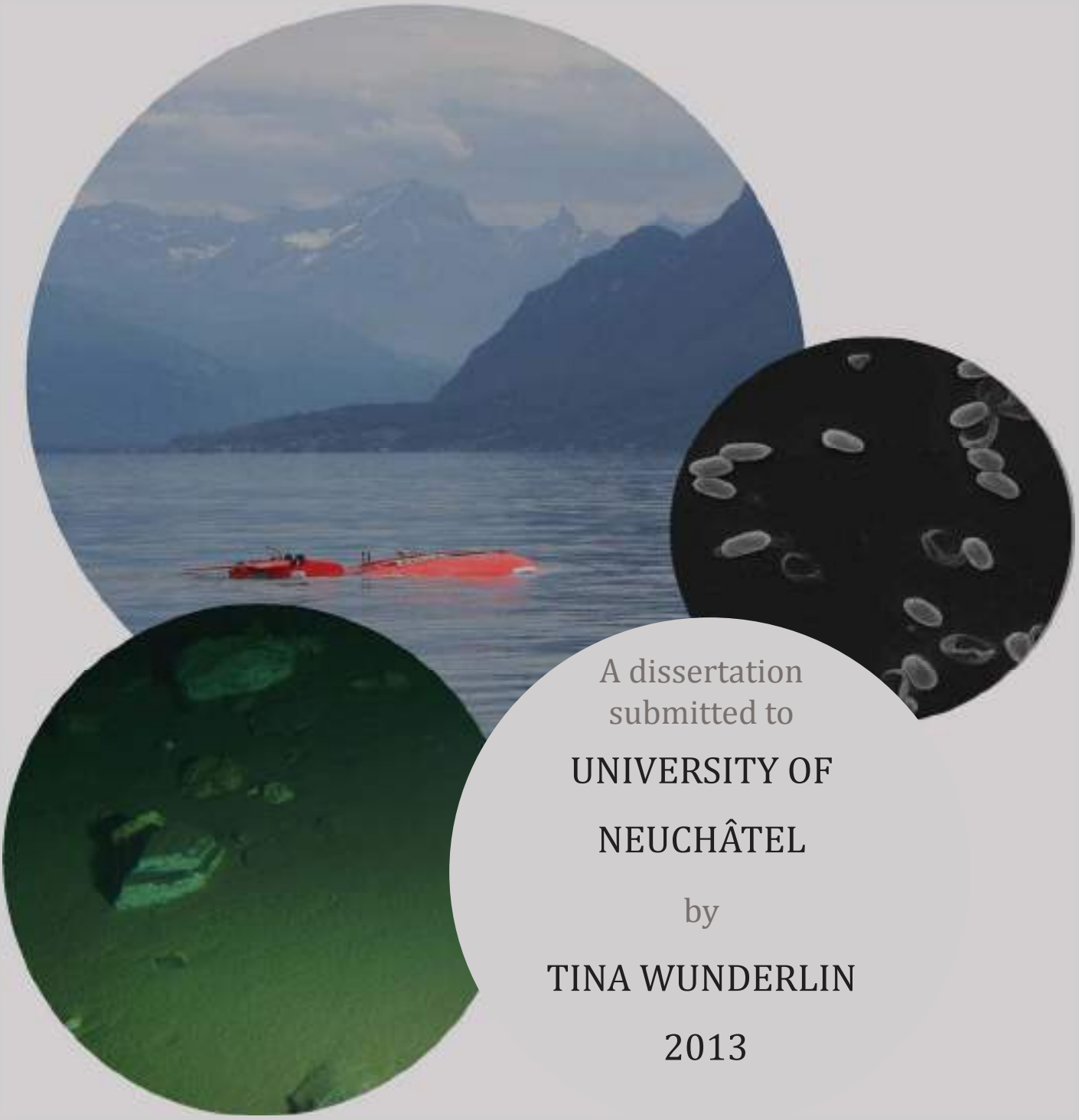
DIVERSITY OF ENDOSPORE-FORMING BACTERIA IN SEDIMENT AS A PROXY FOR ENVIRONMENTAL LAKE HISTORY



A dissertation submitted to

UNIVERSITY OF NEUCHÂTEL

by

TINA WUNDERLIN

2013

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

Madame Tina WUNDERLIN

Titre : Diversity of endospore-forming bacteria in sediment as a proxy for environmental lake history

sur le rapport des membres du jury:

- Prof. ass. Pilar Junier, directrice de thèse, UniNe
- Prof. Edward Mitchell, UniNe
- Dr Michelle Power, Macquarie University, Australie
- Prof. emeritus Judith McKenzie, ETH Zürich
- Dr Helmut Bürgmann, Eawag, Kastanienbaum, Suisse

Neuchâtel, le 10 décembre 2013

Le Doyen, Prof. P. Kropf

Diversity of endospore-forming bacteria in sediment as a proxy for environmental lake history

A dissertation submitted to

University of Neuchâtel, Switzerland
for the degree of
Docteure ès sciences

presented by

Tina Wunderlin
MSc Environ. Sciences ETH

accepted on the recommendation of:

Dr. Pilar Junier, thesis director, University of Neuchâtel
Prof. Edward Mitchell, University of Neuchâtel
Dr. Michelle Power, University of Macquarie, Sydney, Australia
Prof. emeritus Judith McKenzie, ETH Zürich
Dr. Helmut Bürgmann, Eawag, Kastanienbaum

December 2013

Wissenschaft (...) ist die gemeinschaftliche Suche nach wahren, verlässlichen Erkenntnissen über die Welt einschliesslich unserer selbst.
Manfred Spitzer, 2012

Acknowledgements

This thesis would not have been possible without the help and support of many kind, skilled and wonderful people. My profound thanks go to all of them.

I specifically want to thank Thomas Junier for his excellent work in bioinformatics and for his useful comments on this manuscript. Nicole Jeanneret, Ludovic Roussel-Delif, Matthieu Bueche, Sevasti Filippidou, Marion Jaussi, Loïc Sauvain, and Manon Brenier are kindly acknowledged for their contribution to this work.

Juan Pablo Corella, Stéphanie Girardclos, and Jean-Luc Loizeau are acknowledged for the good collaboration, for giving access to the sediment cores and for having provided data and helpful comments on the paleolimnology chapters.

Thanks also go to Saskia Bindschedler and Sevasti Filippidou for their comments on this manuscript, to Ludovic Roussel-Delif for the correction of the abstract in French and to Romain Bessire and the staff at Unine for their support.

I kindly thank my thesis committee: Judith McKenzie, Michelle Power, Helmut Bürgmann, and Edward Mitchell for their fair assessment and their helpful comments to improve this manuscript.

An immense thanks goes to my supervisor Pilar Junier. I have enjoyed every day working with her and I could not have wished for a better relationship. Her scientific and leadership skills, her energy, kind personality and big heart are simply amazing. I hope we can keep the close ties in the future, in personal and professional matters.

Last but not least, my profound thanks go to all my colleagues, my dear friends and loving family who have supported me during these last four years. None of this would have been possible without them.

Contents

List of Figures

List of Tables

Summaries

Summary

Freshwater systems are of high ecological and economical importance but at the same time subjected to anthropogenic pressure such as input of pollutants, overfishing, changes in climate regime, and eutrophication (high nutrient input). For environmental management and the implementation of conservation measures, the dynamics of freshwater ecosystems need to be known and biological reference conditions have to be established against which future changes can be measured. Baseline knowledge about biodiversity and ecosystem responses to environmental perturbations in lakes can be obtained from the sediments that provide an ideal environmental archive of past conditions.

This thesis presents research on the diversity detection of bacterial endospores, and its use as proxy to reconstruct the environmental history of the last 100 years of Lake Geneva at the border of France and Switzerland.

Endospores are resistant structures formed when bacteria are under stress. Once these endospores are deposited in the sediment they remain dormant and serve as natural biological time capsules, archiving the conditions at the time of sedimentation. To infer the diversity of endospores, two specific methods for targeted metagenomics were developed and validated in sediments. Metagenomics is a sequencing approach of the entire genetic pool directly retrieved from an environmental sample. Similarly, targeted metagenomics is based on a targeted genetic pool, for example a sub-community of the sample. Targeted metagenomics increases the sequencing coverage and resolution of detection, circumventing common problems of traditional metagenomics studies. For this work the endospore-forming bacterial community was targeted.

The first targeting method was based on a molecular marker for endospore-forming bacteria, the global transcription regulator of sporulation (*spo0A*). After an optimized DNA extraction method for endospores in sediment, where biomass was separated from the sediment particles (indirect DNA extraction), the *spo0A* gene fragment was amplified and sequenced. With this method, the diversity of the endospore-forming bacteria (vegetative cells and endospores) in sediments was determined.

The second targeting method consisted of a treatment to separate the endospores from vegetative cells, prior to DNA extraction and sequencing. The goal of the treatment was to destroy vegetative cells that are generally more fragile, while leaving the more resistant endospores intact. With this method, the diversity of only the endospores in sediment was detected. The treatment to separate endospores was successful, as shown by an enrichment of endospore-forming bacteria from 10% abundance in the global approach to over 90% abundance in the targeted metagenome. Also the resolution was improved to up to 10-fold increase in detected endospore-forming taxa. The better resolution led to the detection of 34 genera unique to the targeted metagenome, including some supposedly asporogenic groups like *Ethanoligenens* and *Trichococcus* and high numbers of sequences that could be classified to a species level such as *Bacillus longiquaesitum* or *Clostridium bowmanii*.

The application of targeted metagenomics to a sediment core retrieved from Lake Geneva spanning a time period from 1921 to 2010, revealed substantial diversity of endospore-forming bacteria in sediments. The diversity fluctuated significantly in the last 100 years, reflecting the eutrophication period from 1960 to 1990 as well as sulfate metabolism, input of terrestrial organic matter, and specific climate events. The shift in the community composition during eutrophication was linked to a dominance of

anaerobic Clostridia-like members, that reflect anoxic sediment conditions during this time.

The advantage of the treatment is that the communities in vegetative cell state can be differentiated from the communities present as dormant endospores. Using this differentiation we report activity of selected endospore-forming bacteria in sediment, for example members from genus *Clostridium* and *Heliobacterium* at the sediment surface. In contrast, a small fraction of dormant endospores present at high diversity represent the microbial seed bank. This group of bacteria is inactive for long periods of time but selected members can propagate and become dominant if the environmental conditions change to their favour, as was observed in this study for *Desulfotomaculum*, *Sporomusa* or *Brevibacillus*.

The novel targeted metagenomics approaches developed here provide a significant experimental improvement to explore the diversity of endospore-forming bacteria at high resolution. The data provides knowledge on the role of endospore-forming bacteria in freshwater sediment and on freshwater sediment bacteria in general. It is also the first report of metagenomics to reveal the diversity of endospores in sediment and the use of endospores as paleolimnological proxies.

Résumé

Les lacs une grande importance écologique et économique, mais ils sont aussi très vulnérables aux pressions anthropogéniques (pollution; surpêche), aux changements climatiques et à l'eutrophisation (hautes concentrations de nutriments). Pour une gestion environnementale et la mise en place de mesures de conservation, les dynamiques des écosystèmes lacustres doivent être connues et les conditions biologiques de référence doivent être établies afin de mesurer dans le futur la santé écologique des lacs. Les conditions biologiques de références, comme les connaissances de la biodiversité et les réponses des écosystèmes aux perturbations environnementales sont fournies par les sédiments, qui sont des archives idéales des conditions du passé.

Cette thèse présente la recherche sur la détection et la diversité des bactéries sporulantes et leur usage comme indicateurs des conditions écologiques du lac Léman (France-Suisse) au cours des 100 dernières années.

Les endospores sont des structures résistantes qui sont produites par un groupe de bactéries dans des conditions de stress. Les endospores dormantes sont déposées avec le sédiment et sont des capsules biologiques qui reflètent les conditions environnementales au moment de leur sédimentation.

Les approches métagénomiques sont des études de séquençage du métagénome entier qui se trouve dans un échantillon environnemental. Une approche ciblée est réalisée uniquement sur une partie de la communauté (fraction ciblée). Ces approches métagénomiques ciblées augmentent la couverture de séquençage et également la résolution de détection de taxons. Elles résolvent donc les problèmes connus des approches métagénomiques globales. Deux méthodes de métagénomique ciblée ont été développées au cours de cette thèse pour étudier la diversité des bactéries sporulantes dans les sédiments.

La première méthode est basée sur des amorces moléculaires dessinée pour amplifier un fragment du gène *spo0A*, gène spécifique aux bactéries sporulantes codant pour le facteur de transcription de la sporulation. De plus, une méthode optimisée d'extraction d'ADN pour les bactéries sporulantes a été développée. En appliquant ces méthodes, la diversité des cellules végétatives des bactéries sporulantes ainsi que les endospores dans le sédiment peut être étudiée.

La deuxième méthode de métagénomique ciblée est une méthode de traitement avec chaleur et agents chimiques pour détruire les cellules végétatives, qui sont fragiles, comparé aux endospores qui résistent au traitement. Avec cette méthode, la diversité seule des endospores peut-être évaluée. Le traitement pour détruire les cellules végétatives est efficace, 90% des séquences détectées sont classifiées comme bactéries sporulantes. Avec une approche globale, seulement 10% des séquences détectées sont classifiées comme bactéries sporulantes. De plus, la résolution a été augmentée, en détectant jusqu'à 10 fois plus de taxon. La meilleure résolution permet de détecter 34 genres de bactéries sporulantes non révélés avec l'approche globale, dont certain genres qui ont été définis comme étant asporogénique, comme *Ethanoligenens* et *Trichococcus*. D'autre séquences ont pu être classifiées au niveau de l'espèce comme *Bacillus longiquaesitum* ou *Clostridium bowmanii*.

Les méthodes métagénomiques ciblées ont été appliquées à une carotte de sédiment du Lac Léman, qui couvre la période de 1921 à 2010. Une diversité extraordinaire de bactéries sporulantes a été observée dans ces sédiments. Il y a des fluctuations importantes dans la diversité au cours des 100 dernières années. Prioritairement les changements de diversité peuvent être liée à l'eutrophisation du lac de 1960 à 1990, mais

aussi au métabolisme du soufre, la charge de matière organique d'origine terrestre ou à des évènements climatiques. La communauté bactérienne dans la période eutrophe a changé vers une dominance de bactéries anaérobies comme les Clostridia, qui reflètent les conditions anoxiques de la surface du sédiment pendant cette période.

Les avantages du traitement pour séparer les endospores des cellules végétatives sont que ces deux fractions peuvent être analysée individuellement. Avec ces différentiations, des genres actifs dans le sédiment comme *Clostridium* et *Heliobacterium* peuvent être relevés. Par contre, il y a toute une fraction diversifiée d'endospores dormantes qui est présente en faible nombre, jusqu'à ce que les conditions environnementales changent en leur faveur ou elles deviennent dominantes. Des exemples de ce phénomène ont été observés pour les genres *Desulfotomaculum*, *Sporomusa* et *Brevibacillus*.

Les nouvelles méthodes de métagénomique ciblée qui ont été dévéloppées dans le cadre de cette thèse sont des améliorations importantes pour étudier la diversité des bactéries sporulantes à haute résolution. Les données acquises pendant cette recherche amènent de nouvelles informations sur le rôle des bactéries sporulantes dans les sédiments et en général sur les bactéries des sédiments lacustres. Cette étude est aussi la première étude sur la diversité des bactéries sporulantes par des méthodes de séquençage et la première étude d'utilisation des endospores comme indicateur paléolimnologique.

Zusammenfassung

Süsswassersysteme, wie Seen und Flüsse sind ökologisch und ökonomisch von grossem Nutzen. Viele dieser Systeme sind jedoch stark von menschlicher Aktivität wie Schadstoffeintrag, Überfischung, Veränderungen des Klimas sowie Eutrophisierung (hoher Nährstoffgehalt) beeinflusst. Für ein effizientes Umweltmanagement und die Einführung von Massnahmen zum Schutz von Seen braucht es Basiswissen über das Ökosystem sowie Referenzdaten zur Biologie im See. Zukünftige Umweltveränderungen und Änderungen der Biodiversität können dann mit den Referenzdaten abgeglichen werden. Solche Referenzbedingungen der biologischen Aktivität im See, sowie vergangene Veränderungen im System, können anhand von Sedimentdaten erhoben werden, denn Sedimente sind ideale Archive historischer Umweltbedingungen.

Im Rahmen dieser Dissertation wurde die Diversität von sporenbildenden Bakterien ermittelt. Am Beispiel eines Sedimentkerns des Genfersees (Frankreich-Schweiz), der die letzten 100 Jahre abdeckt wurde das Potential von Sporen als Indikatoren für frühere Umweltbedingungen in Seesedimenten untersucht. Endosporen sind resistente Strukturen, die von einer Gruppe von Bakterien gebildet werden, wenn sie Stressbedingungen ausgesetzt sind. Endosporen sedimentieren am Seeboden und bleiben im Sediment in inaktiver Form erhalten. So dienen sie als natürliche Zeitkapseln, die die Bedingungen im See zur Zeit der Sedimentierung wiederspiegeln. Um die Diversität von sporenbildenden Bakterien zu ermitteln wurden zwei verschieden Methoden auf Basis der funktionellen Metagenomik entwickelt.

Megenomik ist die Sequenzanalyse des ganzen genetischen Pools, der in einer Umweltprobe vorhanden ist. Funktionelle Metagnomik, im Gegensatz, fokussiert auf einen Teil des genetische Pools, der anhand von spezifischen funtionellen Genen oder Eigenschaften ausgewählt wird. Mit solchen funktionellen metagenomischen Analysen kann die Abdeckung der zu sequenzierenden Fragmente erhöht werden, was zu einer höheren Auflösung der Artenzusammensetzung führt. In der Forschung für diese Dissertation wurden funktionelle Metagenomik auf die Gruppe von sporenbildenden Bakterien angewendet.

Eine erste Methode, die entwickelt wurde, ist die Anwendung eines molekularen Markers für das Gen *spo0A*, welches für den Transkriptionsfaktor der Sporenbildung codiert. Die DNA der Bakterien von Sedimentproben wurde zuerst mit optimierten Extraktionsmethoden extrahiert und danach wurde ein Fragment des Gens *spo0A* amplifiziert und sequenziert. So konnte die Diversität von sporenbildenden Bakterien (vegetative Zellen sowie Endosporen), ermittelt werden.

Die zweite Methode zur funktionellen Metagenomik, beinhaltete eine Behandlung der Probe mit Hitze und chemischen Agenzien um die vegetativen Zellen in der Probe zu zerstören und dabei die resistenteren Endosporen zu isolieren. Nach der Behandlung wurde die DNA der auf diese Weise isolierten Endosporen extrahiert und einzelne Gene sequenziert. Diese Methode erlaubt die Ermittlung der Diversität der inaktiven Endosporen im Sediment. Die Behandlung zur Isolierung war erfolgreich, was durch die Zunahme der Abundanz von sporenbildenden Bakterien von 10% im Datensatz der globalen metagenomischen Analyse zu 90% im Datensatz der funktionellen metagenomischen Analyse, bewiesen werden konnte. Ausserdem konnte die Auflösung und Sequenzabdeckung in Form der Anzahl erfasster Taxa auf bis zu 10-mal erhöht werden. Durch die bessere Auflösung konnten auch 34 Gattungen von sporenbildenden Bakterien erfasst werden, die in der globalen Metagenomik übersehen werden. Einige dieser Gattungen

gehören zu Gruppen, die bisher als asporogene Arten bekannt waren, wie zum Beispiel *Ethanoligenens* und *Trichococcus*. Auch können einige Sequenzen im Detail bis zur Art bestimmt werden, so zum Beispiel *Bacillus longiquaesitum* oder *Clostridium bowmanii*.

Die Methode der funktionellen Metagenomik wurden dann auf Proben eines Sedimentkerns des Genfersees angewendet, wodurch eine grosse Diversität von sporenbildenden Bakterien ermittelt werden konnte. Der Kern widerspiegelt den Zeitraum von 1921 bis 2010 und beinhaltet auch die Phase der Eutrophisierung von 1960 bis 1990, die mit starken Schwankungen in der Diversität der sporenbildenden Bakterien einhergeht. Generell konnten Veränderungen in der Diversität mit unterschiedlichen Nährstoffgehalten, Schwefelmetabolismus, Eintrag von terrestischem, organischen Material sowie spezifischen Klimaereignissen erklärt werden. Während der eutrophischen Phase des Sees waren grosse Zahlen von Clostridia-verwandten Arten vorhanden, deren anaerobe Lebensweise auf anoxische Bedingungen an der Sedimentoberfläche während dieses Zeitraums hinweist.

Der Vorteil einer Behandlung zur Isolierung von Endosporen ist die individuelle Analyse der aktiven sporenbildenden Fraktion (vegetative Zellen) sowie der inaktiven Endosporen-Fraktion. Mit dieser Differenzierung konnten einige Gattungen als klar aktive Gruppen definiert werden, wie zum Beispiel *Clostridium* und *Heliobacterium*. Auf der anderen Seite war eine inaktive Minderheit mit grosser Diversität vorhanden, die als sogenannte bakterielle Samenbank beschrieben wird. Bakterien dieser Gruppe sind während langer Zeit inaktiv und können dann plötzlich aufwachsen und sich vermehren, wenn die Umweltbedingung für die jeweilige Arten vorteilhaft werden, was hier für die Gattungen *Desulfotomaculum*, *Sporomusa* und *Brevibacillus* der Fall war.

Die hier beschriebenen, neuen Methoden zur funktionellen Metagenomik der Gruppe von sporenbildenden Bakterien, sind wesentliche Verbesserungen zur Erfassungen der genetischen Diverstiät dieser Bakterien. Die Resultate leisten einen substanziellen Beitrag zum Wissen über die Rolle von sporenbildenden Bakterien in Seesedimenten, sowie allgemein über Bakterien in Sedimenten. Ausserdem ist dies die erste Studie der Diversität von sporenbildenden Bakterien mit Hilfe von Sequenziermethoden und die erst Studie zur Nutzung von Endosporen als paleolimnologische Indikatoren.

1 Introduction

1.1 Thesis outline

The present thesis explores the potential of using bacterial endospores as paleoecological indicators for freshwater lake systems. By targeted metagenomics sequencing the diversity of endospore-forming bacteria is revealed and linked to present and past environmental lake conditions. In order to achieve this, developing tailored culture-independent methods was an essential pre-requisite. Two such methods based on targeted metagenomics are described here.

The first chapter contains a general introduction to present the state of the art in the research field of microbial ecology and the research on endospore-forming bacteria. It ends with the description of the scope of the research and the main objectives.

Then, in chapter 2, the first method for targeting endospore-forming bacteria and revealing their diversity is described. This method is based on the sporulation specific gene *spo0A* and molecular primers to sequence this gene and infer the phylogenetic diversity of endospore-forming bacteria.

The second method (chapter 3) describes a treatment applied to a sediment sample to destroy vegetative cells and single out endospores in sediment before subjecting the endospore fraction to DNA extraction and sequencing. Both methods (chapter 2 and 3) were tested on sediment samples from Lake Geneva (Switzerland-France) and Lake Baikal (Russia).

The methods were applied in the two subsequent chapters based on a sediment core from Lake Geneva spanning the time from 1921 to 2010. Chapter 4 established the use of endospores as paleoecological indicators for Lake Geneva and related changing diversity patterns to the lake eutrophication period from 1960 to 1990. In chapter 5, the diversity analysis of endospore-forming bacteria in the sediment was applied in higher resolution and the treatment to select for endospores was added to four samples. With this, the differentiation between vegetative cells and dormant endospores was achieved and could successfully be used to establish the potential of endospores as paleoecological indicators. Finally, a synthesis of all results and perspectives is given in chapter 6.

1.2 Background

Anthropogenic pressure on the environment

The Earths' environment has been subjected to severe anthropogenic pressure since the industrial revolution triggered large demographic increase and fast economic growth (Park 2001). High production and consumption today require unsustainable amounts of minerals, water, food, fibres, and energy and produce large amounts of waste. Human activity is not negligible and has influenced global biogeochemical cycles and climate since the industrial revolution (IPCC Core Writing Team and Pachauri 2007). For example, since the implementation of the process of Haber-Bosch, nitrogen as fertilizers now accounts for over 50% of the global terrestrial nitrogen input (Ducklow 2008). This process has rapidly increased the load of nitrogen on Earth and became one of the most urgent environmental problems of this decade (United Nations Environment Programme (UNEP) 2012). In addition to the input of nitrogen, human activity and industrial processes also release high amounts of other nutrients (phosphorus and carbon), and a variety of metals and organic pollutants into the environment. More space is also required for human activities, resulting in habitat transformation, land degradation or fragmentation. This high demand of material and space, and an ever increasing environmental footprint threatens not only ourselves but the entire Earths' ecosystem (Walther et al. 2002).

Ecosystems are complex systems with large numbers of biological units (biodiversity) and a network of interdependent biological interactions (ecological balance). Changing any parameter in this balanced system, can set in motion feedback mechanisms that throw the ecology entirely out of balance, resulting in a biological state shift (Barnosky et al. 2012). An example of such a biological state shift has been observed around the Antarctic Peninsula where feedback loops due to global warming and melting of sea-ice, the absorbency of heat via radiation, and a reduced albedo effect, have led to magnified warming of the local surface waters. This change of sea-ice cover and higher water temperatures have affected the entire food-web from phytoplankton communities up to marine mammals and penguins (Ducklow et al. 2007). Another scenario of state shifts is the transformation and deforestation of pristine forest land. In addition to destroying the habitat of thousands of species, the removal of photosynthetic trees, essentially turns these regions from net carbon storage sites into carbon emitting sites (Barnosky et al. 2012).

Apart from global-scale scenarios such as climate change or nutrient cycling, numerous environmental impacts have been detected that are confined to regional or local scales. A prominent example of a local scale environmental problem is the eutrophication of freshwater lakes.

The case of lake eutrophication

Freshwater lakes are of high ecological, economical and recreational value (Wilson and Carpenter 1999; Dudgeon et al. 2006), but at the same time, they are heavily

impacted by anthropogenic activity. Lakes have large coastal land to volume ratios, which makes them extremely vulnerable to anthropogenic perturbation on the coastline and in the catchment area. Ecological problems such as water pollution, algal blooms, overfishing, introduction of invasive species, sedimentation or eutrophication have been detected in lakes from many parts of the world (Ogutu-Ohwayo et al. 1997; Madenjian et al. 2002; Le et al. 2010; Anneville et al. 2002; Smith and Schindler 2009).

The word "eutrophication" describes the shift in nutrient state of a water body from oligo- or mesotrophic (low to medium nutrient load) to eutrophic or hypertrophic state (high or very high load of nutrients). The trophic state is generally defined by the nitrogen and phosphorous concentrations. Causes for eutrophication of freshwater lakes are mainly sewage input, run-off of urban or agricultural fertilizers, livestock waste, atmospheric deposition and erosion (United Nations Environment Programme (UNEP) 2012).

Ecological impacts of eutrophication are manifold (Smith and Schindler 2009). The most important impact is linked to an increase in primary production due to the higher nutrient availability. This rise in primary production is usually accompanied by a shift in phytoplankton communities (Anneville et al. 2002). Also high primary production results in large amounts of organic matter in the water column, which increases the activity of microbial degradation of dead organic material. The process of biomineralization of organic material consumes oxygen, leading to a decrease in available oxygen, particularly in the deeper part of the lake (hypolimnion). The resulting anoxia threatens the survival of aerobic organisms, preeminently of fish and zooplankton communities (Vonlanthen et al. 2012; Brede et al. 2009). Furthermore, anoxic sediment can lead to a release of previously trapped phosphorous in the sediment, starting a vicious cycle, without a sink for phosphorous, particularly in iron depleted systems (Smith and Schindler 2009).

Smith and Schindler (2009) conclude that a better understanding of the interaction between nutrient load and biological, physical and chemical variables is needed in order to predict ecosystem impacts of lakes.

Environmental management

Awareness of ecological problems and discourses about sustainability issues have been growing in the last 50 years, among scientists, policy makers, and society as a whole. In consequence, an increasing number of studies have been conducted and conservation policies are regularly implemented. According to the United Nations Environment program, the priority issues to date for conservation management are climate change, freshwater and air quality, chemicals and waste, and biodiversity (United Nations Environment Programme (UNEP) 2012).

For freshwater lakes in Europe, guidelines for environmental management of freshwater systems have been defined under the European Water Framework Directive (European Union (EU) 2000). The objective defined for 2015 is to reach a "high and natural ecological status" of the lakes in Europe. "Natural" in this sense means the state that the lake biology was in, before any anthropogenic impact has occurred. This is challenging as the "natural" reference conditions are difficult

to derive because at the time of anthropogenic impact, regular measurements of lake parameters were not yet taken. Furthermore, as parameters depend on the individual ecosystem studied, it is difficult to unify and define general standards for a good ecological status.

For any ecosystem, the research and management of local environmental impact is challenging, because local heterogeneity makes it difficult to predict ecosystem changes. Even though, global averages of change (i.e. temperature averages) can be measured, they do not necessarily reflect changes on a local level. Also the implementation of these guidelines is still a problematic issue. During the last assessment of the EU in 2007, 40% of the surface water bodies in Europe were at high risk of not reaching the goals set for 2015 (Commission of the European Communities 2007).

Biological Forecasting

One of the main issues in environmental management is the ability to predict and estimate biological stability and ecosystem state shifts in regard to anthropogenic pressure, which has been dubbed "biological forecasting" (Barnosky et al. 2012). The prerequisite for biological forecasting is to know the biodiversity, the ecosystem interactions as well as the stability and potential of recovery of ecological systems. This background knowledge is needed to make reliable predictions on the future behaviour of ecosystems (Ducklow 2008).

Biodiversity

Biodiversity can be defined as the number of biological units in a system. The units can be defined on different levels, the smallest unit is the genetic diversity of a species (number of mutations). Then local level units can be defined, i.e. the number of species in a sample. This is the most common unit, which will be described further down in the section on microbial ecology. Finally, biodiversity units can be measured on a habitat level, i.e. numbers of biological groups within a habitat - or on a global level comprising all of the smaller units (Park 2001).

Biodiversity (for example as species abundance in a sample) and its variation is one of the simplest measures to detect impact of environmental change (Ager et al. 2010). Studying biodiversity however, has inherent challenges, as it depends on the operational definition and it has strong spatial and temporal components. Also, biodiversity is never static, and as such the rate of change in species abundance or adaptation to an environmental perturbation are also measures of biodiversity; as are the interactions among the different biological units of an ecosystem. Such ecosystem interactions include for instance symbiotic relationships between two organisms or metabolic interdependence (e.g. autotrophic and heterotrophic organisms).

To date all organisms can be divided in three groups, called "domains of life". The three domains are the Eukaryotes (organisms containing cells with a nucleus like fungi, plants, protists and all macroorganisms), the Archaea and the Bacteria (Woese et al. 1990). The latter two are frequently but inadequately dubbed the

Prokaryotes. The word pro-karyotes (before nucleus) insinuates the evolutionary development of Bacteria and Archaea into Eukaryotes, which is suggested as incorrect, because Bacteria and Archaea are paraphyletic, meaning they have evolved in parallel (Pace 2009).

Historically, the majority of biodiversity studies have been done on Eukaryotes, in particular macroorganisms that are easy to detect and observe. The vast majority of organisms (based on mass), however are microorganisms defined by their microscopic size (Whitman et al. 1998).

Ecosystem functions

In addition to pure bio-inventories, the conservation of biodiversity is directly linked to the conservation of associated ecosystem functions. Ecosystem service, the provision of functions and services provided that benefit humans, has become a key parameter and main priority for conservationists and environmental managers (World Resources Institute 2005). The goal of conservation science is to decipher the main links and functions of an environmental system so that eventually a monetary value can be attributed to it. If we can translate ecosystem services to actual economic values, the argument for conservation becomes strong.

