPToPlot-1,blastNames$OTU.id)],"Others")

  dataToPlot=t(commun_table_pc[OrdrePlot+1,c(DeltaPToPlot,OtherDeltaPAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table_pc[OrdrePlot+1,OtherDeltaP]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F, ylim=c(0,0.2),
          space=barSpaces)

  dataToPlot=t(commun_table[OrdrePlot,c(DeltaPToPlot,OtherDeltaPAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table[OrdrePlot,OtherDeltaP]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F,
          space=barSpaces,ylim=c(0,5e9))
  legend(x=6,y=5e9, LegendeTxt,fill=couleurs,bty="n",y.intersp=0.75)
#dev.off()



#pdf("Gammaproteo bar_plotMerged.pdf",width=12, height=8)
#  par(mfcol=c(1,2),cex=0.75,oma=c(3,1,1,2))

  couleurs=c(1,grey(3:1/4),2,3,"white")


Gammaproteo=c(which(data[,28]=="c__Gammaproteobacteria"),which(data[,28]=="c__Gammaproteobacteria
;"))
  GammaPToPlot=Gammaproteo[na.omit(match(ToPlot,Gammaproteo))]
  OtherGammaPAbovT2=above_threshold2[na.omit(match(Gammaproteo,above_threshold2))]
  OtherGammaPAbovT2=OtherGammaPAbovT2[-na.omit(match(ToPlot,OtherGammaPAbovT2))]
  OtherGammaP=Gammaproteo[-na.omit(match(c(ToPlot,OtherGammaPAbovT2),Gammaproteo))]

  LegendeTxt=c(LegText[match(GammaPToPlot-1,blastNames$OTU.id)],"Others")

  dataToPlot=t(commun_table_pc[OrdrePlot+1,c(GammaPToPlot,OtherGammaPAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table_pc[OrdrePlot+1,OtherGammaP]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F, ylim=c(0,0.15),
          space=barSpaces)

  dataToPlot=t(commun_table[OrdrePlot,c(GammaPToPlot,OtherGammaPAbovT2)])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table[OrdrePlot,OtherGammaP]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,axisnames=F,
          space=barSpaces,ylim=c(0,5e9))
```

```
  legend(x=6,y=5e9, LegendeTxt,fill=couleurs,bty="n",y.intersp=0.75)
#dev.off()


#pdf("Alphaproteo bar_plotMerged.pdf",width=12, height=8)
#  par(mfcol=c(1,2),cex=0.75,oma=c(3,1,1,2))

  couleurs=c(1,grey(3:1/4),rainbow(12,alpha = 1),"white")

  NotPlotted=ToPlot[-na.omit(match(c(FirmToPlot,BetaPToPlot,DeltaPToPlot,GammaPToPlot),ToPlot))]
  OthersNotPlotted=c(1:length(data[,1]))[-
na.omit(match(c(Firmicutes,Betaproteo,Deltaproteo,Gammaproteo),c(1:length(data[,1]))))]

  LegendeTxt=c(LegText[match(NotPlotted-1,blastNames$OTU.id)],"Others")

  dataToPlot=t(commun_table_pc[OrdrePlot+1,NotPlotted])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table_pc[OrdrePlot+1,OthersNotPlotted]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2, ylim=c(0,0.6),
          space=barSpaces)

  dataToPlot=t(commun_table[OrdrePlot,NotPlotted])
  dataToPlot=rbind(dataToPlot,(colSums(t(commun_table[OrdrePlot,OthersNotPlotted]))/1))
  barplot(dataToPlot,las=2,col=couleurs,
          names.arg=LabelTxt,cex.axis=1.2,cex.names=1.2,
          space=barSpaces,ylim=c(0,3e10))
  legend(x=6,y=3e10, LegendeTxt,fill=couleurs,bty="n",y.intersp=0.75)
dev.off()
```

BLAST assignment of the most representative OTUs

1) Reference database

Download of a 16S rRNA gene sequences database from the ribosomal data project web site (rdp.cme.msu.edu)[1] with the following parameters entered in the « hierarchy browser » :

- ⇨ Selection of Type and Non-type strains
- ⇨ Isolated organisms
- ⇨ Any sizes of the sequences
- ⇨ Good quality sequences
- ⇨ NCBI taxonomy

A file with 326231 sequences was downloaded from the site and transformed in FASTA format by `seqret` command from EMBOSS[2] collection of tools :

```
$ seqret rdp_download_326231.seqs.gen database326231.fasta
```

And transformed in BLAST compatible database :

```
$ makeblastdb –in *.fasta –input_type 'fasta' –dbtype 'nucl' –
  out database326231.dna
```

2) Extracting sequences of the most representative OTUs

- ⇨ Input file : ImportantOTU.csv

OTU id	Taxon.	EFB
228	g__Acidovorax;	Non-EFB
8308	g__Rhodobacter;	Non-EFB
10144	g__Rhodobacter;	Non-EFB
22215	f__Rhodocyclaceae	Non-EFB
25512	o__Bacteroidales;	Non-EFB
25584	f__Microbacteriaceae;	Non-EFB
29201	g__Acidovorax;	Non-EFB
35488	g__KD1-23;	Non-EFB
40337	o__Bacteroidales;	Non-EFB
133	s__kujiense	Non-EFB
32028	f__Desulfobacteraceae;	Non-EFB
39966	f__Desulfobacteraceae;	Non-EFB
32368	g__Dechloromonas;	Non-EFB
35949	s__kujiense	Non-EFB
8756	f__Chromatiaceae;	Non-EFB
16295	f__Rhodocyclaceae;	Non-EFB
20272	f__Sinobacteraceae;	Non-EFB
27474	f__Chromatiaceae;	Non-EFB
29395	f__Sinobacteraceae;	Non-EFB
36587	f__Rhodocyclaceae;	Non-EFB
36968	f__Sinobacteraceae;	Non-EFB
23807	g__Dechloromonas;	Non-EFB
29548	g__Dechloromonas;	Non-EFB
30083	f__Rhodocyclaceae	Non-EFB
39064	f__Rhodocyclaceae;	Non-EFB
42980	f__Rhodocyclaceae	Non-EFB
43524	f__Nitrospinaceae;	Non-EFB
4854	g__Desulfobacterium;	Non-EFB
21715	g__Desulfobacterium;	Non-EFB
29640	f__Rhodocyclaceae	Non-EFB
33317	f__Nitrospinaceae;	Non-EFB
10450	g__WCHB1-05;	Non-EFB
16561	g__T78;	Non-EFB
3868	f__Comamonadaceae	Non-EFB
5924	g__Rhodoferax;	Non-EFB
10175	s__cryaerophilus	Non-EFB
24033	g__Hydrogenophaga;	Non-EFB
32161	s__cryaerophilus	Non-EFB
37534	g__Hydrogenophaga;	Non-EFB
41323	c__Dehalococcoidetes;	Non-EFB
41818	o__Fusobacteriales;	Non-EFB
16706	c__Dehalococcoidetes;	Non-EFB
22516	c__Dehalococcoidetes;	Non-EFB
23695	f__Chromatiaceae;	Non-EFB
42316	p__Bacteroidetes	Non-EFB
312	c__Clostridia	Non-EFB
2708	f__Peptostreptococcaceae;	EFB
9621	g__Clostridium	EFB
11240	g__Clostridium	EFB
11597	c__Clostridia	EFB
12354	g__Clostridium	EFB
12780	g__Clostridium	EFB
13732	g__Clostridium	EFB
14292	g__Clostridium	EFB
16992	f__Peptostreptococcaceae;	EFB
21030	g__Turicibacter;	EFB
24002	g__Trichococcus;	EFB
26646	f__Peptostreptococcaceae;	EFB
28000	f__Peptostreptococcaceae;	EFB
29097	f__Peptostreptococcaceae;	EFB
30885	g__Clostridium	EFB
30989	g__Clostridium	EFB
34982	f__Peptostreptococcaceae;	EFB
42465	g__Turicibacter;	EFB

⇨ Two linux shell scripts for extracting OTUs sequences from « seqs.fna » file

ImportantOTU.sh

```
#!/bin/sh
cat Important_OTU.csv | awk 'FS=";" {print $1}' | awk 'FS="\"" {print
"denovo"$2}' > ImportantOTU.txt

cat seqs_otus.txt | grep -w -f ImportantOTU.txt > SeqID_of_ImpOTU.txt

mkdir Imp_OTU

cd Imp_OTU

cat ../SeqID_of_ImpOTU.txt | awk '{for (i=2; i<=NF; i++) print $i >$1".txt"}'

for FILE in *.txt
do
  echo $FILE," processed" >> ../Imp_OTU_processed.log
  awk 'NR == FNR {list[$1];next} RS=">" {for (item in list) if ($1==item)
      print ">"$0}' $FILE ../../../split_library_output/seqs.fna >>
      $FILE".fasta"; done;
  echo $FILE," done" >>  ../Imp_OTU_processed.log
done
```

ImportantOTU2.sh

```
#!/bin/sh
for FILE in *.txt
do
  if [ -f $FILE".fasta" ]; then
    echo $FILE,"allready done, or in progress\!"
  else
    echo $FILE," processing...",`cat $FILE | grep -c ""`, "entries"
    echo $FILE," processing..." >> ../Imp_OTU_processed.log
    awk 'NR == FNR {list[$1];next} RS=">" {for (item in list) if ($1==item)
        print ">"$0}' $FILE ../../../split_library_output/seqs.fna
        >>$FILE".fasta"
    echo $FILE," done" >>  ../Imp_OTU_processed.log
  fi
done
```

<u>Output :</u> One FASTA file per OTUs

3) BLASTing OTU sequences (*blastOTU.sh*)

```
IFS="."
for FILE in *.txt.fasta; do set -- $FILE; if [ -f $1".fasta" ]; then echo
$1,"allready done, or in progress\!"; else echo "Processing "$1".txt.fasta";
uclust --sort $1.txt.fasta --output seqs_sorted.fasta; uclust --input
seqs_sorted.fasta --uc results.uc --id 0.99; uclust --uc2fasta results.uc --
input $1.txt.fasta --output $1.fasta --types S; time blastn -query $1.fasta -
db ../refDB326231.dna -num_descriptions 1 -num_alignments 1 -num_threads 6 >
$1.blast; fi; done
```

4) BLAST outputs sorting (*sortBlastOTU.sh*)

```
for FILE in *.blast; do awk '{if (($1 == "Query=") || ($1 == ">") || ($1 ==
"Identities")) {print $0;}}' $FILE | awk '{if ($1 == "Query=") {a=NR; print
">",$2;}; if (NR == a+1) {print $2,$3,$4;}; if (NR == a+2) {print $3,$4};}' |
awk 'RS=">" {print $2,$3,$4,$1,$5,$6}' | sort | uniq -w 10 -c | sort >
$FILE.sorted; done
```

Output : One file per OTU containing the number of sequences with the same identifier found in the OTU, the taxonomic affiliation and the identity level of the blast assignation.

Finally a simple command to generate one single text file gathering informations for every OTUs :

```
$ rm sorted.txt ; for i in *.blast.sorted; do echo $i >> sorted.txt; cat $i
  >>sorted.txt; done;
```

Bioinformatic assignment of 454 sequences by BLAST

1) Reference database

Download of a 16S rRNA gene sequences database from the ribosomal data project web site (rdp.cme.msu.edu)[1] with the following parameters entered in the « hierarchy browser » :

- ⇨ Selection of Type strains
- ⇨ Isolated organisms
- ⇨ Any sizes of the sequences
- ⇨ Good quality sequences
- ⇨ NCBI taxonomy

A file with 8911 sequences was downloaded from the site and transformed in FASTA format by `seqret` command from EMBOSS[2] collection of tools :

```
$ seqret rdp_download_8911.seqs.gen database8911.fasta
```

And transformed in BLAST compatible database :

```
$ makeblastdb –in *.fasta –input_type 'fasta' –dbtype 'nucl' –
  out database8911.dna
```

2) BLASTing[3] sequences against database

Starting from demultiplexed sequences files gathered from pyrosequencing (one FASTA file per sample) :

```
$ for File in ./C_pool/SORTING/* ; do time blastn –query
  ${File}/C_pool*_clip.fasta –db database8911.dna –
  num_descriptions 1 –num_alignments 1 –num_threads 10 >
  ${File}/blasthits8911.blast ; done
```

Extraction of species names from the output file :

```
$ cat blasthits8911.blast | grep "^>" > splist.txt
```

Sort and count species :

```
$ awk '{print $4,$5}' splist.txt | sort | uniq –c | sort –r >
  commun.txt
```

Or with sequenc ID :

```
$ awk '{print $2,$4,$5}' splist.txt | sort | uniq –c | sort –r
  > commun2.txt
```

3) Extraction of endospore-forming firmicutes

Starting from a list of endospore forming genera in a text file (spore_genus_list.txt) :