However, ecosystem functions are highly complex and translating them to numbered values is challenging. There are large numbers of interactions and ecosystems are usually extremely dynamic. Also, functions work at different time scales, some of which are difficult to study. For example the breakdown of recalcitrant organic molecules happens at very low rates and is therefore not trivial to detect and follow (Craine et al. 2010).

The black box of microbes

Many studies of biodiversity and ecological interactions have been done on plants, invertebrates and macro-organisms (Walther et al. 2002), however, microorganisms are rarely taken into account. In the synthesis report on economics of ecosystems and biodiversity of the UNEP, microorganisms are only mentioned with respect to waste-water treatment, but not linked to any other ecosystem services (United Nations Environment Programme (UNEP) 2012).

The same issue can be seen in the EU water framework directive, where the indicator groups for the definition of ecological status of freshwater lakes in Europe are algae, daphnia and fish (European Union (EU) 2000, Annex V), while no member from Bacteria or Archaea is taken into account.

For a conclusive analysis of Earths' biodiversity and its functions and to be able to make predictions and implement conservation policies, a collective effort is needed to tap into the black box of microbial ecology. Bodelier states accurately "..., a fundamental problem is that the great majority of microbial diversity is unknown in terms of species identity and associated functional characteristics. Hence, we do not know what to conserve but we also do not know what we have lost." (Bodelier 2011).

Microorganisms are distributed among all three domains of life. For the continuation of this thesis and simplicity, the focus will be on bacterial ecology only.

Importance of bacteria

Research on bacteria is fuelled by three main fields of interest: medicine (threat of pathogens for humans), industry (use of bacteria as agents to produce industrial or food products) and environment (mainly because of bacterial importance for global element cycling). Environmental research is however largely underrepresented in research efforts and funding, compared to medical or industrial microbiology. Nonetheless, significant advances have been made on microbial ecology and ecosystem functioning in the last 30 years.

Microbes have been the only living forms during more than 75% of the history of this planet and are still today the most abundant group of organism in numbers and mass (Ducklow 2008). Global abundance of bacteria is estimated to be 4 to 6 x 10^{30} cells (Whitman et al. 1998). Bacteria and Archaea are also the most important in respect to activity and cumulative mass. Cumulatively, they store 60 to 100% of the total carbon of plants (Whitman et al. 1998). An estimation of 2 x 10^{9} cells in one gram of soil has been done, and these two billion cells are supposedly distributed among a diversity of 8 x 10^{6} individual species (Gans et al. 2005). Some of these calculations have however been disputed to overestimate the richness of this system (Volkov et al. 2006). What is uncontested is that many studies on bacterial diversity have been grossly undersampled (Gilbert and Dupont 2011). And this undersampling is particularly biased against the taxa present at low abundance, the so called rare biosphere (Konopka 2009; Sogin et al. 2006).

1.3 The discipline of microbial ecology

Microbial ecology is the science studying diversity, distribution and environmental functions of microorganisms.

To study biodiversity, two general indices are evaluated, the richness of an ecosystem and the relative abundance of the different groups. In microbial ecology, biodiversity richness is usually defined on a species unit level, i.e. the number of operational taxa (defined by 97% sequence similarity of the 16S rRNA gene). With the two parameters richness and abundance, basic statements about the environment can be made and different systems can be compared. Also, these indices answer questions about occurrence of endemism versus ubiquity and dominant versus rare biosphere (Fuhrman 2009). Microbial ecology has a strong temporal or spatial component, where bacterial diversity and the change of communities is recorded over time or among geographical gradients.

A second important component in microbial ecology is the biological functions of microbes, i.e. their metabolisms, activity, growth rates and other functional parameters of an environment.

Environmental function of bacteria

Bacteria are responsible for most of the biogeochemical cycling, they regulate climate and atmospheric components, purify water and contribute to about half of the primary production (Ducklow 2008). The Earths' nitrogen cycle for example was, before humans started to interfere, solely mediated by microbes (Falkowski et al. 2008). These environmental functions are typical examples of ecosystem services. A list of ecosystem services provided by bacteria is shown in Table 1.1.

Table 1.1: List of ecosystem services provided by bacteria.
Table modified from Ducklow 2008; Bodelier 2011.

Bacterial process	**Ecosystem service**	**Bacterial group**
Mineralization of organic matter	Decomposition, nutrient cycling, water purification, climate regulation	Heterotrophic bacteria
Photosynthesis	Primary production, C sequestration	Phototrophic bacteria
Transformation of specific elements (N, S, F, C, metals)	Nutrient cycling, water purification, climate regulation	Chemo(litho)autotrophic
Production of metabolites (antibiotics, polymers, enzymes)	Production of industrial or pharmaceutical products	Diverse

There are some challenges to link species identity with environmental functions (Bodelier 2011) because of difficulties in species definition, the metabolic versatility of bacteria (same species often have several metabolisms), and the small database of culturable species that we have. And last, it is the challenging to study bacteria because of their microscopic size. Genetic identity is dissociated from functional diversity (Konopka 2009). About 30% of genes in sequenced genomes have not been attributed a function (Konopka 2009). So far, no conclusive relationships (some were positive, others without or negative correlations) between diversity and functional potential have been found (Ducklow 2008).

In order to link ecology to function, an interesting analysis on the ecological coherence of taxa has found that the distance between two communities of different habitats is larger than the difference between two communities of a similar habitat, even though they are spatially far apart (Philippot et al. 2010). The former study has shown that on a phylum level, there are habitat specific associations, for example specific clades of bacteria found mostly in freshwater environments and other clades associated to soil. Habitat-specific communities have also been observed in global metagenomics studies (Tringe et al. 2005; von Mering et al. 2007; Delmont et al. 2011). There are propositions to divide the bacterial community of a particular habitat into two parts (Pedrós-Alió 2006). One part is composed of the dominant members of the community, which are likely the main drivers for the metabolisms in that environment, and the other part are the rare members, a large group of species in low abundance which have very slow or near to zero metabolism . The latter group are part of the so-called seed bank, globally dis-

tributed bacteria in low abundance that can take over the dominant fraction when the conditions turn in their favour.

Four main priorities for the conservation of microorganisms have been defined (Cockell and Jones 2009). These are communities involved in global biogeochemical cycling, communities involved in regional or local cycles (e.g. coral reefs), communities with industrial or medical use, and microorganisms with intrinsic value (aesthetics of microbial mats). Particularly in regard to anthropogenic influence and ecosystem pressure, the resistance or sensitivity of microbial ecosystems to change, their resilience and the stability of the ecological balance are crucial factors. An example of low resilience is the slow re-establishment of phytoplankton communities in Lake Geneva after eutrophication. The phytoplankton community from before the eutrophication event has been re-established today with a lag phase of several years after the lake was back in mesotrophic state (Tadonléké et al. 2009). Predictions are highly challenging, because of the unknown link between environmental function and species. An issue that has been raised is the functional redundancy, meaning that the function does not disappear just because some species do (Bodelier 2011). Many species with the same function usually coexist in an environment so that the main function gets shifted to another member. This aspect of functional redundancy however, is most likely an issue of scale of observation. On a larger scale (several grams of soil, homogeneously mixed lake water) there is probably high functional redundancy and losing species does not influence the ecological function of these particular environments. However on a smaller scale such as a patch of animal skin or along a plant root, there are only a few species, hence the functional redundancy is smaller (Bodelier 2011).

Methods in microbial ecology

For more than a century, the principal methodology used in bacterial ecology consisted of microscopic observation and the isolation and culturing of strains in the laboratory, in order to characterize their phenotypes (physiology and metabolisms). Since Staley and Konopka introduced the "great plate count anomaly" (Staley and Konopka 1985; Amann et al. 1995), revealing that only a small fraction of the microbial community can be cultured in the laboratory, one of the great challenges in environmental microbiology has been the understanding of the diversity and metabolic capabilities of microbes in a culture-independent manner. In the second half of the 20th century, as new methods to reveal the genetic content of microbes started to emerge, this bias was partly overcome by moving into the direction of directly extracting genetic material from environmental samples. With the invention of the polymerase chain reaction (PCR) it became possible to amplify minute amounts of DNA from environmental samples.

Metagenomics

Research using the genetic information directly extracted from an environmental sample is called metagenomics. In the early 1990s, the first studies of environmental genes were done by cloning gene fragments from environmental samples

into Lambda phages (Giovannoni et al. 1990; Schmidt et al. 1991). Both studies were based on the 16S rRNA gene, a good candidate gene for a universal bacterial phylogeny, introduced by Carl Woese (Woese 1987). The rRNA gene is promising, since it is not laterally transferred, conserved and appears in all bacterial cells (Woese 1987; Pace 2009). The species concept in bacterial ecology is since then based on nucleotide similarity at 97% sequence identity of the 16S rRNA gene or 70% DNA-DNA hybridization of the full genome.

The term "metagenomics" was coined by Handelsman et al. (1998), who used the method of cloning DNA fragments extracted from bulk soil with bacterial artificial vectors (BAC) and *Escherichia coli* clones to detect novel biochemical molecules. "Meta" stands essentially for the the collection of genomes (higher level than one individual genome) found in environmental samples.

First sequencing reactions were based on Sanger sequencing, which relied on dideoxy-nucleotides that act as chain terminators in the DNA synthesis (Sanger et al. 1977). The terminated products of different lengths (labelled by radioactive compounds), were then analysed by gel electrophoresis and the gene sequence inferred. Such metagenomics studies were used for the collective study of genomes directly from environmental communities (reviewed in Warnecke and Hugenholtz 2007; Handelsman 2004).

Alternatives were later developed to directly sequence random fragments of total environmental DNA, so called shotgun sequencing (Venter et al. 2004; Tyson et al. 2004). Early shotgun sequencing studies were done using an automated Sanger method, by capillary electrophoresis based on differential fluorescent dyes for the four dideoxynucleotides (Tyson et al. 2004). Such random cloning or shotgun sequencing enabled sequencing random fractions of the genomes or metagenomes. When done in high coverage, the sequence of the entire genome or metagenome could be deciphered.

In 2006, a new method emerged, the first pyrosequencing study. This method was applied to microbial ecology of a deep mine sample (Edwards et al. 2006). Pyrosequencing, in contrast to the Sanger method, is not based on DNA terminators but on the detection of a pyrophosphate release during the DNA synthesis and a subsequent light signal (Ronaghi et al. 1996). The first instruments to use this method were 454 Roche instruments. To date, with this technology approximately 1 million reads of 500-800 bp can be created in one run (Zarraonaindia et al. 2013). Pyrosequencing studies have substantially higher sensitivity and can detect up to two orders of magnitude more taxa than clone libraries or traditional fingerprinting methods like automated ribosomal intergenic spacer analysis (ARISA) (Fuhrman 2009). The sensitivity of the metagenomics method is highly crucial, particularly if a descriptive study of a particular community is done.

In recent years, more and more sequencing studies are being performed with yet another technology from Illumina (Bentley et al. 2008). This method is based on reversible terminator chemistry and a specific dye for each nucleotide. The reaction is a single base extension and the light signal based on the added base is generated after each extension. The current Illumina sequencers provide 2-3 billion reads at approximately 150 bp length per run (Zarraonaindia et al. 2013).

Metagenomics offers an opportunity to reveal the diversity and metabolic ca-

pabilities of the microbial world and numerous metagenomics sequencing studies have been done in a variety of environments such as animal guts (Pope et al. 2010; Warnecke and Hugenholtz 2007), human gut (Gill et al. 2006; Kurokawa et al. 2007), compost (Allgaier et al. 2010), a saline lake (Ferrer et al. 2011), arsenic rich mine drainage (Bertin et al. 2011), acid mine drainage (Tyson et al. 2004), hydrothermal vents (Brazelton and Baross 2009), coral reefs (Dinsdale et al. 2008; Wegley et al. 2007), heavy metal contaminated ground water (Hemme et al. 2010), stratified microbial mats (Kunin et al. 2008), oceans (Rusch et al. 2007; Venter et al. 2004) and marine sediment (Quaiser et al. 2011). In addition, metagenomics has helped to unveil novel microbial metabolisms (Béjà et al. 2000, 2001; Voget et al. 2003; Venter et al. 2004; Ram et al. 2005). In low diversity environments, it was even possible to reconstruct the almost complete genomes of strains (Tyson et al. 2004; Bertin et al. 2011), which is not likely in high diversity samples like soil or ocean water. Nonetheless, bacterial diversity studies have added substantial knowledge on microbial ecosystems, for example the functioning of the ocean microbiology (Gilbert and Dupont 2011).

Molecular analysis of environmental samples have revealed a large diversity of microbes, going far beyond the diversity known based on cultured organisms. There are now entire phyla, for example TM7 (Torf, mittlere Schicht (Rheims et al. 1996)) that are composed of sequences from culture-independent studies, without any isolated member (Pace 2009; Harris et al. 2013; Hugenholtz et al. 2001).

From all past sequencing efforts, a large collection of 16S rRNA gene sequences is now available in databases for the further use of researchers. As of September 2013 over 4.2 million ribosomal RNA gene sequences (16S, 18S, 23S, 28S) are available in the Silva database (Silva, `http://www.arb-silva.de/documentation/release-115`. Release 23.8.2013, accessed 3.9.2013). These sequences are distributed among 63 different phyla in the domain Bacteria. From the Ribosomal Database Project (RDP), on which part of the sequence analysis in this thesis relied, over 2.7 million aligned and annotated sequences are available from the newest release in May 2013 (`http://rdp.cme.msu.edu/misc/rel10info`, released 14.5.2013, accessed 3.9.2013). In addition to ribosomal gene sequences, full genome sequences from over 4000 different bacteria are now available in Genbank (`http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial-taxtree`, released shortly before US Gov. shutdown, accessed 13.10.2013). The analysis of metagenomic data sheds light on community structure, genetic diversity and ecological (functional) role of micro-organisms. Furthermore it enables researchers to screen genomes for new genes and proteins with possible application in biotechnology such as antibiotics or novel enzymes.

Challenges in metagenomics

Together with the explosion in numbers of metagenomics studies, critical views and potential challenges have started to emerge as well. There is a large set of methodological issues, from sampling, to DNA extraction and PCR biases, to issues in sequencing methods. A good overview of potential biases in metagenomics studies in soil samples is given by Lombard (Lombard et al. 2011).

Methodological challenges

For DNA extraction, not all microbial species are equally amenable to the methodological approaches used today, especially considering the richness of morphological and physiological states in which microbes can be found in environmental samples (Morgan et al. 2010). Thus, similar to a percentage of the community not being culturable, we can imagine that a fraction of the genomes is not extractable. This issue is raised as a part of chapter 2 of this thesis. Projects to catalogue diversity of entire microbiomes that involve entire research consortia are under way such as the terragenome project, the global ocean survey, or human microbiome project (Vogel et al. 2009; Turnbaugh et al. 2007; Parthasarathy et al. 2007). An aspect that these projects include are clear standards for techniques and methods and also guidelines to measure and describe metadata, an aspect often neglected in previous studies.

For example large variations between community analysis among different researchers and laboratories have been revealed (Pan et al. 2010).

Phylogenetic relationships

For phylogeny studies, in recent years, the paradigm of the 16S rRNA gene as the best gene to infer phylogeny has been shifting. The rRNA genes are for some analyses too conserved as they do not discriminate between species or strains (Pace 2009). Also, 16S rRNA genes are multi-copy genes, which leads to over-coverage of the dominant members. Still only little is known about the phylogenetic relationships between the different bacterial groups, which is for example shown by the fact that the actual rooting (evolution) of the bacterial tree is unknown (Pace 2009). Other genes (present in single copies) have been proposed for phylogenetic studies, for example the *recA* gene involved in recombination and DNA repair, which is consistent with the ribosomal RNA gene phylogeny (Lloyd and Sharp 1993). Recently, the evolution of bacteria has been studied more closely and based on 25 protein-coding genes in addition to the rRNA gene (Battistuzzi and Hedges 2009).

Issues of coverage

For shotgun metagenomics, we have still only covered a small fraction of the actual diversity. Based on calculations of the number of tera bases of data that a latest technology sequencing instrument can produce and the estimated diversity of ocean water, essentially only 0.00001% of the community can be covered (Gilbert and Dupont 2011). Sequencing costs are still too elevated to sequence the entire metagenome of 1 gram of soil or one ml of seawater (Zarraonaindia et al. 2013; Gilbert and Dupont 2011). Furthermore, the small coverage makes the assembly of genomes in environmental samples almost impossible. Only in low diversity samples (two species) the entire genomes could be assembled (Tyson et al. 2004).

Computational challenges

There are two approaches to analyse metagenomics data. One is the analysis based on reads mapped to known genomes from a database and the second is the analysis based on *de novo* assembly of the metagenome (Scholz et al. 2012). The read-based analysis relies on good quality and quantity genomes in databases that are often incomplete (McHardy et al. 2007). Improvements could be made by automated and streamlined sequence annotation and comprehensively archived sequence data made available to use for other researchers (Raes et al. 2007; Gilbert and Dupont 2011). *De novo* assembly on the contrary is challenging, because it is computationally heavy and is only possible if there is high coverage of the environmental sample (Miller et al. 2010; Scholz et al. 2012). For both methods, metagenomics needs computational resources and specialist bioinformatics support.

With ever larger sequence datasets it will be necessary to develop statistical methods that deal with the issue of richness estimation, coverage and the analysis of the rare biosphere (Quince et al. 2008).

Due to these issues and the biological complexity, the emerging argument is that purely descriptive metagenomics studies will not provide knowledge on the functioning of ecosystems. Some researchers therefore suggested to move away from pure sequencing studies towards more hypothesis driven sequencing and functional studies (Gilbert and Dupont 2011; Jansson and Prosser 2013).

Targeted metagenomics

One solution to some of the above mentioned problems is targeted metagenomics, where a sub-community or single cell is isolated and then subjected to metagenomics sequencing (Suenaga 2012). A schematic overview of metagenomics and targeted metagenomics methods is given in Figure 1.1. Briefly, four different approaches can be made. In shotgun metagenomics the entire DNA pool is randomly sequenced and later re-assembled. In amplicon metagenomics a specific DNA fragment is amplified and sequenced. The targeted shotgun metagenomics study is a random sequencing of a DNA pool that was separated prior to extraction (examples in Table 1.2). Finally, in the targeted amplicon sequencing approach the focus is on one gene of a previously separated fraction of the community.

Today, there are new technologies, by which the genomes of individual cells from environmental samples can be sequenced. By multiple displacement amplification (MDA) enough genomic DNA from a single cell can be generated (Raghunathan et al. 2005). Single cell shotgun sequencing studies of over 200 genomes across diverse phyla, in particular poorly known phyla, have recently been published (Rinke et al. 2013). With this analysis the position of previously uncultured phyla in the tree of life could be determined and overall insights into the phylogeny of bacteria were gained (reviewed by Lasken 2012).

In this thesis, two targeting methods for studying the community of endospore-forming bacteria are described. The first method, amplicon metagenomics, is based on the functional gene *spo0A*, is described in chapter 2 and an application is described in chapter 4. The other targeting method (targeted amplicon metage-

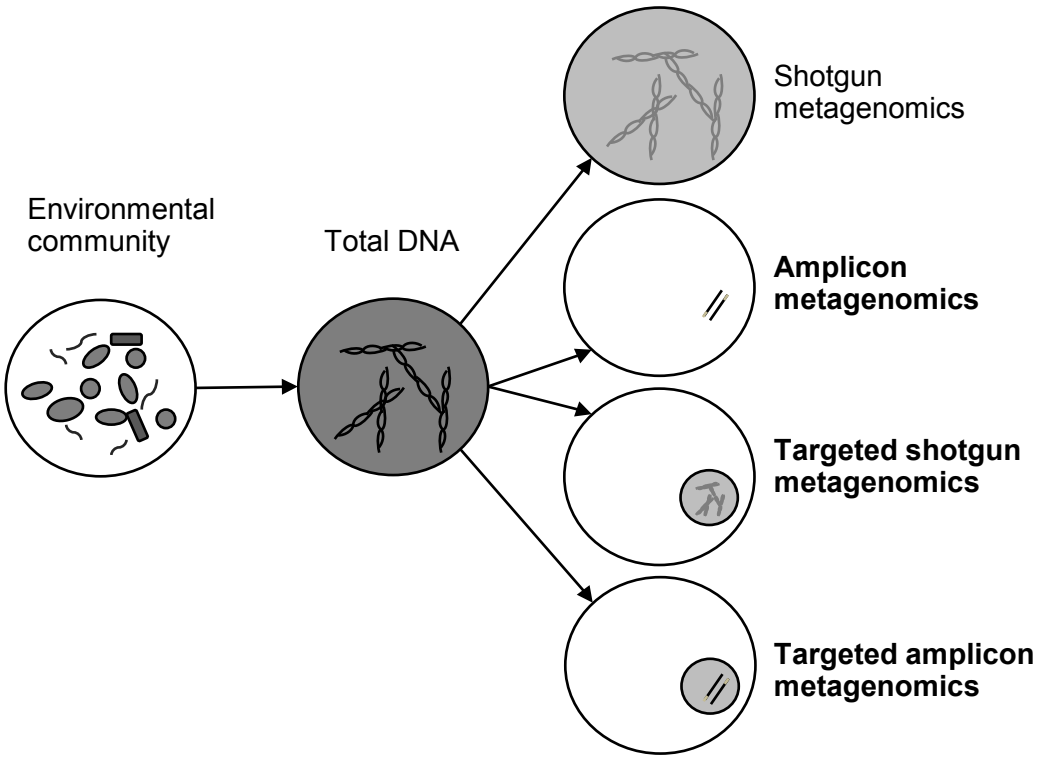


Figure 1.1: Overview of targeted metagenomics methods

nomics) is based on a physical separation of endospores from the vegetative cell pool and subsequent metagenomics analysis of the endospore community. This method is described in chapter 3 and its application on a Lake Geneva sediment core is shown in chapter 5. Some background and state of the art in environmental research about endospore-forming bacteria is given in the subsequent sections of this introduction.

1.4 Endospore-forming bacteria

Endospore-forming bacteria is a group of bacteria classified in the low G-C Gram-positive phylum Firmicutes. Endospores are resistant cell structures that are formed when the cells sense nutrient depletion. The spores can endure environmental stress conditions and germinate back to a vegetative cell when environmental conditions are favourable again. The formation of resting stages or spores is common to several groups of bacteria, but only members of Firmicutes form "true endospores" distinguished by the production of dipicholynic acid (DPA) and their resistance to heat.

Table 1.2: List of examples of targeted metagenomic studies.

Method	**Target group**	**Reference**
Single cell genomics	uncultivated strains	Rinke et al. 2013
Gene targeting	heat shock protein (cpn60)	Hill et al. 2004, Dumonceaux et al. 2006
	dioxygenase gene (bphA1)	Iwai et al. 2010
	butyrate synthesis gene (but)	Vital et al. 2013
	ammonia-oxidizing gene (amoCAB)	Junier et al. 2009b
Enrichment	kimchi, lactic acid fermenters	Jung et al. 2011
Enrichment and stable isotope probing (SIP)	methylotrophic communities	Neufeld et al. 2008
	methylotrophs	Kalyuzhnaya et al. 2008
Enrichment and immunocapturing	mineralization of dissolved organic carbon (dimethylsulphoniopropionate (DMSP) and vanillate)	Mou et al. 2008
Physical (size) separation	picoplankton	Cuvelier et al. 2010
Flow cytometry	Synechococcus	Palenik et al. 2009

Spores or spore-like structures

The formation of resting stages or spores is found among bacteria from the Myxococcales, Actinobacteria, Cyanobacteria and Firmicutes. Most bacteria that can form resting stages are Gram-positive bacteria or those with Gram-positive type cell walls (Madigan et al. 2006). The exception are the Myxococcales, which are true Gram-negative bacteria and members of the class Deltaproteobacteria.

Myxoccocales form aggregated fruiting bodies with resting stages called myxospores (Battistuzzi and Hedges 2009). Myxococcales are biomolecule-degrading bacteria that are ubiquitously distributed on Earth, but principally inhabit soil (Reichenbach 1999). The myxospores are resistant to high temperature and desiccation. However, in contrast to endospores, the myxospores are resistant only to dry heat ant not wet heat (Reichenbach 1999).

The Actinobacteria can form different types of non heat resistant spores. Spore-forming Actinobacteria are the *Streptomyces* and *Micromonospora* (Ensign 1978). They are soil dwelling organisms adapted to fluctuating environmental conditions and porous media. Spores from Actinobacteria do not contain DPA (Ensign 1978). *Actionomyces* spores are mainly built under dry conditions, while humidification induces germination (Ensign 1978). A group formerly classified as Actinobacteria that form true endospores, the *Thermoactionmycetaceae* was recently re-classified and moved to the phylum Firmicutes (Matsuo et al. 2006).

Cyanobacteria can from resting stages (akinetes) triggered by light limitation. They can survive desiccation and cold temperatures but are not heat resistant (Adams and Duggan 1999). Akinetes are thought to provide a survival mechanism in aquatic habitats during winter (Adams and Duggan 1999).

Endospores

As the word endospore indicates, the spore structure is formed within the cell. The endospores are then released into the environment upon death of the vegetative cell and resist extreme conditions of heat, desiccation, chemicals, UV radiation or pressure (Nicholson 2000). In Figure 1.4, a scanning electron microscopy image shows free and mature endospores of *Bacillus subtilis*.

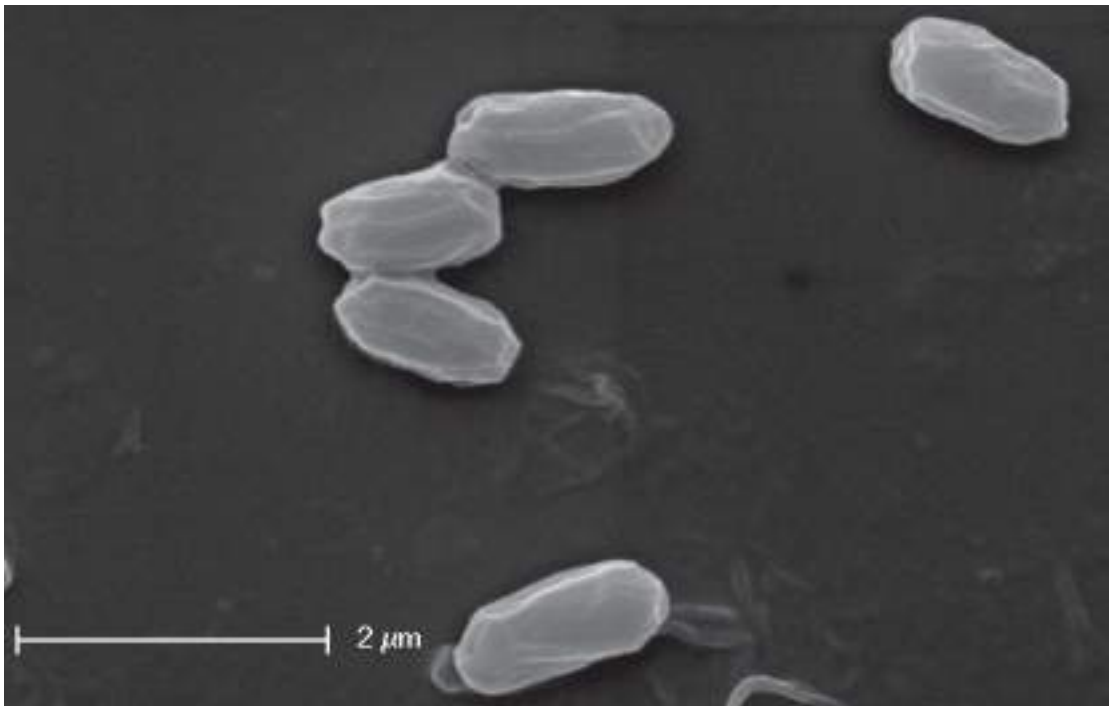


Figure 1.2: Scanning electron microscopy (SEM) image of endospores from *Bacillus subtilis*.

History of endospore research

Bacterial endospores were first described in 1872 by Ferdinand Cohn. He isolated a rod-shaped, aerobic bacterium from hay, with refractile spores. This strain was later named *Bacillus subtilis*. Shortly after, in 1876, Robert Koch described *Bacillus anthracis* and the detailed life cycle involving a vegetative cell state, endospore-formation and re-germination (Zeigler and Perkins 2009; Nicholson 2002). Not long after in 1880, the first anaerobic endospore-former *Clostridium butyricum* was described by Adam Prazmowski in his PhD thesis (Prazmowski 1880 cited in Dürre 2009).

Since then, numerous endospore-forming strains have been detected and described, primarily driven by medical interest of human disease related endospore-formers, but also in chemical industry (production of butyrate, antibiotics and other industrial chemicals) (Nicholson 2002) and the food industry (proxy for

sterilization tests) (Logan 2012). In fact, Firmicutes is today the second most represented phylum in culture collections (Hugenholtz 2002).

There are 843,509 sequences of bacteria from the phylum Firmicutes available from the Silva database as of September 2013 (Silva, `http://www.arb-silva.de/documentation/release-115`). However, most of the datasets of endospore-forming strains are heavily skewed towards strains that are of medical interest (Pace 2009).

Phylogeny

The phylum Firmicutes evolved about 3 billion years ago upon the colonization of land (terrabacteria) (Battistuzzi and Hedges 2009). The anaerobic Clostridia are thought to have diverged from the aerobic or facultative anaerobic Bacilli at about 2.8 billion years, during the Great Oxidation Event (Battistuzzi and Hedges 2009). The evolution as "terrabacteria" explains the physiology of Firmicutes with resistant Gram-positive cell walls and notably endospore-formation for some to escape stress conditions on land such as desiccation, heat or radiation (Battistuzzi and Hedges 2009). The best studied endospore-forming species is *B. subtilis*, which has served as a model organism for Gram-positive bacteria in numerous research laboratories since its discovery. The physiological properties as well as proteins and biochemical mechanisms of cells and endospores are known in detail. *B. subtilis* was among the first bacteria for which the genome was entirely sequenced and annotated (Kunst et al. 1997). Since then, there has been a considerable effort to sequence and annotate various endospore-forming bacteria isolates from different environments (Wu et al. 2005; Hu et al. 2008; Junier et al. 2009a). Firmicutes today comprise three classes, the Bacilli (18 families), the Clostridia (25 families) and the Erysipelotrichi (1 family) (Vos et al. 2006). The phylogenetic relationship of known endospore-forming genera is shown in the cladogram in Figure 1.3. The detailed phylogeny of Firmicutes however is still a point of discussion and taxa are regularly re-named and re-classified. In particular, the taxonomic resolution of the class Clostridia is still not well resolved (Collins et al. 1994). The taxonomy and phylogenetic relationships of endospore-forming bacteria are by far not comprehensive, even though considerable effort has been made to shed light on taxonomy (Collins et al. 1994; Yutin and Galperin 2013). The ability to form endospores is not consistent with the phylogeny, even though there are genera that are largely composed of endospore-forming members such as *Bacillus*, *Paenibacillus*, *Clostridium* or sulphate-reducing *Desulfotomaculum* (Fritze 2004; Galperin et al. 2012). In physiological studies, endospore-formation can in most cases not be ruled out definitely, because it is strongly dependent on culture and sporulation conditions. However, with genome sequencing it has become simpler to rule out endospore-formation or define asporogenic species that still house a majority of the genes necessary for sporulation but do not form endospores (Galperin et al. 2012; Abecasis et al. 2013).