```
$ cd ../C2_21_03 ; cat splist.txt | grep –c “^>“ ; cat
  splist.txt | grep –c –f “../spore_genus_list.txt“
```

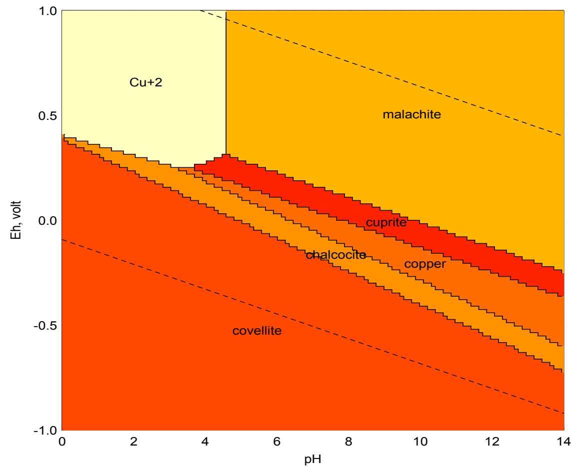
Bibliography :

1. Cole, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37,** D141 –D145 (2009).
2. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16,** 276–277 (2000).
3. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

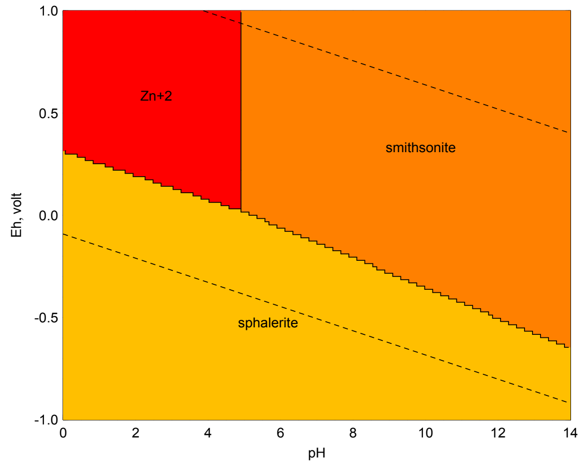
Interesting Eh-Ph diagrams

Diagrams were generated with the library CHNOSZ of the R software.

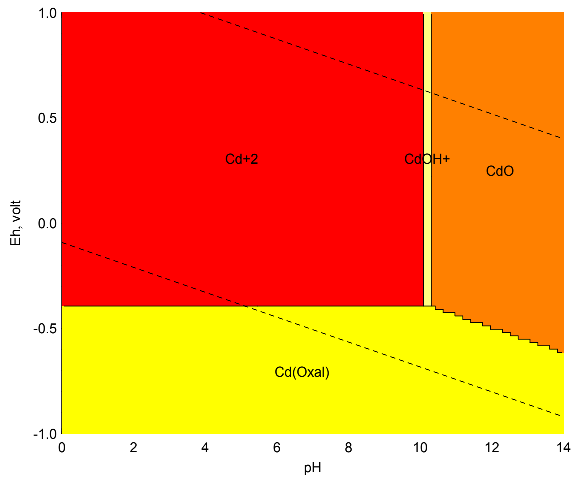
Copper (Cu – S – C – O – H)



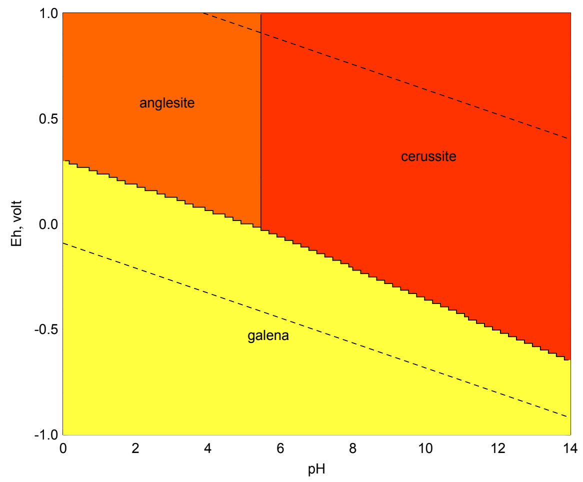
Zinc (Zn – S – C – O – H)



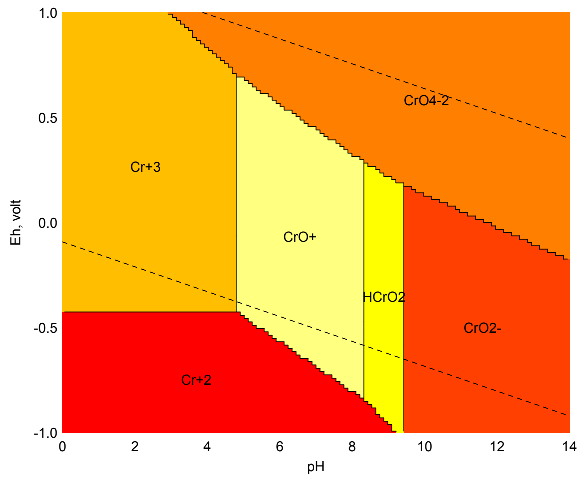
Cadmium (Cd – C – O – H)



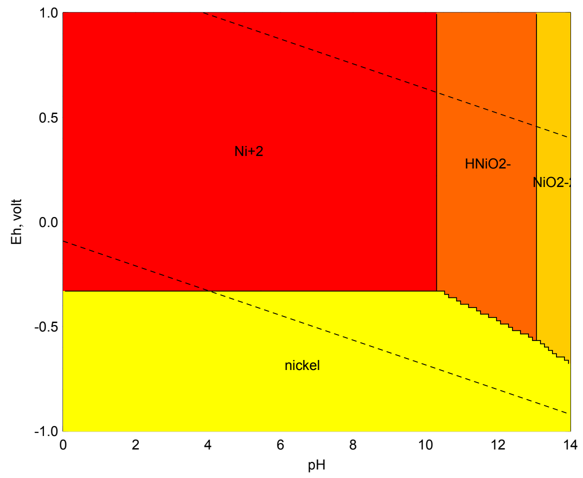
Lead (Pb – S – C – O – H)



Chromium (Cr – C – O – H)



Nickel (Ni – C – O – H)



Precipitation of heavy metals in culture bottles

Part II

Additional Papers

Paper V

In which endospores are assessed as bioenvironmental markers for past eutrophication phenomenon...

Published in Aquatic sciences, online December 10th 2013, 1-14
which should be used for any reference to this work

Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva (Switzerland–France)

Tina Wunderlin · Juan Pablo Corella · Thomas Junier · Matthieu Bueche · Jean-Luc Loizeau · Ste´phanie Girardclos · Pilar Junier

Abstract Measurements of chemical composition and biological parameters of sediment cores are used as proxies for changes in past environmental conditions and more recently the human impact on ecosystem health. In this study, endospore-forming bacteria are proposed as a new biological proxy for such paleoecological reconstructions. A sediment core providing a record for the past 90 years (^{137}Cs and magnetic susceptibility dating) was retrieved from the Rhone Delta of Lake Geneva. X-ray fluorescence was analyzed at a 0.2-cm resolution, while DNA extracts, elemental geochemistry and grain size were obtained at 4-cm intervals. The total number of bacteria and endospore-forming bacteria were quantified by qPCR using the 16S rRNA gene and the endosporulation-specific *spo0A* gene. Furthermore, a *spo0A* fragment was subjected to amplicon sequencing to define OTUs (operational taxonomic units) and the phylogenetic affiliation of the endospore formers. The results showed that despite the fact that the quantity of extracted DNA decreased with the age of the sediment, the abundance of endospore-forming bacteria remained constant. However, the diversity of this group of bacteria changed significantly, reflecting the eutrophication of the lake from 1960 to 1990. The shift in community composition was linked to the dominance of anaerobic clostridia-like endospore formers. This trend has reversed in the last 10 years of the record, suggesting a recovery after perturbation. This study shows that the abundance and diversity of endospore-forming bacteria can be used as proxies to reconstruct lake history. We hereby

This article is part of the special issue "éLEMO – investigations using MIR submersibles in Lake Geneva".

Electronic supplementary material The online version of this article (doi:10.1007/s00027-013-0329-0) contains supplementary material, which is available to authorized users.

T. Wunderlin · T. Junier · M. Bueche · P. Junier (✉)
Laboratory of Microbiology, University of Neuchâtel,
Rue Emile Argand 11, 2000 Neuchâtel, Switzerland
e-mail: pilar.junier@unine.ch

T. Wunderlin
e-mail: tina.wunderlin@unine.ch

T. Junier
e-mail: thomas.junier@unine.ch

M. Bueche
e-mail: matthieu.bueche@unine.ch

J. P. Corella · S. Girardclos
Department of Geology and Paleontology, Institute for Environmental Sciences (ISE), University of Geneva,
1205 Geneve, Switzerland
e-mail: pablo.corella@mncn.csic.es

S. Girardclos
e-mail: Stephanie.Girardclos@unige.ch

J. P. Corella
Museo Nacional de Ciencias Naturales (MNCN/CSIC), Serrano 115bis, 28006 Madrid, Spain

J.-L. Loizeau
Institute F. A. Forel, Earth and Environmental Sciences Section, Faculty of Sciences, University of Geneva, Geneva, Switzerland
e-mail: Jean-Luc.Loizeau@unige.ch

successfully introduce a new strategy for paleoecology that could also be applied to ocean sediments and long sediment cores.

Keywords Endospore-forming bacteria · Paleoecological proxy · Sediment record · Lake Geneva · Human impact · Eutrophication

Introduction

The chemical composition of lake sediments and its biological remains constitute an archive of historical data. Changes in chemical and biological parameters can be used to understand the relationship between ecological disturbances in lake systems and their impact on ecosystem health (Willard and Cronin 2007). Disturbances of lake systems can be physical or chemical. Examples of those are climate changes (rainfall, temperature), fires, storms, modification of land use (anthropogenic deviation of river inlets, deforestation) or pollution, acid deposition, and variations in nutrient input (eutrophication) or soil weath-ering. Many of the former are direct consequences of human activities.

In paleolimnology, changes in community composition of organisms producing identifiable fossilized remnants can be analyzed at different timescales from annual fluctuations to decades or centuries, depending on the resolution and length of a sediment core. The paleoecological record allows posing specific questions, such as: What is the natural variability and frequency of ecosystem differences? Are communities sta-ble over time or do they change according to long-term trends? What are the properties of ecosystems prior to a disturbance and after recovery (Gorham et al. 2001)?

Lake sediments contain a variety of microorganisms that can be used as biological indicators for paleoecological studies (Gorham et al. 2001). To be a good paleoecological proxy, an organism, or at least a recognizable structure linked to it (e.g., pollen grains or siliceous microfossils), must remain unaltered in sediments for long periods of time. This might be the case for bacterial endospores, highly resistant cellular forms produced by certain genera belonging to the Firmicutes (Gram-positive low G+C content bacteria) (Onyenwoke et al. 2004), which are able to survive in a dormant state, with little to no in situ activity, for a long period of time. The potential of dormant bacteria to be used as paleoecological indicators was evaluated by Renberg and Nilsson (1992). They concluded that it was possible to isolate dormant bacteria (*Thermo-actinomyces* spp., Bacilli and Clostridia endospores) from sediment and, by combining the analysis of biochemical and genetic data, to infer past environmental conditions. In addition, an interesting ecological feature within the group of endospore-forming bacteria is that there are aerobic, facultative anaerobic and strictly anaerobic ecotypes. With some exceptions, aerobic types cluster among the class Bacilli and anaerobes cluster mostly in the class Clostridia (Schleifer 2009).

There are few examples of research in paleoecology using endospores. Early reports suggested the presence of viable *Bacillus subtilis* spores in 320-year-old soil samples (Sneath 1962) and of *Thermoactinomyces* spp. in archeo-logical excavations containing plant debris deposited between 85 and 125 AD (Seaward et al. 1976). Spores are commonly found in sediments and isolation and enumer-ation of viable cells and spores from sedimentary archives is an old technique (Renberg and Nilsson 1992). Viable endospores have been isolated from 5,800-year-old (Bar-tholomew and Paik 1966) and 9,000-year-old lake sediments (Renberg and Nilsson 1992). A detailed study of the survival and activity of bacteria in a sediment core of about 7 m deposited over the past 13,000 years in Lake Constance (Rothfuss et al. 1997) shows that below 25 cm all the viable heterotrophic bacteria were present as heat-resistant spores. Counts of viable spores decreased expo-nentially with depth and could not be detected below 6 m (about 8,900-year-old sediment). Furthermore, several publications have shown the isolation of thermophilic endospore-forming bacteria from cold marine sediments (Bartholomew and Paik 1966; Hubert et al. 2010; d e Rezende et al. 2013). The results suggest that endospores are in most cases allochthonous and have been deposited at the time of sedimentation. All the above-mentioned studies used germination and culturing as the approach to establish the presence of viable endospores in the environment. However, culturing is biased toward a small fraction of the community (Staley and Konopka 1985; Amann et al. 1995). In a culture-independent study, the dipicolinic acid content of sediment was quantified to account for endo-spores in the sediment of the North Sea, where endospores have been found to make up to 3 % of the total prokaryotic community (Fichtel et al. 2007). In much older (deeper) sediment cores, the abundance of endospores has been estimated to be as high as the total abundance of vegetative cells (Lomstein et al. 2012). To our knowledge, a culture-independent assessment of endospore-forming community composition in sediments does not exist.

In this study, the community composition of endospore-forming bacteria is assessed as a paleoecological proxy to reconstruct the recent ecosystem history from a sedimentary record of Lake Geneva. Due to a high increase of phosphate release, Lake Geneva has seen a shift from oligotrophic waters toward eutrophication in the late 1960s. Measures to reduce the phosphate input from the early 1970s on have been effective and have decreased the nutrient level to nearly pre-1960 values (Lazzarotto and

Klein 2012). Two sediment cores were retrieved from two inactive canyons in the Rhone Delta. Endospore-forming bacteria were quantified and used as biological markers of changes recorded in the sedimentary record. The contribution of endospore-forming bacteria to microbial communities at different depths was assessed using the gene coding for the sporulation transcription regulator *spo0A*. Environmental amplicons of *spo0A* were sequenced and annotated to determine the community composition of endospore-forming bacteria in the sediment. Here, the changes in the composition of the endospore-forming community were correlated with the chemical and physical characterization of the sediment. A shift in the composition and a sharp decrease of diversity reflected ecosystem changes due to eutrophication.

Materials and methods

Site description

Two Uwitec gravity sediment cores (CAN01, coordinates 559901–139859, 79 m depth, 105 cm; and CAN02, coordinates 559405–140504, 96 m depth, 107 cm) were retrieved in August 2011 using *La Licorne* research vessel (Institute A. Forel, University of Geneva, Switzerland) in two canyons (C1 and C2, Supplementary Figure 1) on the eastern side of the Rhone delta in Lake Geneva (Switzerland), which are inactive Rhone canyons since the river was channeled around 1870 (Sastre et al. 2010). Both canyons constitute paleoreliefs with smooth lateral slopes and without any connection to a modern river, although C1 was likely connected to the local Eau Froide River in the past (Sastre et al. 2010). The cores were stored in a cold room at 4 °C.

Sedimentological description

Measurements of the physical properties every 5 mm were carried out with a Geotek multi-sensor core logger (MSCL) at the ETH Zurich Limnogeology Laboratory. The cores were then split into two lengthwise halves to proceed with the sedimentological description. Pictures were obtained using a digital camera and controlled light conditions. X-ray fluorescence was analyzed in CAN02 core using an AVAATECH XRF core scanner (2000 A, 10 and 30 kV) every 2 mm at the University of Barcelona. CAN01 sediment core was sampled every 4 cm for total carbon (TC) and total nitrogen (Vonlanthen et al. 2012) using an elemental analyzer (Hekatech Euro EA, Germany) and for total inorganic carbon (TIC) using titration coulometry (Coulomat 5015 CO2-Coulometer, Coulometric Inc., USA) at Eawag (Switzerland). Total organic carbon (Gioia et al.) was calculated as the difference between TC and TIC. In addition, grain size distribution was measured with a 4-cm resolution using a Mastersizer 2000 particle-size analyzer (Marlvern instruments Ltd, USA) at the Pyrenean Institute of Ecology (Zaragoza, Spain).

Geochronology

CAN01 was dated by the ^{137}Cs activity method on dry sediment by gamma spectrometry using HPGe well detectors (Ortec, GWL series, USA) at the Institute Forel (University of Geneva, Switzerland). Core correlation with CAN02 sediment core was carried out by visual description, sediment color and texture and by comparing magnetic susceptibility (MS) and density core profiles. Dating was based on the first fallout of ^{137}Cs in 1954–1955, the peak of atmospheric nuclear tests in 1963–1964 and the peak of the Chernobyl accident in 1986. Additional dates were obtained by the correlation of the MS signal with dated sediment cores published previously (Loizeau et al. 1997). Dating was based on a length depth scale and not cumulated sediment mass scales, because porosity did not change drastically with depth; therefore no significant bias was introduced by porosity variations.

DNA extraction

For DNA extraction, sediment core CAN01 was sub-sampled aseptically every 2 cm and samples were stored at −20 °C until processing. The DNA extraction protocol was optimized on preliminary trials with endospores from *B. subtilis*. Endospore preparations of 99 % spores were subjected to successive DNA extractions following protocols for three commercially available kits (Supplementary Figure 2). Between each extraction, the freed DNA was separated from the remaining cell pellet by centrifugation for 5 min at 14,000×*g*. Sediment DNA extractions were finally performed using the MP FastDNA® SPIN kit for soil (MP Biomedicals, Santa Ana, CA, USA) with the following modifications: the sediment was subjected to three repetitive extractions. Briefly, 0.5 g sediment was subjected to in situ cell lysis using bead beating at 50 strokes per second with the TissueLyser LT (QIAGEN, Hilden, Germany) for 10 min. The sample was then centrifuged for 5 min at 14,000×*g* and 900 µl of supernatant fluid containing the initial fraction of free DNA was collected in a separate tube. To the pelleted sediment, lysis buffer was added two additional times for a second and third round of bead beating for 5 min. Each time, between the bead-beating steps the supernatant fluid was collected in a separate tube. The three DNA-containing supernatants were then processed individually for the remaining steps of the extraction protocol following the manufacturer's guidelines. The three purified DNA samples were in the end pooled together and DNA precipitated with 0.3 M Na-acetate and ethanol (99 %) and washed with ethanol (70 %) before re-suspending in sterile water. DNA yield for the pooled extracts was measured with a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using the Quant-iT dsDNA BR assay kit, following the manufacturer's instructions. DNA quality was also verified by agarose gel electrophoresis.

Quantification of total bacteria