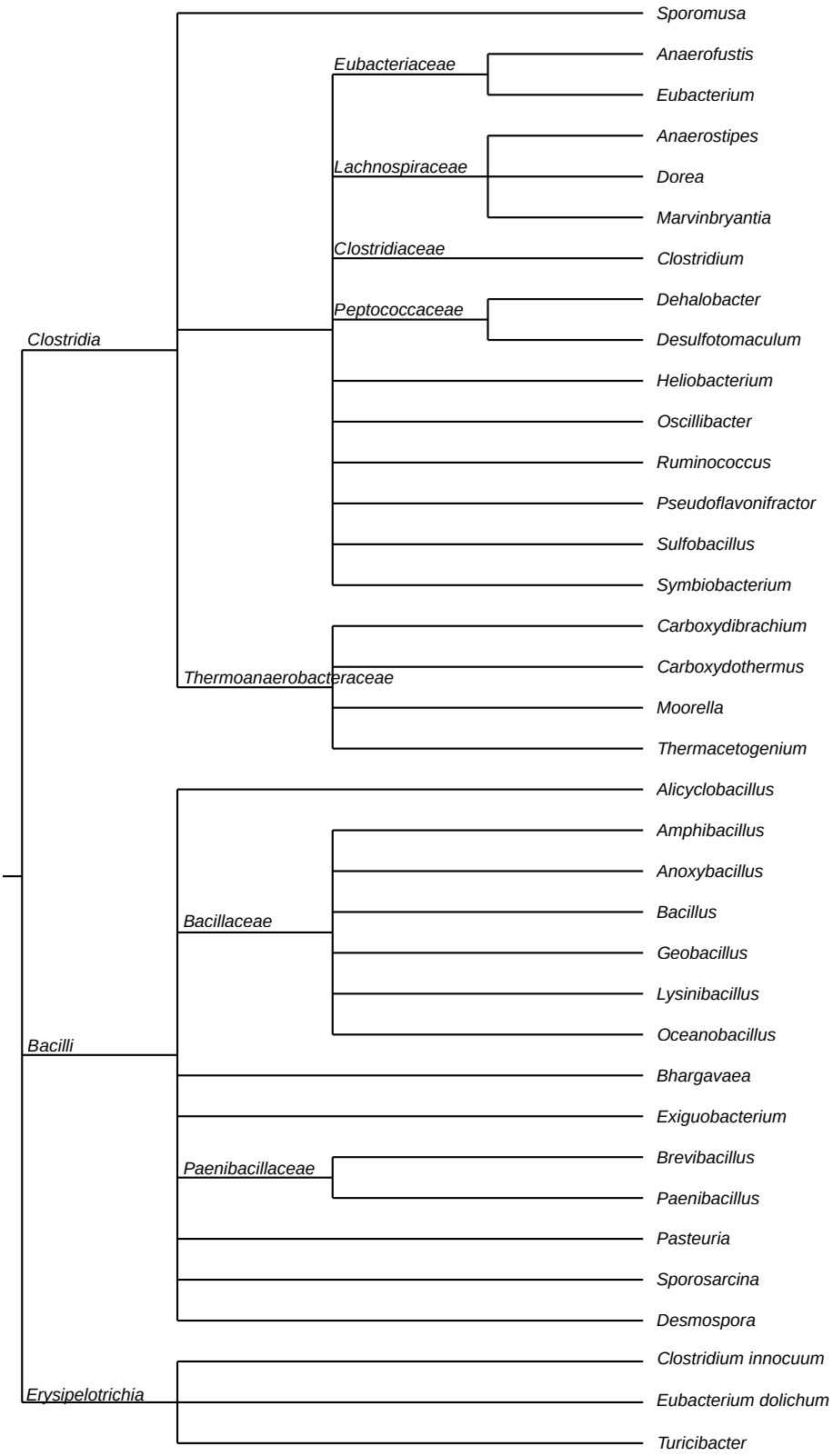


Figure 1.3: Cladogram showing the phylogeny of selected endospore-forming bacteria. Tree built based on NCBI taxonomy (`http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial-taxtree`) and drawn by Thomas Junier, printed with permission.

Metabolisms

Endospore-forming bacteria have a wide metabolic range. In Table 1.3 an overview is given of the different metabolisms found among endospore-forming bacteria. All endospore-forming bacteria are chemotrophs, with the exception of the genus *Heliobacterium*, that grow phototrophically.

Table 1.3: List of metabolisms found among endospore-forming bacteria.

Metabolism	**Oxygen**	**Specifics**	**Genera**
heterotrophic (chemoorgano)	aerobic		*Oceanibacillus* (Lu et al. 2001), *Sporosarcina* (Claus et al. 2006), *Desmospora* (Yassin et al. 2009), *Bhargavaea* (Verma et al. 2012)
heterotrophic (chemoorgano)	aerobic or facultative anaerobic	some N-fixators, nitrate reducers, complex organic matter (cellulose) degradation	*Amphibacillus* (Niimura et al. 1990), *Bacillus* (Claus et al. 2006), *Geobacillus* (Nazina et al. 2001), *Brevibacillus* (Shida et al. 1996), *Paenibacillus* (Shida et al. 1997), *Alicyclobacillus* (Wisotzkey et al. 1992), *Anoxybacillus* (Pikuta et al. 2000)
heterotrophic (chemoorgano)	strict anaerobic		*Eubacterium* (Wade 2006), *Anaerofustis* (Finegold et al. 2004), *Anaerostipes* (Schwiertz et al. 2002), *Dorea* (Taras et al. 2002), *Marvinbryantia* (Wolin et al. 2003), *Dehalobacter* (Holliger et al. 1998), *Turicibacter* (Bosshard et al. 2002)
heterotrophic or autotrophic (organo or lithotroph)	anaerobic	sulfate-reducers, metal-reducers	*Clostridium* (Wiegel et al. 2006), *Desulfotomaculum* (Stackebrandt et al. 1997)
mixotroph (organo or lithotroph)	anaerobic	sulfate reducers	Candidatus *Desulforudis* (Chivian et al. 2008), *Sulfobacillus* (Anderson et al. 2012)
autotroph (litho)	anaerobic	homoacetogens	*Sporomusa* (Möller et al. 1984), *Moorella* (Collins et al. 1994)
autotroph (litho)	anaerobic	CO-utilizer	*Carboxydothermus* (Wu et al. 2005)
phototropic	anoxygenic	also pyruvate fermentation	*Heliobacterium* (Bryantseva et al. 1999)

Sporulation

The genetic background of endospore-formation has been studied extensively in the last 25 years, mainly on the basis of mutants of *B. subtilis*. It was found that the sporulation process involves over 300 gene products (Onyenwoke et al. 2004; Eichenberger et al. 2004) and is steered by a cascade of proteins including sigma factors, each regulating signal molecules of the downstream pathway (Figure 1.4).

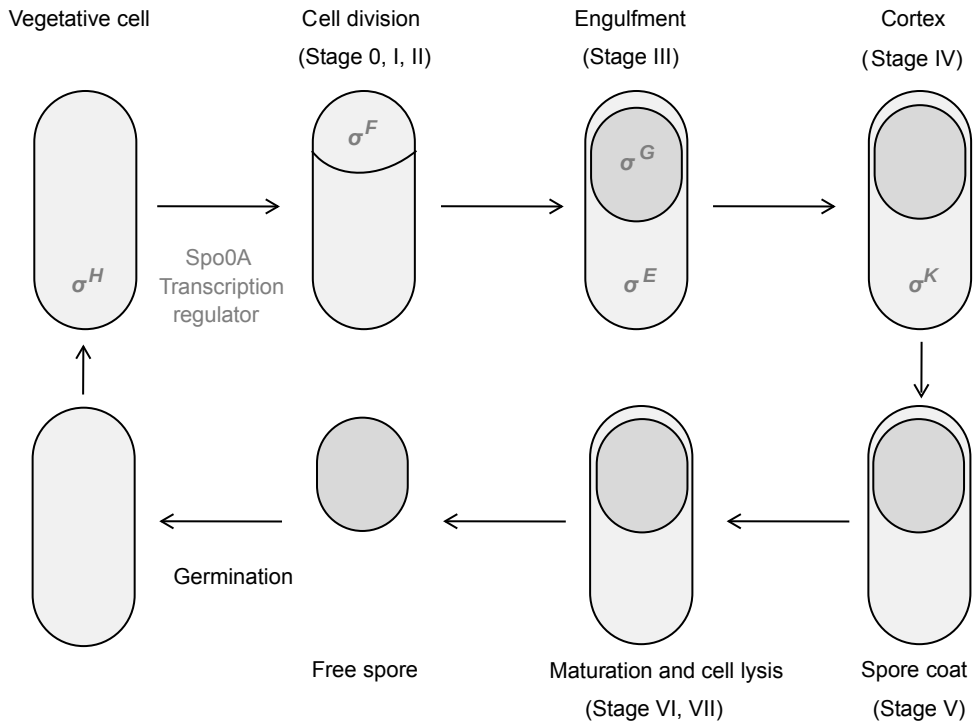


Figure 1.4: Schematic overview of the eight stages in the sporulation process.

The sporulation pathway is commonly divided into eight stages (Stage 0 to VII)(Errington 2003). The sporulation initiation (stage 0) is triggered by starvation and induces a phosphorylation cascade of gene products which eventually induce the transcriptional regulatory function of Spo0A and the sigma factor H (Snyder and Champness 2007). The amount of phosphorylated Spo0A-P in the cell determines the cellular action. At low Spo0A-P levels, synthesis of antibiotics and degradative enzymes is switched on. At higher Spo0A-P levels, several sporulation operons are activated and the cell irreversibly commits to sporulation. The first stage (I) in the sporulation process is the duplication of DNA and an asymmetrical cell-division. At this stage the sigma F and pre-sigma E are synthesized in the prespore. Sigma F activates genes involved in the maturation of the prespore (II) which is then engulfed by the mother cell to become the forespore (stage III). The activated sigma F also regulated synthesis of pre-sigma G, which gets activated in the mother cell during stage III, the engulfment. Activated sigma G in turn regulates activation of sigma K which is involved in the cortex synthesis (V) and final stages of spore maturation (VI) and release of free spore upon death of the mother cell (VII). The alternate activation of sigma factors during the sporulation process ensures a temporal and spatial order throughout the process (Zeigler and Perkins 2009). Certain proteins that build the spore coat and cortex are for example synthesized by the mother cell, while others are transcribed from the forespore DNA. Generally, no marked differences between the mechanisms of sporulation of the analysed members of Bacilli and Clostridia were

detected, with the exception of differences in the phosphorylation mechanisms of Spo0A (Dürre 2009). The duplication of the DNA prior to endospore formation ensures that endospores contain an entire copy of the bacterial genome, which is then surrounded by protective proteins (small acid soluble proteins) forming the core, enabling protection against UV light (Mason and Setlow 1986). The core is then surrounded by a spore cortex, responsible for heat resistance and an outer spore coat (Driks 2003; Henriques and Moran 2007). The inside of the spores is desiccated and stabilized with Ca^{2+}-minerals and DPA, inactivating the intracellular enzymes. Even though much is known about the sporulation pathway, the structure of endospores and their germination, there are still questions about the endospore physiology that remain unanswered. The exact function of DPA in combination with calcium in the spore for example is not known, and also the mechanism of endospore destruction in wet heat has not been studied well (Gould 2006). Also, endospores are considered to be metabolically dormant but they must still be able to monitor their environment, since, upon a favourable change in the surrounding conditions, the spores can germinate back into vegetative cells. The extent of dormancy versus environmental monitoring is subject of remaining open questions. The first report on redox reactions mediated by dormant bacterial endospores was published in 1982, showing that dormant endospores catalyse the oxidation of Mn (I) to Mn (III/IV)-oxides (Rosson and Nealson 1982). These experiments revealed that endospores are capable of altering their surrounding environment. Mn-oxidizing *Bacillus* were later also isolated from different marine sites (de Vrind et al. 1986; Francis et al. 2002; Dick et al. 2006). Furthermore, research was conducted on metal redox reactions by endospores of *Desulfotomaculum reducens* for uranium and iron (Junier et al. 2009a). Endospores have been shown to be resistant to desiccation, UV radiation, toxic chemicals, high salinity and other extreme conditions (Nicholson and Law 1999). It is assumed that the structure of spore coat and cortex as well as intraspore proteins, allows the spore to survive for extended periods of time. Viable endospores have been isolated from sediment cores with an estimated age of 5,800 years (Bartholomew and Paik 1966). The oldest spore-forming isolates are claimed to come from a primary salt crystal of 250 million years (Vreeland et al. 2000). And another spore-forming species was claimed to be isolated from a 25-40 million year old amber inclusion (Cano et al. 1994). Even though large efforts were done to exclude contamination, doubts remain and we cannot be sure that neither the endospores nor the crystal or minerals itself were in fact this old (Hazen and Roedder 2001; Fischman 1995). As an ecological trait, sporulation can be considered a survival advantage, since it preserves the genetic information through unfavourable environmental conditions in which non sporulating bacteria cannot survive. Endospore formation is therefore an elaborate mechanism for resistance and dormancy for extremely long periods of time and in extreme environments, which makes endospores interesting models for ecological questions and to study life in extreme environments. However, apart from medical and industrially important strains, few endospore-forming strains have been studied from an ecological viewpoint (Chivian et al. 2008).

1.5 Endospores in the environment

Isolation

Since endospores are highly resistant structures and provide a meaning for living bacteria to conserve their genetic information through environmental stress, they are assumed to play a significant role in the total microbial mass of extreme environments. Historically, upon development of sterile working conditions and culture media, endospore-forming bacteria were widely studied by isolation, culturing and subsequent morphological and physiological descriptions. Endospore-forming strains have been isolated in numerous studies and from almost every environment (Table 1.4).

Microscopy

In the early days of environmental research in microbiology, endospore-forming bacteria were studied by microscopy, for example of hay infusions (Cohn 1872 cited in Zeigler and Perkins 2009). Under phase-contrast bacterial endospores are easily distinguished, as bacterial cells are dark, while the light refracting endospores are phase bright (Figure 1.5).

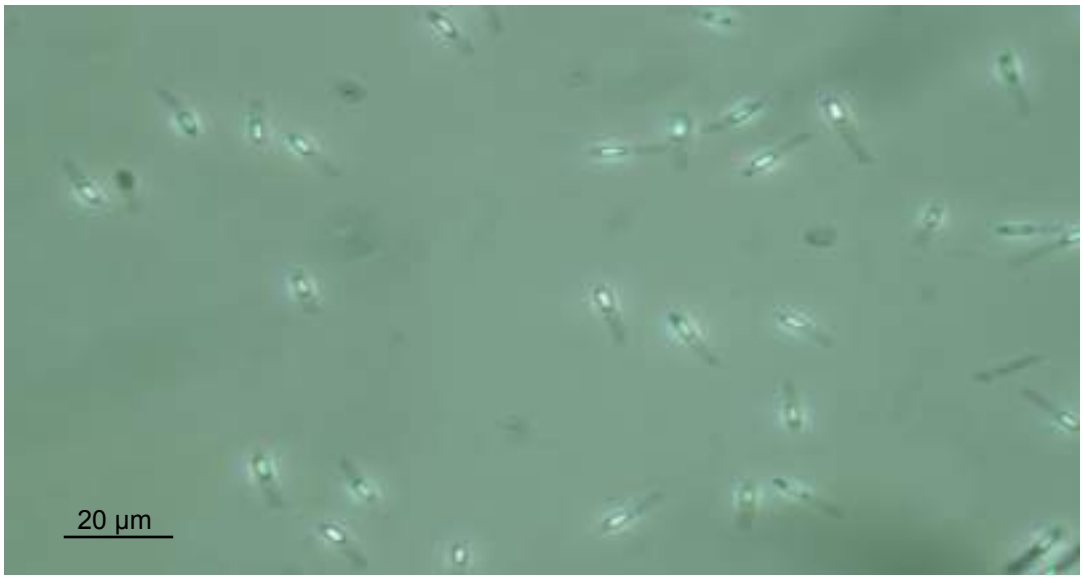


Figure 1.5: Microscopy image under phase contrast of *Paenibacillus alvei*. Cells are dark and endospores phase-bright.

Sporulation in environmental samples can be confirmed by the Schaeffer-Foulton staining (Mormak and Casida 1985), where cells and endospores are differentiated by two different colours (Figure 1.6).

The first targeting of spores directly in environmental samples was done by specific fluorescent antibodies for endospores and vegetative cells of *B. subtilis* and *Bacillus cereus* (Hill and Gray 1967). Tagging of endospores with fluorescent

Table 1.4: List of strains isolated from different environments.

Isolate	Environment	Reference
Alkaliphilus metalliredigens	heavy metal contaminated soils	(Ye et al. 2004)
Alicyclobacillus acidocaldarius	acid hot spring	(Darland and Brock 1971)
Bacillus spp.	hydrothermal sediments	(Dick and Tebo 2010)
Bacillus tusciae, *Kyrpidia tusciae*	geothermal area	(Bonjour and Aragno 1984; Klenk et al. 2011)
Bacillus schlegelii, *Hydrogenibacillus schlegelii*	volcanic areas, geothermal, lake sediments and air samples, geothermal Antarctic soil	(Bonjour et al. 1988; Kampfer et al. 2012)
Desulfotomaculum and Candidatus *Desulforudis audaxviator*	geothermal plants in Southern Germany	(Alawi et al. 2011)
Desulfotomaculum	cold Arctic ocean sediments	(Hubert et al. 2010; de Rezende et al. 2013)
Bacillus	oligotrophic granite rock	(Fajardo-Cavazos and Nicholson 2006)
Halobacillus sp., *Bacillus* sp. *Paenibacillus* sp. and *Paucisalibacillus* sp.	salt-attacked stone chapel	(Ettenauer et al. 2010)
Ureibacillus thermosphaericus	air	(Fortina et al. 2001)
Bacillus subtilis, *Bacillus pumilus*	extreme arid desert soils	(Lester et al. 2007)
Bacillus spp.	marine sediments	(Gontang et al. 2007)
Bacillus fumarioli	geothermal, Antarctica	(Logan et al. 2000)
Pelotomaculum, *Sulfobacillus*, *Paenibacillus*, *Thermaerobacter*, *Moorella*	hot spring 55℃ (Boiling springs Lake)	(Wilson et al. 2008)
Bacillus stearothermophilus	ocean basin core	(Bartholomew and Paik 1966)

antibodies was applied to forest soil, where the majority of *Bacillus* cells were found associated to organic matter particles in soil (Siala et al. 1974). This study is particularly interesting as it differentiated between cells and endospores and found that in the C-horizon of the soil about 80% of the community was in form of endospores, while in the A-horizon the proportion of endospores was lower (33% of the community). It was also the first study to reveal activity of *Bacillus* species in temperate soil, mostly linked to decaying organic material (leaves, roots), which contradicted the previous idea of *B. subtilis* able to thrive only at mesophilic temperature ranges.

By fluorescence microscopy, endospores can be detected via complexation of DPA with terbium, upon which the endospores become fluorescent (Rosen et al. 1997; Fell et al. 2001). Using this method, endospores have been quantified in samples, such as soil (Brandes Ammann et al. 2011), air (Li et al. 2008) and sediment (Fichtel et al. 2007, 2008). In sediment, the highest abundance was found at sediment depth of 50 cm, where 6 x 10^6 spores/g sediment were estimated (three orders of magnitude higher than most probable number quantification) (Fichtel

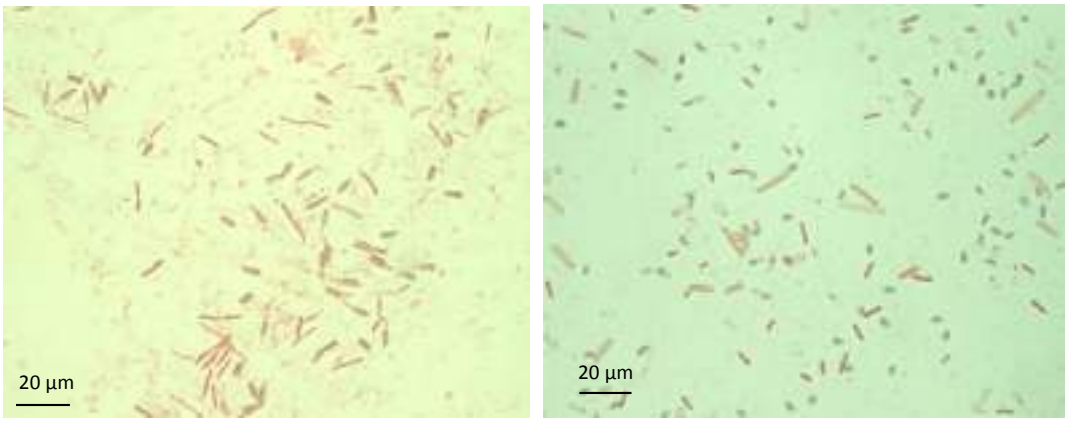


Figure 1.6: Microscopy image of stained cells and spores using the Schaeffer-Foulton-stain. Culture on the left is *Paenibacillus alvei* and on the right is a culture of *Bacillus subtilis*. Cells are red and endospores are blue.

et al. 2007). In a longer sediment core, it was shown that endospores made 1% of the total community at 50 cm depth and 10% of the community at 5 m depth (Fichtel et al. 2008).

Flow cytometry for the quantification of endospores has been used to distinguish between vegetative cells and endospores with fluorescent dyes (Comas-Riu and Vives-Rego 2002; Cronin and Wilkinson 2010). However, the studies are based on pure cultures of *Paenibacillus polymyxa* or *B. cereus* and do not include environmental samples.

Some bacteria (genus *Desulfotomaculum*) were only detected after heating the sample to germinate endospores in the sample. This implies that these bacteria are present only in the form of endospores and not as active vegetative cells (Hubert et al. 2010).

Molecular methods

In molecular ecology, specific primers for the 16S rRNA gene tailored to *Bacillus* and related species have been developed. With this, the diversity of aerobic endospore-forming bacteria in soils was detected and compared among different arable or grassland soils (Garbeva et al. 2003). Group specific primers for *Bacillus* taxa based on the 16S rRNA gene have also been developed (Wu et al. 2006; Mühling et al. 2008).

Together with the emergence of molecular methods, probes for sporulation genes (*ssp*, *dpaA/B*, *spo0A*) were developed to distinguish between endospore-forming bacteria and asporogenic species using PCR and southern hybridization methods (Brill and Wiegel 1997). The *ssp* gene encodes the small acid soluble proteins that help protect the genome inside the endospores, primarily against UV damage (Fairhead et al. 1993). The *dpaA/B* gene is involved in the synthesis of DPA, making the endospore heat resistant (Daniel and Errington 1993). As mentioned above, the gene *spo0A* codes for the transcription factor regulating the

onset of the sporulation pathway, and has been shown to be homologous among the Bacilli and Clostridia (Brown et al. 1994). The nucleotide sequence of the *spo0A* from *B. subtilis* is known since 1985 (Kudoh and Ikeuchi 1985). It encodes a transcription factor regulating over 500 downstream genes (Molle et al. 2003). In an extensive study of over 600 genomes one ortholog of *spo0A* was found in all analysed endospore-forming genomes (46) and in one non-endospore-forming genome (*Synthrophobacter*)(Traag et al. 2013). Based on the *spo0A* gene, a quantitative real-time PCR method has been developed to quantify endospores in milk powder (Rueckert et al. 2006). This assay focused on 6 thermophilic *Geobacillus* strains that routinely appear as milk contaminants. The primers have however not been tested on other strains or environmental samples. By cloning and sequencing of the metagenome of copper contaminated marine sediment, 20% of the detected community was affiliated to Bacilli and Clostridia (Besaury et al. 2013). In an exhaustive 16S rRNA sequencing study of bacterial soil communities across seven different soil types from hot desert, cold desert, forests, tundra, and prairie, the abundance of Firmicutes ranged from below 0.5% to a maximum of 3.7% of the total community (Fierer et al. 2012). In acidic soil on the contrary, members of Firmicutes (mostly *Bacillus* spp.) have been shown to make up almost half of the entire active community, as detected by temperature gradient gel electrophoresis (TGGE) and clone libraries (Felske et al. 1998).

Based on whole genome sequencing, new species of endospore-forming bacteria (e.g. *Carboxydothermus hydrogenoformans* Z-2901) and new metabolic functions connected to bacterial endospore-formers have been repeatedly reported in the last years, such as uranium reduction (Junier et al. 2009a), chemolithoautorophy (Chivian et al. 2008), and carbon monoxide utilization (Wu et al. 2005). These numbers are significantly different from each other, which can be due to environmental differences or due to methodological bias. It is very difficult to compare abundance of endospore-forming bacteria across different publications as the DNA extraction methods are diverse and heavily influence the detection of endospores in samples.

Ecological studies with endospores

Ecological questions about the distribution and origin of endospore-forming strains can be answered by looking at their physiology and by the distinction between vegetative cells and endospores. Isolates from cold marine sites have been shown to be closely related to strains in petroleum-associated point sources or geothermal seeps in the Arctic Ocean (Hubert et al. 2010). Thermophilic endospore-forming isolates were also obtained as the only isolates from temperate ocean floors (Bartholomew and Paik 1966). The isolates (growing only at 65 °C aerobically and on low salinity nutrient agar) are most likely of allochthonous origin, as they are not adapted to the cold marine environment (Bartholomew and Paik 1966). This study was the first to show endospores in old sediments (5,800 years) and claim that they are fossil records, deposited at the time of sedimentation (Bartholomew and Paik 1966).

Long-range transport of endospore-forming bacteria has been shown in several publications (Hubert et al. 2010; de Rezende et al. 2013; Marchant et al. 2011). Marchant and collegues (in Marchant et al. 2011 have found thermophilic geobacilli in rainwater and air samples and have also shown possible low growth rates in cold soils (4℃) (Marchant et al. 2011). Their results suggest that this group of bacteria is globally transported and can ensure survival in suboptimal conditions by very low growth rates or long term resistance due to sporulation. Airborne isolate of thermophilic *Bacillus thermosphaericus* species were found in Finland (Andersson et al. 1995). This strain was later re-classified to *Ureibacillus thermosphaericus*, a group of aerobic, thermophilic strains (originally isolated from soil) able to use urea as substrate (Fortina et al. 2001).

One of the main difficulties when studying endospore-forming microorganisms is the inability to specifically target this group within a complex microbial community. Most extraction protocols are not designed to retrieve DNA from spores, because they are not harsh enough to break spores open (Ricca et al. 2004). An analysis of DNA extraction methods of publications of soil communities has indeed found a positive correlation between abundance of *Bacillus* species and increased physical cell lysis such as bead-beating (Ricca et al. 2004). Considerable effort was put into the optimization of extraction methods for endospore-forming bacteria and other hard to lyse cells (Dineen et al. 2010; Kuske et al. 1998). Dineen and colleagues showed that the FastDNA spin kit for soil (MP Biomedicals) resulted in the highest DNA yield from *B. cereus* spore preparations (Dineen et al. 2010). However, endospore-forming bacteria have strong heterogeneity in lysis affectivity. Different groups of thermophilic bacteria differ in their lysis potential from lysozyme (not effective on *Geobacillus stearothermophilus* and *Brevibacillus brevis*, effective on *Bacillus licheniformis*, *Bacillus coagulans*, and others), while sodium dodecyl sulphate was ineffective for the *B. brevis* and *G. stearothermophilus*, but not the others (De Bartolomeo et al. 1991). Little is known about the actual diversity of endospore-forming bacteria in environmental samples. By isolation only a limited diversity can be revealed, and with molecular methods based on the entire bacterial community, few endospore-forming species are detected, particularly if they only confine to a few percent of the community. Therefore, until today detailed information about the diversity of endospore-forming bacteria has only been revealed in environmental samples where endospores correspond to a dominant fraction of the community. Such an example is a eutrophic lake sediment sample where endospore-forming strains of *Bacillus*, *Brevibacillus*, and *Exiguobacterium* were retrieved by denaturing gradient gel electrophoresis (DGGE) and subsequent sequencing of excised gel-bands (Zhao et al. 2008). They corresponded to the most dominant bacteria in the samples (based on band intensity). Nonetheless, the endospores found in natural samples have often been overlooked and therefore substantial effort is still needed to comprehensively assess the metabolic and ecological diversity of endospore-forming bacteria in the environment.

1.6 Research objectives

The general aim of this research was to develop methods for targeted metagenomics sequencing on bacterial endospores and to analyse the diversity of endospore-forming bacteria in sediment. Specific objectives were:

To find a marker gene for sporulation and design molecular primers for targeting endospore-forming bacteria in environmental samples.

To develop a method to kill the vegetative cells in an environmental sample and thereby isolate the more resistant endospores.

To optimize DNA extraction methods for endospores in sediment.

To apply the targeted metagenomics methods to lake sediment samples and analyse the diversity of endospore-forming bacteria.

To compare the diversity of endospores in sediment with results of sediment geochemistry and to find a link between the community composition and environmental lake history.

1.7 References

Abecasis, A. B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J. B., and Henriques, A. O. (2013). A genomic signature and the identification of new sporulation genes. *J Bacteriol*, 195(9):2101–15.

Adams, D. G. and Duggan, P. S. (1999). Tansley review no. 107. Heterocyst and akinete differentiation in cyanobacteria. *New Phytologist*, 144(1):3–33.

Ager, D., Evans, S., Li, H., Lilley, A. K., and van der Gast, C. J. (2010). Anthropogenic disturbance affects the structure of bacterial communities. *Environ Microbiol*, 12(3):670–8.

Alawi, M., Lerm, S., Vetter, A., Wolfgramm, D. M., Seibt, D. A., and Würdemann, H. (2011). Diversity of sulfate-reducing bacteria in a plant using deep geothermal energy. *Grundwasser*, 16(2):105–112.

Allgaier, M., Reddy, A., Park, J. I., Ivanova, N., D'Haeseleer, P., Lowry, S., et al. (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One*, 5(1):e8812.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev*, 59(1):143–69.

Anderson, I., Chertkov, O., Chen, A., Saunders, E., Lapidus, A., Nolan, M., et al. (2012). Complete genome sequence of the moderately thermophilic mineral-sulfide-oxidizing Firmicute *Sulfobacillus acidophilus* type strain (NAL(T)). *Stand Genomic Sci*, 6(3):1–13.

Andersson, M., Laukkanen, M., Nurmiaho-Lassila, E.-L., Rainey, F., Niemelä, S., and Salkinoja-Salonen, M. (1995). *Bacillus thermosphaericus* sp. nov. a new thermophilic ureolytic *Bacillus* isolated from air. *Syst Appl Microbiol*, 18(2):203–220.

Anneville, O., Ginot, V., Druart, J.-C., and Angeli, N. (2002). Long-term study (1974–1998) of seasonal changes in the phytoplankton in Lake Geneva: a multi-table approach. *J Plankton*, 24(10):993–1008.

Barnosky, A. D., Hadly, E. A., Bascompte, J., Berlow, E. L., Brown, J. H., Fortelius, M., et al. (2012). Approaching a state shift in earth's biosphere. *Nature*, 486(7401):52–8.

Bartholomew, J. W. and Paik, G. (1966). Isolation and identification of obligate thermophilic sporeforming Bacilli from ocean basin cores. *J Bacteriol*, 92(3):635–8.

Battistuzzi, F. U. and Hedges, S. B. (2009). A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol*, 26(2):335–43.

Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., et al. (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, 289(5486):1902–1906.

Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., and DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, 411(6839):786–789.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218):53–59.

Bertin, P. N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., et al. (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J*, 5(11):1735–1747.

Besaury, L., Marty, F., Buquet, S., Mesnage, V., Muyzer, G., and Quillet, L. (2013). Culture-dependent and independent studies of microbial diversity in highly copper-contaminated Chilean marine sediments. *Microb Ecol*, 65(2):311–324.

Bodelier, P. L. E. (2011). Toward understanding, managing, and protecting microbial ecosystems. *Front Microbiol*, 2:80.

Bonjour, F. and Aragno, M. (1984). *Bacillus tusciae*, a new species of thermoacidophilic, facultatively chemolithoautotrophic hydrogen oxidizing sporeformer from a geothermal area. *Arch Microbiol*, 139(4):397–401.

Bonjour, F., Graber, A., and Aragno, M. (1988). Isolation of *Bacillus schlegelii*, a thermophilic, hydrogen oxidizing, aerobic autotroph, from geothermal and nongeothermal environments. *Microb Ecol*, 16(3):331–337.

Bosshard, P. P., Zbinden, R., and Altwegg, M. (2002). *Turicibacter sanguinis* gen. nov., sp. nov., a novel anaerobic, gram-positive bacterium. *Int J Syst Evol Microbiol*, 52(4):1263–1266.

Brandes Ammann, A., Lle, L., and Brandl, H. (2011). Detection of bacterial endospores in soil by terbium fluorescence. *Int J Microbiol*, 2011:435281.

Brazelton, W. J. and Baross, J. A. (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J*, 3(12):1420–1424.

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., and Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proc Natl Acad Sci USA*, 106(12):4758–4763.

Brill, J. A. and Wiegel, J. (1997). Differentiation between spore-forming and asporogenic bacteria using a PCR and southern hybridization based method. *J Microbiol Methods*, 31(1–2):29–36.

Brown, D. P., Ganova-Raeva, L., Green, B. D., Wilkinson, S. R., Young, M., and Youngman, P. (1994). Characterization of *spo0a* homologues in diverse *Bacillus* and *Clostridium* species identifies a probable DNA-binding domain. *Mol Microbiol*, 14(3):411–426.

Bryantseva, I., Gorlenko, V., Kompantseva, E., Achenbach, L., and Madigan, M. (1999). *Heliorestis daurensis*, gen. nov. sp. nov., an alkaliphilic rod-to-coiled-shaped phototrophic Heliobacterium from a Siberian Soda Lake. *Arch Microbiol*, 172(3):167–174.

Cano, R. J., Borucki, M. K., Higby-Schweitzer, M., Poinar, H. N., Poinar, G O, J., and Pollard, K. J. (1994). *Bacillus* DNA in fossil bees: an ancient symbiosis? *Appl Environ Microbiol*, 60(6):2164–2167.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science*, 322(5899):275–278.

Claus, D., Fritze, D., and Kocur, M. (2006). Genera related to the genus *Bacillus—Sporolactobacillus*, *Sporosarcina*, *Planococcus*, *Filibacter* and *Caryophanon*. In Dr Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, page 631–653. Springer US.

Cockell, C. S. and Jones, H. L. (2009). Advancing the case for microbial conservation. *Oryx*, 43(04):520.

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., et al. (1994). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol*, 44(4):812–826.

Comas-Riu, J. and Vives-Rego, J. (2002). Cytometric monitoring of growth, sporogenesis and spore cell sorting in *Paenibacillus polymyxa* (formerly *Bacillus polymyxa*). *J Appl Microbiol*, 92(3):475–481.

Commission of the European Communities (2007). *Commission Staff Working Document "Towards Sustainable water mangement in the European Union"*. Commission of the European Communities.

Craine, J. M., Fierer, N., and McLauchlan, K. K. (2010). Widespread coupling between the rate and temperature sensitivity of organic matter decay. *Nat Geoscience*, 3(12):854–857.

Cronin, U. and Wilkinson, M. (2010). The potential of flow cytometry in the study of *Bacillus cereus*. *J Appl Microbiol*, 108(1):1–16.

Cuvelier, M. L., Allen, A. E., Monier, A., McCrow, J. P., Messié, M., Tringe, S. G., et al. (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci USA*, 107(33):14679–14684.

Daniel, R. A. and Errington, J. (1993). Cloning, DNA sequence, functional analysis and transcriptional regulation of the genes encoding dipicolinic acid synthetase required for sporulation in *Bacillus subtilis*. *J Mol Biol*, 232(2):468–483.

Darland, G. and Brock, T. D. (1971). *Bacillus acidocaldarius* sp.nov., an acidophilic thermophilic spore-forming bacterium. *J Biochem*, 67(1):9–15.

De Bartolomeo, A., Trotta, F., La Rosa, F., Saltalamacchia, G., and Mastrandrea, V. (1991). Numerical analysis and DNA base compositions of some thermophilic *Bacillus* species. *Int J Syst Bacteriol*, 41(4):502–509.

de Rezende, J. R., Kjeldsen, K. U., Hubert, C. R., Finster, K., Loy, A., and Jorgensen, B. B. (2013). Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J*, 7(1):72–84.

de Vrind, J. P. M., de Vrind-de Jong, E. W., de Voogt, J.-W. H., Westbroek, P., Boogerd, F. C., and Rosson, R. A. (1986). Manganese oxidation by spores and spore coats of a marine *Bacillus* species. *Appl Environ Microbiol*, 52(5):1096–1100.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., et al. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol*, 77(4):1315–24.

Dick, G. J., Lee, Y. E., and Tebo, B. M. (2006). Manganese(II)-oxidizing *Bacillus* spores in Guaymas Basin hydrothermal sediments and plumes. *Appl Environ Microbiol*, 72(5):3184–3190.

Dick, G. J. and Tebo, B. M. (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol*, 12(5):1334–1347.

Dineen, S. M., Aranda, R. t., Anders, D. L., and Robertson, J. M. (2010). An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *J Appl Microbiol*, 109(6):1886–96.

Dinsdale, E. A., Pantos, O., Smriga, S., Edwards, R. A., Angly, F., Wegley, L., et al. (2008). Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One*, 3(2):e1584.

Driks, A. (2003). The dynamic spore. *Proc Natl Acad Sci USA*, 100(6):3007–3009.

Ducklow, H. (2008). Microbial services: challenges for microbial ecologists in a changing world. *Aquat Microb Ecol*, 53(1):13–19.

Ducklow, H. W., Baker, K., Martinson, D. G., Quetin, L. B., Ross, R. M., Smith, R. C., et al. (2007). Marine pelagic ecosystems: the West Antarctic Peninsula. *Philos Trans R Soc Lond B Biol Sci*, 362(1477):67–94.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., et al. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev*, 81(2):163–182.

Dumonceaux, T. J., Hill, J. E., Hemmingsen, S. M., and Van Kessel, A. G. (2006). Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. *Appl Environ Microbiol*, 72(4):2815–2823.

Dürre, D. R. (2009). The genus *Clostridium*. In *Practical Handbook of Microbiology, Second Edition*, volume 25. Taylor and Francis Group.

Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., et al. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*, 7:57.

Eichenberger, P., Fujita, M., Jensen, S. T., Conlon, E. M., Rudner, D. Z., Wang, S. T., et al. (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol*, 2(10):e328.

Ensign, J. C. (1978). Formation, properties, and germination of actinomycete spores. *Annu Rev Microbiol*, 32(1):185–219.

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

Ettenauer, J., Sterflinger, K., and Piñar, G. (2010). Cultivation and molecular monitoring of halophilic microorganisms inhabiting an extreme environment presented by a salt-attacked monument. *Int J Astrobiol*, 9(01):59–72.

European Union (EU) (2000). *EUR-Lex - 32000 L0060 - EN*. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.

Fairhead, H., Setlow, B., and Setlow, P. (1993). Prevention of DNA damage in spores and in vitro by small, acid-soluble proteins from *Bacillus* species. *J Bacteriol*, 175(5):1367–1374.

Fajardo-Cavazos, P. and Nicholson, W. (2006). *Bacillus* endospores isolated from granite: close molecular relationships to globally distributed *Bacillus* spp. from endolithic and extreme environments. *Appl Environ Microbiol*, 72(4):2856–2863.

Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *Science*, 320(5879):1034–1039.

Fell, N. F., Pellegrino, P. M., and Gillespie, J. B. (2001). Mitigating phosphate interference in bacterial endospore detection by Tb dipicolinate photoluminescence. *Anal Chim Acta*, 426(1):43–50.

Felske, A., Wolterink, A., Lis, R. V., and Akkermans, A. D. L. (1998). Phylogeny of the main bacterial 16S rRNA sequences in Drentse A grassland soils (The Netherlands). *Appl Environ Microbiol*, 64(3):871–879.

Ferrer, M., Guazzaroni, M.-E., Richter, M., García-Salamanca, A., Yarza, P., Suárez-Suárez, A., et al. (2011). Taxonomic and functional metagenomic profiling of the microbial community in the anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo, Central Spain). *Microb Ecol*, 62(4):824–837.

Fichtel, J., Koster, J., Rullkotter, J., and Sass, H. (2007). Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol Ecol*, 61(3):522–32.

Fichtel, J., Köster, J., Rullkötter, J., and Sass, H. (2008). High variations in endospore numbers within tidal flat sediments revealed by quantification of dipicolinic acid. *Geomicrobiol J*, 25(7-8):371–380.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et al. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci USA*, 109(52):21390–21395.

Finegold, S. M., Lawson, P. A., Vaisanen, M.-L., Molitoris, D. R., Song, Y., Liu, C., and Collins, M. D. (2004). *Anaerofustis stercorihominis* gen. nov., sp. nov., from human feces. *Anaerobe*, 10(1):41–45.

Fischman, J. (1995). Have 25-million-year-old bacteria returned to life? *Science*, 268(5213):977.

Fortina, M. G., Pukall, R., Schumann, P., Mora, D., Parini, C., Manachini, P. L., and Stackebrandt, E. (2001). *Ureibacillus* gen. nov., a new genus to accommodate *Bacillus thermosphaericus* (Andersson et al. 1995), emendation of *Ureibacillus thermosphaericus* and description of *Ureibacillus terrenus* sp. nov. *Int J Syst Evol Microbiol*, 51(Pt 2):447–455.

Francis, C. A., Casciotti, K. L., and Tebo, B. M. (2002). Localization of Mn(II)-oxidizing activity and the putative multicopper oxidase, *MnxG*, to the exosporium of the marine *Bacillus* sp. strain SG-1. *Arch Microbiol*, 178(6):450–456.

Fritze, D. (2004). Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. *Phytopathology*, 94(11):1245–1248.

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature*, 459(7244):193–199.

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., and Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol*, 14(11):2870–90.

Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, 309(5739):1387–1390.

Garbeva, P., Veen, J. A. v., and Elsas, J. D. v. (2003). Predominant *Bacillus* spp. in agricultural soil under different management regimes detected via PCR-DGGE. *Microb Ecol*, 45(3):302–316.

Gilbert, J. A. and Dupont, C. L. (2011). Microbial metagenomics: beyond the genome. *Annu Rev Mar Sci*, 3:347–371.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, 312(5778):1355–1359.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., and Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345(6270):60–63.

Gontang, E. A., Fenical, W., and Jensen, P. R. (2007). Phylogenetic diversity of Gram-positive bacteria cultured from marine sediments. *Appl Environ Microbiol*, 73(10):3272–82.

Gould, G. W. (2006). History of science– Spores. *J Appl Microbiol*, 101(3):507–513.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev*, 68(4):669–685.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*, 5(10):R245–9.

Harris, J. K., Caporaso, J. G., Walker, J. J., Spear, J. R., Gold, N. J., Robertson, C. E., et al. (2013). Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J*, 7(1):50–60.

Hazen, R. M. and Roedder, E. (2001). Biogeology. How old are bacteria from the permian age? *Nature*, 411(6834):155–6.

Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., et al. (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J*, 4(5):660–672.

Henriques, A. O. and Moran, Charles P, J. (2007). Structure, assembly, and function of the spore surface layers. *Annu Rev Microbiol*, 61:555–588.

Hill, I. R. and Gray, T. R. G. (1967). Application of the fluorescent-antibody technique to an ecological study of bacteria in soil. *J Bacteriol*, 93(6):1888–1896.

Hill, J. E., Penny, S. L., Crowell, K. G., Goh, S. H., and Hemmingsen, S. M. (2004). cpnDB: a chaperonin sequence database. *Genome Res*, 14(8):1669–1675.

Holliger, C., Hahn, D., Harmsen, H., Ludwig, W., Schumacher, W., Tindall, B., et al. (1998). *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetra- and trichloroethene in an anaerobic respiration. *Arch Microbiol*, 169(4):313–321.

Hu, X. M., Fan, W., Han, B., Liu, H. Z., Zheng, D., Li, Q., et al. (2008). Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related *Bacillus* species. *J Bacteriol*, 190(8):2892–2902.

Hubert, C., Arnosti, C., Bruchert, V., Loy, A., Vandieken, V., and Jorgensen, B. B. (2010). Thermophilic anaerobes in arctic marine sediments induced to mineralize complex organic matter at high temperature. *Environ Microbiol*, 12(4):1089–104.

Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biol*, 3(2):Reviews0003.

Hugenholtz, P., Tyson, G. W., Webb, R. I., Wagner, A. M., and Blackall, L. L. (2001). Investigation of candidate division TM7, a recently recognized major lineage of the domain bacteria with no known pure-culture representatives. *Appl Environ Microbiol*, 67(1):411–419.

IPCC Core Writing Team and Pachauri, R. (2007). Climate change 2007: Synthesis report. Technical report, IPCC, Geneva.

Iwai, S., Chai, B., Sul, W. J., Cole, J. R., Hashsham, S. A., and Tiedje, J. M. (2010). Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *ISME J*, 4(2):279–285.

Jansson, J. K. and Prosser, J. I. (2013). Microbiology: The life beneath our feet. *Nature*, 494(7435):40–41.

Jung, J. Y., Lee, S. H., Kim, J. M., Park, M. S., Bae, J.-W., Hahn, Y., et al. (2011). Metagenomic analysis of kimchi, a traditional korean fermented food. *Appl Environ Microbiol*, 77(7):2264–2274.

Junier, P., Frutschi, M., Wigginton, N. S., Schofield, E. J., Bargar, J. R., and Bernier-Latmani, R. (2009a). Metal reduction by spores of *Desulfotomaculum reducens*. *Environ Microbiol*, 11(12):3007–3017.

Junier, P., O.S., K., Junier, T., Ahn, T., Imhoff, J., and Witzel, K. (2009b). Community analysis of betaproteobacterial ammonia-oxidizing bacteria using the *amoCAB* operon. *Appl Microbiol Biotechnol*, 83(1):175–188.

Kalyuzhnaya, M. G., Lapidus, A., Ivanova, N., Copeland, A. C., McHardy, A. C., Szeto, E., Salamov, A., et al. (2008). High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol*, 26(9):1029–1034.

Kampfer, P., Glaeser, S. P., and Busse, H.-J. (2012). Transfer of *Bacillus schlegelii* to a novel genus and proposal of *Hydrogenibacillus schlegelii* gen. nov., comb. nov. *Int J Syst Evol Microbiol*, 63(Pt 5):1723–1727.

Klenk, H.-P., Lapidus, A., Chertkov, O., Copeland, A., Del Rio, T. G., Nolan, M., et al. (2011). Complete genome sequence of the thermophilic, hydrogen-oxidizing *Bacillus tusciae* type strain (T2) and reclassification in the new genus, *Kyrpidia* gen. nov. as *Kyrpidia tusciae* comb. nov. and emendation of the family *Alicyclobacillaceae* (da Costa and Rainey, 2010). *Stand Genomic Sci*, 5(1):121–134.

Konopka, A. (2009). What is microbial community ecology? *ISME J*, 3(11):1223–30.

Kudoh, J. and Ikeuchi, T. (1985). Nucleotide sequences of the sporulation gene *spo0A* and its mutant genes of *Bacillus subtilis*. *Proc Natl Acad Sci USA*, 82(9):2665–2668.

Kunin, V., Raes, J., Harris, J. K., Spear, J. R., Walker, J. J., Ivanova, N., et al. (2008). Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Mol Syst Biol*, 4:198.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657):249–256.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., et al. (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res*, 14(4):169–181.

Kuske, C. R., Banton, K. L., Adorada, D. L., Stark, P. C., Hill, K. K., and Jackson, P. J. (1998). Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl Environ Microbiol*, 64(7):2463–72.

Lasken, R. S. (2012). Genomic sequencing of uncultured microorganisms from single cells. *Nat Rev Micro*, 10(9):631–640.

Le, C., Zha, Y., Li, Y., Sun, D., Lu, H., and Yin, B. (2010). Eutrophication of lake waters in China: Cost, causes, and control. *J Environ Manage*, 45(4):662–668.

Lester, E. D., Satomi, M., and Ponce, A. (2007). Microflora of extreme arid Atacama Desert soils. *Soil Biol Biochem*, 39(2):704–708.

Li, Q., Dasgupta, P. K., and Temkins, H. K. (2008). Airborne bacterial spore counts by terbium-enhanced luminescence detection: pitfalls and real values. *Environ Sci Technol*, 42(8):2799–2804.

Lloyd, A. T. and Sharp, P. M. (1993). Evolution of the *recA* gene and the molecular phylogeny of bacteria. *J Mol Evol*, 37(4):399–407.

Logan, N. A. (2012). *Bacillus* and relatives in foodborne illness. *J Appl Microbiol*, 112(3):417–429.

Logan, N. A., Lebbe, L., Hoste, B., Goris, J., Forsyth, G., Heyndrickx, M., et al. (2000). Aerobic endospore-forming bacteria from geothermal environments in northern Victoria Land, Antarctica, and Candlemas Island, South Sandwich archipelago, with the proposal of *Bacillus fumarioli* sp. nov. *Int J Syst Evol Microbiol*, 50(5):1741–1753.

Lombard, N., Prestat, E., van Elsas, J. D., and Simonet, P. (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol Ecol*, 78(1):31–49.

Lu, J., Nogi, Y., and Takami, H. (2001). *Oceanobacillus iheyensis* gen. nov., sp. nov., a deep-sea extremely halotolerant and alkaliphilic species isolated from a depth of 1050 m on the Iheya Ridge. *FEMS Microbiol Lett*, 205(2):291–297.

Madenjian, C. P., Fahnenstiel, G. L., Johengen, T. H., Nalepa, T. F., Vanderploeg, H. A., Fleischer, G. W., et al. (2002). Dynamics of the Lake Michigan food web, (1970-2000). *Can J Fish Aquat Sci*, 59(4):736–753.

Madigan, M. T., Martinko, J. M., and Brock, T. D. (2006). *Brock Mikrobiologie*. Pearson Studium, München, Germany.

Marchant, R., Banat, I. M., and Franzetti, A. (2011). Thermophilic bacteria in cool soils: metabolic activity and mechanisms of dispersal. In Fontaneto, D., editor, *Biogeography of Microscopic Organisms*, page 43–57. Cambridge University Press, Cambridge, UK.

Mason, J. M. and Setlow, P. (1986). Essential role of small, acid-soluble spore proteins in resistance of *Bacillus subtilis* spores to UV light. *J Bacteriol*, 167(1):174–178.

Matsuo, Y., Katsuta, A., Matsuda, S., Shizuri, Y., Yokota, A., and Kasai, H. (2006). *Mechercharimyces mesophilus* gen. nov., sp. nov. and *Mechercharimyces asporophorigenens* sp. nov., antitumour substance-producing marine bacteria, and description of *Thermoactinomycetaceae* fam. nov. *Int J Syst Evol Microbiol*, 56(Pt 12):2837–2842.

McHardy, A. C., Martín, H. G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007). Accurate phylogenetic classification of variable-length DNA fragments. *Nat Methods*, 4(1):63–72.

Miller, J. R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6):315–327.

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S., and Losick, R. (2003). The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol*, 50(5):1683–1701.

Morgan, J. L., Darling, A. E., and Eisen, J. A. (2010). Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS One*, 5(4):e10209.

Mormak, D. A. and Casida, L. E. (1985). Study of *Bacillus subtilis* endospores in soil by use of a modified endospore stain. *Appl Environ Microbiol*, 49(6):1356–1360.

Mou, X., Sun, S., Edwards, R. A., Hodson, R. E., and Moran, M. A. (2008). Bacterial carbon processing by generalist species in the coastal ocean. *Nature*, 451(7179):708–711.

Möller, B., Oßmer, R., Howard, B. H., Gottschalk, G., and Hippe, H. (1984). *Sporomusa*, a new genus of gram-negative anaerobic bacteria including *Sporomusa sphaeroides* spec. nov. and *Sporomusa ovata* spec. nov. *Arch Microbiol*, 139(4):388–396.

Mühling, M., Woolven-Allen, J., Murrell, J. C., and Joint, I. (2008). Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J*, 2(4):379–392.

Nazina, T. N., Tourova, T. P., Poltaraus, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., et al. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*. *Int J Syst Evol Microbiol*, 51(Pt 2):433–46.

Neufeld, J. D., Chen, Y., Dumont, M. G., and Murrell, J. C. (2008). Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol*, 10(6):1526–1535.

Nicholson, W. L. (2002). Roles of *Bacillus* endospores in the environment. *Cell Mol Life Sci*, 59(3):410–416.

Nicholson, W. L. and Law, J. F. (1999). Method for purification of bacterial endospores from soils: UV resistance of natural Sonoran desert soil populations of *Bacillus* spp. with reference to *B. subtilis* strain 168. *J Microbiol Methods*, 35(1):13–21.

Niimura, Y., Koh, E., Yanagida, F., Suzuki, K.-I., Komagata, K., and Kozaki, M. (1990). *Amphibacillus xylanus* gen. nov., sp. nov., a facultatively anaerobic spore-forming xylan-digesting bacterium which lacks cytochrome, quinone, and catalase. *Int J Syst Evol Microbiol*, 40(3):297–301.

Ogutu-Ohwayo, R., Hecky, R. E., Cohen, A. S., and Kaufman, L. (1997). Human impacts on the African Great Lakes. *Environ Biol Fishes*, 50(2):117–131.

Onyenwoke, R. U., Brill, J. A., Farahi, K., and 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.

Pace, N. R. (2009). Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev*, 73(4):565–76.

Palenik, B., Ren, Q., Tai, V., and Paulsen, I. T. (2009). Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol*, 11(2):349–359.

Pan, Y., Bodrossy, L., Frenzel, P., Hestnes, A.-G., Krause, S., Lüke, C., et al. (2010). Impacts of inter- and intralaboratory variations on the reproducibility of microbial community analyses. *Appl Environ Microbiol*, 76(22):7451–7458.

Park, C. C. (2001). *The environment: principles and applications*. Routledge, London; New York.

Parthasarathy, H., Hill, E., and MacCallum, C. (2007). Global ocean sampling collection. *PLoS Biol*, 5:e83.

Pedrós-Alió, C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol*, 14(6):257–263.

Philippot, L., Andersson, S. G. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., and Hallin, S. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol*, 8(7):523–529.

Pikuta, E., Lysenko, A., Chuvilskaya, N., Mendrock, U., Hippe, H., Suzina, N., et al. (2000). *Anoxybacillus pushchinensis* gen. nov., sp. nov., a novel anaerobic, alkaliphilic, moderately thermophilic bacterium from manure, and description of *Anoxybacillus flavitherms* comb. nov. *Int J Syst Evol Microbiol*, 6(Pt 50):2109–2117.

Pope, P. B., Denman, S. E., Jones, M., Tringe, S. G., Barry, K., Malfatti, S. A., et al. (2010). Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc Natl Acad Sci USA*, 107(33):14793–14798.

Prazmowski, A. (1880). *Untersuchungen ueber die Entwicklungsgeschichte und Fermentwirkung einiger Bacterian-Arten.* PhD thesis, Universität Leipzig.

Quaiser, A., Zivanovic, Y., Moreira, D., and López-García, P. (2011). Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J*, 5(2):285–304.

Quince, C., Curtis, T. P., and Sloan, W. T. (2008). The rational exploration of microbial diversity. *ISME J*, 2(10):997–1006.

Raes, J., Foerstner, K. U., and Bork, P. (2007). Get the most out of your metagenome: computational analysis of environmental sequence data. *Curr Opin Microbiol*, 10(5):490–498.

Raghunathan, A., Ferguson, H. R., Bornarth, C. J., Song, W., Driscoll, M., and Lasken, R. S. (2005). Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol*, 71(6):3342–3347.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., et al. (2005). Community proteomics of a natural microbial biofilm. *Science*, 308(5730):1915–1920.

Reichenbach, H. (1999). The ecology of the myxobacteria. *Environ Microbiol*, 1(1):15–21.

Rheims, H., Rainey, F. A., and Stackebrandt, E. (1996). A molecular approach to search for diversity among bacteria in the environment. *J Indust Microbiol*, 17(3-4):159–169.

Ricca, E., Henriques, A. O., and Cutting, S. M. (2004). *Bacterial Spore Formers: Probiotics And Emerging Applications*. Horizon Bioscience.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459):431–437.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*, 242(1):84–89.

Rosen, D. L., Sharpless, C., and McGown, L. B. (1997). Bacterial spore detection and determination by use of terbium dipicolinate photoluminescence. *Anal Chem*, 69(6):1082–1085.

Rosson, R. A. and Nealson, K. H. (1982). Manganese binding and oxidation by spores of a marine *Bacillus*. *J Bacteriol*, 151(2):1027–1034.

Rueckert, A., Ronimus, R. S., and Morgan, H. W. (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(3):220–230.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol*, 5:e77.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Biotechnology*, 24(12):104–108.

Schmidt, T. M., DeLong, E. F., and Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol*, 173(14):4371–4378.

Scholz, M. B., Lo, C. C., and Chain, P. S. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol*, 23(10):9–15.

Schwiertz, A., Hold, G. L., Duncan, S. H., Gruhl, B., Collins, M. D., Lawson, P. A., et al. (2002). *Anaerostipes caccae* gen. nov., sp. nov., a new saccharolytic, acetate-utilising, butyrate-producing bacterium from human faeces. *Syst Appl Microbiol*, 25(1):46–51.

Shida, O., Takagi, H., Kadowaki, K., and Komagata, K. (1996). Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. *Int J Syst Evol Microbiol*, 46(4):939–46.

Shida, O., Takagi, H., Kadowaki, K., Nakamura, L. K., and 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):289–98.

Siala, A., Hill, I. R., and Gray, T. R. G. (1974). Populations of spore-forming bacteria in an acid forest soil, with special reference to *Bacillus subtilis*. *J General Microbiol*, 81(1):183–190.

Smith, V. H. and Schindler, D. W. (2009). Eutrophication science: where do we go from here? *Trends Ecol Evol*, 24(4):201–207.

Snyder, L. and Champness, W. (2007). *Molecular genetics of bacteria*. ASM Press, Washington, D.C.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA*, 103(32):12115–12120.

Stackebrandt, E., Sproer, C., Rainey, F. A., Burghardt, J., Päuker, O., and 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, J. T. and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*, 39:321–46.

Suenaga, H. (2012). Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol*, 14(1):13–22.

Tadonléké, R. D., Lazzarotto, J., Anneville, O., and Druart, J.-C. (2009). Phytoplankton productivity increased in Lake Geneva despite phosphorus loading reduction. *J Plankton Res*, 31(10):1179–1194.

Taras, D., Simmering, R., Collins, M. D., Lawson, P. A., and Blaut, M. (2002). Reclassification of *Eubacterium formicigenerans* (Holdeman and Moore 1974) as *Dorea formicigenerans* gen. nov., comb. nov., and description of *Dorea longicatena* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol*, 52(2):423–428.

Traag, B. A., Pugliese, A., Eisen, J. A., and Losick, R. (2013). Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. *J Bacteriol*, 195(2):253–60.

Tringe, S. G., Mering, C. v., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science*, 308(5721):554–557.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature*, 449(7164):804–810.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978):37–43.

United Nations Environment Programme (UNEP) (2012). Global Environment Outlook 5. Technical report, United Nations Environment Programme (UNEP).

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667):66–74.

Verma, P., Pandey, P. K., Gupta, A. K., Seong, C. N., Park, S. C., Choe, H. N., et al. (2012). Reclassification of *Bacillus beijingensis* (Qiu et al. 2009) and *Bacillus ginsengi* (Qiu et al. 2009) as *Bhargavaea beijingensis* comb. nov. and *Bhargavaea ginsengi* comb. nov. and emended description of the genus *Bhargavaea*. *Int J Syst Evol Microbiol*, 62(Pt 10):2495–2504.

Vital, M., Penton, C. R., Wang, Q., Young, V. B., Antonopoulos, D. A., Sogin, M. L., et al. (2013). A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community. *Microbiome*, 1(1):8.

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., van Elsas, J. D., et al. (2009). TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature*, 7(4):252–252.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E., and Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol*, 69(10):6235–6242.

Volkov, I., Banavar, J. R., and Maritan, A. (2006). Comment on "Computational improvements reveal great bacterial diversity and high metal toxicity in soil". *Science*, 313(5789):918a–918a.

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., et al. (2007). Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science*, 315(5815):1126–1130.

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., et al. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482(7385):357–62.

Vos, P., Garrity, G., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., et al. (2006). *Bergey's Manual of Systematic Bacteriology - Volume 3: The Firmicutes*.

Vreeland, R. H., Rosenzweig, W. D., and Powers, D. W. (2000). Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature*, 407(6806):897–900.

Wade, W. G. (2006). The genus *Eubacterium* and related genera. In Dr Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, page 823–835. Springer US.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., et al. (2002). Ecological responses to recent climate change. *Nature*, 416(6879):389–95.

Warnecke, F. and Hugenholtz, P. (2007). Building on basic metagenomics with complementary technologies. *Genome Biol*, 8(12):231.

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., and Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environ Microbiol*, 9(11):2707–2719.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA*, 95(12):6578–83.

Wiegel, J., Tanner, R., and Rainey, F. A. (2006). An introduction to the family *Clostridiaceae*. In Dr Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, page 654–678. Springer US.

Wilson, M. A. and Carpenter, S. R. (1999). Economic valuation of freshwater ecosystem services in the United States: 1971-1997. *Ecol App*, 9(3):772–783.

Wilson, M. S., Siering, P. L., White, C. L., Hauser, M. E., and Bartles, A. N. (2008). Novel archaea and bacteria dominate stable microbial communities in North America's largest hot spring. *Microb Ecol*, 56(2):292–305.

Wisotzkey, J. D., Jurtshuk, P, J., Fox, G. E., Deinhard, G., and Poralla, K. (1992). Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidocaldarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov. *Int J Syst Bacteriol*, 42(2):263–269.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev*, 51(2):221–271.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*, 87(12):4576–4579.

Wolin, M. J., Miller, T. L., Collins, M. D., and Lawson, P. A. (2003). Formate-dependent growth and homoacetogenic fermentation by a bacterium from human feces: Description of *Bryantella formatexigens* gen. nov., sp. nov. *Appl Environ Microbiol*, 69(10):6321–6326.

World Resources Institute, W. (2005). Ecosystems and human well-being: General synthesis. Technical report, World Resources Institute, Washington, D.C.

Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., et al. (2005). Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet*, 1(5):e65.

Wu, X.-Y., Walker, M. J., Hornitzky, M., and Chin, J. (2006). Development of a group-specific PCR combined with ARDRA for the identification of *Bacillus* species of environmental significance. *J Microbiol Methods*, 64(1):107–119.

Yassin, A. F., Hupfer, H., Klenk, H.-P., and Siering, C. (2009). *Desmospora activa* gen. nov., sp. nov., a thermoactinomycete isolated from sputum of a patient with suspected pulmonary tuberculosis, and emended description of the family *Thermoactinomycetaceae* (Matsuo et al. 2006). *Int J Syst Evol Microbiol*, 59(3):454–459.

Ye, Q., Roh, Y., Carroll, S. L., Blair, B., Zhou, J., Zhang, C. L., and Fields, M. W. (2004). Alkaline anaerobic respiration: isolation and characterization of a novel alkaliphilic and metal-reducing bacterium. *Appl Environ Microbiol*, 70(9):5595–5602.

Yutin, N. and Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced Clostridia. *Environ Microbiol*, 15(10):2631–2641.

Zarraonaindia, I., Smith, D. P., and Gilbert, J. A. (2013). Beyond the genome: community-level analysis of the microbial world. *Biol Philos*, 28(2):261–282.