Quantification of bacterial DNA in sediment extracts was carried out by real-time quantitative PCR of the V3 region of the 16S rRNA gene with primers 338f and 520r (Ovreå°s et al. 1997). The qPCR mix contained 0.5 ng of DNA template, 0.3 μM of each primer and 10 μL of QuantiTect SYBR® Green PCR Kit (QIAGEN). Total reaction volume of 20 μL was reached with PCR-grade water. The qPCR was run with a Rotor-GeneTM 6000 instrument (QIAGEN) with the following program: enzyme activation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 5 s, annealing at 55 °C for 15 s and extension at 72 °C for 20 s. Thresholds (Th), Ct values and derivatives of melting curves were determined using Rotor-Gene 6 software. All the qPCR reactions were run in three technical replicates. For quantification, three independent plasmid standards series with 300 to 3×10^6 gene copies/μL of the 16S rRNA gene of an environmental clone were included.

Quantification of endospore-forming bacteria

Quantification of *spo0A* gene was done as mentioned above for the 16S rRNA gene, but with the primers *spo0A*655f and *spo0A*923r (Bueche et al. 2013). The qPCR mix contained 5 ng DNA template, 0.76 μM of each primer and 1 × QuantiTect SYBR® Green PCR Kit. Total reaction volume of 20 μL was reached with PCR-grade water. The program differed by an annealing temperature of 52 °C for 30 s and extension at 72 °C for 30 s. All qPCR reactions were run in three technical replicates. For quantification, three independent plasmid standards series with 30 to 3×10^5 gene copies/μL o f *spo0A* gene of *B. subtilis* were included.

Determination of OTUs of endospore-forming bacteria

Degenerate primers amplifying a 602 bp fragment of the *spo0A* gene (Wunderlin et al. 2013) were used for determination of the phylogenetic affiliation (OTUs) of endospore-forming bacteria. PCR reactions were performed with 0.5 ng DNA template, 1× reaction buffer (TaKaRa Bio, Shiga, Japan), 3 mM $MgCl_2$, 1 0 μg bovine serum albumin (BSA; New England Biolabs, Ipswich, MA, USA), 1 U of the proofreading Ex Taq Polymerase (TaKaRa), 200 μM of each dNTP and 1 μM of each primer in a total reaction volume of 50 μl, completed with PCR-grade water. Negative controls (1 μl PCR-grade water) and positive controls (1 ng *Paenibacillus alvei* DNA template) were included in all reactions. Reactions were done with the Arktik Thermo Cycler (Thermo Fisher Scientific, Vantaa, Finland) with the following temperature program: initial denaturation at 94 °C for 5 min; then 10 cycles of denaturation at 94 °C for 30 min, touchdown annealing starting at 55 °C with decrease of 0.3 °C per cycle for 30 s and elongation at 72 °C for 1 min, followed by 30 cycles of denaturation at 94 °C for 30 min, annealing at 52 °C for 30 s and elongation at 72 °C for 1 min; and a final extension at 72 °C for 5 min. Amplified fragments were sent for barcode amplicon sequencing with Roche GS FLX + (Eurofins MWG Operon, Ebersberg, Germany). Sequences were binned according to their barcode and the corresponding sample.

A size distribution with the entire set of sequences was computed. The length distribution of the amplicons showed a peak at around 600 bp (95 % of the sequences), corresponding to the expected fragment size. Sequences shorter or larger than that were removed for further analysis. The remaining sequences were then curated to establish operational taxonomic units (OTUs). Briefly, the amplicons were subjected to the following steps: removing of duplicates, denoising (removing sequences containing sequencing errors) and removing chimeras. The remaining sequences were clustered with the UCLUST algorithm (Edgar 2010). Putative OTUs were defined based on over 97 % nucleotide sequence identity (uclust default parameters) in the same way as commonly done for the 16S rRNA gene.

A classifier to define the genus affiliation of the OTUs for the *spo0A* amplicons was developed in analogy to the naı¨ve Bayesian classifier used by the Ribosomal Database Project (RDP) (Wang et al. 2007). Classification was based on a training set of 238 Spo0A sequences representing all 17 genera of Firmicutes available from the European Bioinformatics Institute (EMBL-EBI) database (Kanz et al. 2005). The genera included *Alicyclobacillus, Alkaliphilus, Bacillus, Brevibacillus, Caldicellulosiruptor, Clostridium, Desulfitobacterium, Desulfotomaculum, Eubacterium, Exiguobacterium, Geobacillus, Halanaerobium, Paenibacillus, Ruminococcus, Sulfobacillus, Thermoanaerobacter* and *Thermoanaerobacterium*. Unfortunately, classification down to the species level was not possible because the reference data required for a reliable assignment is currently unavailable and biased toward medically relevant species.

Statistical analysis and display

The distribution of the most abundant OTUs (over 40 sequences per OTU) per depth was analyzed using the heatmap function in R (R 2012). Correlations with depth were calculated using linear or exponential regression with the program SigmaPlot 12.0 (Systat Software, San Jose, CA, USA). Significance of differences of parameters before and after 1960 was calculated using Mann–Whitney rank sum test or Student's *t* test, when data were normally distributed. Correlations between community structure (only OTUs shared between at least two samples) and environmental parameters were determined by canonical correspondence analysis (CCA) with the program R using the package vegan (Oksanen et al. 2007) and BiodiversityR (Kindt and Coe 2005). For the CCA, total phosphorus values (mean weighted concentrations in µg/L measured at the center of the lake at station SHL2) were retrieved from Lazzarotto and Klein (2012); www.cipel.org, Annex 1, p. 46.). Phosphorus values before 1957 were assumed to be constant (average values from 1957 to 1959).

Results

Sedimentary facies

The description of CAN01 and CAN02 sediment revealed unique sedimentary facies corresponding to hemipelagic sediments as previously shown in Loizeau (1991) and Corella et al. (2011). This background sedimentation consisted of alternating triplets of (1) mm-thick organic debris layers (2) calcite-rich white laminae and (3) mm- to cm-thick allochthonous detrital layers mostly transported within the river plume and dispersed as interflow. Grain size profile (Fig. 1) revealed that sediments were fine silts with a mean grain size of 17 µm, although an interval between 15 and 42 cm with coarser material (mean grain size 26 µm) can be seen. The lack of turbidities along the sediment core revealed the absence of underflow processes during the last decades in these inactive canyons and makes this emplacement adequate to carry out the paleoecological reconstruction proposed in this study.

Chronology of the sedimentary sequence

^{137}Cs activity in core CAN01 ranged from 7 to 297 Bq kg^{-1} (Fig. 1). The lack of ^{137}Cs in the sediment from 44-cm depth downwards suggests that deposition below this depth predates AD 1954. Two well-defined peaks were found at 37 and 19.5-cm depth (149 and 247 Bq kg^{-1}, respectively), most probably corresponding to the 1963–1964 atmospheric nuclear tests maximum fallout and the 1986 Chernobyl accident. According to these two peaks, sedimentation rate since 1963 has been quite stable at 0.77 $\pm$ 0.05 cm year^{-1}. As there is no ^{137}Cs signal below 44 cm, MS measurements were used to date the lower part of the core. MS shows a peak at a depth of 65.5 cm, which can be correlated with peak # 8, dated to 1943 $\pm$ 1.4, from a previous study in the Rhone delta area (Loizeau et al. 1997). This suggests a higher sedimentation rate before 1950 corresponding to $\sim$1.83 cm year^{-1}, similar to the increase observed in the more distal area of the delta (Loizeau et al. 1997). Assuming constant sedimentation rate, the dating of the bottom of the core can be extrapolated to $\sim$1920 AD. Correlation between both sediment cores enables estimating similar sedimentation rates for CAN02 (Fig. 1).

Sediment geochemistry

The XRF data shown in Fig. 1 reflect biogeochemical relations in the lake and its catchment area (Corella et al. 2011). The downcore XRF profile in the studied sediment core (CAN02) revealed a large limnological change in the lake in the years from 1953 to 1991 (15–47-cm sediment depth) (shaded area in Fig. 1), with a decrease in K and Ti and a significant increase in Ca above 47 cm (1953)($p = < 0.001$, Whitney rank sum test). The Fe/Mn ratio fluctuated strongly in this time period and displayed significantly higher values since 1953 ($p = < 0.001$, Whitney rank sum test). Organic matter content in the sediment was relatively low with TOC values ranging from 0.4 to 2.2 % and C/N ratio from 10.9 to 42.4. The data of TIC, TOC as well as C/N ratio reflect the shift in lake conditions around 1960, after which TIC is significantly higher (average of 2.1 %, $p = 0.007$, Whitney rank sum test); TOC is also significantly higher (average 1.17 %, $p = 0.002$) and C/N ratio is significantly lower (average 14.4, $p = < 0.001$).

DNA yields and quantification of total bacteria

The DNA extraction method was optimized to target endospores as well as vegetative cells. Three different commercial DNA extraction kits were tested on an endospore preparation of *B. subtilis* to decide which was the most efficient in extracting DNA from endospores. The results show that a mechanical disruption including three rounds of extended bead beating is needed to obtain DNA from hard-to-break structures such as spores (Supplementary Figure 2).

Using the modified extraction method, the obtained DNA ranged from 0.8 to 16.4 µg g^{-1} sediment (Fig. 2a). The highest DNA yield was not obtained at the top, but at 5-cm sediment depth and two other peaks could be iden-tified at depths of 17 and 43 cm. Throughout the entire depth of the core, there is a significant exponential decay in DNA (correlation $r^2 = 0.89$, $p = < 0.001$).

The number of 16S rRNA gene copies (Fig. 2b) ranged from 8.2 $\times$ 10^9 copies g^{-1} sediment in the upper part of the core (at a depth of 5 cm) to 3.8 $\times$ 10^8 copies g^{-1}

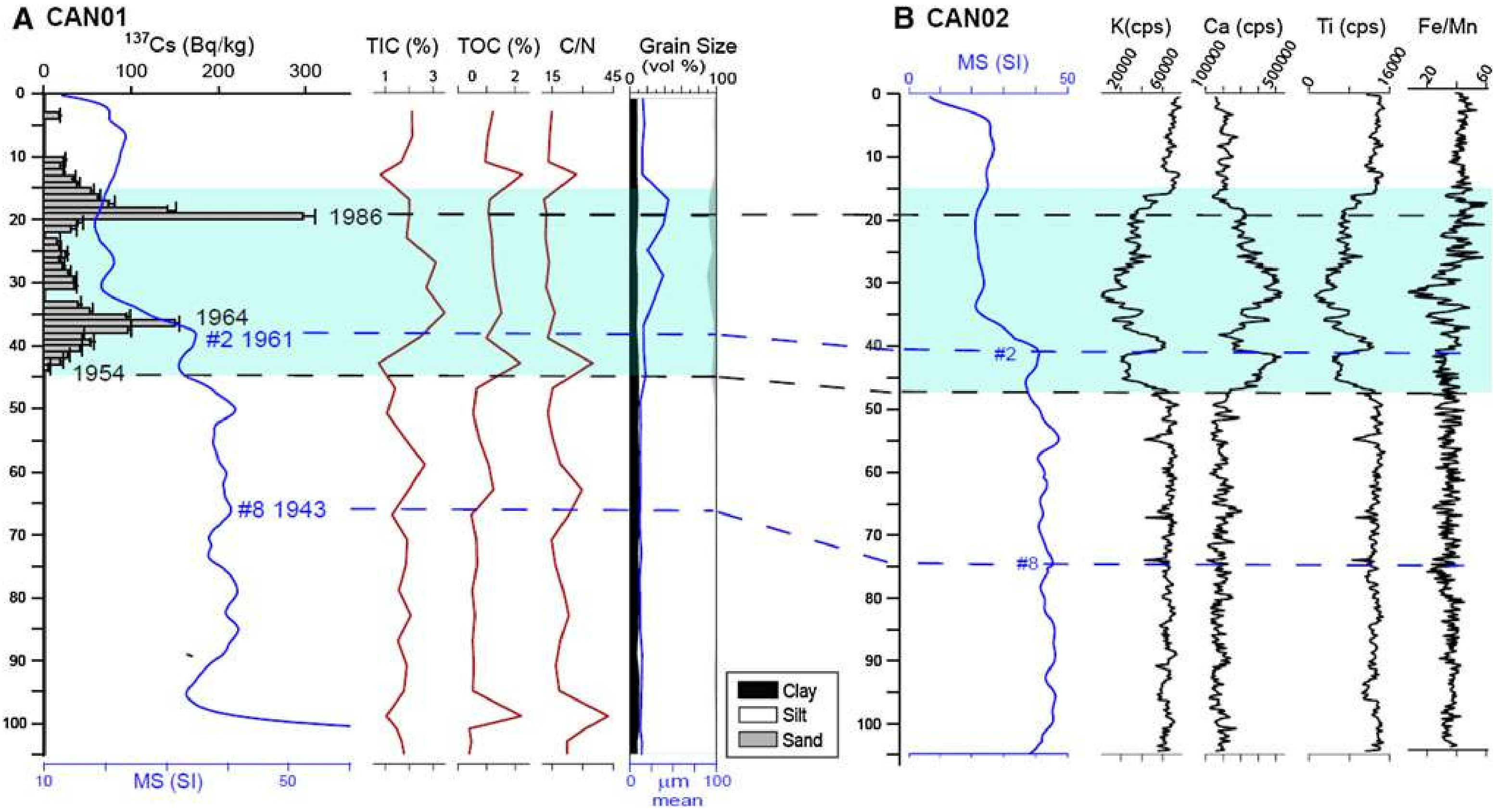


Fig. 1 Sediment cores analyzed in this study. **a** From *left to right*: ^{14}Cs activity and magnetic susceptibility (MS) profiles and geochemical profiles of sediment core CAN01 (*TIC* total inorganic carbon; *TOC* total organic carbon; *C/N* organic carbon/nitrogen ratio), grain size (*blue line* indicates the mean grain size). **b** From *left to right*: magnetic susceptibility (MS) and XRF profiles in cps (counts per second) (K, Ca, Ti, Fe/Mn ratio) in core CAN02. *Blue lines* correspond to correlation horizons based on sedimentological and magnetic susceptibility profiles. *Colored area* corresponds to a period with a large limnological change (color figure online)

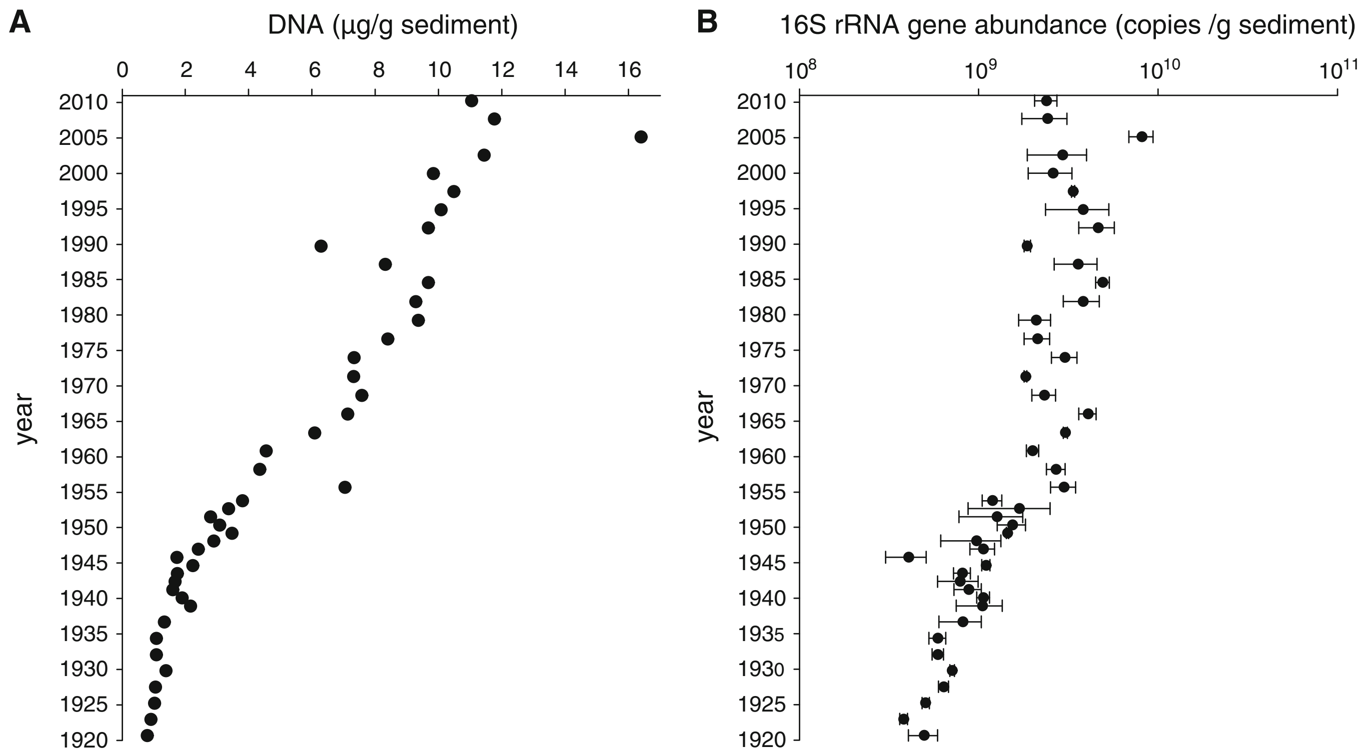


Fig. 2 **a** DNA content (µg g^{-1} sediment) with depth corresponding to calendar year timeline. **b** Distribution of 16S rRNA gene abundance (copies g^{-1} sediment) with depth. *Error bars* correspond to independent technical qPCR replicates

sediment at a depth of 99 cm. The decrease with depth can be fitted to an exponential decay ($r^2 = 0.56$, $p = <0.001$). The results for the 16S rRNA gene abundance from the core could be divided into three sections, with each showing a different trend. In the top section of the sediment core, there is no clear trend of decreasing or increasing gene abundance with depth, although the variation between samples is considerably higher than below. At 5 cm, coinciding with the high value of DNA yield, the highest 16S rRNA gene numbers of 8.2×10^{9} copies g^{-1} sediment were measured, which was substantially higher than all other values. In the middle section (between 15 and 47 cm), there was a trend toward an increasing abundance when moving upward, although there was some variation between samples, especially in the upper half of this section. From 47 cm downward, gene abundance was relatively constant, but one order of magnitude lower than those above.

Quantification of endospore-forming bacteria

Endospore-forming bacteria in the sediments were quantified by counting the copies of the *spo0A* gene. The pattern of *spo0A* gene abundance with depth (Fig. 3a) followed a different trend than the pattern from the 16S rRNA gene. Gene abundance ranged from 5.4×10^{3} copies g^{-1} sediment at the bottom of the sediment core to 1.9×10^{5} copies g^{-1} at the top. The highest value did not coincide with the sample that had the highest 16S rRNA gene abundance. There is a weak negative correlation between *spo0A* gene abundance and depth (regression $r^2 = 0.14$, $p = 0.01$). When splitting the core in the three sections mentioned above, the pattern looked different for each segment. In the deepest segment of the sediment core (50–106 cm), no overall trend could be seen. There were two samples with higher values than the others in this section (at 87-cm depth and in the area between 65 and 75 cm). In the middle section of the core (15–50 cm), there was high variation between the samples and, contrary to the 16S rRNA gene numbers, there was a slight decreasing trend when moving upward in the core. Finally, in the top section of the core (0–15 cm), the variation between samples was high, but overall a trend could be seen shifting toward increasing values at the top of the core. The highest value of the *spo0A* gene abundance was found at a depth of 3 cm.

The ratio between the 16S rRNA gene and the *spo0A* gene abundance was in the range of 0.7×10^{-3} % to 8×10^{-3} % and averaged 3.2×10^{-3} % (Fig. 3b). The ratios were significantly higher in the bottom section of the core (below 50 cm) than in the upper half (Student's *t* test; $p = <0.001$).

Determination of phylotypes of endospore-forming bacteria

A partial sequence of the *spo0A* gene from a selection of samples was amplified and subsequently sequenced. The samples were selected (1) to represent depths spanning through the entire core and (2) from interesting points where trends were shifting or values were particularly high or low. The sequences of *spo0A* were then used to determine the phylotypes and abundance of endospore-forming bacteria. After curation of the sequences, 5,144 sequences could be clustered into 552 OTUs. The results showed that despite that the quantity of extracted DNA decreased with the age of the sediment, the richness of endospore-forming

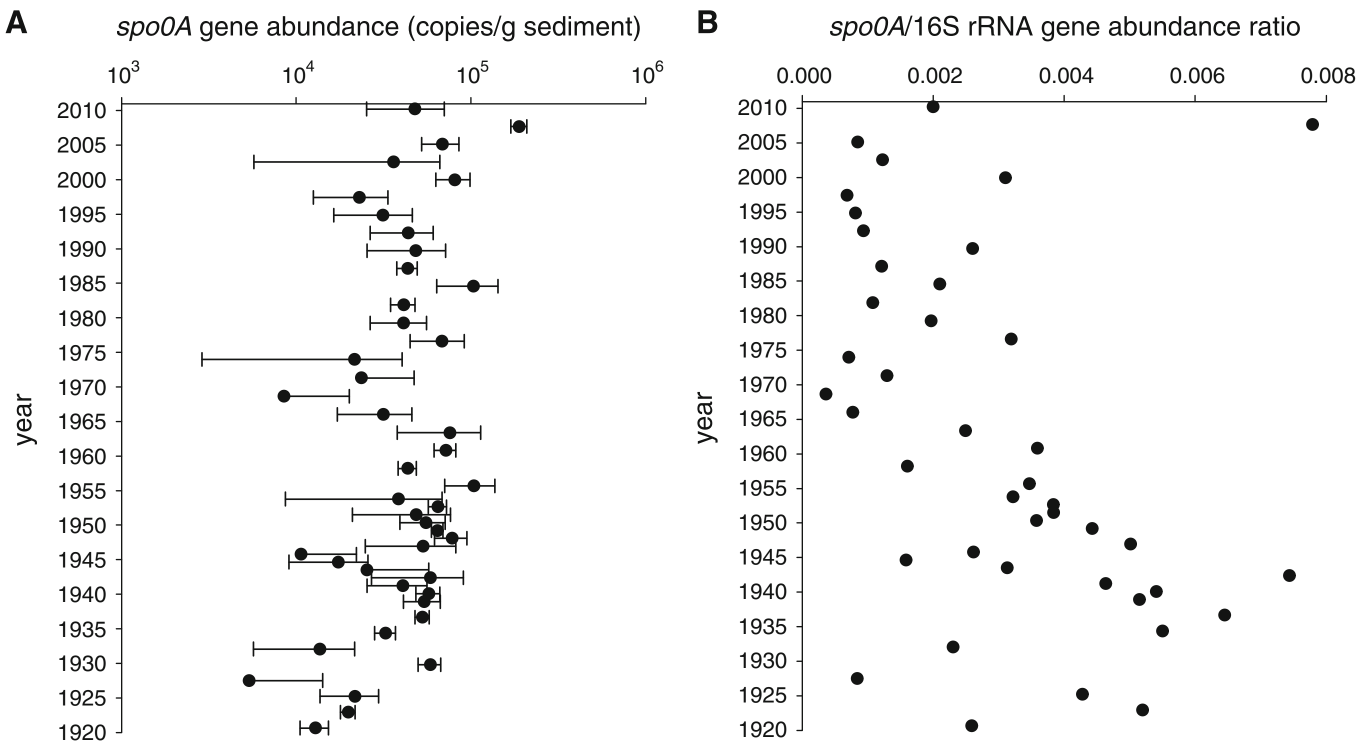


Fig. 3 **a** Depth distribution of *spo0A* gene abundance (copies g^{-1} sediment). *Error bars* correspond to independent technical qPCR replicates. **b** Ratio of 16S rRNA/*spo0A* gene abundance

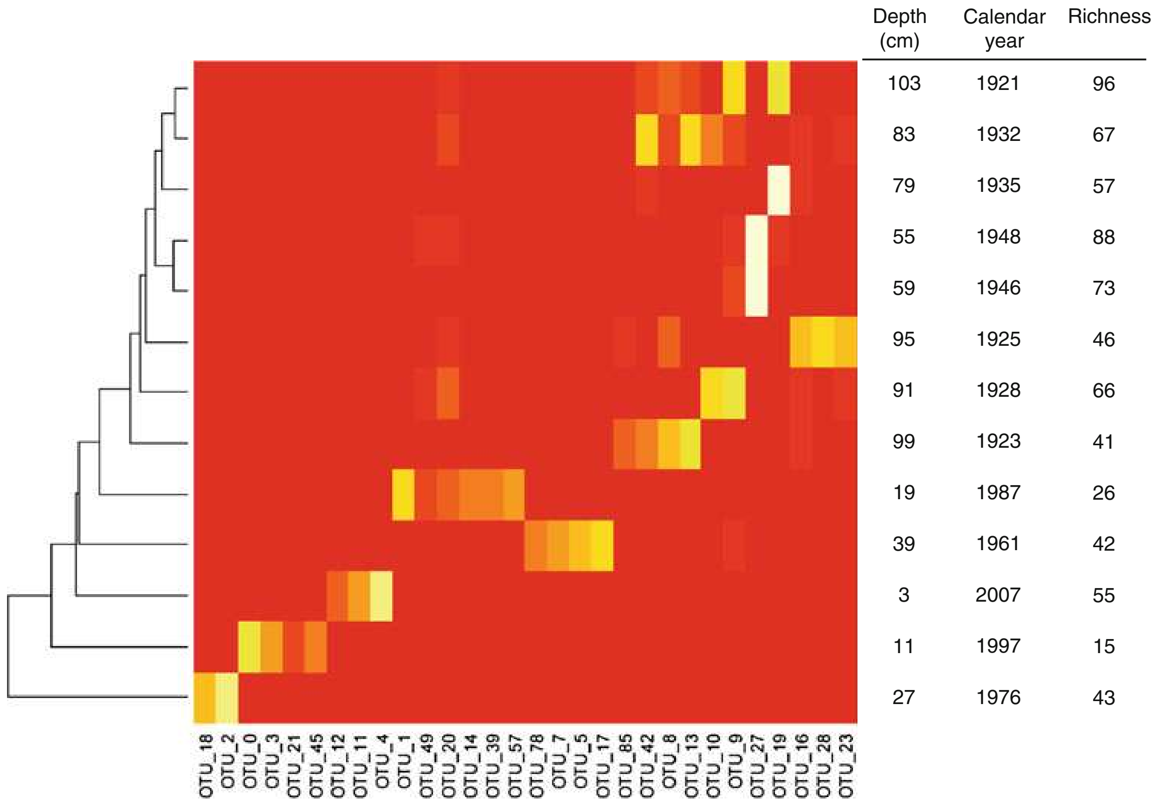


Fig. 4 Heatmap of the most dominant phylotypes obtained from *spo0A* sequences (>40 sequences per OTU) with clustering of samples from different depths according to similarities in community structure (OTU presence and abundance in the *vertical axis*). The frequency of OTUs is indicated by the color with *red shades* standing for no sequence and *yellow* indicating a high number of sequences. The depth, corresponding dating and number of OTUs (Richness) are shown on the *right* (color figure online)

bacteria did increase; richness was significantly less in the samples from depth 11, 19, 27 and 39 cm (*t* test, $p = 0.007$) with a mean of 31.5 OTUs per sample than in the other samples with a mean of 65.4 OTUs per sample. Richness ranged from 15 (at 11-cm depth) to 96 OTUs at 103-cm depth (Fig. 4). The majority (471) of the total identified phylotypes (554) only appeared as a single copy sequence. The heatmap (Fig. 4) shows the relationship of the most abundant phylotypes, with representatives that have >40 sequences per OTU. The lighter the color in the heatmap, the more sequences per OTU were found. The samples from the top of the core (1961–2007) clustered separately from the older half of the core (1921–1948) and have distinct OTUs that are only found at one depth. On the contrary, in the lower half of the core a number of domi-nant phylotypes were detected at several depths (OTUs Nr. 8, 9, 13, 20, 42) (Fig. 4).

To determine the community structure, OTUs were classified into genera on the basis of Spo0A protein sequence belonging to a known training dataset of 17 endospore-forming genera. This resulted in the classification of 552 OTUs into seven genera (*Alkaliphilus* 4 OTUs, *Bacillus* 76, *Brevibacillus* 5, *Clostridium* 104, *Desulfitobacterium* 12, *Desulfotomaculum* 7 and *Paenibacillus* 112). The remaining 232 OTUs could only be classified as far as the classes Bacilli or Clostridia (data not shown). The number of individual sequences per genus and samples are displayed in Fig. 5. Contrary to the trend of richness, the abundance values in the upper section of the core (years 1997, 1987, 1976 and 1961) with an average of 464 sequences per sample is significantly higher than in the rest of the samples with an average abundance of 363.3 sequences per sample (*t* test; $p = 0.01$). On the contrary, the number of genera present in those samples is significantly reduced (*t* test; $p = 0.008$). The mean number of genera for years 1997, 1987, 1976 and 1961 was 2.7, while other samples had on average 4.6 genera per sample. Between years 1961 and 1987, the reduction in richness coincided with an increase in the representation of anaerobic ecotypes represented by the genera *Clostridium* (years 1987 and 1976) and *Desulfitobacterium* (year 1961), and a reduction in the representation of *Paenibacillus*, in particular in 1976. In the year 1997, the reduction in richness was due to the dominance of the genus *Bacillus*.

Correlation between physico-chemical and microbiological parameters

The main changes in biological parameters could be seen in the core section of 15–47 cm, where also the chemical and physical parameters showed greatest variation.

Correlation of both types of parameters is shown as a biplot of a canonical correspondence analysis (CCA), based on the abundance dataset (containing all OTUs with more than five sequences) of all samples and the parame-ters Ca, K, Ti, Fe, Mn, Fe/Mn, depth and total P (Fig. 6). The community composition of endospore-forming bacte-ria from the samples of 1976, 1987, 1997 and 2007 shows divergence from the cluster of samples from the older part of the core (1921–1948). The direction of the divergence (particularly for sample 1976) is on the same axis as the significant contribution of Fe/Mn ($p = 0.01$), P ($p = 0.01$) and Ca ($p = 0.03$). The ordination of Ca is directly opposite of K, Ti, Fe and Mn. The direction of highest

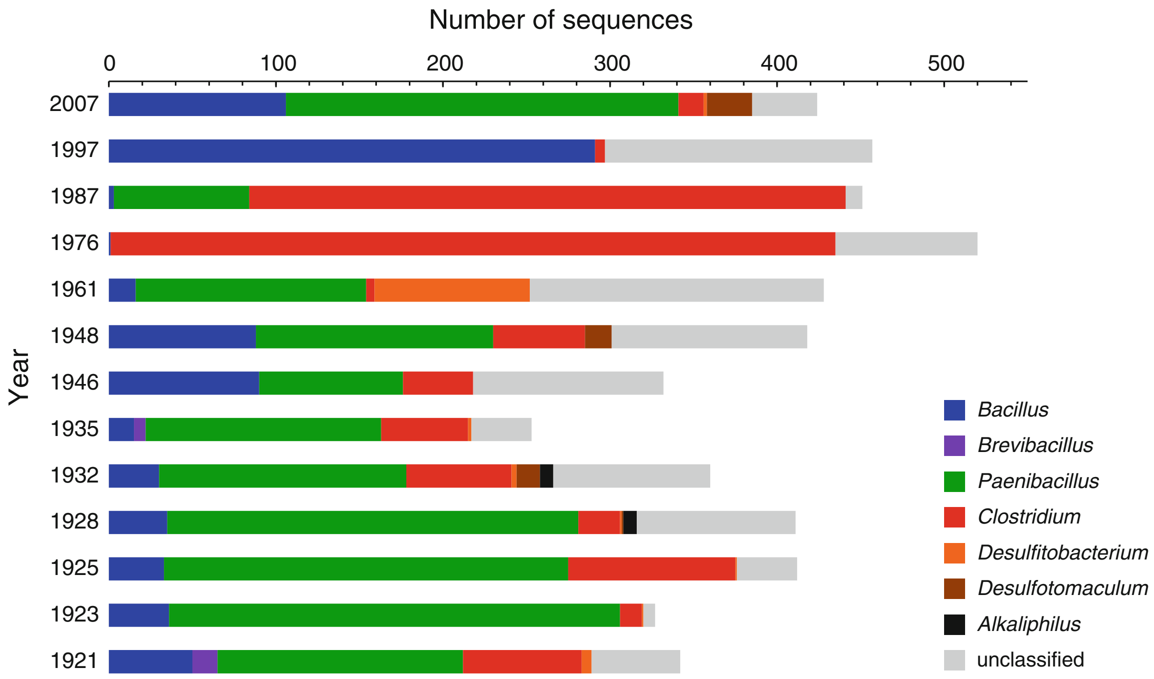


Fig. 5 Endospore-forming community structure represented by *spo0A* sequences and their genus affiliation at different depths. The length of the *bar* corresponds to total sequence abundance. Different genera are depicted in different *colors*. *Gray bars* are sequences that could not be classified to genus level (unclassified) but corresponded to Bacilli or Clostridia (color figure online)

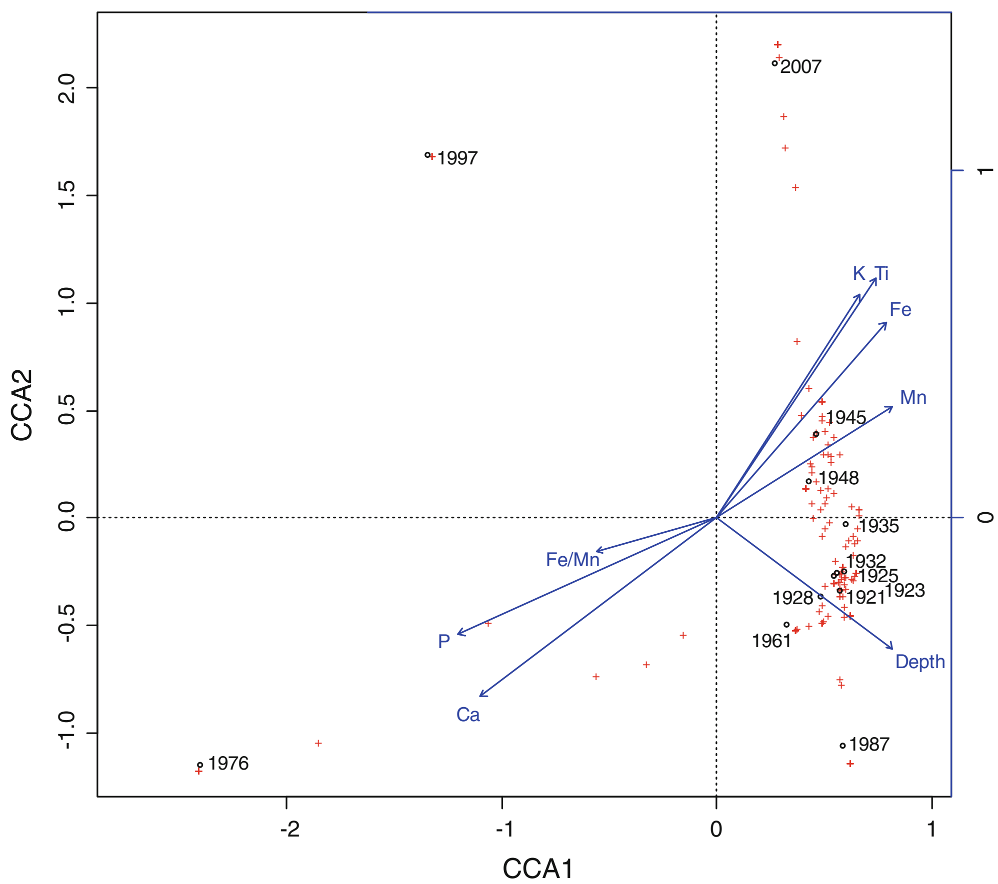


Fig. 6 Canonical correspondence analysis (CCA) of the community matrix with all OTUs composed of a minimum of five sequences and the environmental matrix with values of depth, total P, Ca, K, Ti, Fe, Mn and Fe/Mn as constraining parameters. The total P (μg L^{-1}) values are average yearly values from the Grand Lac (CIPEL, 2012). *X*-axis explained 15.3 % and *Y*-axis explained 15.0 % of importance. The positions of samples are depicted with *black circles* and the positions of individual OTUs with a *red plus* (color figure online)

variance (tip of the arrow) of the ratio of Fe/Mn is close as for total P values, suggesting a strong correlation between these two parameters.

Discussion

From a paleoecological point of view, the sedimentary record studied here is relatively recent, with the bottom of the core being assigned to AD 1920. Nonetheless, the results show that nucleic acids in the sediment decrease exponentially with time, so does the number of bacteria assessed as 16S rRNA gene abundance, which is significantly reduced with depth. Decrease of DNA with sediment depth is more prominent than the decrease in 16S rRNA gene abundance, which can be explained by the contribution of eukaryotic DNA to the total yield. Quantification of bacteria cannot directly be correlated with DNA yields in this sediment core. However, at the bottom of the core, there are still 3.8×10^8 copies per gram sediment of the 16S rRNA gene detectable. Considering that bacteria can house up to 15 copies of the 16S rRNA gene per cell

(Klappenbach et al. 2000), this accounts for 25 × 10^6 cells per gram sediment of an age of about 90 years. However, sediments with such high numbers of bacterial cells most likely house a large diversity of species and the communities are too complex to analyze for trends over time. Neither DNA nor molecular quantification and community assessment of total bacteria can therefore serve as ecological proxies in this sediment core.

Endospores have been considered as ideal candidates to survive in the sediment for long periods of time and to potentially serve as paleolimnological proxies (Renberg and Nilsson 1992). Endospores are very resistant structures, built with the purpose of protecting its enclosed DNA from degradation (Nicholson et al. 2000). In this study, we confirm this hypothesis with the constant if not increasing quantification of the *spo0A* gene abundance with sediment depth. Indeed, the decreasing trend observed in DNA and 16S rRNA gene abundance with depth of the sediment core is independent of the *spo0A* gene abundance determined throughout the sediment core, as seen by the ratio of *spo0A* and 16S rRNA gene increasing with depth (Fig. 3b). The ratio between *spo0A* and 16S rRNA genes is low with an average of 0.003 %. By assuming an average of 15 16S rRNA gene copies per cell (Klappenbach et al. 2000), roughly 1 in 2,000 bacterial cells is an endospore-forming bacteria. Targeting such a small fraction of the bacterial community results in reduced complexity and higher resolution. Because of low complexity and the stability of DNA retrievable from this group of bacteria over time, the endospore-forming bacterial community is a promising proxy for paleoecology.

The composition of the endospore-forming community at different depths of the sediment reflect prominent shifts in the community in the years between 1961 and 1997. These shifts correspond to reduced species richness, while total abundance of endospore-forming bacteria increased during the same time period. In samples from 1976 to 1987, the members from the genus *Clostridium* are most abundant, and in the sample from 1997 members from *Bacillus* are most abundant. The classification of the OTUs into these groups is important as it can be linked to some of the ecological features known for cultured species, espe-cially concerning aerobic or anaerobic growth conditions. The genera *Clostridium* (Schleifer 2009), *Desulfotomacu-lum* (Stackebrandt et al. 1997) and *Desulfitobacterium* (Utkin et al. 1994) are strict anaerobic heterotrophic bac-teria. Bacteria from genera *Bacillus* (Schleifer 2009) and *Brevibacillus* (Shida et al. 1996), on the contrary, grow aerobically, while *Paenibacillus* (Shida et al. 1997) i s composed of members that grow as facultative anaerobes. Therefore, shifts in community composition for endospore-forming bacteria from 1961 to 1987 appear to be linked to decrease availability of oxygen in the sediment or at its surface at the moment of sedimentation.

Three issues should be considered for the interpretation of these results. The first one is the specificity of the *spo0A* as molecular marker to target endospore-forming bacteria. The second is the origin of endospore-forming bacteria (i.e., autochthonous versus allochthonous) and the extent to which endospore-forming communities reflect the environmental conditions at the moment of sedimentation. The third issue is the highly dynamic nature of sediments and the potential changes in the paleoecological signature that will result from in situ microbial activity.

Specificity of the *spo0A* gene in an important issue. This gene is conserved among all endospore-forming species known so far (Galperin et al. 2012; Abecasis et al. 2013; Traag et al. 2013), and is notably absent from exospore-forming groups such as *Streptomyces* spp. and Mycobacteria (Abecasis et al. 2013), as well as in the developmental cascade at the origin of the fruiting bodies in *Myxococcus* (Kroos 2007). A previous study has shown similar sequences to *spo0A* being present in nonspore-forming species, even though at low similarity scores (Onyenwoke et al. 2004). However, a more recent profile analysis of *spo0A* in 626 genomes has found only one putative ortholog in a non-endospore-forming genome (Traag et al. 2013). This suggests that *spo0A* can be considered as a specific molecular marker for endospore-forming bacteria. In addition, all of our attempts to amplify nonspore-form-ing strains with our *spo0A* primer have been negative (data not shown). In addition, although not all the OTUs could be assigned to the genus level because of a lack of references, a significant score and *e* value for *spo0A* were obtained for all the curated sequences, and a classification into the classes Bacilli or Clostridia was achieved.

Regarding the origin of the endospore-forming bacteria, previous studies based on culturing of specific endospore-forming species from lake or marine sediments (Bartholomew and Paik 1966; Robles et al. 2000; Hubert et al. 2010) have suggested that endospores are in most cases allochthonous and have been deposited at the time of sedimentation followed by little in situ activity. In the present study, this might not be the case. The high C/N ratio shown in Fig. 1, with values close to 19, underlines a mixed origin of terrestrial and aquatic organic matter (Meyers 2003). However, the fact that DNA yields are high at the same depth that have peaks in C/N ratio (13 and 43 cm) imply terrestrial input of organisms at these same time points, potentially due to floods. The quantity of endospore formers at these depths does not correlate with the C/N ratio. This means that likely the majority of the endospore-forming community in the sediments is from an autochthonous source and is less influenced by transport from rivers. An example for this could be members from *Paenibacillus*, facultative anaerobes known to hydrolyze complex carbohydrates such as chitin (Shida et al. 1997). Both properties of *Paenibacillus* indicate potential advantages to actively thrive at the sediment surface, which is confirmed by their dominance in the majority of samples. The difference between previous reports and the results obtained in the present study could be due to the biases introduced by culturing, which are omitted in the molecular approach taken here since DNA can also be recovered from non-viable spores of both, autochthonous and allochthonous origin.

With respect to the influence of the in situ activity in the paleoecological signature of the microbial community, it is not clear if endospore-forming bacteria grow actively on the lake bottom or in the sediment. From the TOC values (top sediment at 1.23 %) that stay stable through the core, it can be implied that the available carbon is quickly turned over and that in the sediment the carbon available for heterotrophic growth is very small. Also, a recent study from Lake Geneva has shown ATP to strongly decrease in the top few centimeters of the sediment to almost zero at 10 cm, suggesting little microbial activity below this depth (Thevenon et al. 2011). When conditions get oligotrophic and in particular if the carbon is low, endosporulation sets in due to starvation (Hageman et al. 1984). Based on this, our current interpretation is that burial of endospores happens for the most part at this stage.

A way to distinguish between active and inactive members of the community could be the differentiation between vegetative cells and endospores. However, the current techniques to quantify endospores [e.g., dipicolinic acid content as in (Fichtel et al. 2007)] are destructive and thus identification cannot be coupled to quantification. Further studies including differentiation between cells and endospores and a characterization of the metabolic properties of these endospore formers could give additional insights into this issue.

Signals of ecosystem changes in the period between 1960 and 1990 are also reflected in some chemical and physical parameters of the sediment cores in this study. The sediment cores are composed of silt-sized hemipelagic sediment, suggesting no direct influence by the Rhone River hyperpycnal flows. But as most of the particle input to Lake Geneva, including clay- and silt-sized minerals, is due to the Rhone River and as the general Lake Geneva circulation deflects stratified inflows to the right, the overall sediment signature at the sampling site mostly reflects changes in the lake's catchment. However, authigenic calcite production is mainly influenced by the lake's productivity, which adds an in situ signature to the sediment record. The results of the elemental analysis in the sediment core (CAN02) of this study therefore mainly reflect changes in runoff, autogenic production in the water column or changes at the sediment–water interface.

K and Ti are known to be linked to allochtonous input by runoff. As K and Ti counts follow an inverse trend to grain size, they seem to be related to fine sediment, as seen in Corella et al. (2011). Ca counts profile shows an opposite trend to K and Ti, and is thus interpreted as mainly due to endogenic productivity and larger-sized particles. Even though Ca and the published values for total P (measured at the center of the lake over 20 different depths) are linked, when looking at our CCA it is difficult to disentangle endogenic Ca precipitation due to increased primary production from the inputs due to changes in runoff. However, recent studies show the link between photosynthetic microorganisms and the formation of low-Mg calcite in freshwater (Plee et al. 2008; Pacton et al. 2012), favoring the hypothesis of a link between Ca and eutrophication. Fe and Mn depth profiles are directly linked to K and Ti counts in the CCA biplot; they are therefore also inter-preted as influenced by allochthonous inputs. Neither of the above single elements can therefore clearly be linked to changes solely due to lake eutrophication.

On the contrary, the Fe/Mn ratio, frequently used as an indicator of the redox condition in lake hypolimnia (Koinig et al. 2003; Corella et al. 2012), point to changes linked to eutrophication. An increase in Fe/Mn ratio may reflect hypoxia at the sediment water interface due to preferential resolubilization of Mn over Fe linked to differences in redox kinetics. Mn is more easily reduced and transported away from the sediment, therefore depleting the Mn in respect to the residual Fe content in the sediment (Schaller and Wehrli 1996). At site CAN02, the increase in the Fe/Mn ratio since 1960 (Fig. 1) can be associated with hypoxia at the sediment water interface. A decrease in the oxygen content at lake bottom conditions has also been connected to lower C/N ratio and higher TOC content due to higher primary production in the water column (Corella et al. 2011). These three parameters have significantly changed since 1960 in this sediment record, coinciding with the shift in the structure of anaerobic endospore-forming bacterial communities and the dominance of anaerobic genera such as *Clostridium*.