Zeigler, D. R. and Perkins, J. B. (2009). The genus *Bacillus*. In *Practical Handbook of Microbiology, Second Edition*, volume 24. Taylor and Francis Group.

Zhao, J., Krishna, V., Moudgil, B., and Koopman, B. (2008). Evaluation of endospore purification methods applied to *Bacillus cereus*. *Sep Purif Technol*, 61(3):341–347.

2 Stage 0 sporulation gene A (*spo0A*) as a molecular marker to study diversity of endospore-forming Firmicutes

---

This chapter is based on the published article:
Tina Wunderlin, Thomas Junier, Ludovic Roussel-Delif, Nicole Jeanneret, Pilar Junier (2013). Stage 0 sporulation gene A (*spo0A*) as a molecular marker to study diversity of endospore-forming Firmicutes. *Environmental Microbiology Reports* 2013 Dec;5(6):911-24. doi:10.1111/1758-2229.12094.

Abstract

In this study we developed and validated a culture-independent method for diversity surveys to specifically detect endospore-forming Firmicutes. The global transcription regulator of sporulation (*spo0A*) was identified as a gene marker for endospore-forming Firmicutes. To enable phylogenetic classification, we designed a set of primers amplifying a 602 bp fragment of *spo0A* that we evaluated in pure cultures and environmental samples. The amplification was positive for 35 strains from 11 genera, yet negative for strains from *Alicyclobacillus* and *Sulfobacillus*. We also evaluated various DNA extraction methods because endospores often result in reduced yields. Our results demonstrate that procedures utilizing increased physical force improve DNA extraction. An optimized DNA extraction method on biomass pre-extracted from the environmental sample source (indirect DNA extraction) followed by amplification with the aforementioned primers for *spo0A*, was then tested in sediments from two different sources. Specifically, we validated our culture-independent diversity survey methodology on a set of 8338 environmental *spo0A* sequences obtained from the sediments of Lakes Geneva (Switzerland-France) and Baikal (Russia). The phylogenetic affiliation of the environmental sequences revealed a substantial number of new clades within endospore-formers. This novel culture-independent approach provides a significant experimental improvement that enables exploration of the diversity of endospore-forming Firmicutes.

2.1 Introduction

Endospore-formers are Gram-positive bacteria from the phylum Firmicutes, although not all species in this phylum can form endospores. In culture collections Firmicutes represent the second most abundant bacterial phylum known (Klenk and Goker 2010). For the endospore-forming species, the resilient outer cortex of the endospores and the small acid-soluble proteins stabilizing their DNA (Driks 2002; Onyenwoke et al. 2004; Yudkin and Clarkson 2005) facilitate their dispersion and capacity to colonize every habitat on Earth (Staley and Gosink 1999). Indeed, endospore formers have been found in a wide range of environments on Earth's surface and subsurface (Nicholson et al. 2000; Nicholson 2002). Although molecular biology techniques have greatly contributed to the general comprehension of microbial diversity, investigation of the diversity of endo- and exospore-forming bacteria remains problematic and warrants improved methodology. In a recent phylogenetic assessment of microbial communities in a diverse set of environments, a surprisingly small number of known microbial groups containing spore-formers was observed (von Mering et al. 2007). Although the frequency of endospore-formation in Firmicutes varies significantly among four different environments, one explanation for their underrepresentation in genomic analyses is that spores can resist the protocols used for extracting DNA from vegetative cells.

Previous studies have identified a number of common genetic elements for endospore formation (Arcuri et al. 2000; Onyenwoke et al. 2004; Paredes et al. 2005;

Dubey et al. 2009). Additionally, recent work in comparative genomics yielded a comprehensive set of the genetic elements involved in forming a minimal sporulation core (Galperin et al. 2012; Abecasis et al. 2013). However, this information has not yet been translated into the development of specific molecular markers for diversity surveys of endospore-forming Firmicutes in environmental samples.

The aim of this study was to develop a culture-independent approach to reveal the diversity of endospore-forming Firmicutes. To achieve this, we identified a functional marker for endospore formation from the genes involved in the sporulation pathway. Furthermore, since the suitability of different DNA extraction methods emerged as a potential caveat for the detection of endospore-formers, the primer design was complemented by experiments testing different DNA extraction methods on cultures and on lake sediment samples. As a final step, to target the endospore-forming fraction of the bacterial community and discover its diversity, we amplified and sequenced the sporulation gene *spo0A* directly from DNA extracted from sediments. To the best of our knowledge, this is the first environmental set of *spo0A* sequences, whose subsequent analysis reveals a large diversity of endospore-forming bacteria.

2.2 Materials and Methods

Preparation of cells and endospores for DNA extraction test

Cell cultures of *Paenibacillus alvei* and *Bacillus subtilis* were grown in nutrient broth (NB) and *Lactococcus lactis* subsp. *lactis* in DSM medium 92 at 30℃. Endospore preparations of *P. alvei* and *B. subtilis* were obtained with Schaeffer sporulation medium (Schaeffer et al. 1965), by vigorous shaking until cultures were composed of >90% spores. Numbers of cells and endospores were determined microscopically using a Neubauer counting chamber. Cell and endospore preparations were then centrifuged at 6000 x g for 10 min, re-suspended in phosphate-buffered saline solution (PBS) to a density of 10^8 cells/ml and 10^9 endospores/ml, respectively. Four ml of this preparation was collected by centrifugation at 10,000 x g for 10 min, and pellet was stored at -20℃ until DNA extraction.

DNA extraction of cells and endospores

DNA extraction with a repetitive protocol was performed using the MP FastDNA Spin Kit for soil (MP Biomedicals, Santa Ana, CA, USA). First round of cell lysis was achieved by bead-beating at 50 strokes/sec with the TissueLyser LT (QIAGEN, Hilden, Germany) for 10 min. The sample was then centrifuged and 900 µl of supernatant fluid (containing free DNA) was collected in a new tube. Lysis buffer was again added to the sample pellet before subjecting it to a second round of bead-beating for 5 min. After centrifugation and supernatant fluid collection in new tube, the cell lysis procedure was repeated a third time. The three supernatants were then processed separately following the standard protocol. DNA yield was measured with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) us-

ing the Quant-iT dsDNA BR assay kit, following the manufacturer's instructions. DNA extraction yield is given in percent of cumulated DNA quantity.

DNA extraction from sediment samples

Sediment samples from Lake Geneva (sampling in June 2011) and Lake Baikal (sampling August 2010) were retrieved using a push-corer. Upon return to the surface, the core fractions two-to-seven cm were immediately sub-sampled in the center using sterile cut-open syringes. Samples were then stored at -20 ℃ until DNA extraction. DNA from sediment samples was extracted using three different protocols:

Protocol 1. Standard extraction with *in situ* lysis in 0.5 g sediment using the MP FastDNA SPIN Kit for Soil, following the manufacturer's instructions.

Protocol 2. Repeated extraction using MP FastDNA SPIN kit with the following modifications: 0.5 g sediment was subjected to three repetitive extractions with *in situ* lysis using bead-beating at 50 strokes per second with the Tissue Lyser LT (QIAGEN) for 10 min. The sample was then centrifuged and 900 µl of supernatant fluid was collected in a separate tube. Lysis buffer was again added to the samples before subjecting to a second round of bead-beating for 5 min, then centrifuged and supernatant fluid collected. This procedure was repeated a third time. The three supernatants were then processed separately following the standard protocol. Finally, the three extractions were pooled and DNA precipitated with 0.3 M Na-acetate and ethanol (99%) and washed with ethanol (70%) before being re-suspended in sterile water.

Protocol 3. Indirect extraction, separating biomass from sediment particles prior to lysis. Three grams of sediment were homogenized with 15 ml of dispersing agent (1% Na-hexa-meta-phosphate) using an Ultra-Turrax homogenizer (IKA) at 15,500 rpm for two minutes to separate cells from the sediment matrix. Coarse particles were then removed from the slurry by centrifugation at 20 x g for 1 min, and the supernatant (containing the cells) was collected on a nitrocellulose membrane of 0.2 µm pore size (Whatman, Dassel, Germany). The cell separation step was then repeated. Filters were immediately frozen in liquid nitrogen and stored at -80 ℃. DNA was then extracted directly from the membrane with Protocol 2. DNA yield was measured with a Qubit 2.0 Fluorometer using the Quant-iT dsDNA BR assay kit, following the manufacturer's instructions. DNA quality was verified by agarose gel electrophoresis and by spectrophotometer absorbance at 260 and 230 nm using NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA).

Quantification of total bacteria

Quantification of bacterial DNA in sediment extracts was performed by real-time quantitative PCR of the V3 region of the 16S rRNA gene with primers 338f and 520r (Ovreas et al. 1997). The qPCR mix contained 1 µL of 10-fold diluted DNA template (1.3 to 8.4 ng/µL), 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 instru-

ment (QIAGEN) with the program: enzyme activation at 95 ℃ for 5 min, 40 cycles of denaturation at 95 ℃ for 5 sec, annealing at 55 ℃ for 15 sec and extension at 72 ℃ for 20 sec. Thresholds (Th), Ct values, and derivatives of melting curves were determined using Rotor-Gene 6 software. All extracts were analysed in triplicate. For quantification three independent plasmid standards series with 300 to 3,000,000 gene copies/µL of the 16S rRNA gene of an environmental clone were included.

Test of *spo0A* primer on pure strains

Degenerate primers amplifying a 602 bp sequence of the *spo0A* gene were designed for this study. The detailed description of the design is available in supplementary information. Primer sequences of the forward primer was (spo0A166f)(5'-GATATHATYATGCCDCATYT-3') and the reverse primer (spo0A748r)(5'-GCNACCATHGCRATRAAYTC-3'). The primers were tested on DNA extracts from pure strains (Figure 2.1) obtained with the innuPREP Bacteria DNA Kit (Analytik Jena, Jena, Germany). PCR reactions were performed with 0.5 ng DNA template, 1 x reaction buffer (TaKaRa Bio, Shiga, Japan), 3 mM MgCl2, 10 µg bovine serum albumin (BSA; New England Biolabs, Ipswich, MA, USA), 1 U of 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. Reactions were performed in an Arktik Thermo Cycler (Thermo Fisher Scientific, Vantaa, Finland) with the following temperature program: initial denaturation at 94 ℃ for 5 min; then 10 cycles of denaturation at 94 ℃ for 30 sec, touchdown annealing starting at 55 ℃ with decrease of 0.3 ℃ per cycle for 30 sec and elongation at 72 ℃ for 1 min; followed by 30 cycles of denaturation at 94 ℃ for 30 sec, annealing at 52 ℃ for 30 sec and elongation at 72 ℃ for 1 min; and a final extension at 72 ℃ for 5 min. The efficiency of amplification results (correct fragment size and band intensity) were verified by running the products on a 1% agarose gel, stained in 3 x GelRed bath (Biotium, Hayward, USA).

Quantification of endospore-forming bacteria

Quantification of *spo0A* gene was performed as mentioned above for the 16S rRNA gene but with the primers spo0A655f and spo0A923r (Bueche et al. 2013). The qPCR mix contained 1 µL of 10-fold diluted DNA sample (1.3 to 8.4 ng/µL), 0.76 µM of each primer and 1 x QuantiTect SYBR Green PCR Kit (Invitrogen). Total reaction volume of 20 µL was reached with PCR-grade water. The program differed in an annealing at 52 ℃ for 30 sec and extension at 72 ℃ for 30 sec. For quantification three independent plasmid standards series with 30 to 300,000 gene copies/µL of *spo0A* gene of *B. subtilis* were included.

Sequencing of the *spo0A* gene amplicons

One extract from each DNA extraction protocol from both sediments (Lake Geneva and Baikal) was subjected to amplicon sequencing of the *spo0A* gene. The primers

(spo0A166f and spo0A748r) described above were used for amplification. PCR reactions were performed with 0.5 ng DNA template, 1 x reaction buffer (TaKaRa, Shiga, Japan), 3 mM MgCl2, 10 µg BSA, 1 U of 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 run on an Arktik Thermo Cycler (Thermo Fisher Scientific) with the following temperature program: initial denaturation at 94 ℃ for 5 min; then 10 cycles of denaturation at 94 ℃ for 30 sec, touchdown annealing starting at 55 ℃ with decrease of 0.3 ℃ per cycle for 30 sec and elongation at 72 ℃ for 1 min; followed by 30 cycles of denaturation at 94 ℃ for 30 sec, annealing at 52 ℃ for 30 sec and elongation at 72 ℃ for 1 min; and a final extension at 72 ℃ for 5 min. Amplified sediment samples were sent for barcode amplicon sequencing with Roche GS FLX+ (Eurofins MWG Operon, Ebersberg, Germany).

Sequence analysis

Sequences were binned according to their barcode and the corresponding sample origin (Lake and DNA extraction protocol), and filtered according to Phred (Ewing and Green 1998) quality score (minimum of 30) and length. *De novo* operational taxonomic units (OTUs) from the curated sequences were defined with the pick-otus.py program (QIIME package) using the Uclust method applying a cut-off of 97% nucleotide identity, based on the definition of OTUs applied for the 16S rRNA gene (Caporaso et al. 2010). Centroids from each OTU cluster were used to build cladograms with Newick Utilities (Junier and Zdobnov 2010). OTUs from the samples from the indirect extraction method were then compared using BLASTp (Altschul et al. 1997) against a Spo0A database containing all the sequences in InterPro (Mulder et al. 2002) in order to identify the closest Spo0A sequence belonging to a known genus.

All metagenomic sequences were submitted to Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under Accession Numbers SRR870694, SRR870695, SRR870696, SRR870698, SRR870699, and SRR870700.

2.3 Results and Discussion

Identification of molecular markers for endospore-forming bacteria

From an initial dataset of 59 genome sequences of endospore-forming Firmicutes, including both finished (48) and draft (11) genomes, 27 genomes (Supplementary Table 2.5) were selected for the search of common orthologous sporulation genes. Redundant species were reduced to a single species representation to avoid over fitting to specific species. Furthermore, the amount of genomes was reduced given that we observed a large variation in the number of annotated sporulation-associated genes in the 59 initial genomes. Part of this variation was explained

by annotation problems in the uncompleted genomes. For example, when the distribution of sporulation-associated genes was analysed in the 59 genomes, two peaks: one around 60 genes and a second at 190 genes were detected (Supplementary Figure 2.4). However, using only well-annotated "finished" genomes, there was a shift in the distribution towards 80-90 minimal genes. Therefore, to avoid any exclusion of orthologs by annotation errors, only finished genomes with more than 60 sporulation-related genes were considered. This minimal number of genes coincides with recent results suggesting that around 60 protein-coding genes are essential for sporulation in Bacilli and Clostridia (Galperin et al. 2012).

We selected these 27 genomes (Supplementary Table 2.5) in order to create a balance between Bacilli (12) and Clostridia (15) and to prevent biases due to phylogenetic distribution (Supplementary Table 2.5). These genomes originated from diverse habitats including soil (7), freshwater (2), sediment (2), clinical samples (7), deep surface habitats (3), hot springs (3), and others (3). The majority of the genomes (20) correspond to mesophilic microorganisms; six are thermophilic and one psychrotolerant. Additionally, one of the mesophilic species (*Alkaliphilus metalliredigens* QYMF) was reported to be both halophile and alkalophile.

Orthology groups were delineated based on best reciprocal BLASTP hits on the annotated sporulation genes from the 27 genomes. Each sequence in the set was BLASTPed against all sequences except those of the same species (thus avoiding paralogs). The best hit in each species was retained, and sequence pairs that were each other's best match were defined as best reciprocal hits (BRHs). Putative orthology groups were defined using the algorithm used by OrthoDB (Kriventseva et al. 2008). In this manner, six orthologous genes (*spo0A*, *spoIVB*, *spoVAC*, *spoVAD*, *spoVT* and *gpr*) were found to be common and highly conserved among endospore-forming Firmicutes (Supplementary Table 2.6). All six genes are part of the core sporulation gene set that seems to be indispensable for sporulation, appearing in both classes of endospore-forming Firmicutes: the Clostridia and Bacilli (Galperin et al. 2012). A phylogenetic reconstruction based on the concatenated sequences of these 6 genes was similar to the phylogeny inferred from the 16S rRNA gene (Supplementary Figure 2.5). In particular, the phylogeny showed a clear separation between Bacilli and Clostridia. Based on the analysis of the phylogeny and conservation profile of the individual genes (Supplementary Figures 2.6 and 2.7), *spo0A* was chosen as a molecular marker. The phylogenetic reconstruction based on Spo0A sequences alone (Supplementary Figure 2.6) was consistent with the phylogeny based on the core genes and the 16S rRNA gene (Supplementary Figure 2.5), and supports a recent report on the separation of the *B. subtilis* and *Bacillus cereus* clades (Bhandari et al. 2013). Additionally, the conservation profile showed two highly conserved regions flanking a highly variable region covering 300 bp (Supplementary Figure 2.6). The stage 0 sporulation gene A (*spo0A*) is the master regulator of sporulation. No convincing homolog of *spo0A* has been found outside the Firmicutes (Brill and Wiegel 1997; Onyenwoke et al. 2004). A recent profile analysis of *spo0A* on 626 genomes found one putative orthologous sequence for each of the 46 endospore-forming genomes and one ortholog in a single non-endospore forming genome (Traag et al. 2013). The ability of some Firmicutes species to form endospores has not yet been experimen-

tally confirmed, even though they contain the *spo0A* gene, and are thus defined as asporogenic (Galperin et al. 2012). Some asporogenic species might have truly lost the trait of sporulation in the course of evolution, but they still conserve the response regulator gene as a relic of this. Such species could trigger false positives when using *spo0A* as a functional marker. Conversely, based on the analysis of the *spo0A* gene in some asporogenic strains, it could be that some are actual endospore-formers, however the phenotype has not been observed (Abecasis et al. 2013). Since *spo0A* is one of the best-studied sporulation genes, it is often annotated automatically, leading to a rapidly growing database of *spo0A* sequences. Although the risk for false positive exists, *spo0A* can nonetheless be considered an ideal candidate as a molecular marker to target endospore-forming Firmicutes in environmental samples.

Design and validation of *spo0A* primers

Degenerate primers for diversity studies based on the *spo0A* gene were designed. The *spo0A* genes of the 27 genomes previously mentioned were aligned using CLUSTALW (Thompson et al. 1994) and scanned for conserved regions. Seven degenerate forward primer regions and 10 reverse primer regions were defined (Supplementary Tables 2.7 and 2.8). As a first screen these primers were tested in all combinations. Based on specificity, amplification efficiency and fragment length, the primer sequences *spo0A*166f (5'-GATATHATYATGCCDCATYT-3') and *spo0A*748r (5'-GCNACCATHGCRATRAAYTC-3') were selected (Figure 2.1). Using these primers, a sequence of 602 bp length could be amplified.

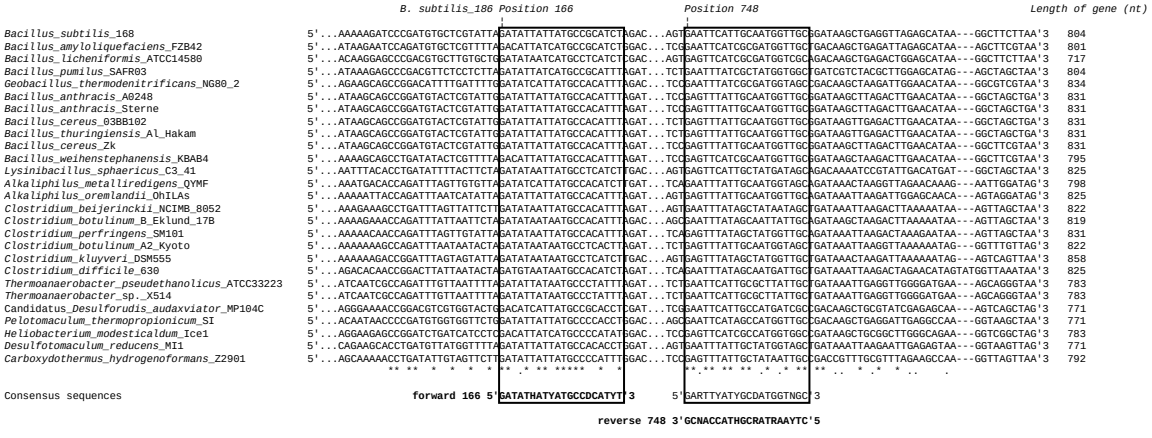


Figure 2.1: Alignment of 27 *spo0A* gene sequences encompassing regions used for primer design. The position of the primers is indicated on top of the figure using the *spo0A* gene from *Bacillus subtilis* as reference. The annealing sites are marked by two squares. The consensus sequence is shown underneath (in bold). For the reverse primer, the reverse complement sequence is indicated. The degenerate positions in the primers are shown by the letters H (A, C, or T), Y (C or T), D, (A, G, or T), R (A or G), N (A, C, G, or T).

To validate the *spo0A* primers, amplification efficacy was determined using a collection of 53 pure cultures (Table 2.1). The cultured strains corresponded mainly to the class Bacilli and in particular to *Bacillus* spp., with a few strains from other genera such as *Anoxybacillus*, *Brevibacillus*, *Geobacillus*, *Halobacillus*, *Lysinibacillus*, and P*aenibacillus*. A PCR product with the correct size (602 bp) was obtained in 35 out of 43 endospore-forming bacterial cultures belonging to nine different genera. From the strains tested, some did not yield a PCR amplicon (e.g. *Brevibacillus brevis* or *Geobacilllus themoparaffinovorans*), but overall the primers demonstrated good coverage. Three strains from the genera *Alicyclobacillus* and one *Sulfobacillus* strain were also included but did not amplify with the primers. The match between the primer sequences and the *spo0A* gene sequence in two available genomes for these genera (Supplementary Figure 2.8) revealed that in the case of *Alicyclobacillus acidocaldarius* TC41 there are five mismatches with the forward primer, four of which are found in the 3' region that could impair annealing and amplification. However in the case of *Sulfobacillus acidophilus*, one mismatch with each primer was observed, and thus the failure of the amplification is surprising. A subsequent inhibition test suggests that the lack of amplification was probably due to a chemical remnant from the culture medium (data not shown). Only three Clostridia strains could be tested, and two of them gave a positive amplification signal (*Clostridium pasteurianum* and *Desulfotomaculum reducens*). None of the 10 non-endospore formers amplified with the primers. The non-endospore formers included three exospore-formers (*Actinobacteria*), one non-endospore-forming Firmicute (*Lactococcus lactis* subsp. *lactis*) and six members from outside the Firmicutes (five *Proteobacteria* and one *Bacteroidetes*).

Table 2.1: Specificity test for the amplification of *spo0A* using the primers spo0A166f and spo0A748r. A PCR product of the correct size (602 bp) is indicated by a "+" sign, no PCR product is indicated by "-". A total of 53 strains were analyzed, of which 43 corresponded to endospore-forming bacteria.

Genus	Species	Optimal growth (℃)	Endospores (y/n)	Amplification of *spo0A*
Alicyclobacillus	*acidocaldarius*	55	y	-
Alicyclobacillus	*tolerans*	50	y	-
Alicyclobacillus	sp.	50	y	-
Anoxybacillus	sp.	70	y	+
Anoxybacillus	sp.	55	y	+
Bacillus	*aquimaris*	25	y	+
Brevibacillus	*brevis*	30	y	-
Bacillus	*cereus*	30	y	+
Bacillus	*cereus* var. *mycoïdes*	24	y	+
Bacillus	*horikoshii*	25	y	+
Bacillus	*jeotgali*	45	y	+
Bacillus	*licheniformis*	45	y	+
Bacillus	*macerans*	30	y	+
Bacillus	*niabensis*	20	y	+
Bacillus	*niacini*	45	y	+
Bacillus	*oceanisediminis*	45	y	+
Bacillus	*pallidus* T	60	y	+
Bacillus	*polymyxa*	30	y	+
Bacillus	*pumilus*	30	y	+
Bacillus	*selenatarsenatis*	37	y	+
Bacillus	*stearothermophilus*	55	y	+
Bacillus	*subtilis*	30	y	+
Bacillus	*thermoglucosidasius*	65	y	+
Bacillus	*thermoruber*	45	y	+
Bacillus	*thuringiensis*	45	y	+
Bacillus	*tusciae*	55	y	-
Bacillus	*firmus*	20	y	+
Bacillus	*vietnamensis*	37	y	+
Bacillus	sp.	45	y	+
Brevibacillus	*agri*	30	y	+
Brevibacillus	*formosus*	30	y	+
Clostridium	*pasteurianum* T	37	y	+
Clostridium	sp.	30	y	-
Desulfotomaculum	*reducens*	NA	y	+
Geobacillus	sp. A14	50	y	+
Geobacillus	*thermoleovorans*	50	y	+
Geobacillus	*thermoparaffinovorans*	70	y	-
Geobacillus	sp.	65	y	+
Halobacillus	*trueperi*	70	y	+
Lysinibacillus	*sphaericus*	30	y	+
Lysinibacillus	sp.	30	y	+
Paenibacillus	*alvei*	30	y	+
Sulfobacillus	*acidophilus*	50	y	-
Actinomyces	sp.	24	n	-
Escherichia	*coli*	37	n	-
Hymenobacter	*daecheongensis*	30	n	-
Lactococcus	*lactis* subsp. *lactis*	37	n	-
Streptomyces	*grisechromogenes*	30	n	-
Streptomyces	sp. nu40	30	n	-
Stenotrophomonas	*rhizophila* SMPG9	20	n	-
Comamonas	sp. lb15	20	n	-
Delftia	sp. S17	20	n	-
Pseudomonas	*fluorescens* NBRC12568	20	n	-

Comparison of DNA extraction methods on cells and endospore preparations

We conducted experiments testing different DNA extraction methods in order to examine and optimize the extractability of DNA from endospores. We used the commercially available FastDNA Spin kit for soil (MP Biomedicals), previously shown to produce high DNA yields and a relatively good phylogenetic distribution from soil samples and low biomass samples from the deep biosphere (Webster et al. 2003; Dineen et al. 2010). The use of commercially available products increases repeatability and standardization of the extraction procedure.

A DNA extraction protocol composed of three repetitive extraction cycles (to increase the total mechanical disruption by bead-beating time) was first tested on cell cultures of *L. lactis* subsp. *lactis*, a non spore-forming Firmicute, and on cell and endospore preparations of *P. alvei* and *B. subtilis* (Figure 2.2). The total, cumulative yield of DNA isolated from cell cultures (normalized to 10^8 cells) was 198.8 ±72.8 ng for *L. lactis* subsp. *lactis*, 497.6 ±36.9 ng for *B. subtilis* and 1402.3 ±254.8 ng for *P. alvei*. The yields from endospore preparations were significantly lower (p=0.002) with 86.5 ±3.3 ng for *B. subtilis* and 83.1 ±2.0 ng for *P. alvei*. Over three consecutive extraction cycles, the quantity of isolated DNA increased considerably; the total yield could almost be doubled when adding a second and third round of extraction. This was especially true for endospore preparations, where the overall percentage of the total DNA isolated after the first extraction step was significantly lower (average 57.8 ±5.8%, p=0.009) than that of the vegetative cells (average 76.0 ±13.6%). After the second extraction, the percentage of isolated DNA from endospores was still significantly lower (average 81.1 ±4.1%, p=0.039) than from cells (average 92.3.0 ±9.7%). This result agrees with previous studies showing that Gram-positive cells or endospores only lyse with harsh physical methods (More et al. 1994; Zhou et al. 1996; Kuske et al. 1998). In our experience, three successive cycles is the best balance of time, cost, and overall DNA yield and quality of the DNA extract.

Test of DNA extraction methods on environmental samples

Different DNA extraction protocols were then tested on sediment samples collected during a research campaign with the MIR manned submersibles in Lake Baikal (Russia) and Lake Geneva (Switzerland) (pap of sampling location in Supplementary Figure 2.14 and image of sediment core in Supplementary Figure 2.15). Three DNA extraction protocols were tested, all based on the MP Fast DNA Spin Kit for Soil. Protocol 1 (standard) corresponded to a standard extraction with *in situ* lysis in 0.5 g sediment following the manufacturer's instructions. Protocol 2 (repeated) also corresponded to *in situ* lysis in 0.5 g sediment, but with three sequential extractions, as was used for cells and endospores (see Figure 2.2). In Protocol 3 (indirect) the biomass was separated from sediment particles prior to lysis.

For the environmental samples, the DNA yields and humic acid contamination (determined by absorbance ratio at 260/230 nm) of the different protocols varied

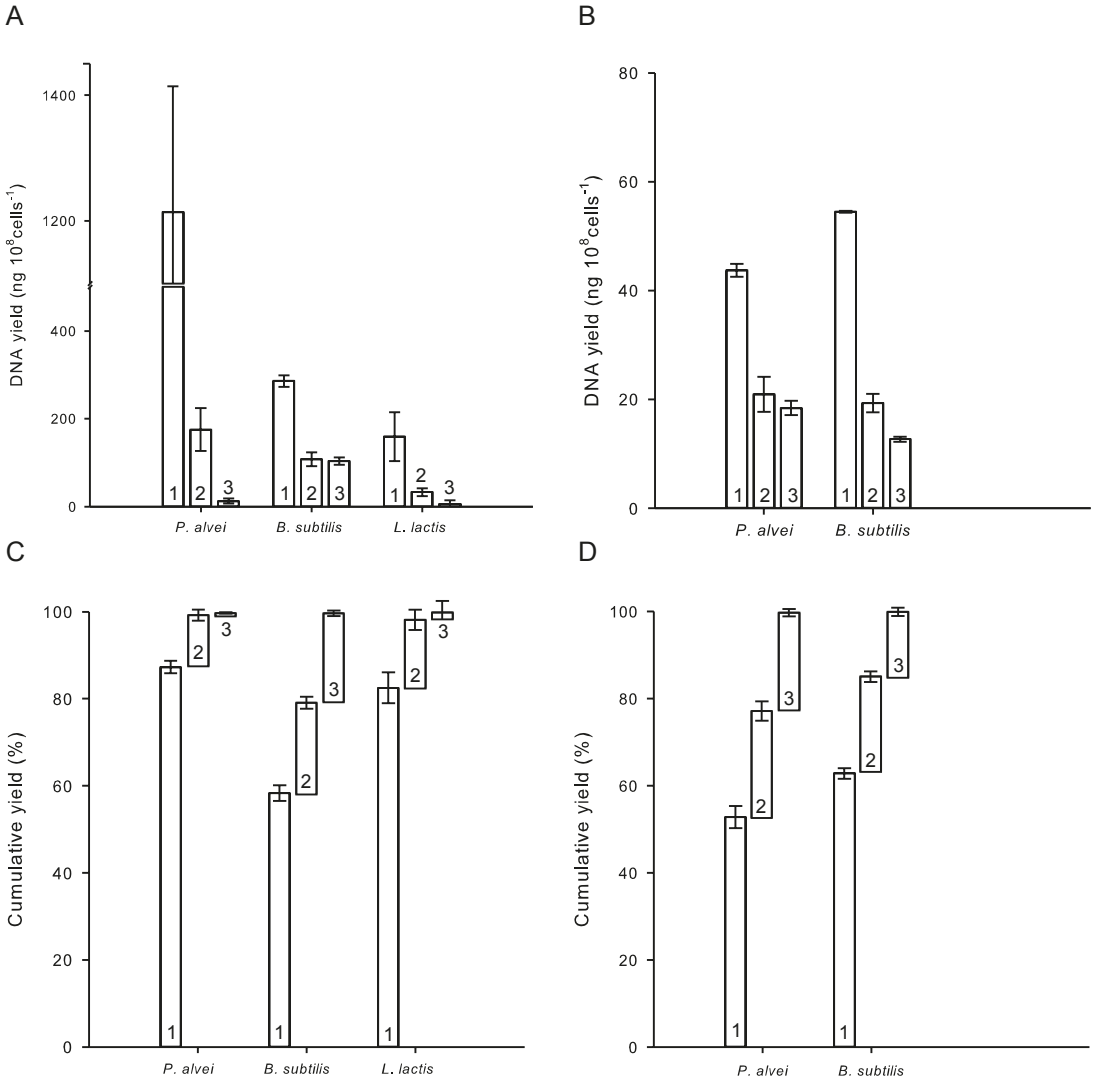


Figure 2.2: DNA extraction yields (ng/10^8 cells or ng/10^9 endospores; top panels) and cumulative yield (in percentage of the total; lower panels) obtained for each sequential extraction steps (1, 2 and 3) for cell cultures (A and C) and endospore preparations (B and D).