If the changes in the community structure of endospore-forming bacteria and the redox proxies in the sediment during the years 1961–1997 are indeed a record of the variation in oxygen availability, evidence for this environmental fluctuation should also be found in the history of this lake. In Lake Geneva, in the last 100 years, anthropogenic pressures such as the release of nutrients via wastewater or agricultural runoff have had an important impact on the ecosystem. The lake has been closely monitored since 1957 and data are made publicly available by the 'Commission Internationale de la Protection des Eaux du Le´man' (CIPEL) in the form of yearly published reports (http://www.cipel.org), with the most recent one released in 2012 (Lazzarotto and Klein 2012). Long-term trends show a steady increase of total phosphorus since 1957 with a peak in 1979. These values, together with phosphate data since 1970, indicate a shift in the trophic status of the lake from oligotrophic to eutrophic taking place in the late 1960s. The system has since recovered, even though total phosphorus levels are still double the values before 1960. Eutrophication of Lake Geneva is one of the environmental disturbances with the best ecological record (Anneville and Pelletier 2000; Gerdeaux and Perga 2006; Thevenon et al. 2012). High nutrient levels increased primary production, which together with warm winters and successive incomplete mixing of the lake, resulted in low oxygen levels in deeper waters. Long-term lake bottom hypoxia (delimited as <4 mg O_2 L^{-1}) have been registered in the deepest part of the lake (300 m) in the years from 1973 to 1978 and 1986 to 1998, by CIPEL (Lazzarotto and Klein 2012). Important changes have been seen in the composition of pelagic primary and secondary producers as a consequence of higher phosphorus concentrations as well as a warming climate (Molinero et al. 2006). Changes at higher trophic levels had also occurred such as the extinction of whitefish (Vonlanthen et al. 2012).

Interestingly, the most recent part of the record showed a trend toward the recovery of endospore-forming communities, although indicating a delay between the decrease of water nutrients and the response of the community. The community composition in the most recent sample (2007) is changing back to a similar diversity that was found in samples dating from 1928 to 1932, demonstrating an intrinsic resilience of the system for these bacteria. This is demonstrated by eight OTUs that are shared between the samples from 2007 and samples prior to 1948. In samples from the years 1976, 1987 and 1997, none, three and one OTU, respectively, are shared with samples from before 1948. This result is interestingly opposite to *Daphnia* population evolution in unproductive Swiss lakes, where eutrophication led to partly irreversible species changes (Rellstab et al. 2011).

In conclusion, the research presented here breaks new ground for the use of specific groups of bacteria as proxies for changes found in sedimentary records. In the past, other microbial groups have been used in freshwater paleoecology. A good example of this is the study of siliceous microfossils from diatoms, which are preserved in sediments as amorphous biogenic silica (Hobbs et al. 2010). However, the dissolution of diatoms within lake sediments may compromise the interpretation of the sediment record (Ryves et al. 2006). In this sense, bacterial endospores have the advantage of being biological structures with the specific role of resisting environmental stress for long time periods (Nicholson et al. 2000), and therefore can be expected to remain unaltered within the sediment record. In addition, nowadays, with new molecular techniques and the possibility to directly sequencing metagenomes from environmental samples, the discovery and use of even older endospores as paleoecological proxies is plausible, since viability will no longer be a major issue. Although issues related to the dynamic nature of sediments and the origin and in situ activity of endospore-forming bacteria need to be studied further, this study is a proof of concept that endospore-forming community reflects changes in oxygen conditions in this lake.

Acknowledgments This publication is part of the international, interdisciplinary research project ELEMO (http://www.elemo.ch) to investigate the deep waters of Lake Geneva using two Russian MIR submarines. Funding for this study was provided by the ''Fondation pour l'Etude des Eaux du Le´man'' (FEEL; in particular by the Ferring Pharmaceuticals St Prex). Additional funding for the work described in this paper was provided by Swiss National Science Foundation grant No. 31003A-132358/1. We are grateful for the support. We thank the Russian MIR crew members (www.elemo.ch/mir-team) for their excellent performance and the SAGRAVE team who provided and operated the platform from which the dives were carried out. We also thank Ulrich Lemmin and Jean-Denis Bourquin for project coordination and Samuel Arey for dive planning. The service of Mikhail Kranoperov (Russian Honorary Consulate) as liaison is greatly appreciated. The authors acknowledge Philippe Arpagaus and Angel Arantegui for their support during the fieldwork and laboratory analyses. The authors thank the two anonymous reviewers for their valuable comments for improving this manuscript.

References

Abecasis AB, Serrano M, Alves R, Quintais L, Pereira-Leal JB, Henriques AO (2013) A genomic signature and the identification of new sporulation genes. J Bacteriol 195(9):2101–2115

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59(1):143–169

Anneville O, Pelletier JP (2000) Recovery of Lake Geneva from eutrophication: quantitative response of phytoplankton. Archiv für hydrobiologie 148(4):607–624

Bartholomew JW, Paik G (1966) Isolation and identification of obligate thermophilic sporeforming bacilli from ocean basin cores. J Bacteriol 92(3):635–638

Bueche M, Wunderlin T, Roussel-Delif L, Junier T, Sauvain L, Jeanneret N et al (2013) Quantification of endospore-forming Firmicutes by quantitative PCR with the functional dene spo0A. Appl Environ Microbiol 79(17):5302–5312

Corella JP, Amrani A, Sigro J, Morellon M, Rico E, Valero-Garces B (2011) Recent evolution of Lake Arreo, northern Spain: influences of land use change and climate. J Paleolimnol 46(3):469–485

Corella JP, Brauer A, Mangili C, Rull V, Vegas-Villarrubia T, Morellon M et al (2012) The 1.5-ka varved record of Lake Montcortès (southern Pyrenees, NE Spain). Quatern Res 78(2):323–332

de Rezende JR, Kjeldsen KU, Hubert CR, Finster K, Loy A, Jorgensen BB (2013) Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. ISME J 7(1):72–84

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19):2460–2461

Fichtel J, Koster J, Rullkotter J, Sass H (2007) Spore dipicolinic acid contents used for estimating the number of endospores in sediments. FEMS Microbiol Ecol 61(3):522–532

Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ (2012) Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. Environ Microbiol 14(11):2870–2890

Gerdeaux D, Perga ME (2006) Changes in whitefish scales 13C during eutrophication and reoligotrophication of subalpine lakes. Limnolo Oceanogr 51:772–780

Gorham E, Brush GS, Graumlich LJ, Rosenzweig ML, Johnson AH (2001) The value of paleoecology as an aid to monitoring ecosystems and landscapes, chiefly with reference to North America. Environ Rev 9(2):99–126

Hageman JH, Shankweiler GW, Wall PR, Franich K, McCowan GW, Cauble SM et al (1984) Single, chemically defined sporulation medium for *Bacillus subtilis*: growth, sporulation, and extracellular protease production. J Bacteriol 160(1):438–441

Hobbs WO, Lalonde SV, Vinebrooke RD, Konhauser KO, Weidman RP, Graham MD et al (2010) Algal-silica cycling and pigment diagenesis in recent alpine lake sediments: mechanisms and paleoecological implications. J Paleolimnol 44(2):613–628

Hubert C, Arnosti C, Bruchert V, Loy A, Vandieken V, Jorgensen BB (2010) Thermophilic anaerobes in Arctic marine sediments induced to mineralize complex organic matter at high temperature. Environ Microbiol 12(4):1089–1104

Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K et al (2005) The EMBL nucleotide sequence database. Nucleic Acids Res 33:D29–D33

Kindt R, Coe R (2005) Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre Eastern and Central Africa Program

Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microbiol 66(4):1328–1333

Koinig A, Shotyk W, Lotter A, Ohlendorf C, Sturm M (2003) 9000 years of geochemical evolution of lithogenic major and trace elements 75 in the sediment of an alpine lake: the role of climate, vegetation, and landuse. J Paleolimnol 30(3):307–320

Kroos L (2007) The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. Annu Rev Genet 41:13–39

Lazzarotto J, Klein A (2012) Physical–chemical changes in the waters of Lake Geneva (major elements). Rapport de la Commission Internationale pour la Protection des Eaux du Léman contre la Pollution; Campagne 2011. Lausanne, Switzerland

Loizeau J-L (1991) La sédimentation récente dans le delta du Rhône, Léman: processus et évolution. Université de Genève, Geneva

Loizeau J-L, Dominik J, Luzzi T, Vernet J-P (1997) Sediment core correlation and mapping of sediment accumulation rates in Lake Geneva (Switzerland, France) using volume magnetic susceptibility. J Great Lakes Res 23(4):391–402

Lomstein BA, Langerhuus AT, D'Hondt S, Jorgensen BB, Spivack AJ (2012) Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature 484(7392):101–104

Meyers PA (2003) Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. Org Geochem 34(2):261–289

Molinero JC, Anneville O, Souissi S, Balvay G, Gerdeaux D (2006) Anthropogenic and climate forcing on the long-term changes of planktonic rotifers in Lake Geneva, Europe. J Plant Res 28(3):287–296

Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P (2000) Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol Mol Biol Rev 64(3):548–572

Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M (2007) The vegan package. Community Ecol Package 10(01):2008

Onyenwoke RU, Brill JA, Farahi K, Wiegel J (2004) Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). Arch Microbiol 182(2–3):182–192

Ovreås L, Forney L, Daae FL, Torsvik V (1997) Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl Environ Microbiol 63(9):3367–3373

Pacton M, Ariztegui D, Wacey D, Kilburn MR, Rollion-Bard C, Farah R et al (2012) Going nano: a new step toward understanding the processes governing freshwater ooid formation. Geology 40(6):547–550

Plee K, Ariztegui D, Martini R, Davaud E (2008) Unravelling the microbial role in ooid formation–results of an in situ experiment in modern freshwater Lake Geneva in Switzerland. Geobiology 6(4):341–350

R TC (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rellstab C, Keller B, Girardclos S, Anselmetti FS, Spaak P (2011) Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing *Daphnia* in unproductive lakes. Limnol Oceanogr 56(1):292–302

Renberg I, Nilsson M (1992) Dormant bacteria in lake sediments as paleoecological indicators. J Paleolimnol 7(2):125–135

Robles S, Rodriguez JM, Granados I, Guerrero MC (2000) Sulfite-reducing clostridia in the sediment of a high mountain lake (Laguna Grande, Gredos, Spain) as indicators of fecal pollution. Int Microbiol 3(3):187–191

Rothfuss F, Bender M, Conrad R (1997) Survival and activity of bacteria in a deep, aged lake sediment (Lake Constance). Microb Ecol 33(1):69–77

Ryves DB, Battarbee RW, Juggins S, Fritz SC, Anderson NJ (2006) Physical and chemical predictors of diatom dissolution in freshwater and saline lake sediments in North America and West Greenland. Limnol Oceanogr 51:1355–1368

Sastre V, Loizeau J-L, Greinert J, Naudts L, Arpagus P, Anselmett F et al (2010) Morphology and recent history of the Rhone River Delta in Lake Geneva (Switzerland). Swiss J Geosci 103(1):33–42

Schaller T, Wehrli B (1996) Geochemical-focusing of manganese in lake sediments: an indicator of deep-water oxygen conditions. Aquat Geochem 2(4):359–378

Schleifer KH (2009) Phylum XIII. Firmicutes Gibbons and Murray 1978, 5 (Firmacutes [sic] Gibbons and Murray 1978, 5). In: Vos P, Garrity G, Jones D et al (eds) Bergey's Manual® of systematic bacteriology. Springer, New York, pp 19–1317

Seaward MR, Cross T, Unsworth BA (1976) Viable bacterial spores recovered from an archaeological excavation. Nature 261(5559): 407–408

Shida O, Takagi H, Kadowaki K, Komagata K (1996) Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. Int J Syst Bacteriol 46(4):939–946

Shida O, Takagi H, Kadowaki K, Nakamura LK, Komagata K (1997) Transfer of *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*. Int J Syst Bacteriol 47(2):289–298

Sneath PH (1962) Longevity of micro-organisms. Nature 195: 643–646

Stackebrandt E, Sproer C, Rainey FA, Burghardt J, Pauker O, Hippe H (1997) Phylogenetic analysis of the genus *Desulfotomaculum*: evidence for the misclassification of *Desulfotomaculum guttoideum* and description of *Desulfotomaculum orientis* as *Desulfosporosinus orientis* gen. nov., comb. nov. Int J Syst Bacteriol 47(4):1134–1139

Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol 39:321–346

Thevenon F, Graham ND, Herbez A, Wildi W, Pote J (2011) Spatio-temporal distribution of organic and inorganic pollutants from Lake Geneva (Switzerland) reveals strong interacting effects of sewage treatment plant and eutrophication on microbial abundance. Chemosphere 84(5):609–617

Thevenon F, Adatte T, Wildi W, Pote J (2012) Antibiotic resistant bacteria/genes dissemination in lacustrine sediments highly increased following cultural eutrophication of Lake Geneva (Switzerland). Chemosphere 86(5):468–476

Traag BA, Pugliese A, Eisen JA, Losick R (2013) Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. J Bacteriol 195(2):253–260

Utkin I, Woese C, Wiegel J (1994) Isolation and characterization of *Desulfitobacterium dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. Int J Syst Bacteriol 44(4):612–619

Vonlanthen P, Bittner D, Hudson AG, Young KA, Muller R, Lundsgaard-Hansen B et al (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. Nature 482(7385):357–362

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73(16):5261–5267

Willard DA, Cronin TM (2007) Paleoecology and ecosystem restoration: case studies from Chesapeake Bay and the Florida Everglades. Front Ecol Environ 5(9):491–498

Wunderlin T, Junier T, Roussel-Delif L, Jeanneret N, Junier P (2013) Stage 0 sporulation gene A (*spo0A*) as a molecular marker to study diversity of endospore-forming Firmicutes. Environ Microbiol Rep 5(6):911–924

Paper VI

In which endospores are used as a proxy for environmental history of Lake Geneva...

Endospores in sediment as a proxy for 20th century environmental lake history

Tina Wunderlin[1], Thomas Junier[1], Matthieu Bueche[1], Juan Pablo Corella[2,3], Jean-Luc Loizeau[4], Stéphanie Girardclos[2] and Pilar Junier[1*]

[1]Laboratory of Microbiology, University of Neuchâtel, 2000 Neuchâtel (Switzerland)

[2]Institute for Environmental Sciences, Earth and Environmental Sciences section, University of Geneva, 1205 Geneva (Switzerland)

[3]Museo Nacional de Ciencias Naturales (MNCN/CSIC), Serrano 115bis, 28006 Madrid (Spain)

[4]Institute F. A. Forel, Earth and Environmental Sciences section, Faculty of sciences, University of Geneva, 1205 Geneva (Switzerland)

[*] Corresponding author. Mailing address: Laboratory of Microbiology, Institute of Biology, University of Neuchâtel, CH - 2000 Neuchâtel

Phone: +4132 7182244

Fax: +4132 7182231

E-mail: pilar.junier@unine.ch

co-authors e-mail:

tina.wunderlin@unine.ch	+41 327182313
thomas.junier@unine.ch	+41 32 718 2244
matthieu.bueche@unine.ch	+41 32 718 2242
pablo.corella@mncn.csic.es	+34 917452500 ext. 980603
Jean-Luc.Loizeau@unige.ch	+41 22 379 0319
Stephanie.Girardclos@unige.ch,	+41 22 379 0786 or +41 22 379 6617

Running title: Bacterial endospores as paleolimonological indicators

Key words: paleolimnology, sediment, activity, diversity,

Abstract

In this study we show that bacterial endospores buried in the sediment can be used as natural time capsules to unveil lake history. We report that variations in community structure, determined by pyrosequencing, can be linked to lake eutrophication, sulphate metabolism, input of terrestrial organic matter, and climatic events. A treatment for the separation of endospores from vegetative cells and the sequencing of the endospore fraction allows to differentiate the effect of disturbance on active versus dormant populations. This additional information is key to define the "seed bank" component of the microbial pool and to demonstrate its importance in response to changes in environmental conditions. From the analysis, endospore-formers emerge as a microbial group that can be used as a paleoecological proxy for ecosystem prospection in aquatic systems.

Freshwater lakes are heavily subjected to anthropogenic pressure, threatening water supply and the health of aquatic ecosystems all over the world[1, 2]. This calls for measures for ecosystem management of freshwater bodies to be implemented[3]. However, an inventory of conditions prior to disturbance (i.e. reference conditions) and in response to it, are required. Lake sediments provide an ideal environmental archive of these former reference conditions[4]. Sediment cores have been analyzed in paleoenvironmental reconstructions to understand the relationship between environmental disturbance and its impact on the ecology of aquatic ecosystems[5]. In paleolimnological studies the range of natural fluctuations, anthropogenic pressure and resilience of the aquatic system are inferred from the combined record of physicochemical parameters and biological indicators[6]. Typically, the latter correspond to organisms producing identifiable fossilized structures (e.g. pollen grains or siliceous microfossils) that must remain unaltered in the sediments to be analysed at different time-scales[4]. The potential of bacteria as paleolimnological indicators is high considering their abundance and diversity (both phylogenetic and metabolic)[7]. However, the comprehensive detection of bacterial communities in a sedimentary record is challenging, especially because of the degradation of DNA with age[8]. Therefore most paleolimnological studies targeting bacteria have been performed using fossil pigments[9, 10]. The use of bacterial resting states, and in particular of endospores, is an alternative[11]. Bacterial endospores are highly resistant cellular forms produced by certain genera of the phylum Firmicutes (Gram-positive low G+C content bacteria), able to survive in a dormant state for long periods of time with little to no *in situ* activity[12]. Biochemical and genetic data from revived endospores buried in sediments has been used to reconstruct past environmental conditions[13, 11, 14]. This demonstrated, for example, the dispersal of metabolically inactive thermophiles in cold marine sediments[13, 15, 16]. However, culturing studies are notoriously biased[17, 18], and therefore culture-independent alternatives have been developed. In one of such studies, the quantification of dipicolinic acid (an endospore-specific biomarker) suggested that endospores make up to 3% of the total prokaryotic community in sediments of the North Sea[19]. In much older (deeper) sediment cores, the abundance of endospores has been estimated to be as high as the total abundance of vegetative cells[20]. Recently, an alternative culture-independent molecular approach based on the amplification and sequencing of the *spo0A* gene, found in all and only in endospore-formers[21, 22], has been developed[23, 24]. Using this method the applicability of endospore-forming bacteria as proxies to lake eutrophication was shown[25].

Here we analyzed the variations in the structure of endospore-forming bacterial communities in a lake sediment core and establish that changes in the community structure can be used as a proxy for the lake history. For this, Lake Geneva was selected as case study, as it is a prime example of the impact of human activities on freshwater ecosystem health. This peri-alpine lake is the largest lake in Western Europe and its surroundings are heavily populated, with three major cities relying on the lake for drinking water, fishery, and recreational activities. Due to increasing anthropogenic pressure, Lake Geneva has experienced an eutrophication period, starting in the early 1960s, followed by re-oligotrophication after the implementation of measures to decrease nutrient input[26].

We have analyzed a sediment core of 103 cm length from Lake Geneva (Supplementary Figure 1), for which a detailed sedimentological description is available[25]. Biological samples for DNA extraction and analysis of endospore-forming bacterial communities were taken every four cm. Community structure was determined by sequencing of the *spo0A* gene[24]. The methodological approach is analogous to the one used for the 16S rRNA gene, commonly applied to study bacterial communities[27]. However, the advantage of focusing on this specific bacterial group is the vast amount of metabolic information available thanks to a large collection of culturable representatives[28].

Results and Discussion

The community structure based on OTUs affiliated to 20 endospore-forming genera is shown in Figure 1 (total and percentage values are provided in Supplementary Table 1). Some OTUs (maximum 8 %) could not be classified to genus, but to family level or remained as unclassified bacteria. The major genera were *Paenibacillus* (over all samples 45.7 %), *Clostridium* (25.5 %) and *Bacillus* (9.4 %). Based on the community structure (Figure 1), the sediment core was divided into four sections. Section I comprises the lower half of the core (103-55 cm) with presence of three major genera *Bacillus* (1.2-23.9%), *Paenibacillus* (19.4-84.8%), and *Clostridium* (3.0-29.7%), as well as a minor contribution of groups like *Desmospora* (0-22.5%), *Sporomusa* (0-39.7%), and *Moorella* (0-6.8%). The Simpson index in these samples indicates relatively high diversity with a mean of 0.56, and a range from 0.27 (99 cm) to 0.74 (59 cm) (Supplementary Table 2). The exception to the constant community pattern was the sample from 87 cm depth, where the relative abundance of OTUs affiliated to *Bacillus* (1.2%) was reduced, and instead other genera such as *Sporomusa* (6.8%) and *Moorella* (39.7%) increased. A clear change in diversity at 51 cm depth (Simpson index changing from 0.72 to

0.31) and a shift in community composition marked the beginning of Section II (51-39 cm). In this section, *Paenibacillus* is the most dominant genus with up to 82 % relative abundance in the sample at 51 cm depth (onset of Section II), while the representatives from *Bacillus* have entirely disappeared (<0.1%). The communities in sediments from Section II fluctuated importantly as seen in the variation of the Simpson index. The variation was marked by the alternated dominance of *Clostridium* (47 cm), *Paenibacillus* (39 and 51 cm), or *Desulfotomaculum* (43 cm). The high abundance of *Desulfotomaculum* and *Moorella* in three samples of this section (39, 43, 47 cm) is surprising, considering that these genera are only minor community members in all other samples. In Section III (35-11 cm) the diversity was generally low (average Simpson index of 0.34) and *Paenibacillus* and *Clostridium* clearly dominated the community. In the extreme case (sample from 27 cm depth), the Simpson index was 0.01 and the community was composed of 99.5 % *Clostridium*. The community at 11 cm depth was peculiar because of the dominance of *Brevibacillus*, which was not common at other depths. In the last Section (IV) with samples from 7 cm to 1 cm depth, the community was again more diverse (Simpson index 0.56 and 0.76) and the percentage of *Bacillus* increased (25.1 and 21.0 %). The results from the most recent sediment (1 cm) showed high diversity and evenness and the most abundant genus was *Heliobacterium* (35.1%). However, this sample should be interpreted with care because of the substantially smaller number of sequences obtained (Supplementary Table 2).

The core used in this study has been dated using radiometric techniques (^{137}Cs) and the magnetic susceptibility profile of the sediment core[25]. Five time-points could be inferred: 1943 (peak in magnetic susceptibility), 1954 (beginning of the ^{137}Cs signal), 1964 (peak of ^{137}Cs activity from atomic weapon test fallout) and 1986 (2nd ^{137}Cs activity peak from the Chernobyl accident fallout) as well as 2011, the year of sediment coring. An age-depth model was built by linear interpolation between the time points. The ages for the different samples analyzed were inferred by extrapolation downwards (Supplementary Table 2), in order to correlate endospore-forming bacterial community changes to the historical record. Based on this age model, Section I covers a time period from 1921 to 1950, Section II 1952 to 1961, Section III 1964 to 1997 and Section IV 2002 to 2011.

Since 1957, the water quality of Lake Geneva has been monitored. The marked changes in the endospore-forming bacterial communities from Section I to Section II of the sediment core are consistent with the acceleration of industrial activity and intensification of agriculture after the end of the Second World War, which is reflected by the surge in abundance of *Paenibacillus*. This is followed by an increase in abundance of sulfate-reducing endospore-formers such as *Desulfotomaculum* or *Moorella* at 39 and 43 cm depth. This dominance suggests a change in the content of sulfate either in the water column or in the sediment, which could be due to rapid increase in the burning of fossil fuels in the 1950s and 60s[29, 30]. The peak of prevalence of *Desulfotomaculum* and *Moorella* is correlated with an increase in sulfur registered in the sediment from 43 to 51 cm depth[25], while the variation in the prevalence of groups such as *Bacillus, Paenibacillus* or *Clostridium* is independent of this parameter (Supplementary Figure 2). A significant increase in phosphorus and nitrogen is documented for the water column from 1964 to 1997 (35-11 cm depth)[26], which corresponded once more to the dominance of *Paenibacillus* and *Clostridium*. The turn back to lower levels of nutrients after 2002 is reflected in the most diverse communities from 7 to 1 cm depth, which can also be seen in a redundancy analysis (Supplementary Figure 3). In addition to the nutrient record in the water column, an analysis of carbon (organic and inorganic) and total Nitrogen was carried out in the sediment. This analysis shows a correlation between an increase in the abundance of *Brevibacillus* at 11 and 87 cm depth, with peaks in TOC and C/N ratio registered in the sediment (Figure 2). The dominance of *Brevibacillus,* a common metal-resistant bacterium isolated from soils[31, 32], is thus related to allochthonous input of terrestrial organic material as opposed to aquatic organic matter[33].

The question arises as to whether the changes in the community structure reflect the abundance of endospore-forming bacteria present in vegetative cell state or as inactive endospores. To answer this, selected samples were subjected to a treatment to remove all vegetative cells from the sample prior to DNA extraction and *spo0A* amplicon sequencing. The treatment was applied to four samples corresponding to sediment deposited at 1 cm, 27 cm, 87 cm, and 103 cm depth. The samples from 1 and 103 cm corresponded to the oldest and most recent periods, expected to give the most distinct image between the active and dormant community. The samples from 87 and 27 cm were selected because they corresponded to unique communities that could not be entirely explained by the factors analyzed previously. The community composition in the treated sediments (inactive endospores) was then compared to community in the untreated sediment samples (cells and endospores) (Figure 3) (total and percentage values are provided in Supplementary Table 3).

The untreated and inactive communities were very similar in the oldest sediment sample (103 cm) showing the dominance of *Paenibacillus* (64.1%) and to a lesser extent of *Bacillus* (15.5%), *Desmospora* (6.2%), and *Clostridium* (5.1%). A similar community was found in the endospore fraction of 87 cm sample, with the exception of a higher abundance of *Desulfotomaculum* (8.4%) and the presence of *Sporomusa* (3.3%) and *Moorella* (1.3%). In contrast, the community from the untreated sediment deposited in 87 cm was entirely different. The *Clostridium* (28.2%), *Moorella* (6.8%), *Heliobacterium* (1.6%) and particularly *Sporomusa* (39.7%) abundances were substantially higher, while *Paenibacillus* (19.4%)*,* *Desulfotomaculum* (0.4%), and *Bacillus* (1.2%) were highly reduced. These results signal a change in the water column and the sediment at the time of the deposition of the sample at 87 cm depth, which fundamentally affected the active endospore-forming community. To support this interpretation, we compared the active communities of the samples before (91 cm) and after (83 cm) (Supplementary Figure 4). In both samples, the community was similar to the active community from 103 cm and the endospore communities at 103 and 87 cm depth. At 87 cm the vegetative community shifted towards high abundance of acetogenic *Sporomusa* and fermentative

Clostridium, both anaerobic bacteria previously isolated from lake sediments[34, 35]. *Sporomusa* are homoacetogens that require CO_2 and H_2 for growth and thus are frequently reported as growing in syntrophic association[36]. In temperate regions, acetogens are often out-competed by methanogens and sulfate-reducing bacteria for the use of H_2, and therefore acetogenesis appears to be a particularly significant process in cold anoxic environments[37, 38]. Therefore, a possible explanation for the development of *Sporomusa* at 87 cm could be favorable environmental conditions of anoxia at the sediment-water interface and colder conditions due to an abrupt temperature decrease registered in the winter of 1929 that generated little mixing in the water column and partial lake freezing[39, 40]. Clearly, the effect was pronounced on the active fraction of the community, as the specific analysis of the endospore-fraction did not reveal an increase in *Sporomusa*.