(Table 2.2). The DNA yield after the repeated extractions (protocol 2 and 3) was lower than after the standard method, particularly for the indirect extraction (protocol 3). Cell lysis alone is therefore not the determining factor for DNA yield. Lower yields could be due to increased adsorption of DNA to clay particles when bead-beating for longer times (Frostegard et al. 1999), due to disturbance of DNA-silica binding from co-extracted humic acids or salts, or due to the exclusion of specific morphological groups by the biomass separation procedure used in the indirect extraction protocol. Overall, DNA extracts from sediments of Lake Geneva were less contaminated with humic acids, visually obvious given the color of the extract. The DNA extracted from sediment of Lake Baikal had lower purity (brownish color) and lower quantity.

Table 2.2: Comparison of DNA extraction protocols on sediment samples from Lake Geneva and Lake Baikal. DNA was extracted with standard (protocol 1) or modified extraction methods (protocol 2 (repetitive); protocol 3 (indirect)). DNA yield in ng/g sediment. Gene abundances were normalized to 1 ng DNA and 1 g sediment.

Sample	Protocol	DNA yield	A260/230	Abundance 16S rRNA gene (*10^4)	Abundance *spo0A* gene	Ratio *spo0A*/16S rRNA gene (%)	Sequences	Curated sequences	OTUs
Lake Geneva	1	7872	1.5	25.9 ± 2.3	147.8 ±9.8	0.06	2349	804	91
	2	3888	1.3	20.4 ± 1.6	144.2± 24.1	0.07	2703	926	13
	3	2240	1.43	68.6± 1.7	428.8 ± 27.1	0.06	4210	1590	409
Lake Baikal	1	2480	0.84	17.5 ± 0.2	339.4± 30.3	0.19	3737	1470	212
	2	2176	0.87	24.3 ± 2.6	611.0 ± 41.3	0.25	4759	1720	71
	3	616	0.89	38.0 ± 1.9	1414.7 ± 56.3	0.37	4089	1828	289

Gene abundances of the 16S rRNA gene and the *spo0A* gene were then determined (Bueche et al. 2013). We observed an inverse correlation between DNA yields and gene copy numbers of 16S rRNA and the *spo0A* genes. There was a 2.6-fold increase of 16S rRNA genes and 2.9-fold increase of detection of *spo0A* gene copy numbers in the extract from Lake Geneva with the indirect method when compared to the standard protocol. In extracts from sediment of Lake Baikal the increase was even more prominent; 2.2-fold for the 16S rRNA genes and 4.2-fold for *spo0A*. Copy numbers of extracts from the multi-cycle protocol were always intermediate. The percentages of *spo0A* genes relative to 16S rRNA genes were constant for samples from the same sediment, independently of the extraction protocol (Lake Geneva sediment 0.063 ± 0.005% and Lake Baikal sediment 0.27 ± 0.075%).

In summary, extracts from the indirect protocols had substantially better amplification despite relatively lower DNA yields. This is most likely due to reduced co-extraction of contaminants that could inhibit the downstream PCR. The same effect is observed with direct *in situ* DNA extraction that provides high yields but lower purity (Leff et al. 1995), often requiring high dilution of the extracts in order to avoid amplification inhibition due to contaminants (Dineen et al. 2010).

Application of the *spo0A* primer on environmental samples

A fragment of 602 bp was amplified from the environmental samples with the *spo0A* primers spo166f and spo748r, and sequenced to assess the diversity of endospore-forming Firmicutes. A total of 21,847 sequences were obtained from the six samples. After curation of amplicon size and quality, a total of 8,338 sequences with an average length of 625 bp remained and were translated to their amino acid sequence and checked for STOP-less open reading frames (ORFs). The numbers of *spo0A* sequences varied substantially among the different extraction protocols as well as between the two sediments (Table 2.2). We observed varying amplification efficiencies depending on sediment type, purity of extract and community composition, among others. In this study, the amount of final *spo0A* sequences was greatest in both sediments (Lakes Geneva and Baikal), in the samples prepared with the indirect extraction protocol.

Phylogenetic distribution of *spo0A* sequences

De novo operational taxonomic units (OTUs) from the curated sequences were defined applying a cut-off of 97% nucleotide identity. This OTU assignment threshold has not yet been experimentally validated for *spo0A*, and therefore is based solely on current knowledge defining bacterial species based on the 16S rRNA gene. The number and distribution of the OTUs varied with the extraction method (Table 2.2). For Lake Geneva sediment, a 4-fold increase in the number of OTUs was obtained in the samples extracted with the indirect protocol over the standard protocol. In sediment from Lake Baikal the differences were smaller, with repeated extraction representing about 25% of the number of OTUs from the indirect extraction. All defined OTUs were then displayed with a phylogenetic tree

(Supplementary Figures 2.20 and 2.21). The branches were collapsed according to the extraction method. Overall, the indirect method is the most promising, revealing entire clusters that do not appear in data from the other two extraction protocols. This result confirms previous research where different community profiles are detected when comparing direct or indirect DNA extraction protocols on the same soil sample (Delmont et al. 2011). Additionally, successive extractions can result in a shift in the community composition (Feinstein et al. 2009), as was observed here. However, in the case of endospore-forming Firmicutes, repetitive extractions (protocol 2) from the same sediment sample produced a poor representation of the community with groups that are either not represented at all (e.g. Bacilli in Lake Geneva, yellow branches in Supplementary Figure 2.20) or underrepresented (e.g. Geobacilli in Lake Baikal, yellow branches in Supplementary Figure 2.21). Separation of cells from the sediment matrix prior to DNA extraction requires additional laborious and time-consuming steps. However, with respect to the time and cost of downstream processes (sequencing, analysis and data storage), it was worth increasing the effort of applying an indirect and repetitive DNA extraction method, in particular for endospore-forming Firmicutes prone to be underrepresented, as shown in this study. The closest Spo0A sequence belonging to a known genus was then searched for each OTU from the samples extracted with the indirect method (Figure 2.3 for the most abundant genera and Table 2.3 for details on Lake Geneva and Table 2.4 on Lake Baikal community). In both samples (Lake Geneva and Lake Baikal), the OTUs could be assigned to sequences belonging to both classes; the Clostridia and Bacilli. The detection of both classes of endospore-forming Firmicutes with the primers for *spo0A* supports the results of the primer validation in pure cultures. Furthermore, we can detect a broader range of endospore-forming Firmicutes, including strains that were not present in the pure cultures. This included one additional genus of Bacilli (*Solibacillus* spp. detected in Lake Geneva) and 11 additional genera of Clostridia. Moreover, the detection of 48 OTUs related to *Sulfobacillus* spp. in Lake Baikal supports the analysis of the annealing sites (Supplementary Figure 2.3) and PCR inhibition due the composition of the culture medium to grow *S. acidophilus*. For Lake Geneva the OTUs could be assigned to 48 groups according to the closest Spo0A sequence (Table 2.3). The most abundant genera were *Geobacillus* (30.2% of all sequences), *Clostridium* (25.9%), and *Bacillus* (23.1%), *Paenibacillus* (7.8%), and *Brevibacillus* (6.9%). *Lysinibacillus*, *Alkaliphilus*, *Thermincola*, *Desulfosporosinus* were between 0.6 and 1.4%, and the remaining 1.8% of the sequences were assigned to 7 other genera (Figure 2.3).

The distribution of the groups within clusters in the cladogram was verified (Supplementary Figure 2.20). For some of the species the position in the cladogram was consistent with the affiliation by BLAST. For example the groups J, B, D, E, I, G, H, A, F, C, and M clustered together within Bacilli. Likewise, all the groups consisting of Spo0A related to Clostridia clustered together (clusters 27 to 60). However, OTUs in large groups (i.e. groups B, D, AF, C, T, V, Y, and Z) appeared distributed in several clusters in the cladogram. Surprisingly, several clusters within the groups AF, L, AG and AH, which were affiliated to *Geobacillus*, *Brevibacilllus* or *Paenibacillus*, formed a third branch (indicated as "undefined" in

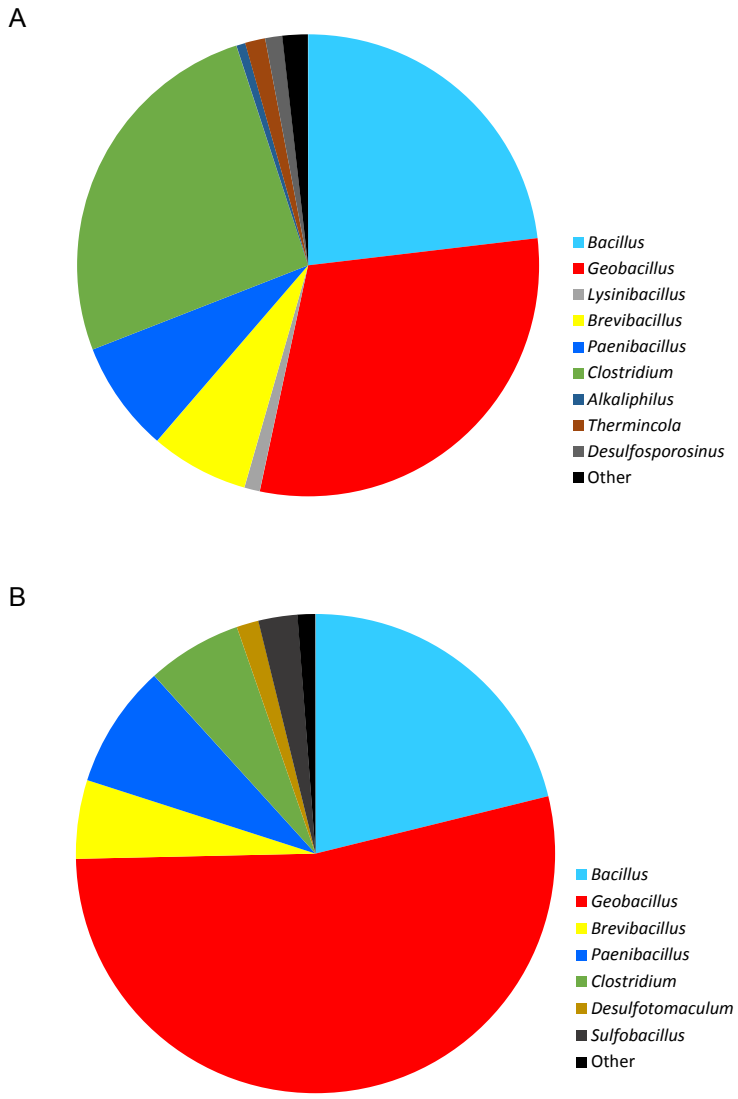


Figure 2.3: Distribution of Spo0A OTUs from the indirect DNA extraction method classified into genera Lake Geneva (A). Lake Baikal (B).

the cladogram) more closely related to Clostridia than to Bacilli. For these groups the identity levels of Spo0A were in some cases very low (down to 24%). Finally, the groups U (cluster 34), L (cluster 58), and Q (cluster 30) likely reflect annotation errors in the reference sequences, since they were placed within consistent clusters from a different phylogenetic affiliation.

In samples from Lake Baikal 35 groups were assigned (Table 2.4). The most abundant genus was *Geobacillus* with 53.5%, followed by 21% corresponding to the genus *Bacillus*. Contrary to Lake Geneva (25.9%), the genus *Clostridium* was poorly represented (6.4%). The remaining composition consisted of *Paenibacillus* (8.3%), *Brevibacillus* (5.3%), *Sulfobacillus* (2.6%), and *Desulfotomaculum* (1.5%), with the final 1.2% of the sequences assigned to 4 other genera (Figure 2.3). Clusters 1 to 28 corresponded to different OTUs for which the closest relative species

belongs to the class Bacilli.

As for Lake Geneva, groups containing a large number of OTUs did not cluster together (e.g. groups B and E). OTUs related to species from the class Clostridia corresponded to clusters 35 to 53, although the closest relative to group AN (cluster 35) is likely a wrongly annotated bacillus-like Spo0A sequence. The grouping of some OTUs related to *Geobacillus* (group AF), *Brevibacillus* (cluster L), and *Paenibacillus* (AH) species was closer to Clostridia than to Bacilli (indicated as "undefined" in Supplementary Figure 2.21). Obtaining *spo0A* sequences directly from the environment opens the possibility of studying the patterns of distribution of endospore-forming Firmicutes.

Although the *spo0A* reported here represent the first environmental sequences reported in literature, already an interesting pattern could be observed for the two sediments studied. In both sediments, *Geobacillus* represented the dominant group. Members of the genus *Geobacillus* have been traditionally isolated from environments with high temperatures, as part of the community of thermophilic Firmicutes growing with temperature optima ranging from 45 to >70°C (Nazina et al. 2001). According to this, a previous study characterizing the community of Gram-positive bacteria in marine sediments at an intermediate depth (500m) between the sediments studied here (284 and 1597 m deep), found a diverse community of Actinobacteria and Firmicutes, but no isolate was affiliated with Geobacilli (Gontang et al. 2007). In contrast, various species of *Geobacillus* have been isolated from cold soils (Marchant et al. 2011) and several publications have shown the isolation of thermophilic endospore-forming Firmicutes from cold marine sediments (Bartholomew and Paik 1966; Hubert et al. 2010; de Rezende et al. 2013). These results suggest that endospores are in most cases allochthonous and have been deposited at the time of sedimentation, but several metabolic activity tests indicate that these microorganisms do not thrive in temperatures below 20°C. While the activity and the origin of the Geobacilli found in the present study were not assessed, it is an aspect that will be further studied. In contrast to Geobacilli, the Clostridia sequence abundance differs greatly between the two sediments studied. While Clostridia represented 26% of the sequences in Lake Geneva sediments, in Lake Baikal their abundance was only 6.4%. An interesting ecological feature within the group of endospore-forming Firmicutes is that there exist three ecotypes: aerobic, facultative anaerobic, and strictly anaerobic. With some exceptions, aerobic types cluster among the class Bacilli and anaerobes cluster mostly in the class Clostridia (Schleifer 2009). We have found a correlation between an increase in the abundance of Clostridia and lake eutrophication (see Chapter 4 of this manuscript) or the pollution associated with treated wastewater disposal (Sauvain et al. 2013) in other areas of Lake Geneva. We postulate that the larger fraction of Clostridia found in the sediment of Lake Geneva reflects an increasing effect of human activities there, however this needs to be verified.

The distribution and affiliation of the environmental *spo0A* sequences also raised some questions regarding the taxonomy of endospore-forming Firmicutes. The amino acid sequence identities for *Alkaliphilus*, *Thermincola*, *Desulfotomaculum* and *Desulfitobacterium*, to name only a few, are considerably lower (in the range of 70%) than the identities for most of the well-known *Bacillus* and *Clostrid-*

ium species (mostly between 80 up to 100%). This difference could be due to underrepresentation of the former genera in the databases. The taxonomical distribution of the environmental *spo0A* sequences could also reveal problems with the annotation, or more importantly, the potential detection of a yet unknown group of endospore-forming bacteria.

Even though Firmicutes are the second most abundant bacterial phylum in terms of culture representatives (Klenk and Goker 2010), many of the environmental *spo0A* OTUs obtained in this study were only distantly related to reference strains. Therefore a significant effort will be required in order to evaluate the diversity of endospore-forming Firmicutes in environmental samples, including a precise characterization of species belonging to the undefined clusters related to *Geobacillus*, *Brevibacillus* and *Paenibacillus*.

Here, we demonstrate how an improved DNA extraction protocol increases the diversity of endospore-forming Firmicutes retrieved from environmental samples. This is a clear example of how specific methods must be considered by those in the microbial community where traditional molecular microbial ecology methods are inadequate. We designed and validated a primer set for the *spo0A* gene that is specific for endospore-forming bacteria, thus enabling detection of endospore-forming Firmicutes by molecular methods. Environmental sequencing of this gene has opened, for the first time, a window into the diversity of endospore-forming bacteria by culture-independent methods. Additionally, using a targeted sequencing approach for a functional sub-group, the higher resolution and sequence coverage revealed a very diverse community and potentially uncharacterized groups of endospore-forming Firmicutes. Future studies using other environmental samples will likely clarify the environmental relevance and biogeographical distribution patterns of endospore-forming Firmicutes in nature.

Table 2.3: Identification of the closest relative for *spo0A* sequences from Lake Geneva extracted with protocol 3 (indirect method) by BLAST. The identity range is listed as % of amino acid similarity. E-values are the minimum values for each closest relative. Seq = Sequence counts, NDT= not displayed in the cladogram (Supplementary Figures 2.20)

Genus	Species	Group	Seq	Id %	E-value	Cluster
Bacillus	*Bacillus pseudofirmus* OF4	J	7	79-80	1.00E-115	19
	Bacillus methanolicus	B	244	75-90	3.00E-134	2,4,6,7
	Bacillus megaterium	AI	1	99	2.00E-138	NDT
	Bacillus cereus	U	1	81	1.00E-117	34*
	Bacillus cereus subsp. *cytotoxis* NVH 391-98	D	28	76-93	5.00E-136	5,10,18,22
	Bacillus thuringiensis	E	10	84	2.00E-121	9,23
	Bacillus mycoides	I	39	80-100	4.00E-148	17
	Bacillus atrophaeus 942	G	4	86	2.00E-126	15
	Bacillus subtilis	H	1	99	5.00E-150	16
	Bacillus amyloliquefaciens	A	3	75-85	9.00E-120	1,21
	Bacillus licheniformis	AJ	4	75-100	6.00E-149	NDT
	Bacillus anthracis	AV	2	25	3.00E-01	NDT
	Bacillus pumilus	AW	4	99-100	1.00E-151	NDT
	Bacillus pumilus SAFR-032	AX	15	98-99	1.00E-149	NDT
	Bacillus weihenstephanensis BAB4	AY	1	98	8.00E-144	NDT
Geobacillus	*Geobacillus thermoleovorans*	AZ	1	93	1.00E-108	NDT
	Geobacillus thermodenitrificans NG80-2	F	5	98-100	6.00E-144	12
	Geobacillus sp.Y412MC10	AF	429	60-93	9.00E-137	61,64,66,69,71,73,75
	Geobacillus sp. WCH70	C	41	70-94	1.00E-138	3,8,11,13,14
Anoxybacillus	*Anoxybacillus flavithermus* DSM 21510 / WK1	AU	2	78-87	3.00E-124	NDT
Lysinibacillus	*Lysinibacillus sphaericus* C3-41	K	17	68-82	1.00E-120	20
Solibacillus	*Solibacillus silvestris* tLB046	BB	2	86-97	5.00E-146	NDT
Brevibacillus	*Brevibacillus laterosporus*	L	93	24-93	1.00E-134	24,26,58,63,65,68
	Brevibacillus brevis 47	M	15	69-93	1.00E-137	25
Paenibacillus	*Paenibacillus polymyxa*	AG	89	65-97	7.00E-147	62,67,70
	Paenibacillus polymyxa E681	BA	3	98	3.00E-146	NDT
	Paenibacillus mucilaginosus KNP414	AH	29	84-90	2.00E-130	72,74
	Paenibacillus sp. JDR-2	AM	1	90	1.00E-131	NDT
Clostridium	*Clostridium clariflavum* DSM 19732	T	179	67-93	4.00E-138	33,35,37,39

Table 2.3: continued

Genus	Species	Group	Seq	Id %	E-value	Cluster
	Clostridium thermocellum	V	149	57-89	1.00E-132	36,38,40
	Clostridium cellulolyticum DSM 5812	W	42	59-89	2.00E-143	41
	Clostridium botulinum	Y	11	68-84	4.00E-127	42,44,48,50,52,55,59
	Clostridium botulinum B Eklund 17B	BC	5	79-80	1.00E-122	NDT
	Clostridium novyi NDT	BD	3	80-84	1.00E-126	NDT
	*Clostridium ljungdahlii*DSM 13528	Z	11	77-86	1.00E-131	43,46,49,51
	Clostridium perfringens	AA	3	100	4.00E-155	47
	Clostridium sporogenes	AB	5	75-76	1.00E-112	53,54
Alkaliphilus	*Alkaliphilus metalliredigens* QYMF	AC	3	63-76	1.00E-109	56
	Alkaliphilus oremlandii OhILAs	AD	7	66-84	2.00E-125	57
Thermincola	*Thermincola potens* JR	N	22	69-72	7.00E-99	27
Desulfotomaculum	*Desulfotomaculum acetoxidans* DSM 771	O	4	68-69	1.00E-96	28
	Desulfotomaculum kuznetsovii DSM 6115	AE	1	66	7.00E-90	60
Syntrophobotulus	*Syntrophobotulus glycolicus* DSM 8271	R	2	75	1.00E-112	31
Desulfosporosinus	*Desulfosporosinus orientis* DSM 765	P	19	71-81	4.00E-125	29
Desulfitobacterium	*Desulfitobacterium hafniense*	BE	4	31-75	2.00E-110	NDT
Heliobacterium	*Heliobacterium modesticaldum* ATCC 51547	S	6	75-77	4.00E-110	32
Thermosediminibacter	*Thermosediminibacter oceani* DSM 16646	Q	6	40	5.00E-01	30
Thermaerobacter	*Thermaerobacter marianensis* DSM 12885	BF	1	54	6.00E+00	NDT
	Abundance		1567			
	Richness		48			

Table 2.4: Identification of the closest relative for spo0A sequences from Lake Baikal extracted with protocol 3 (indirect method) by. BLAST. The identity range is listed as % of amino acid similarity. E-values are the minimum values for each closest relative. Seq = Sequence counts, NDT= not displayed in the cladogram (Supplementary Figure 2.20)

Genus	Species	Group	Seq	Id %	E-value	Cluster
Bacillus	*Bacillus methanolicus*	B	171	77-90	4.00E-130	6,7,9,10,12,28
	Bacillus cellulosilyticus ATCC 21833	AN	1	66	3.00E-93	35*
	Bacillus megaterium	AI	1	84	5.00E-119	1
	Bacillus megaterium DSM 319	AK	1	94	8.00E-113	3
	Bacillus cereus	U	3	81-84	6.00E-119	NDT
	Bacillus cereus subsp. *cytotoxis* NVH 391-98	D	35	77-85	8.00E-120	NDT
	Bacillus thuringiensis	E	55	81-86	5.00E-123	5,14
	Bacillus mycoides	I	94	77-84	4.00E-122	13
	Bacillus subtilis	H	1	29	6.00E+00	NDT
	Bacillus amyloliquefaciens	A	1	29	8.00E-02	NDT
	Bacillus licheniformis	AJ	23	65-75	4.00E-108	2
Geobacillus	*Geobacillus* sp. Y412MC10	AF	930	61-90	2.00E-132	18,21,23,27,29,31
	Geobacillus sp. WCH70	C	47	73-94	3.00E-137	4,8,11
Brevibacillus	*Brevibacillus laterosporus*	L	60	72-80	5.00E-115	15,30,32,34
	Brevibacillus brevis 47	M	37	78-93	2.00E-136	16
Paenibacillus	*Paenibacillus polymyxa*	AG	8	74-85	4.00E-121	25
	Paenibacillus mucilaginosus KNP414	AH	118	71-88	1.00E-125	17,20,24,26
	Paenibacillus sp. JDR-2	AM	26	58-91	1.00E-131	22
Clostridium	*Clostridium clariflavum* DSM 19732	T	5	72-81	4.00E-116	51
	Clostridium thermocellum	V	5	52-79	2.00E-114	52
	Clostridium cellulolyticum DSM 5812	W	4	87-90	4.00E-143	53
	Clostridium botulinum	Y	21	73-84	5.00E-125	44,46,49
	Clostridium botulinum Kyoto / Type A2	AT	1	75	1.00E-110	NDT
	Clostridium ljungdahlii DSM 13528	Z	49	69-89	3.00E-137	43,45
	Clostridium sporogenes	AB	18	74-76	9.00E-113	48
	Clostridium haemolyticum	AL	1	48	2.00E-27	19
	Clostridium butyricum	AR	4	79	7.00E-120	42
	Clostridium kluyveri DSM 555	AS	9	83-84	1.00E-126	47
	Alkaliphilus oremlandii OhILAs	AD	4	26-77	4.00E-114	50

Table 2.4: continued

Genus	Species	Group	Seq	Id %	E-value	Cluster
Thermincola	*Thermincola potens* JR	N	5	68-71	3.00E-98	NDT
Candidatus *Desulforudis*	Candidatus *Desulforudis audaxviator* MP104C	AQ	8	70-72	3.00E-98	39
Desulfotomaculum	*Desulfotomaculum acetoxidans* DSM 771	O	27	69-70	7.00E-99	38,40
Desulfosporosinus	*Desulfosporosinus orientis* DSM 765	P	5	79-81	3.00E-124	41
Sulfobacillus	*Sulfobacillus acidophilus* TPY	AO	10	65-69	3.00E-99	36
	Sulfobacillus acidophilus DSM 10332	AP	38	62-70	5.00E-101	37
	Abundance			1827		
	Richness			35		

Acknowledgements

We acknowledge the elemo project (`elemo.ch`) for the sampling campaigns using the MIR submersibles. We are grateful for the help of the MIR team and colleagues from the elemo project as well as the help from Manon Brenier in the laboratory. We thank Sevasti Filippidou and Matthieu Bueche for the inhibition test in *Sulfobacillus acidophilus*. Furthermore, we thank Patricia Siering from Humboldt State University for providing endospore-forming strains for the validation test. We would like to thank Thomas Petty from University of Geneva for his comments. This work was supported by Swiss National Science Foundation grant No. 31003A-132358/1.

2.4 Supplementary information

Search for molecular markers for endospore-forming bacteria

Complete and draft genome sequences of 59 endospore-forming isolates were downloaded from the Comprehensive Microbial Resource (CMR, `http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage`, data release 24.0) and Integrated Microbial Genomes (IMG, `http://img.jgi.doe.gov/cgi-bin/w/main`, data release 3.0) websites. Protein and nucleotide sequences of spore-related genes were obtained by search for role category/function "sporulation and germination" (CMR) and "sporulating" (IMG). Additional information on all retrieved genomes was obtained from the GenBank database (`www.ncbi.nlm.nih.gov/genome`). The number of sporulation genes was obtained from the genome annotations. The number of annotated sporulation genes from each of the 59 endospore-forming genomes were retrieved and visualized with a frequency plot (Figure 2.4). In the top panel (a), showing frequency of gene numbers across all 59 genomes, a peak of highest frequency around 60 sporulation genes occurs. In the lower panel (b), the frequency of sporulation genes from the 27 finished genomes is shown, with a peak of highest gene frequency at 90 sporulation genes per genome. This shows that the annotation of the finished genomes is substantially improved, so finally only the 27 finished genomes were selected for this study (listed in Table 2.5).

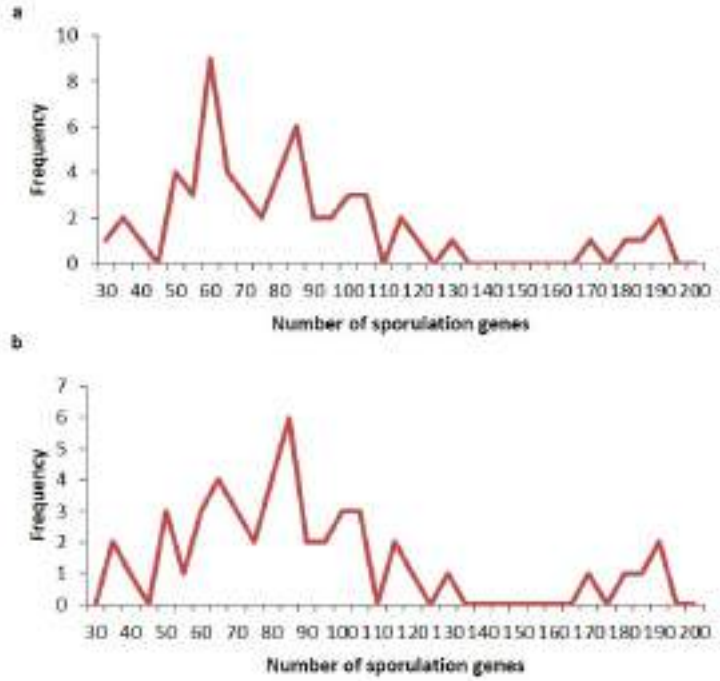


Figure 2.4: Frequency of number of sporulation genes per genome of the 59 (finished and draft) endospore-forming genomes (a) and a subset of 27 finished genomes (b).

Table 2.5: List of genome sequences from the 27 endospore-forming Firmicutes used in this study. Clas= taxonomical classification; B= Bacilli; C= Clostridia; T°= temperature range; M= mesophile; T= thermophile; P= psychrophile; H= hyperthermophile; Sp. Genes= number of sporulation-related genes. CDS= coding DNA sequences, GC Perc= percent GC.

Name	Taxon ID	Clas.	Isolation	Temp	Genes	CDS	rRNA	GC Perc	Sp. genes	Reference
Bacillus amyloliquefaciens FZB42	326423	B	Soil	M	3814	3696	30	0.46	111	(Chen et al. (2007))
Bacillus anthracis A0248	592021	B	Human isolate	M	5418	5291	33	0.35	190	Unpublished
Bacillus anthracis Sterne	260799	B	Soil	M	5521	5287	33	0.35	82	Unpublished
Bacillus cereus 03BB102	572264	B	Human blood	M	5767	5621	42	0.35	183	Unpublished
Bacillus cereus Zk	288681	B	Wab of a dead zebra carcass	M	5134	5323	39	0.35	75	(Han et al. (2006))
Bacillus licheniformis ATCC 14580 (Novozymes)	279010	B	Soil	M	4420	4196	21	0.46	69	(Rey et al. (2004))
Bacillus pumilus SAFR-032	315750	B	Soil	M	3823	3729	21	0.41	117	(Gioia et al. (2007))
Bacillus subtilis 168	224308	B	X-ray irradiated strain	M	4298	4106	30	0.44	129	(Kunst et al. (1997))
Bacillus thuringiensis Al Hakam	412694	B	Severe human tissue necrosis	M	5050	4798	42	0.35	102	(Challacombe et al. (2007))
Bacillus weihenstephanensis KBAB4	315730	B	Soil	P	5983	5831	42	0.35	115	Unpublished
Geobacillus thermodenitrificans NG80-2	420246	B	Water in oil reservoir formation	T	3642	3471	30	0.49	101	(Feng et al. (2007))
Lysinibacillus sphaericus C3-41	444177	B	Mosquito breeding site	M	4887	4771	31	0.37	96	(Hu et al. (2008))

Table 2.5: continued

Name	Taxon ID	Clas.	Isolation	Temp	Genes	CDS	rRNA	GC Perc	Sp. genes	Reference
Alkaliphilus metalliredigens QYMF	293826	C	Borax leachate ponds	M	5016	4801	31	0.37	83	Unpublished
Alkaliphilus oremlandii OhILAs	350688	C	Freshwater USA	M	2951	2836	26	0.36	75	Unpublished
Candidatus *Desulforudis audaxviator* MP104C	477974	C	Fracture water from a borehole	M	2293	2239	6	0.61	63	(Chivian et al. (2008))
Carboxydothermus hydrogenoformans Z-2901	246194	C	Hot swamp	H	2707	2645	12	0.42	61	(Wu et al. (2005))
Clostridium beijerinckii NCIMB 8052	290402	C	Freshwater, Soil	M	5290	5100	43	0.3	62	Unpublished
Clostridium botulinum A2 Kyoto-F	536232	C	Infant botulism	M	3978	3878	20	0.28	98	Unpublished
Clostridium botulinum B Eklund 17B	508765	C	Marine sediments	M	3639	3527	34	0.27	82	Unpublished
Clostridium difficile 630	272563	C	Clinical isolate	M	3983	3777	32	0.29	63	(Sebaihia et al. (2006))
Clostridium kluyveri DSM 555	431943	C	Mud of a canal in Delft	M	4073	3913	20	0.32	77	(Seedorf et al. (2008))
Clostridium perfringens SM101	289380	C	Soil	M	2748	2578	30	0.28	67	(Shimizu et al. (2002))
Heliobacterium modesticaldum Ice1	498761	C	Hot spring microbial mats and volcanic soil	T	3142	3001	30	0.57	82	(Sattley et al. (2008))
Pelotomaculum thermopropionicum SI	370438	C	Thermophilic anaerobic sludge	T	3018	2920	6	0.53	66	Unpublished

Table 2.5: continued

Name	Taxon ID	Clas.	Isolation	Temp	Genes	CDS	rRNA	GC Perc	Sp. genes	Reference
Thermoanaerobacter pseudethanolicus ATCC 33223	340099	C	Thermal springs	T	2363	2291	16	0.35	76	Unpublished
Thermoanaerobacter sp. X514	399726	C	Deep subsurface sample	T	2467	2397	16	0.35	79	Unpublished
Desulfotomaculum reducens MI-1	349161	C	Cr-contaminated marine sediment	M	3324	3324	18	0.42	83	(Junier and Zdobnov (2010))

As described in the methods section (Chapter 2), bi-directional BLAST was done on the sporulation genes of the 27 genomes listed in Table 2.5 in order to find genes that are common among all endospore-forming bacteria and conserved. The search resulted in six orthologous genes that appeared in all analysed genomes and that were highly conserved (Table 2.6).