The active community at 27 cm depth was very different from any other community. This sample was almost entirely dominated by *Clostridium* (99.5%). In the endospore fraction, *Clostridium* was still dominant but to a lesser extent (68.3%). In that fraction, a larger diversity was found (Simpson index changing from 0.01 in the active versus 0.97 in the endospore fraction; Supplementary Table 4), with community members among the genera *Paenibacillus* (12.1%), *Bacillus* (9.1%), *Desmospora* (2.5%), *Desulfotomaculum* (2.1%), *Geobacillus* (1%), *Sporomusa* (1%), *Brevibacillus* (0.9%), and *Heliobacterium* (0.7%). The dominance of *Clostridium* suggests anoxia at the sediment-water interface at the time of burial of the sediment now at 27 cm, even though the endospore community was largely different from the active fraction. In addition, the endospore community at 27 cm was also very different from the endospore-fraction at 1 cm and 87 cm depth, alluding to a variation in the communities at sediment surface before burial started and sporulation had set in. In addition to high P values, which can signify higher eutrophication, the year 1976 (corresponding to 27 cm depth) also had the lowest annual discharge for the period between 1915 and 1994[41] which would lead to lesser reoxygenation of hypolimnetic lake water and deep-water anoxia. Anoxia at the lake sediments-water interface can be independently inferred from the Fe/Mn ratio in sediments[42], which is consistent with the Fe/Mn ratio measured in this section of the core[25].

In the most recent sediment sample (1 cm) the communities composed by active cells and endospores and the purely endospore fraction were highly different. In the active cell community, *Heliobacterium* (35.1%), *Paenibacillus* (24.3%), and *Clostridium* (21.6%) made up the majority of all sequences. In the endospore community the diversity was higher (Simpson index of 0.98), with the most abundant genus being *Desulfotomaculum* (41.9%). Also *Bacillus* (13.2%), *Paenibacillus* (12.2%), and a small abundance of *Sporomusa* (4.8%), *Desmospora* (2.3%), and *Heliobacterium* (1.1%) were observed. In addition, this was the only sample harboring *Carboxydothermus* (2%). The dominance of *Heliobacterium* in the top sediment sample (1 cm) is interesting because this genus comprises anaerobic phototrophs associated more commonly to terrestrial environments (e.g. rice paddy soils), rather than aquatic habitats[43]. In our samples however, members of *Heliobacterium* were present in the vegetative cells fraction and not in the endospore fraction, suggesting that they are indeed active at the sediment-water interface in the sediment studied here.

In all the samples a small fraction of the community (10-19% except for 1 cm) corresponded to a series of low abundance groups that were constantly found in the endospore fraction. Even though the community composition fluctuated in the last 100 years, particularly in respect to the dominant members, this fraction of the inactive endospore community remained stable. Those genera correspond to the "seed bank" that has been described in other studies[44, 45], and which represents a series of endospores that get deposited in the sediment with little to no activity during most of the record. The seed bank was composed of 23 genera representing an overall abundance of 19.7% (Figure 3). Within this fraction genera such as *Desulfotomaculum, Desmospora, Geobacillus*, *Sporomusa*, and *Brevibacillus* were present. The importance of this seed bank becomes evident as a source of specialized groups whose abundances varied drastically depending on the environmental conditions. This was clearly the case of *Desulfotomaculum* at 43 cm depth and 1 cm, *Brevibacillus* at 11 cm, or *Sporomusa* and *Morella* at 87 cm depth. Additionally the seed bank concept is supported by the resilience of the system (rapid re-establishment of the community at 83 cm after the event at 87 cm depth). Seed bank of endospores that have been transported from allochthonous sources have been previously studied in cold marine sediments[16] but not from freshwater lakes.

Although the origin of the endospores (allochthonous or autochthonous) cannot be ascertained conclusively, a BLAST search of indicator groups such as *Bacillus thuringiensis*, whose endospores are commercially used as insecticides in agriculture and thus would suggest influence of run-off material, gave negative results. Therefore, we conclude that endospore-forming bacteria are promising as paleolimnological proxies because they are most likely autochthonous to the sediment. By combining the diversity analysis of the vegetative cell fraction together with the dormant endospore fraction, we obtained additional information allowing to reconstruct past environmental history of a model lake. This promising culture-independent approach does not only solves issues of low bacterial cultivability and degradation of DNA, but also has the advantage of focusing on a specific group (endospore-formers) for which a vast amount of metabolic information is available. Moreover, the principle of using endospore-forming bacteria as biological proxies can be easily applied to other freshwater systems, marine sediments, ice cores, and deeper sediments spanning longer timescales.

Methods

Collection of environmental samples

Two sediment cores were retrieved with a gravity corer (UWITEC, Mondstein, Au) in August 2011 in an inactive canyon (C1) on the eastern side of the Rhone delta in Lake Geneva (Switzerland) (CAN01, coordinates 559901-139859, 79 m depth, 105 cm). The core was stored at 4 ºC immediately after retrieval. Core CAN01 was sub-sampled aseptically every 2 cm for biological analysis and samples were stored at -20°C until processing.

Treatment of sediment

For analysis of the fraction of inactive endospores in the sediment, samples from selected depths (1, 27, 87 and 103 cm) were subjected to a treatment that destroys the vegetative cells. The treatment is described in detail elsewhere (Wunderlin et al. *submitted*). Briefly, 0.5 g sediment was dispersed in 900 µl tris-EDTA buffer and incubated in a waterbath at 65°C for 20 min. After cooling the samples down for 5 min at at 4°C, 100 µl of lysozyme was added to reach a final concentration of 2 mg/ml and incubated at 37°C for 1 h and 120 rpm shaking. After that 250 µl of sodium hydroxide (NaOH) and 250 µl of sodium dodecyl sulphate (SDS) were added to final concentrations of 0.5 N and 1% respectively. The mix was incubated at room temperature for 1 h with low agitation (100 rpm). The samples were then filtered onto 0.5 µm pore size nitrocellulose membrane and washed with sterile physiological solution to remove the detergents. DNase mix of 400 µl containing water, 1 x reaction buffer and 2 µl DNase (New England Biolabs, Ipswich, MA, US) was then added directly onto the membrane and left standing for digestion during 15 min. The digested extracellular DNA in the sample was then washed off with physiological solution by continuous filtration. The membrane (containing sediment particles and endospores) was then removed from the apparatus and stored at -20°C until DNA extraction.

DNA extraction

DNA was extracted using a modified protocol of the MP FastDNA® SPIN kit for soil (MP Biomedicals, Santa Ana, CA, USA). The modification consisted of an addition of three repetitive extraction cycles on 0.5 g of sediment as starting material as described elsewhere[24]. The three final DNA extracts were pooled together and DNA was precipitated with 0.3 M Na-acetate and ethanol (99%) and washed with ethanol (70%) before re-suspended in sterile water. DNA yield for the pooled extracts was measured with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using the Quant-iT dsDNA BR assay kit, following the manufacturer's instructions.

spo0A amplicon sequencing

Degenerate primers (spo0A166Af and spo0A748Ar) amplifying a 602 nucleotide fragment of the *spo0A* gene[24] were used to amplify *spo0A* fragments directly from the sediment extracts. PCR reactions were performed in 50 µl final volume adding 0.5 ng DNA template in 1 x reaction buffer (TaKaRa Bio, Shiga, Japan), 3 mM $MgCl_2$, 10 µg bovine serum albumin (BSA; New England Biolabs), 1 U of proof-reading Ex Taq Polymerase (TaKaRa), 200 µM of each dNTP and 1 µM of each primer, completed with PCR-grade water. Negative controls (1 µl PCR-grade water) and positive controls (1 ng *Paenibacillus alvei* DNA template) were included in all reactions. Reactions were done with the Arktik Thermo Cycler (Thermo Fisher Scientific, Vantaa, Finland) with the following temperature program: initial activation at 94°C for 5 min, then 10 cycles of denaturation (94°C; 30 s), touchdown annealing from 55°C to 52°C (-0.3°C/cycle) for 30s and elongation (72°C; 1 min); followed by 30 cycles of denaturation (94°C; 30 s), annealing (55°C; 45s) and elongation (72°C; 1 min) and final extension (72°C; 5 min). PCR products were then purified with MultiScreen PCRµ96 plate (Merck Millipore, Darmstadt, Germany) using a vacuum pump and eluted in 40 µl water. The purified samples were then sent to Eurofins MWG Operon (Ebersberg, Germany) for barcode amplicon sequencing with Roche GS FLX+.

All metagenomic sequences were submitted to GenBank Sequene Read Archive under BioProject number PRJNA206067 and accession numbers SRR1011318 to SRR1011346.

Determination of OTUs of endospore-forming bacteria

Sequences were first binned according to their barcode and the corresponding sample. For quality filtration, the nucleotide sequences were translated to amino acid sequences, based on ORF detection. The amino acid sequences were then compared to a Gribskov-style protein profile of Spo0A[46] that was built based on 27 known Spo0A sequences obtained from the CMR database. Filtration was applied as function of the profile score and profile alignment length, which separates true negatives from true positives. For determination of operational taxonomic units (OTUs), the nucleotide sequences were clustered at 97% sequence identity using uclust[47] in the same way as commonly done for the 16S rRNA gene. The centroid of each OTU cluster was then used to find the closest relative in GenBank by translated protein BLAST (BLASTx).

References

1. Smith, V.H., Tilman, G.D. & Nekola, J.C. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ Pollut* **100**, 179-196 (1999).
2. Smith, V.H. Eutrophication of freshwater and coastal marine ecosystems: a global problem. *Environ Sci Pollut R Int* **10**, 126-139 (2003).
3. European Union (EU) in Official Journal L 327 , 22/12/2000 P. 0001 - 0073; (2000).
4. Gorham, E., Brush, G.S., Graumlich, L.J., Rosenzweig, M.L. & Johnson, A.H. The value of paleoecology as an aid to monitoring ecosystems and landscapes, chiefly with reference to North America. *Environ Rev* **9**, 99-126 (2001).
5. Willard, D.A. & Cronin, T.M. Paleoecology and Ecosystem Restoration: Case Studies from Chesapeake Bay and the Florida Everglades. *Front Ecol Environ* **5**, 491-498 (2007).
6. Thevenon, F., Adatte, T., Wildi, W. & Pote, J. Antibiotic resistant bacteria/genes dissemination in lacustrine sediments highly increased following cultural eutrophication of Lake Geneva (Switzerland). *Chemosphere* **86**, 468-476 (2012).
7. Nealson, K.H. SEDIMENT BACTERIA: Who's there, what are they doing, and what's new? *Annu Rev Earth Pl Sc* **25**, 403-434 (1997).
8. Dell'Anno, A. & Corinaldesi, C. Degradation and turnover of extracellular DNA in marine sediments: ecological and methodological considerations. *Appl Environ Microbiol* **70**, 4384-4386 (2004).
9. Gorham, E., Brush, G.S., Graumlich, L.J., Rosenzweig, M.L. & Johnson, A.H. The value of paleoecology as an aid to monitoring ecosystems and landscapes, chiefly with reference to North America. *Environ Rev* **9**, 99-126 (2001).
10. Dreßler, M., Hübener, T., Görs, S., Werner, P. & Selig, U. Multi-proxy reconstruction of trophic state, hypolimnetic anoxia and phototrophic sulphur bacteria abundance in a dimictic lake in Northern Germany over the past 80 Years. *J Paleolimnol* **37**, 205-219 (2007).
11. Renberg, I. & Nilsson, M. Dormant bacteria in lake sediments as palaeoecological indicators. *J Paleolimnol* **7**, 127-135 (1992).
12. Onyenwoke, R.U., Brill, J.A., Farahi, K. & Wiegel, J. Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). *Arch Microbiol* **182**, 182-192 (2004).
13. Bartholomew, J.W. & Paik, G. Isolation and identification of obligate thermophilic sporeforming bacilli from ocean basin cores. *J Bacteriol* **92**, 635-638 (1966).
14. Rothfuss, F., Bender, M. & Conrad, R. Survival and activity of bacteria in a deep, aged lake sediment (Lake Constance). *Microb Ecol* **33**, 69-77 (1997).
15. Hubert, C. et al. Thermophilic anaerobes in Arctic marine sediments induced to mineralize complex organic matter at high temperature. *Environ Microbiol* **12**, 1089-1104 (2010).
16. de Rezende, J.R. et al. Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J* **7**, 72-84 (2013).
17. Amann, R.I., Ludwig, W. & Schleifer, K.H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**, 143-169 (1995).
18. Staley, J.T. & Konopka, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**, 321-346 (1985).
19. Fichtel, J., Koster, J., Rullkotter, J. & Sass, H. Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol Ecol* **61**, 522-532 (2007).
20. Lomstein, B.A., Langerhuus, A.T., D'Hondt, S., Jorgensen, B.B. & Spivack, A.J. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* **484**, 101-104 (2012).
21. Abecasis, A.B. et al. A genomic signature and the identification of new sporulation genes. *J Bacteriol* **195**, 2101-2115 (2013).
22. Galperin, M.Y. et al. Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol* **14**, 2870-2890 (2012).
23. Bueche, M. et al. Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene spo0A. *App Environ Microbiol* **79**, 5302-5312 (2013).

24. Wunderlin, T., Junier, T., Roussel-Delif, L., Jeanneret, N. & Junier, P. Stage 0 sporulation gene A as a molecular marker to study diversity of endospore-forming Firmicutes. *Environ Microbiol Rep* **5**, 911-924 (2013).
25. Wunderlin, T. et al. Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva (Switzerland-France). *Aq Sci* DOI 10.1007/s00027-013-0329-0 (2013).
26. Lazzarotto, J. & Klein, A. in Rapport de la Commission Internationale pour la Protection des Eaux du Léman contre la Pollution (CIPEL); Campagne 2011. (ed. CIPEL) (Lausanne, Switzerland; 2012).
27. Pace, N.R. Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* **73**, 565-576 (2009).
28. Hugenholtz, P. Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3** (2002).
29. Lavanchy, V.M.H., Gaeggeler, H.W., Schotterer, U., Schwikowski, M. & Baltensperger, U. Historical record of carbonaceous particle concentrations from a European high-alpine glacier (Colle Gnifetti, Switzerland). *J Geophys Res-Atmos* **104**, 21227-21236 (1999).
30. Mylona, S. Sulphur dioxide emissions in Europe 1880–1991 and their effect on sulphur concentrations and depositions. *Tellus B* **48**, 662-689 (1996).
31. Vivas, A. et al. Symbiotic efficiency of autochthonous arbuscular mycorrhizal fungus (*G. mosseae*) and *Brevibacillus* sp. isolated from cadmium polluted soil under increasing cadmium levels. *Environ Pollut* **126**, 179-189 (2003).
32. de Oliveira, E.J., Rabinovitch, L., Monnerat, R.G., Passos, L.K.J. & Zahner, V. Molecular characterization of *Brevibacillus laterosporus* and its potential use in biological control. *App Environ Microbiol* **70**, 6657-6664 (2004).
33. Meyers, P.A. Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. *Org Geochem* **34**, 261-289 (2003).
34. Breznak, J.A. in The Prokaryotes. (eds. Prof. Dr. M Dworking, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt) 991-1001 (Springer US, 2006).
35. Wiegel, J., Tanner, R. & Rainey, F.A. in The Prokaryotes. (eds. Prof. Dr. M Dworking, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt) 654-678 (Springer US, 2006).
36. Cord-Ruwisch, R. & Ollivier, B. Interspecific hydrogen transfer during methanol degradation by Sporomusa acidovorans and hydrogenophilic anaerobes. *Archives of Microbiology* **144**, 163-165 (1986).