Table 2.6: Orthologous genes and their function found to be common and conserved among all endospore-forming genomes used in this analysis. Protein lengths indicated in amino acids (aa) for *Bacillus subtilis* obtained from Stragier and Losick 1996.

Name	Gene symbol	Function	Length (aa)
Stage 0 sporulation protein A	*spo0A*	Global transcription regulator for sporulation	267
Spore protease	*gpr*	Degradation of the small acid-soluble spore proteins (SASPs) during germination	368
Stage V sporulation protein T	*spoVT*	Global regulator activated by sigma G	178
Stage IV sporulation protein B	*spoIVB*	Protease that activates processing of the pro-sigma K factor	425
Stage V sporulation protein AD	*spoVAD*	Potential transmembrane protein with unknown function	150
Stage V sporulation protein AC	*spoVAC*	Potential transmembrane protein with unknown function	338

A comparative phylogenetic analysis of 16S rRNA gene sequences and the six conserved genes (*spo0A*, *spoIVB*, *spoVAC*, *spoVAD*, *spoVT* and *gpr*) for the 27 endospore-forming genomes was done (Figure 2.5). The concatenated six genes as well as the 16S rRNA genes for the 27 genomes, were aligned with MAFFT (Katoh et al. 2005) or Muscle (Edgar 2004) using default parameters. Multiple-FastA alignments were then converted to Phylip format with the seqret program from the EMBOSS package (Rice et al. 2000) and phylogenies constructed from the Phylip-formatted alignments with PhyML (Guindon and Gascuel 2003). Default parameters were used, except the following: JTT+ substitution model for proteins and GTR+ model for nucleic acids; 4 classes of substitution rate categories; estimation of the shape parameter, proportion of invariants, and transition/transversion ratios (for nucleotides). Trees were processed (re-rooted, topology extracted, and plotted) with the Newick Utilities (Junier and Zdobnov 2010). Bootstrap values (percentage over 1000 samplings) are shown at the nodes of the trees. The two trees are highly similar, demonstrating that a similar phylogenetic relationship of the strains can be seen when the sporulation genes are used than when the 16S rRNA gene is used to construct the phylogeny. Particularly the division between clades of *Bacillus* species and clades of *Clostridium* species was evident.

Sporulation genes *spo0A* and *gpr*

The gene coding for the sporulation transcription factor *spo0A* was analysed more in detail. A phylogeny of the 27 species based on the *spo0A* (Figure 2.6, top

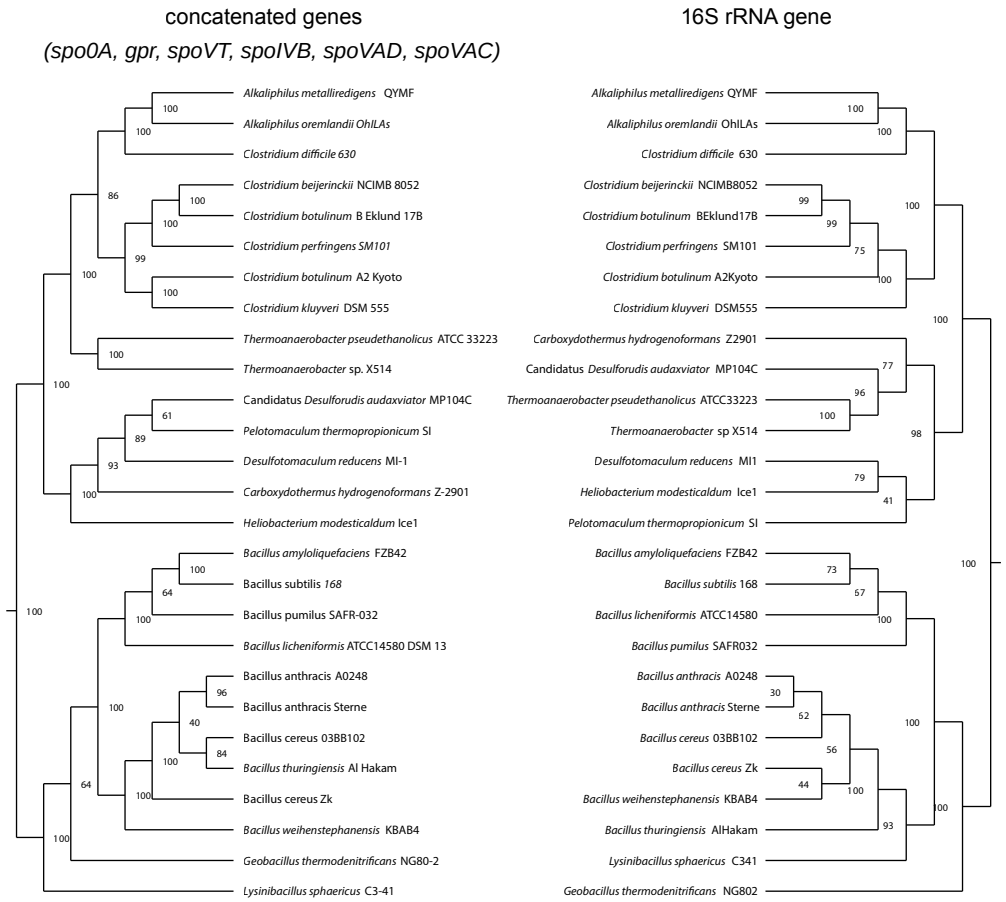


Figure 2.5: Comparative phylogenetic analysis of 16S rRNA gene sequences and the six concatenated sporulation genes (*spo0A*, *spoIVB*, *spoVAC*, *spoVAD*, *spoVT* and *gpr*) for the 27 spore-forming Firmicutes. Bootstrap values (percentage over 1000 samplings) are shown at the nodes of the trees.

panel) and a conservation profile of the Spo0A protein sequence was drawn (Figure 2.6, lower panel). The same figures are also displayed for the germination gene (*gpr*) (Figure 2.7). Conservation plots were made with the plotcon program from EMBOSS. This is a program that computes a weighted average of the similarity scores for all residue pairs in each sliding-window mode. The default window size of 4 amino acid residues was used for the conservation plot. The phylogeny of the 27 endospore-forming strains is similar when built with the *spo0A* gene and with the 16S rRNA gene. In particular the division of clades between *Bacillus*, *Clostridium* and extremophile species (*Desulfotomaculum reducens*, *Pelotomaculum thermopropionicum*, and others) is present in both trees. The conservation profile reveals regions of the gene that are highly conserved and other regions that are more variable. The information on conserved versus variable regions is necessary for the development of a successful molecular marker, where primers are designed in two conserved regions flanking a variable region that provides enough variance to deduce the phylogenetic relationships. Based on the stable phylogeny

of *spo0A* and *gpr* alignments and the presence of conserved and variable regions in the gene, the two genes (*spo0A* and *gpr*) were chosen for the design of molecular primers.

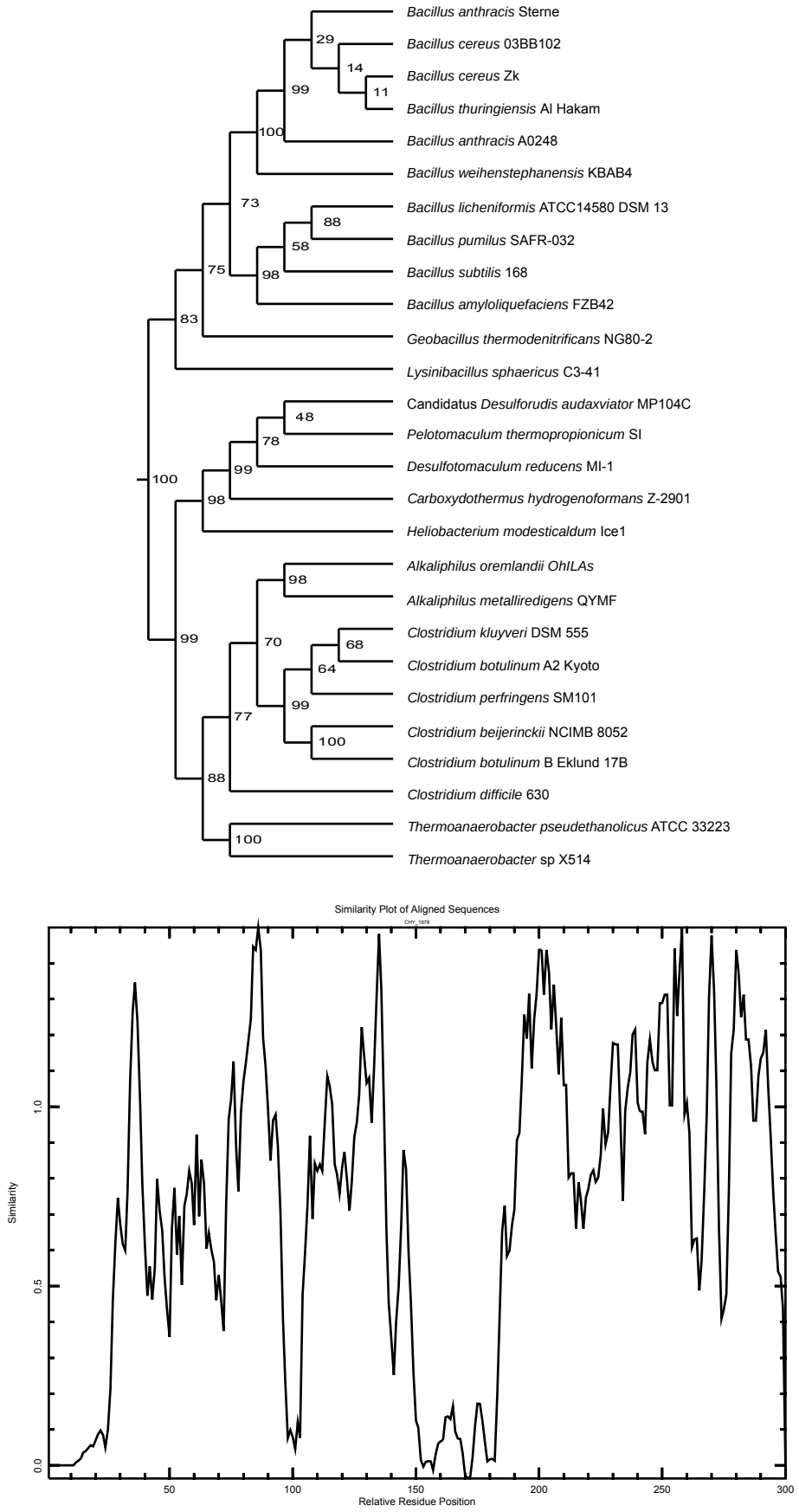


Figure 2.6: Phylogenetic reconstruction (top panel) and conservation profile (lower panel) for sequences of the stage 0 sporulation protein Spo0A.

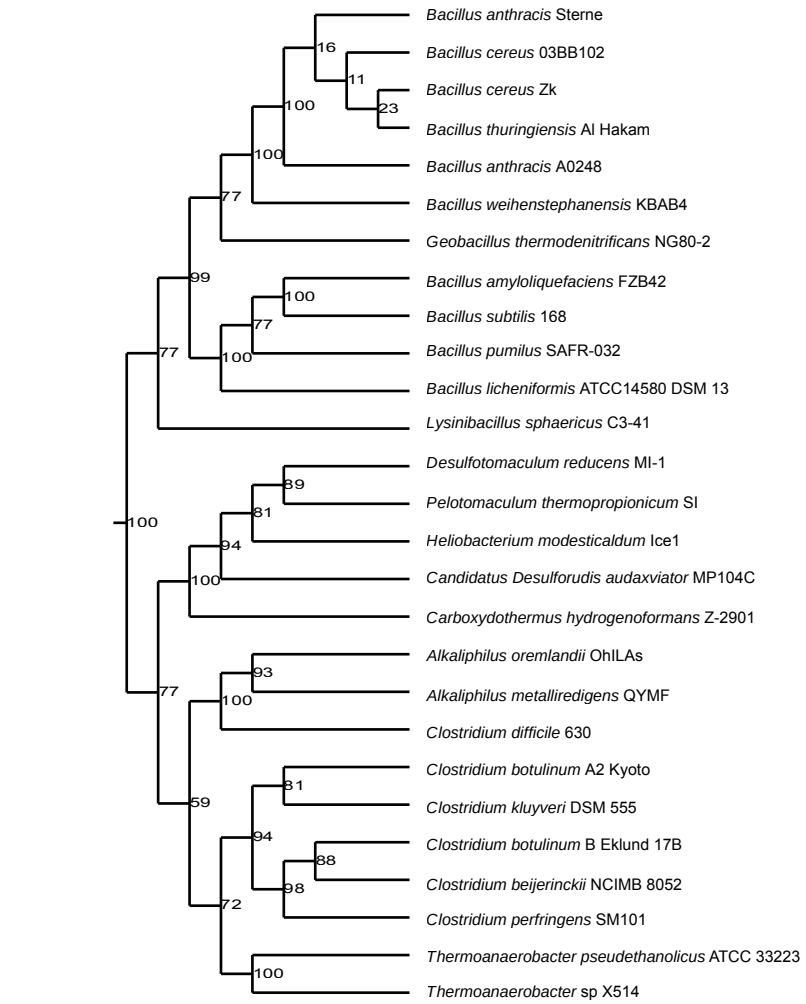


Figure 2.7: Phylogenetic reconstruction (top panel) and conservation profile (lower panel) for the germination protease gene *gpr*.

Primer design

Alignments of the *spo0A* from the 27 strains in Table 2.5 were scanned for highly conserved 20 nucleotide regions. Several potential forward and reverse primers were chosen, synthesized and then tested in different combinations on a selection of culture strains (Table 2.7). As a first screening the primer pairs were tested on representatives of three groups of endospore-forming bacteria, the Bacilli (*Bacillus subtilis*), the Clostridia (*Clostridium pasteurianum*) and representatives of the extremophiles (*Sulfobacillus acidophilus*, *Geobacillus thermoleovorans*, *Alicyclobacillus acidocaldarius*). Also, two non endospore-forming bacteria (*Escherichia coli*, *Lactococcus lactis* subsp. *lactis*) and a sediment sample were tested. The same procedure of alignment, selection of potential primers and amplification tests was done for the *gpr* gene. Primer pairs and results of amplification tests for gpr are shown in Table 2.8.

All strains for the amplification tests were grown in liquid cultures at 30 ℃ in nutrient broth (*Bacillus subtilis*, *E. coli*) (Biolife, Milano, I), in medium DSM 92 (*Lactococcus lactis* subsp. *lactis*) (`www.dsmz.ch`), in medium ST1 (*Clostridium pasteurianum*) and in medium in FeSo and 50 ℃ (*Sulfobacillus acidophilus*), in medium PTYG at pH 5 (*Geobacillus thermoleovorans*) and pH 2.9 (*Alicyclobacillus acidocaldarius*). The medium DSM 92 contained 30 g/L trypticase soy broth (TSB) and 3 g/L yeast extract. The medium FeSo contained 0.25 g/L TSB, 1.25 g/L $(NH_4)_2SO_4$, 0.5 g/L $MgSO_4$ x 7 H_2O, 90 mg/L $FeSO_4$ x 7 H_2O and 1.75 mg/L potassium tetrathionate. The medium PTYG contained 0.25 g/L peptone, 0.25 g/L triptone, 0.5 g/L yeast extract, 0.5 g/L glucose, 0.6 g/L $MgSO_4$ x 7 H_2O, 70 mg/l $CaCl_2$ 2 x H_2O, 10mM $MgCl_2$ and 2.5 mM $CaCl_2$. DNA was extracted from all strains using InnuPrep Bacterial DNA kit (Analytik Jena, Jena, Germany) from 2 ml culture in exponential growth phase, centrifuged to pellet at 10,000 rpm for 5 min. DNA was quantified with NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA).

PCR reactions for initial screening contained 1x PCR buffer (New England Biolabs (NEB), Ipswich, USA), 0.2 mM dNTPs, 0.2 µM of each primer and 1 unit Taq polymerase (NEB) and 1 ng DNA. Program ran at 94 ℃ for 5 min for initial denaturation, then 35 cycles of 94 ℃ for 30 sec, annealing temperature of 52 ℃ for 45 sec and 72 ℃ for 90 sec. If amplification results were negative, reactions were repeated by adding 1 x of bovum serum albumin (BSA; NEB) and 1% formamide and by adding a touchdown program over 10 cycles, diminishing the annealing temperature by 0.5 ℃ per cycle before continuing the remaining 25 cycles at the actual annealing temperature.

The primer pairs 3 and 7 of the *spo0A* gene were the most efficient primer pairs with successful amplification in all three groups of endospore-forming bacteria and no amplification of the negative controls. Primer pair 3 also amplified fragments from the sediment sample. Furthermore, this primer pair amplifies a gene fragment of 602 nucleotides which is sufficient in length for inference of phylogeny. This primer pair was finally chosen as the primer for the gene *spo0A*. The primer sequences and regional alignment is shown in Figure 2.1 in the main results of this chapter.

Table 2.7: Combination of potential forward and reverse primers for *spo0A* gene and amplification test conducted with representative strains from three target groups (TG) groups of endospore-forming bacteria. B: Bacilli (*Bacillus subtilis*) C: Clostridia (*Clostridium pasteurianum*) E: extremophiles (*Sulfobacillus acidophilus, Geobacillus thermoleovorans, Alicyclobacillus acidocaldarius*) N (non endospore-forming bacteria, *Escherichia coli, Lactococcus lactis* subsp. *lactis*), S (sediment sample). pos: positive amplification, neg: no amplification, NA: not analysed.

	forward	reverse	forward sequence	reverse sequence	length (nt)	TG	Dimer	B	C	E	N	S
1	sp0A166Af	sp0A493Ar	GATATHATYATGCCDCATYT	CCTTTAATATGAGCRGGBAC	347	B		NA	NA	NA	pos	NA
2	sp0A166Af	sp0A652Ar		TCCATGCNACTTCAATNGC	505	B		NA	NA	NA	NA	NA
3	sp0A166Af	sp0A748Ar		GCNACCATHGCRATRAAYTC	602	B		pos	pos	pos	NA	pos
4	sp0A166Af	sp0A589r		ATTTYTTBGCRATDTCNGGATA	445	B		pos	neg	pos	NA	NA
5	sp0A259Af	sp0A493Ar	GCVTTYGGNCARGAAGATGT	CCTTTAATATGAGCRGGBAC	254	B		NA	NA	NA	pos	NA
6	sp0A259Af	sp0A589r			352	B		pos	neg	pos	pos	NA
7	sp0A259Af	sp0A748Ar			509	B		pos	pos	pos	NA	NA
8	sp0A493Af	sp0A652Ar	GTVCCYGCTCATATTAAAGG		178	B		pos	pos	neg	NA	NA
9	sp0A493Af	sp0A748Ar			275	B		pos	neg	pos	NA	NA
10	sp0A28Bf	sp0A478Br	GCWGATGATAAYAARGADTT	GCTGGNACDCCWAYYTCATG	470	C		neg	pos	neg	NA	pos
11	sp0A28Bf	sp0A625Br		CTTATNGCTCTTTCHACTCT	617	C		NA	NA	NA	NA	NA
12	sp0A28Bf	sp0A750Br		TTATCWGCWAYMATHGCWAT	742	C		neg	neg	neg	NA	NA
13	sp0A169Bf	sp0A478Br	GATATWATAATGCCHCATY		329	C	yes	pos	pos	neg	NA	NA
14	sp0A169Bf	sp0A625Br			476	C	yes	NA	NA	NA	NA	NA
15	sp0A169Bf	sp0A750Br			601	C	yes	NA	NA	NA	NA	NA
16	sp0A478Bf	sp0A625Br	CATGARRTWGGHGTNCCAGC		167	C		NA	NA	NA	pos	NA
17	sp0A478Bf	sp0A750Br			292	C		neg	neg	neg	NA	NA
18	sp0A34Cf	sp0A451Cr	GATGAYAACMRVGADTTYTG	GCBGGHACGCCCATYTSARTG	437	E		NA	NA	NA	pos	NA
19	sp0A34Cf	sp0A619Cr		TCCCAGGCCAGYTCRATGGCATG	608	E		NA	NA	NA	NA	NA
20	sp0A34Cf	sp0A721Cr		TCRGCNAYYATRGCDATRAA	707	E		neg	neg	pos	NA	pos
21	sp0A172Cf	sp0A451Cr	GAYATYATYATGCCVCAYHTSG		299	E		NA	NA	NA	pos	NA
22	sp0A172Cf	sp0A619Cr			467	E		neg	neg	neg	NA	NA
23	sp0A172Cf	sp0A721Cr			569	E		NA	NA	NA	NA	NA
24	sp0A451Cf	sp0A619Cr	CAYTSARATGGGCGTDCCVGC		191	E		neg	neg	neg	NA	NA
25	sp0A451Cf	sp0A721Cr			290	E		NA	NA	NA	pos	NA

Table 2.8: Combination of potential forward and reverse primers for *gpr* gene and amplification test conducted with representative strains from three target groups (TG) of endospore-forming bacteria. B: Bacilli (*Bacillus subtilis*) C: Clostridia (*Clostridium pasteurianum*) E: extremophiles (*Sulfobacillus acidophilus*, *Geobacillus thermoleovorans*, *Alicyclobacillus acidocaldarius*) N (non endospore-forming bacteria, *Escherichia coli*, *Lactococcus lactis* subsp. *lactis*). pos: positive amplification, neg: no amplification, NA: not analysed.

	Forward	Reverse	Forward sequence	Reverse sequence	lenghtTG (nt)		DimerB	C	E	N
1	gpr361Af	gpr499Ar	AAYGTVACSCCNGAYGCGCT	TCGCTYGTTTCAATNCCKGT	158	B		neg	pos	NA
2	gpr431Af	gpr655Ar	TGCAGCCKGAAARYGTRSA	ARTTCYTTWCKYTTATTSCC	244	B		neg	neg	NA
3	gpr499Af	gpr1010Ar	ACMGGNATTGAAACRAGCGA	GCCATATCNTCDATRAANG	530	B		NA	NA	pos
4	gpr361Af	gpr1010Ar			668	B		NA	NA	pos
5	gpr300Bf	gpr530Br	GTRGTWGGDCTTGGNAAYTGG	CTNGAAGCHARDGCATCWA	249	C		NA	NA	pos
6	gpr300Bf	gpr682Br		GTWGCTGCATCWACNACNGT	402	C	x	NA	NA	pos
7	gpr150Cf	gpr440Cr	GGNAACTATRTNACHATTGAAGC	ATTTCDGCCGTTTCWATKCCKG	312	E		NA	NA	pos
8	gpr290Cf	gpr440Cr	TDGGHAACTGGAAYGCHAC		172	E		neg	neg	NA
9	gpr440Cf	gpr800Cr	CMGGMATWGAAACGGCHGAAA	TCDATTTCYTTNGGNGTNA	379	E		NA	NA	pos
10	gpr150Cf	gpr800Cr	GGNAACTATRTNACHATTGAAGC		669	E		neg	neg	NA
11	gpr290Cf	gpr800Cr	TDGGHAACTGGAAYGCHAC		529	E		NA	NA	pos

The Table 2.1 of the main results section in this chapter, reports, that for some strains such as *Sulfobacillus acidophilus* and *Alicyclobacillus acidocaldarius* the *spo0A* primers did not amplify any gene fragments. To find an explanation for this problem, the full *spo0A* sequences of *S. acidophilus* and *A. acidocaldarius* were aligned against the *spo0A* sequence of *B. subtilis*. The alignment was done with ClustalW (`http://www.ebi.ac.uk/Tools/msa/clustalw2`. The regions corresponding to the annealing sites of the forward and the reverse primer are shown in Figure 2.8. The alignment for *S. acidophilus* has one mismatch in the forward and one mismatch in the reverse primer sequence, which does not explain the lack of amplification. The alignment for *A. acidocaldarius* however has 5 mismatches for the forward and 1 mismatch for the reverse primer sequences. Five mismatches are likely too many positions in order for the primer to anneal properly to the sequence, to the ineffective amplification of the spo0A primers for *A. acidocaldarius* can be explained.

Figure 2.8: Alignment of *spo0A* gene of *Sulfobacillus acidophilus* and *Alicyclobacillus acidocaldarius* Tc41 against *spo0A* of *Bacillus subtilis* 168. The two regions shown correspond to the forward primer spo0A166f (left) and the reverse primer spo0A748r (right) described in this study. Stars indicate 100% identity. The exclamation points highlight mismatches with the primer sequence.

Optimization of *spo0A* primer

The PCR reaction with the *spo0A* primers was then optimized with sediment samples from Lake Geneva and Lake Baikal, described in the main methods section of this chapter. Precisely a gradient PCR was done on the two samples to define the best annealing temperature (Figure 2.9). The gradient PCR was conducted with 1 x PRC buffer (TaKaRa), 0.2 mM dNTPs, 1 µM of each primer, 1 unit Taq Polymerase, 1 x BSA. One ng of DNA from the sediment samples was used as template. The program was run at 94℃ for 5 min for initial denaturation, 40 cycles of 94℃ for 30 sec, 30 sec at specific annealing temperature (see image headers in Figure 2.9) and 72℃ for 1 min, followed by a final extension time of 5 min at 72℃. The annealing temperature 53.3℃ (white squares in Figure 2.9) was chosen as the optimal, because it results in high amplification efficiency, seen as an intense band at the correct size of 602 bp (same position as the band in the positive control), without too many unspecific bands in both samples.

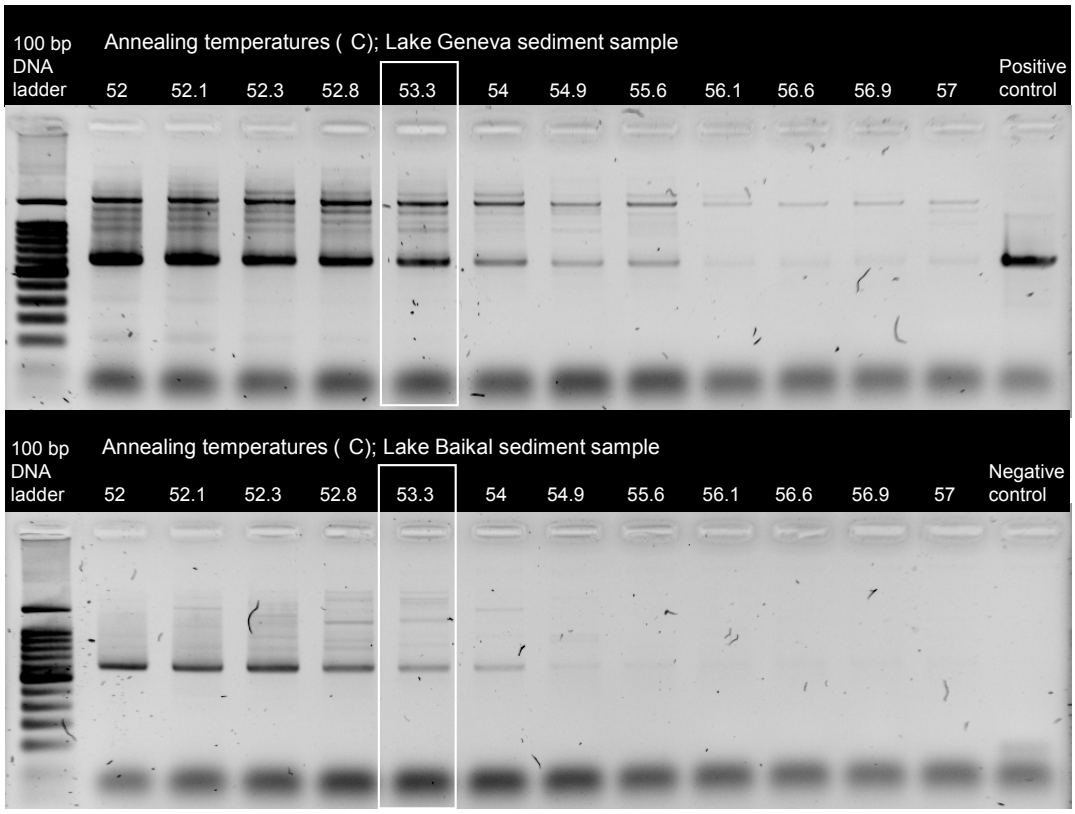


Figure 2.9: Electrophoresis gel image of gradient PCR to find the best annealing temperature for *spo0A* primer (spo0A166f-748r). The positive control is a DNA extract from *Paenibacillus alvei*.

Tests of DNA extractions

A range of tests have been performed on culture strains, endospore preparations and sediment samples in order to arrive at the optimized DNA extraction method described in the methods section of this chapter. Generally, the DNA extraction tests were optimized for efficient lysis of spores and for high DNA yield. All DNA extraction methods applied were based on commercially available kits, so that repeatability and standardization was ensured. The first tests were done with the MoBio Powersoil DNA extraction kit. It was a trial on cultures of *P. alvei*, *B. tusciae*, *G. thermoparaffinovorans* and *S. acidophilus* strains, all endospore-forming bacteria in order to check if DNA yield would increase or decrease when a repeated protocol was used. Results are displayed in Figure 2.10. All strains were grown in liquid medium to exponential phase and 2 ml were centrifuged at 6,000 x g for 10 min to pellet the cells. DNA was extracted from cell pellets with MoBio Powersoil DNA extraction kit (MoBio, Carlsbad, CA, USA) according to manufacturer's instructions except the following modification. After the bead-beating and centrifugation step, the supernatant (containing free DNA) was collected in a 2 ml collection tube. To the residual cell pellet, PowerBead solution was again added and the bead-beating step was repeated. After the consequent centrifugation, the

supernatant (containing 2nd fraction of free DNA) was collected in a separate 2 ml collection tube. PowerBead solution was again added to the residual pellet and the bead-beating and centrifugation was done for a third time. The supernatant containing the 3rd fraction of free DNA was then again collected in a separate 2 ml collection tube. This protocol resulted in 3 supernatants which were then handled as three separate samples for the remaining downstream protocol. The concentration of the three repeated extractions was quantified using NanoDrop 2000 (ThermoFisher Scientific).