37. Kotsyurbenko, O.R. et al. New species of psychrophilic acetogens: *Acetobacterium bakii* sp. nov., *A. paludosum* sp. nov., *A. fimetarium* sp. nov. *Arch Microbiol* **163**, 29-34 (1995).
38. Simankova, M.V. et al. *Acetobacterium tundrae* sp nov, a new psychrophilic acetogenic bacterium from tundra soil. *Arch Microbiol* **174**, 440-447 (2000).
39. Maurer, J. Der strenge Winter und die große Seegefroerne Februar-Maerz 1929 in der Schweiz. *Das Wetter* **46** (1929).
40. Hendricks-Franssen, H.J. & Scherrer, S.C. Freezing of lakes on the Swiss plateau in the period 1901–2006. *Int J Climatol* **28**, 421-433 (2008).
41. Loizeau, J.-L. & Dominik, J. Evolution of the Upper Rhone River discharge and suspended sediment load during the last 80 years and some implications for Lake Geneva. *Aq Sci* **62**, 54-67 (2000).
42. Naeher, S., Gilli, A., North, R.P., Hamann, Y. & Schubert, C.J. Tracing bottom water oxygenation with sedimentary Mn/Fe ratios in Lake Zurich, Switzerland. *Chem Geol* **352**, 125-133 (2013).
43. Asao, M. & Madigan, M.T. Taxonomy, phylogeny, and ecology of the heliobacteria. *Photosynth Res* **104**, 103-111 (2010).
44. Caporaso, J.G., Paszkiewicz, K., Field, D., Knight, R. & Gilbert, J.A. The Western English Channel contains a persistent microbial seed bank. *ISME J* **6**, 1089-1093 (2012).
45. Lennon, J.T. & Jones, S.E. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Rev Microbiol* **9**, 119-130 (2011).
46. Gribskov, M., McLachlan, A.D. & Eisenberg, D. Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci USA* **84**, 4355-4358 (1987).

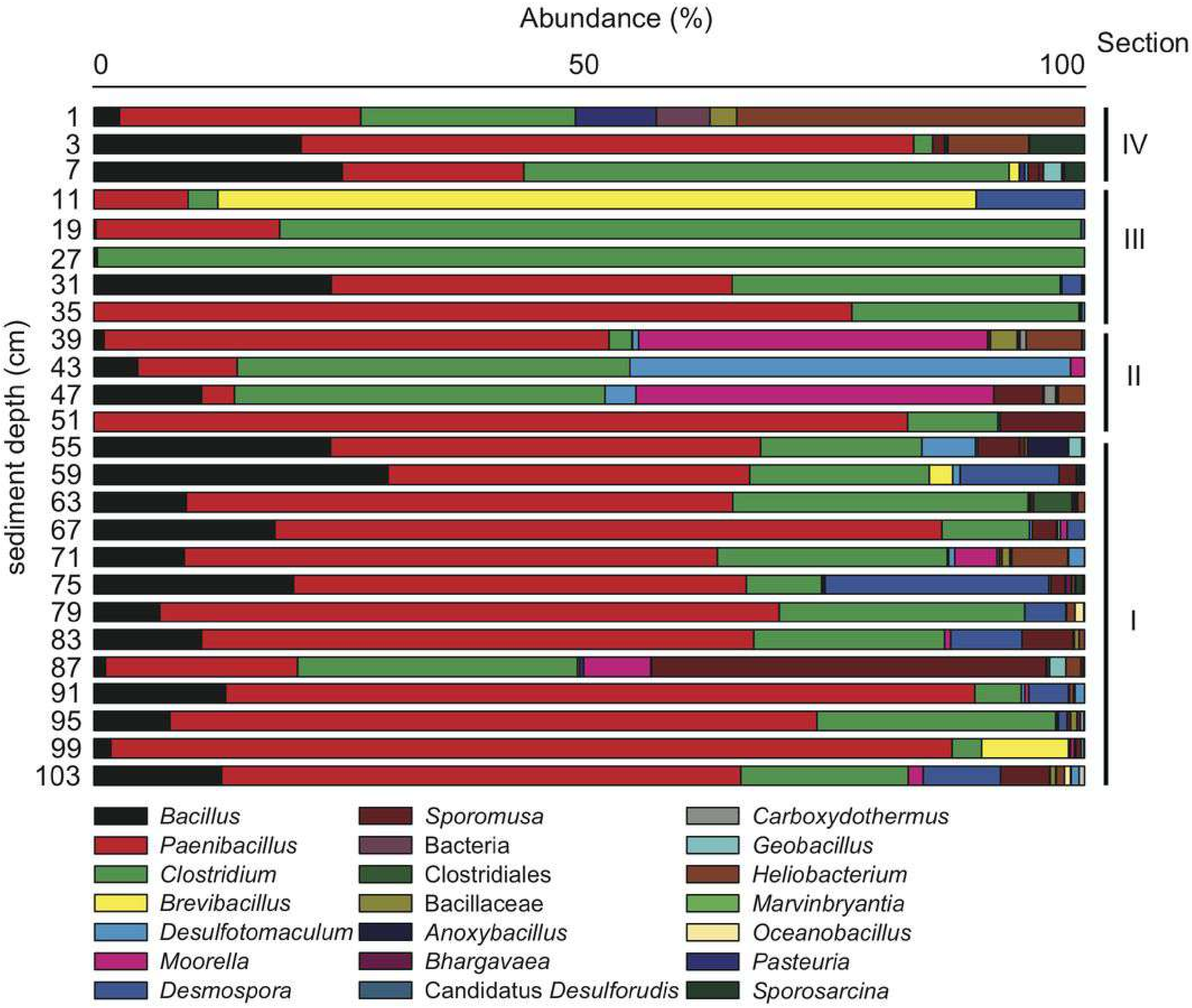
47. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).

Figure Legends:

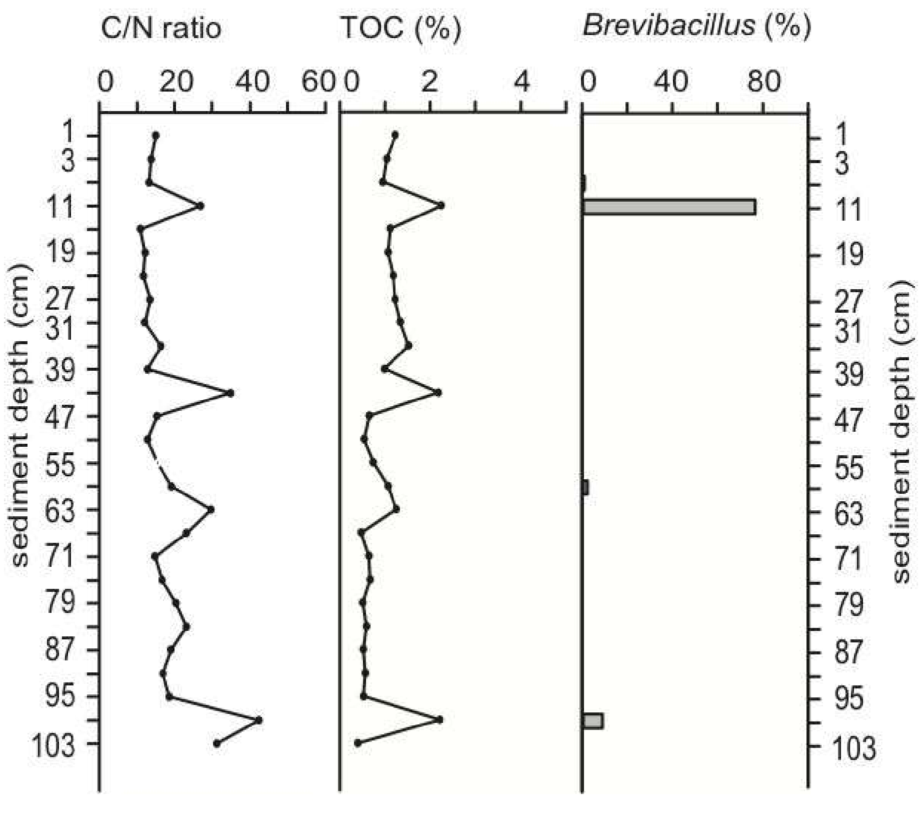
Figure 1. Distribution of endospore-forming genera with sediment depth. The *spo0A* sequences were clustered into OTUs at 97% identity and classified to genus level. The cumulative frequency (in number of sequences) of OTUs belonging to the same genus was added to calculate the percentage values represented for each genera.

Figure 2. Depth profile showing the correlation of C/N ratio and total organic carbon (TOC) with the relative abundance of *Brevibacillus*.

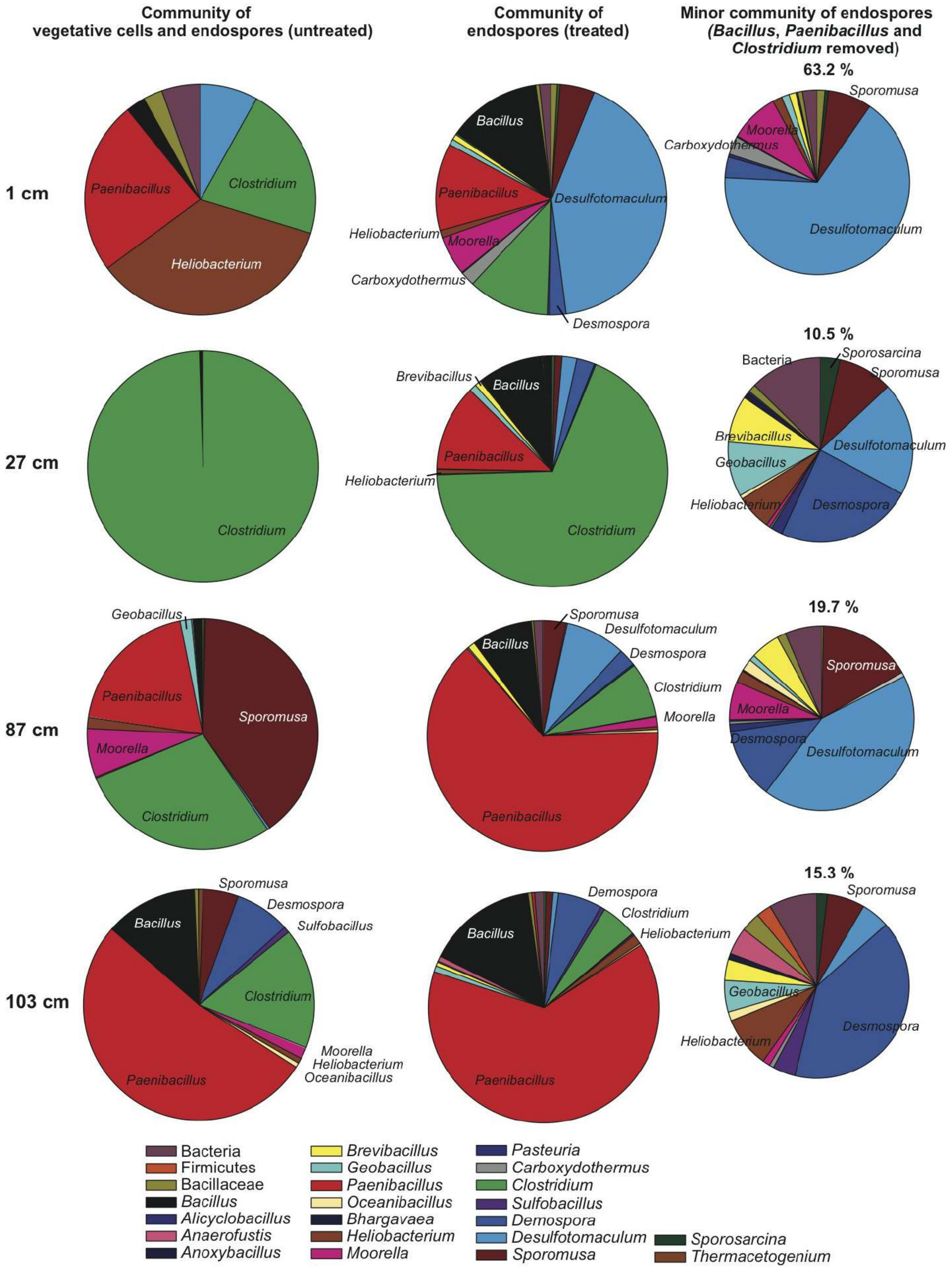
Figure 3. Pie charts displaying the community composition of the total (untreated samples; left column), endospore (treated sample; middle column), and seed bank fractions (minor community treated; left column) of endospore-forming bacteria analyzed at 1 cm, 27 cm, 87 cm and 103 cm depth. The cumulative percentage of the minor fraction (all except most abundant genera being *Bacillus, Clostridium, Paenibacillus*) is shown on top of the respective pie chart.



Abundance (%)
0
50
100
Section
sediment depth (cm)
1
3
7
11
19
27
31
35
39
43
47
51
55
59
63
67
71
75
79
83
87
91
95
99
103
IV
III
II
I
Bacillus
Paenibacillus
Clostridium
Brevibacillus
Desulfotomaculum
Moorella
Desmospora
Sporomusa
Bacteria
Clostridiales
Bacillaceae
Anoxybacillus
Bhargavaea
Candidatus Desulforudis
Carboxydothermus
Geobacillus
Heliobacterium
Marvinbryantia
Oceanobacillus
Pasteuria
Sporosarcina



C/N ratio
TOC (%)
Brevibacillus (%)
0 20 40 60
0 2 4
0 40 80
sediment depth (cm)
1
3
11
19
27
31
39
47
55
63
71
79
87
95
103



Community of vegetative cells and endospores (untreated)
Community of endospores (treated)
Minor community of endospores (Bacillus, Paenibacillus and Clostridium removed)
1 cm
27 cm
87 cm
103 cm
63.2 %
10.5 %
19.7 %
15.3 %
Paenibacillus
Clostridium
Heliobacterium
Bacillus
Desulfotomaculum
Moorella
Carboxydothermus
Desmospora
Sporomusa
Brevibacillus
Bacteria
Sporosarcina
Geobacillus
Sulfobacillus
Oceanibacillus
Demospora
Bacteria
Firmicutes
Bacillaceae
Bacillus
Alicyclobacillus
Anaerofustis
Anoxybacillus
Brevibacillus
Geobacillus
Paenibacillus
Oceanibacillus
Bhargavaea
Heliobacterium
Moorella
Pasteuria
Carboxydothermus
Clostridium
Sulfobacillus
Demospora
Desulfotomaculum
Sporomusa
Sporosarcina
Thermacetogenium

Paper VII

In which endosporulation is thought to be a reaction to multiple environmental stress factors rather than an unique one...

Survival strategy meets classic ecological theories: the case of endospore-forming bacteria diversity and abundance in extreme environments.

Authors: S. Filippidou, M. Bueche, T. Wunderlin, T. Junier, L. Sauvain, L. Roussel- Delif, N. Jeanneret, C. Dorador, V. Molina, A. Ioannidou, G. Vargemezis, D. Johnson, P. Junier

Metagenomic data are not only helpful in descriptive studies of diverse ecosystems but have recently been used to reveal relationships between environmental data and patterns of microbial dispersal in extreme environments. For instance, in geothermal environments, recent studies have shown that temperature, and to a less extend pH, dictates prevalence of specific bacterial and archaeal groups. However, this raises the question of how (and if) this trend can be extrapolated to specific microbial groups within the communities.

Endospore-forming bacteria, while in spore state, can survive adverse conditions and disperse over large distances. Thus, they can be found in different environments: from psychrophilic to thermophilic ones, from natural springs to highly contaminated ecosystems and from acidic to alkaline substrates. Based on their ubiquitous distribution, we hypothesize that the co-existence of multiple stressing environmental parameters, rather than temperature or pH alone, favors the relative abundance of endospore- forming bacteria and influences their diversity in extreme environments. These parameters are temperature, pH, salinity, high concentration of heavy or trace metals, UV radiation, and limited carbon and energy sources.

In order to address this hypothesis we have held a worldwide sampling campaign, including 23 sites with diverse environmental characteristics such as high temperature, salinity, depth, altitude, low and high pH, and contamination with heavy and trace metals. In total, 126 samples have been collected and analyzed. DNA was extracted using an extraction method that enables the extraction of DNA from hard to break microbial cell walls or spore coats. 16S rRNA gene (total bacteria) and sporulation transcriptional factor *spo0A* gene (endospore-specific) copy numbers were then estimated by quantitative PCR (qPCR). Moreover, community composition based on 16S rRNA and *spo0A* gene amplicon sequencing, was analysed in 12 samples representing environments with multiple, single or no stressing factors.

The quantification shows that endospore formers are indeed favored when stressing factors, and more significantly a combination of those, are present in the environment. Our data demonstrate that in sites where two or more environmental stressors are present, endospore-forming bacteria are highly prevalent (ratio median = 26.46%). When there is a single stressor, endospore-forming bacteria are favored but not high in abundance (ratio median = 2.79%), while in non-extreme environments they make up only a small fraction of the microbial community (ratio median = 0.079%).

In comparison to the whole-community, in which bacterial diversity is controlled by a single environmental stressor, (e.g. temperature in geothermal sites or trace metal concentration in polluted sediments), at the scale of endospore- forming bacteria, diversity is influenced by a combination of stressors. These observations suggest that, in extreme environments endospore-formers, behave as a ‘system within a system’ and are influenced by their environment in a different way than the whole bacterial community.

Part III

Curriculum Vitae

Matthieu**Bueche**

Microbiologist

contact

University of Neuchâtel
Emile-Aragand 11
2000 Neuchâtel
Switzerland

++41 (0)32 718 2242
++41 (0)77 402 5339

matthieu.bueche@unine.ch
www2.unine.ch/lamun

languages

French: Native language
English: Fluent
German: Basic
Italian: Basic

informatic

♥ R, Qiime, linux, C++, MS-Office

molecular biology

DNA extraction, PCR, qPCR, DGGE, CE-SSCP, Cloning

microbiology

aerobe and anaerobe cultivation, continuous culture

analytical

Flow cytometry, ICP-OES, colorimetry, laser granulometry, epifluorescence microscopy

education

2011–2014 **PhD thesis** in Biology — University of Neuchâtel, Switzerland
Spore-forming bacteria as indicators of pollution in sediments of Lake Geneva
This thesis explored the idea that endospore-forming bacteria are more resistant to heavy-metal pollution than others bacterial taxa. The results showed that they are effectively predominant in the most polluted environments.

2008–2010 **Master** of Biogeosciences — University of Neuchâtel, Switzerland
Final research project performed at the Laboratory of Biotechnology of the Environment at the INRA of Narbonne, France
Facteurs influençant la dynamique de la diversité des communautés microbiennes

2005–2008 **Bachelor** of Biology — University of Neuchâtel, Switzerland

1994–1997 **Engineer** in Electronics — Engineering school of St-Imier, Switzerland

1990–1994 **CFC** in Electronics — School of crafts of St-Imier, Switzerland

training

Nov. 2013	**Write a review, write now!**	Neuchâtel, Switzerland
Nov. 2013	**Introduction to bioinformatics (5 days)**	Neuchâtel, Switzerland
October 2012	**Intro. to chromatography and mass spectrometry for biologists**	Neuchâtel, Switzerland
August 2012	**Field course in geomicrobiology - a joint course with CUSO**	Graubünden, Switzerland
July 2012	**Practical course on metagenomics for biodiversity**	Madrid, Spain
June 2012	**Fungi-bacteria interactions - a joint symposium with CUSO**	Neuchâtel, Switzerland
Sept. 2011	**Microbial diversity (bacteria, fungi, protists), joint with CUSO**	Neuchâtel, Switzerland
Sept. 2011	**Aboveground and below-ground community ecology**	Neuchâtel, Switzerland
June 2011	**Minisymposium in metagenomics**	Neuchâtel, Switzerland

experience

2002–2005 **THEATRE POPULAIRE ROMAND** La Chaux-de-Fonds, Switzerland
Régisseur général
- Accueil technique des artistes, créations son et lumière
- Entretient de l'infrastructure technique de la maison

2001–2002 **ETEL MOTION TECHNOLOGY SA** Môtiers, Switzerland
Research Engineer
- Développement de l'électronique de gestion pour une moteur linéaire
- Présentations en séminaires de formation interne

1998-2001 **ENGINEERING SCHOOL OF ST-IMIER** St-Imier, Switzerland
Assistant in the laboratory of electronics
- Projets électroniques internes, administration informatique, réparations
- Aide aux étudiants dans leurs travaux de laboratoire, enseignement

teaching

2013-2014 **Master thesis supervisor** Supervision of the master thesis research project of Coralie Flury
Microbial community in the sediments of Lake Kastoria (Greece), a step towards a better understanding of eutrophication

2013 **APP project** Mini-research project for the 3rd year bachelor student in biology
Microbiological quality of water - Classical and modern FCM approach

2012 **APP project** Mini-research project for the 3rd year bachelor student in biology
Substrates evolution during fermentation - Beer making

2012 **Practicals in microbiology** Practicals for the 2nd year bachelor student in biology

2011-2012 **Master thesis supervisor** Supervision of the master thesis research project of Loïc Sauvain
Link between spore-forming bacteria and pollution

2011 **Practicals in microbiology** Practicals for the 2nd year bachelor student in biology

2000 **Diploma project supervision** Supervision of the end-study research project for one group of two electronics engineers

awards

2012 **ICHMET 2012** 16th International Conference on Heavy Metals in the Environment, Rome, Italy
Best Student Poster Award

2010 **Master in Biogeosciences** University of Neuchâtel, Switzerland
magma cum laude

2008 **Bachelor in Biology** University of Neuchâtel, Switzerland
cum laude

1994 **CFC in Electronics** School of crafts of St-Imier, Switzerland
Distinction for the CFC in Electonics

grants

2012 **COST program** COST internat. workshop on NGS and metagenomics, Madrid, Spain
Travel and accommodation expenses covered by COST program

2010 **Master in Biogeosciences** University of Neuchâtel, Switzerland
Prix Louis Paris

1994 **CFC in Electronics** School of crafts of St-Imier, Switzerland
Prix de la société des anciens étudiants

communication skills

July 2012 **Oral Presentation** COST internat. workshop on NGS and metagenomics, Madrid, Spain
Presented the bioinformatic treatment of 454-pyrosequencing data by blastn

August 2012 **Oral Presentation** Geo-Microbiology Filed Course, Davos, Switzerland
Diversity of some filamentous bacteria

October 2012 **Oral Presentation** Comparative genomics course, Lausanne, Switzerland
Presented the bioinformatic treatment of 454-pyrosequencing data by QIIME

Nov. 2012 **Oral Presentation** 10th Swiss geoscience meeting, Bern, Switzerland
Presented the first results of my Ph.D. thesis in which the link between endospore-forming Firmicutes and heavy and trace metals in lake sediments is highlighted

May 2012 **Poster** Annual Ph.D. students meeting, Neuchâtel, Switzerland
Presented the outline of my Ph.D. thesis project

June 2012 **Poster** Joint annual meeting SSI-SSHH-SSM-SSTMP, St-Gallen, Switzerland
Presented the design pipeline of quantitative PCR primers for the *spo0A* gene

Sept. 2012 **Poster - *Awarded*** ICHMET 2012, Rome, Italy
Presented the first results of my Ph.D. thesis in which the link between endospore-forming Firmicutes and heavy and trace metals in lake sediments is highlighted

interests

professional: microbial ecology, environmental issues, bioremediation technologies, biogeochemical cycling of the elements, aquatic microbiology, bioinformatic, data analysis
personal: do-it-your-self, biking, hiking, electronics, theatre, botany

publications

articles in peer-reviewed journals

Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva (Switzerland–France)
Tina Wunderlin, JuanPablo Corella, Thomas Junier, Matthieu Bueche, Jean-Luc Loizeau, Stéphanie Girardclos, Pilar Junier
Aquatic Sciences *(Dec. 2013) pp. 1–14. 2013*

Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria
Loïc Sauvain, Matthieu Bueche, Thomas Junier, Matthieu Masson, Tina Wunderlin, Roxane Kohler-Milleret, Elena Gascon Diez, Jean-Luc Loizeau, Mary-Lou Tercier-Waeber, Pilar Junier
Aquatic Sciences *(Oct. 2013) pp. 1–14. 2013*

Quantification of Endospore-Forming Firmicutes by Quantitative PCR with the Functional Gene *spo0A*

Matthieu Bueche, Tina Wunderlin, Ludovic Roussel-Delif, Thomas Junier, Loic Sauvain, Nicole Jeanneret, Pilar Junier

Applied and Environmental Microbiology *79.17 (2013) pp. 5302–5312. 2013*

international peer-reviewed conferences/proceedings

Endospore-forming bacteria as an indicator of pollution in sediments of Lake Geneva

M. Bueche, L. Sauvain, T. Wunderlin, T. Junier, M. Tercier, J.-L. Loizeau, M. Masson, E. Gascon, P. Junier

E3S Web of Conferences, 2013

in preparation

Tina Wunderlin, Thomas Junier, Juan Pablo-Corella, Mattieu Beuche, Jean-Luc Loizeau, Stephanie Girardclos, Pilar Junier. “Endospores in Sediment as a Proxy for 20th Century Environmental Lake History”. Submitted to Proceedings of the National Academy of Sciences.

Matthieu Bueche, Loïc Sauvain, Pilar Junier. “Endosporulation as a survival strategy into heavy and trace metal contaminated sediments”. To be submitted to Environmental Science and Technology.

Sevasti Filippidou, Matthieu Bueche, Tina Wunderlin, Thomas Junier, Ludovic Roussel-Delif, Nicole Jeanneret, Christina Dorador, V. Molina, A. Ioanidou, G. Vergemezis, Pilar Junier. “Diversity and abundance of Endospore-forming bacteria influenced by the co-existence of stressors in extreme environments”. To be submitted to ISME journal.