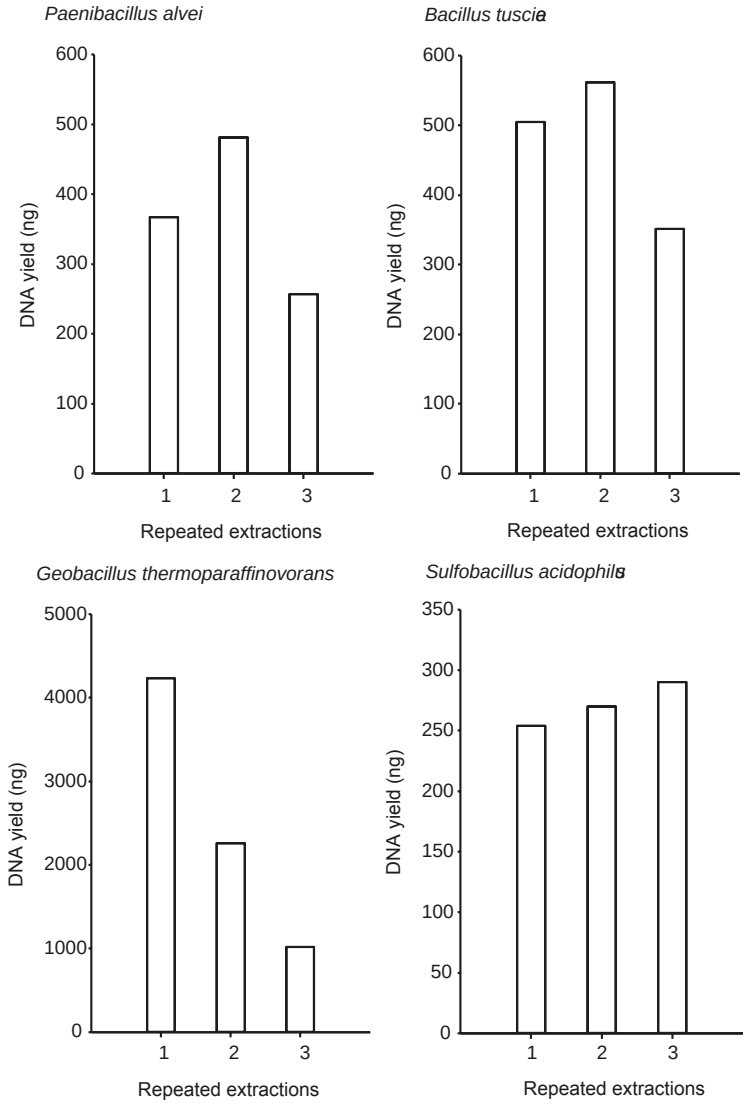


Figure 2.10: DNA yield of three repeated extractions done on endospore-forming strains using MoBio Power Soil DNA extraction kit.

The subsequent experiment was done to test different lysis instruments during the bead-beating step. The protocols were based on the MoBio Power Soil DNA extraction kit and conducted on endospore preparations of *P. alvei* and *L.*

sphaericus (Figure 2.11). The endospore preparations were composed of 99% endospores, verified by phase-contrast microscopy. The DNA extraction was done following the standard manufacturer's protocol but with either the FastPrep FP120 (Savant, Farmingdale, NY, USA) bead-beating machine at 6.5 m/sec for 30 sec (black bars in figure) or with the TissueLyzer LT (QIAGEN) at 3,000 rpm for 10 min (white bars in figure). The DNA yield was quantified using NanoDrop 2000 (ThermoFisher Scientific). The results show that even though the bead-beating with the TissueLyzer is less rigorous (albeit longer in time) the DNA yield obtained with the TissueLyzer is higher for both endospore strains. A less harsher bead-beating, but extended in time (10 min) seems to yield more DNA.

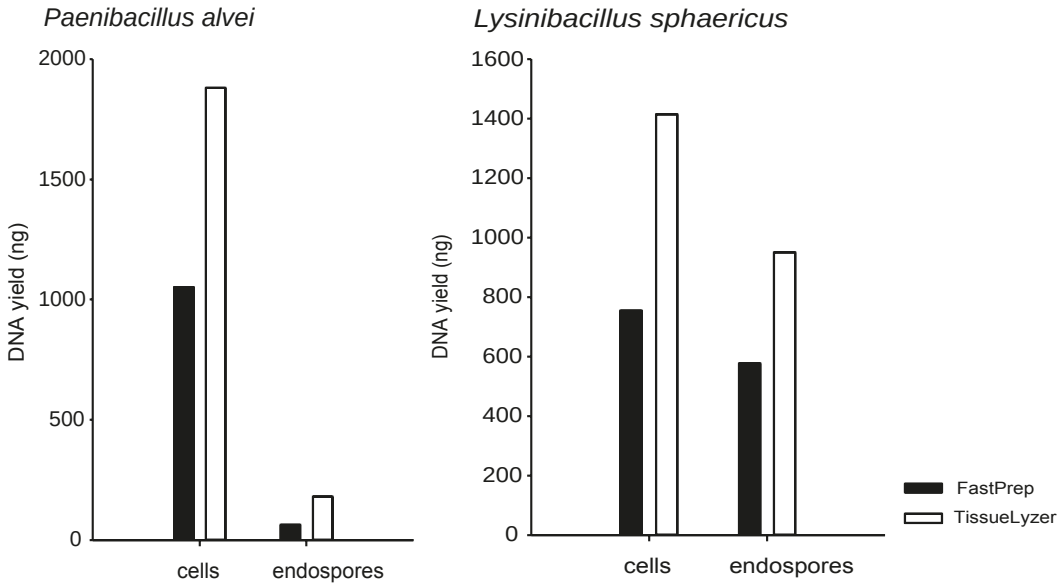


Figure 2.11: Comparison of DNA yields with MoBio PowerSoil extraction kit and two different bead-beating instruments.

Three different DNA extraction kits were then compared. The tests were done on endospore preparations of *B. subtilis*. For all three kits, the protocol was modified as to include four repetitive bead-beating steps. Results are shown as a gel-electrophoresis image in Figure 2.12. The band intensity is positively correlated with DNA yeld. In addition, in the gel-electrophoresis image, the extent of degradation of the genomic DNA can be verified.

The endospore preparations were composed of >90% endospores, verified by phase-contrast microscopy. The three kits used were a) PowerSoil DNA kit (MoBio), b) FastDNA kit for soil (MP Biomedicals) and c) InnuPrep Bacteria DNA kit (Analytik Jena). All extractions were done with 4 sequential bead-beating rounds as described above. The standard protocol of the InnuPrep Bacteria DNA kit does not contain a bead-beating step (standard cell lysis method is lysozyme). Here, bead-beating was added by using the lysis matrix provided by the MP FastDNA Spin kit for soil. All bead-beating steps were done with the TissueLyzer LT (Qiagen) at 3,000 rpm for 10 min. Two µl of DNA extract were loaded onto 1%

agarose gel with 2 µl of loading dye (0.25% bromphenol blue and 40% sucrose). Electrophoresis was done at 80V for 40 min. Gel was stained in 3 x GelRed bath (Biotium, Hayward, USA). DNA ladder of 1 kilobase was added (Promega, Madison, USA). The residual DNA extracts were afterwards pooled (for each kit individually) and precipitated at -20 ℃ in absolute ethanol and Na-acetate (0.3 M) and quantified with Qubit Fluorometer and Qubit ds DNA kit (Invitrogen, Carlsbad, USA). The DNA yield from the PowerSoil kit was 30.2 ng, the yield from FastDNA Spin kit for soil was 1,300 ng and the yield from InnuPrep kit was below the detection limit (<0.5 ng).

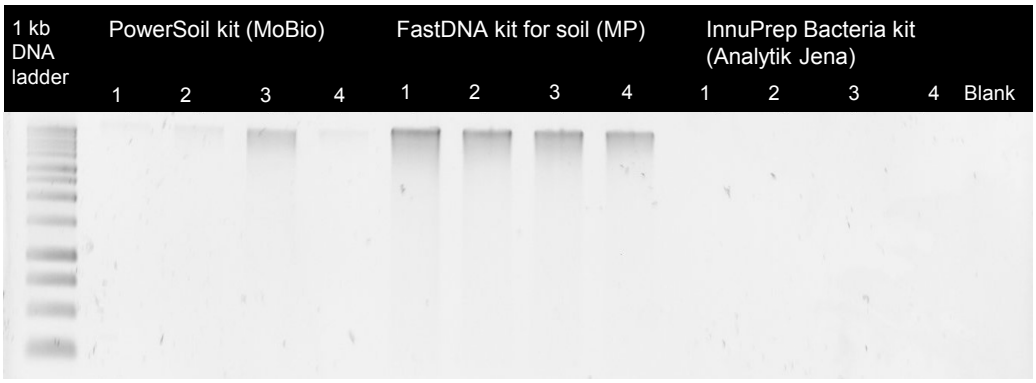


Figure 2.12: Comparison of different DNA extraction kits on an endospore preparations of *B. subtilis*. The numbers (1, 2, 3 and 4) correspond to the respective extraction cycle. The extraction kits are mentioned on top of the lanes.

Quantification of cells and endospores

To compare the results of the DNA extraction tests described in the main results of this chapter (Figure 2.3), the DNA yields had to be normalized to 10^8 cells and 10^9 endospores, respectively. For this, the cell and endospore preparations were previously quantified using microscopic counting chamber (with Neubauer ruling). The quantification is shown in Figure 2.13. This data is important, since it shows the repeatability of the counting method by depicting the standard deviation of quantification in triplicates.

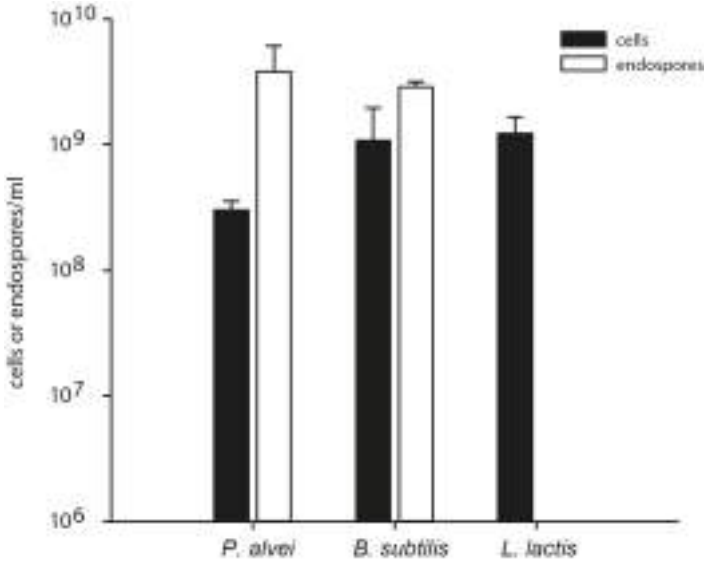


Figure 2.13: Quantification of cells and endospore preparations prior to DNA extraction tests. Error bars show standard deviations of triplicates.

Sediment samples

The sediment samples described in this study have been retrieved from Lake Geneva and Lake Baikal. The exact locations of sampling are shown in Figure 2.14 and two images of the sediment are shown in Figure 2.15.

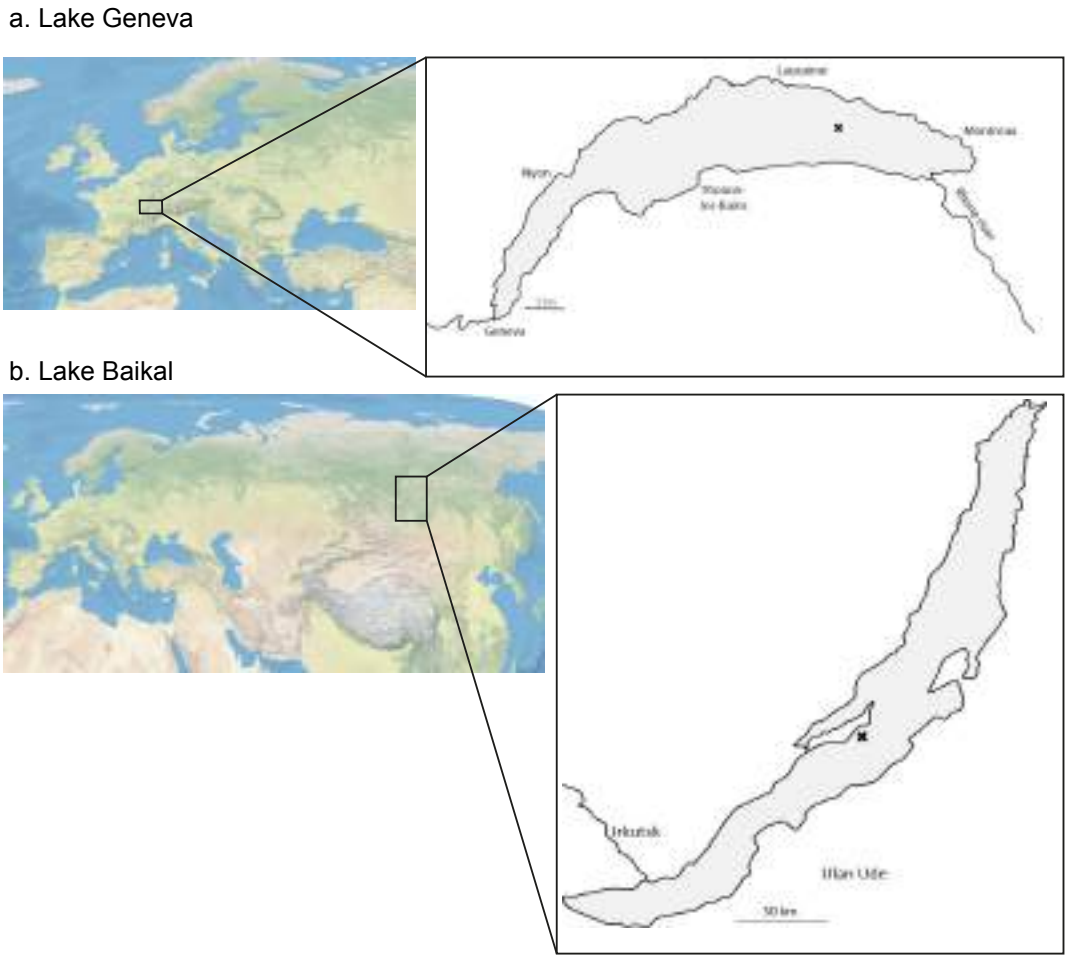


Figure 2.14: Map indicating the position of the sampling sites in Lake Geneva and Lake Baikal.

a. Lake Geneva

b. Lake Baikal

Figure 2.15: Pictures of sediments from Lake Geneva and Lake Baikal.

DNA extraction of sediments

In order to improve DNA extraction from sediment samples, an indirect extraction protocol was considered. Such an indirect approach, consists of the separation of sediment biomass from the mineral particles prior to DNA extraction. Different methods to separate biomass from particles were tested (Table 2.9). These tests are based on autoclaved sediment samples that were spiked with endospores from *P. alvei*. All samples were first homogenized in Na-hexametaphosphate using the Ultra-Turrax homogenizer (IKA). The slurries were then subjected to different separation methods listed in Table 2.9. The efficiency of separation was either verified by microscopic observation of the supernatant and pellets, or by enumeration of spores using the method of colony forming units (CFU). Some samples were also subjected to DNA extraction using MP FastDNA Spin kit for soil (MP Biomedicals) and DNA yield quantification using NanoDrop (Thermo Fisher Scientific). The best method to separate biomass from sediment particles was low-speed centrifugation at 20 x g for 1 min, which removed enough sediment particles and at the same time leaving the majority of the biomass in the supernatant as can be seen by the highest number of CFUs detected in the supernatant (20,000,000 CFU/ml). The efficiency of the density gradient centrifugation of nicodenz or sucrose was verified by microscopic observation of the individual density layers. In both approaches, cells or endospores were always detected in the sediment pellet in large numbers. The density centrifugation methods were therefore not efficient to separate biomass from sediment particles. The method of centrifugation at 20 x g for 1 min was finally selected as the best separation of biomass and sediment particles.

Table 2.9: Test of methods to separate biomass from sediment particles in a homogenized slurry of sediment and Na-hexametaphosphate. The efficiency of the separation was verified by quantification of colony forming units (CFU/ml), DNA yield (ng/g wet weight) or microscopic observation. NA: not analysed.

Separation method	Fraction	CFU/ml	DNA	Microscopy
Centrifugation 1 x g for 5 min	pellet	3,000,000	21.2	NA
Centrifugation 10 x g for 1 min	pellet	710,000	<1	NA
Centrifugation 20 x g for 1 min	pellet	0	<1	NA
	supernatant	20,000,000	2000	NA
Nicodenz 1% density gradient centrifugation	cell layer	NA	NA	high density biomass
	pellet	NA	NA	sediment and cells/spores
Sucrose 1% density gradient centrifugation	cell layer	NA	NA	medium density biomass
	pellet	NA	NA	sediment and biomass

Two different DNA extraction kits (MP FastDNA Spin kit for soil and MoBio

Power Soil DNA kit) were tested as a direct and indirect protocol on four different samples (three sediments and a microbial mat). DNA yield (measured with NanoDrop) and purity is given as absorbance ratio at 260/230 nm was measured on all extracts (Table 2.10). The direct method was applied to 0.5 g sediment, directly added to lthe ysis tube. The indirect methods consisted of separation of biomass with above described method using homogenization with Na-hexametaphosphate and centrifugation. The supernatant containing the biomass was then concentrated by filtration through a nitrocellulose filter of 0.2 µm poresize. DNA from the filter was then extracted with MP Fast DNA Spin kit for soil or MoBio Power Soil DNA kit. Both methods consisted of three repeated bead-beating steps. As can be seen in Table 2.10, the MoBio PowerSoil DNA kit yielded in lower DNA quantities for Lake Baikal and Lake Geneva samples. For Lake Loclat and the Lirima (microbial mat) sample, the yields were higher but the purity in these extracts was lower. In the two extracts with MP FastDNA Spin kit for soil, the direct extraction yielded higher DNA quantity and better purity.

Table 2.10: DNA extraction with three different methods on sediment samples from Lake Geneva (1-6 cm depth), Lake Baikal (2-7 cm), Lake Loclat (10-15 cm) and Lirima 1 (microbial mat).

Extraction		Lake Baikal		Lake Loclat		Lake Geneva		Lirima 1	
		ug	260/230	ug	260/230	ug	260/230	ug	260/230
MP FastDNA kit	direct (0.5 g)	3.9	1.2	1	1.1	8.9	2	1.9	1.8
MP FastDNA kit	indirect (3 g)	1.5	0.9	0.5	0.9	1.5	1.4	2.9	1.9
MoBio PowerSoil kit	indirect (3 g)	0.6	1.1	1.7	0.8	1.1	1.7	1.8	2.7

The DNA of the three repeated extractions with MP Fast DNA Spin kit for soil was then loaded onto agarose gel. Gel-electrophoresis image is shown for the four sediment samples (Figure 2.16). Two µl of the extracts were loaded on 1% agarose gel and run at 80 V for 40 min. The image shows that in the third extraction step there is still substantial amounts of DNA released from the sample. Also, except for Lake Geneva 1 and Lirima 1 and 2, the DNA has not been sheared during the bead-beating step (shearing can be observed by a smearing of the band).

Purification tests

As noted above with the ratio between absorbance at 230 and 260 nm (Table 2.10), the purity of some of the extracts was not very good. Contaminants such as humic acids or salts can impact the downstream analysis of PCR and sequencing. It was therefore necessary to test different purification possibilities. One such purification method is the drop dialysis, where the DNA extract is loaded as a drop onto a filter membrane of 0.25 µporesize (Merck Millipore, Billerica, USA). This membrane is placed floating onto the surface of sterile water. Once the sample is loaded, the drop gets purified via osmosis as contaminants move through the membrane into the water below. The DNA however is too large in size and cannot traverse the

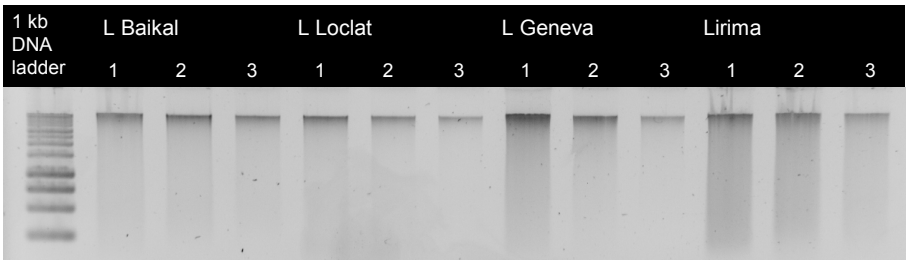


Figure 2.16: Gel-electrophoresis image of the indirect and 3 sequential DNA extraction test with MP FastDNA Spin kit for soil. The numbers represent the individual sequential extracts.

membrane. Two DNA extracts from Lake Geneva and Lake Baikal sediments (extracted with indirect method and MP Fast DNA Spin kit for soil) were purified with drop dialysis and DNA concentration was measured before and after dialysis using the NanoDrop spectrophotometer (Table 2.11). The DNA concentrations have decreased, since the drop has increased in volume, purify however has not improved significantly.

A second method of purification was tested based on gel purification. The DNA extract was loaded onto a 1% agarose gel and run by gel-electrophoresis. The genomic DNA band was then excised and purified with Wizard SV gel and PCR clean-up system (Promega, Madison, USA). DNA concentrations were then measured with NanoDrop spectrophotometer (Table 2.11).

Table 2.11: Concentrations of DNA in extracts before, after drop dialysis and after gel purification. NA: not analysed.

Sample	original (ng/µl)	dialysed (ng/µl)	gel purified (ng/µl)
Lake Geneva	72.2	39.5	NA
Lake Baikal	40	27.3	3.1

A PCR test with samples from Lake Geneva and Lake Baikal sediment was conducted to assess the PCR inhibition due to impurity of DNA extracts (Figure 2.17). PCR amplification of 16S rRNA gene using primers gm3f and gm4r was tested on the original (orig.) extract and the dialysed (dial.) DNA extracts and on both samples spiked with 1 ng of *P. alvei* DNA (+orig. and +dial.). The positive control (+) was 1 ng *P. alvei* DNA. PCR reactions contained 1.25 units of Taq DNA Polymerase (Roche, Mannheim, D), 1 x reaction buffer, 0.2 mM dNTPs, 0.2 µM primers, 1 x BSA and 2 ng sample template (and 1 ng *P. alvei* template for +). The program was a touchdown program for 10 cycles at 94 ℃ for 30 sec denaturation, annealing at 56 ℃ (-0.5 ℃ cycle) for 45 sec and elongation at 72 ℃ for 1 min, followed by 25 cycles of denaturation, annealing at 51 ℃ for 45 sec and elongation at 72 ℃ for 1 min. The amplification of the spiked templates was severely inhibited in all samples. There is no reduction of inhibition in the dialysed samples, rendering the dialysis step useless.

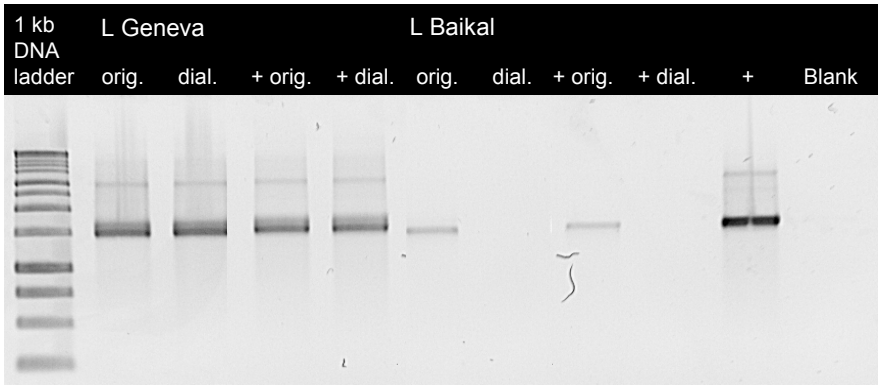


Figure 2.17: PCR inhibition test with original and drop dialysis purified DNA extracts from Lake Geneva and Lake Baikal sediment. The original (orig.) and dialysed (dial.) samples were added as well as a spiked (with *Paenibacillus alvei* original (+ orig.) and a spiked dialysed (+ dial.) sample. The + is the positive control.

The original and purified extracts of Lake Baikal were also quantified with quantitative PCR (qPCR) of 16S rRNA gene adding 1 ng DNA template, as well as a 10 x and a 50 x dilution of the extracts (Figure 2.18). The quantification with qPCR reflects the amplification bias due to contaminants in the extracts. The 16S rRNA gene copy numbers were normalized to the total DNA yield of each sample. The results show that the reduction of DNA during the purification step (due to dilution or loss of DNA in the purification membranes) is substantially higher than the reduction in quantification because of qPCR inhibition of contaminants.

Proof of DNA extraction method

Based on all above mentioned tests, the decision was made to use a DNA extraction protocol based on the MP Fast DNA Spin kit for soil and to not apply any downstream DNA purification procedures. As described in the main results of this chapter, three extraction methods were compared. Protocol 1 (standard manufacturer's method), protocol 2 (three sequential extraction steps) and protocol 3 (indirect and sequential extraction steps). The methods were tested on sediment samples from Lake Geneva and Lake Baikal and the results are shown in Table 2.2 of the main results of this chapter. Here, the electrophoresis gel image of the PCR amplification of 16S rRNA gene and the spo0A gene are displayed to show amplification efficiency and ensure the correct size of the fragments (Figure 2.19). PCR for the 16S rRNA gene was done in a reaction volume of 50 µl as described above but with 0.5 ng DNA template. The PCR for *spo0A* gene amplification was done in reaction volumes of 50 µl, containing water, 1 unit TaKaRa Ex Taq HS polymerase in 1 x buffer (TaKaRa), with 0.2 mM dNTPs, 1 µM of primer spoA166f and spoA748r, 1x BSA and 0.5 ng DNA template. The program consisted of 10 cycles of touchdown with 94 ℃ desaturation for 30 sec, 55 ℃ annealing for 30 sec (-0.3 ℃ /cycle) and 72 ℃ elongation for 1 min, followed by 30 regular cycles at 52 ℃ annealing temperature. For the gel image, 2 µl of PCR product were loaded

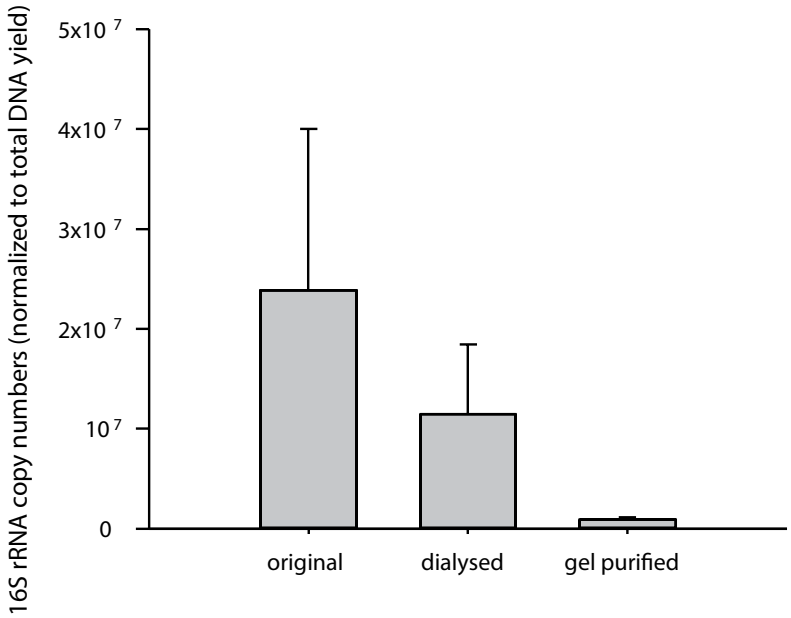


Figure 2.18: Amplification test by qPCR of originial (unpurified), dialysed and gel purified DNA extracts of Lake Baikal. Gene abundance is normalized to the total DNA yield. Error bars are standard deviation of three measurement (1 ng DNA, 10 x and 50 x dilution).

onto 1% agarose, together with 2 µloading dye and then run for 40 minutes at 80 V. The gel was stained in GelRed bath for 30 min, before image was taken.

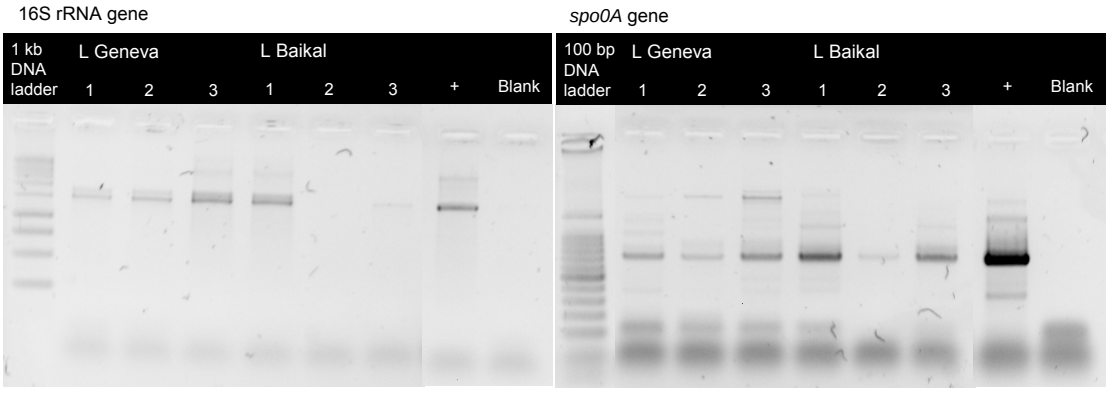


Figure 2.19: Gel-electrophoresis image of amplified sediment samples (16SrRNA gene on the left and *spo0A* gene on the right). The DNA was extracted with three different extraction methods (protocols 1, 2 and 3.)

Identification of *spo0A* sequences

Sediment samples from Lake Geneva and Lake Baikal were amplified with primers spoA166f and spoA748r as described in the main methods section of this chapter. The amplified samples were then purified on a MultiScreen Millipore 96-well plate (Merck Millipore, Billerica, MA, USA) with a vacuum pump and then re-eluted in 40 µl PCR-grade water. The *spo0A* gene amplicons were sent to Eurofins MWG Operon (Ebersberg, D) for barcode amplicon sequencing with Roche GS FLX+ (Roche Diagnostics, Fishers, IN, USA). From the sequences clustered into OTUs, one representative sequence of each OTU was used to build the phylogeny for Lake Geneva (Figure 2.20) and Lake Baikal sediments (Figure 2.21). Phylogenies were constructed from Phylip-formatted alignments with PhyML (Guindon, 2003), using default parameters. The trees were re-rooted, condensed according to DNA extraction protocol, and displayed with the Newick utilities (Junier and Zdobnov 2010). Each branch represents a cluster of OTUs of >97% sequence similarity. The branches are coloured according to extraction protocol with sequences extracted from protocol 1 (blue), protocol 2 (yellow) and protocol 3 (red) coloured branches. The trees display the substantially better sequencing result when DNA is extracted with protocol three. The most ineffective DNA extraction protocol is protocol one (yellow) branches, which has significantly fewer sequences. More important than the lower number of OTUs in extracts from protocol 2 is there unequal distribution. Using this protocol in comparably few clusters would be detected.

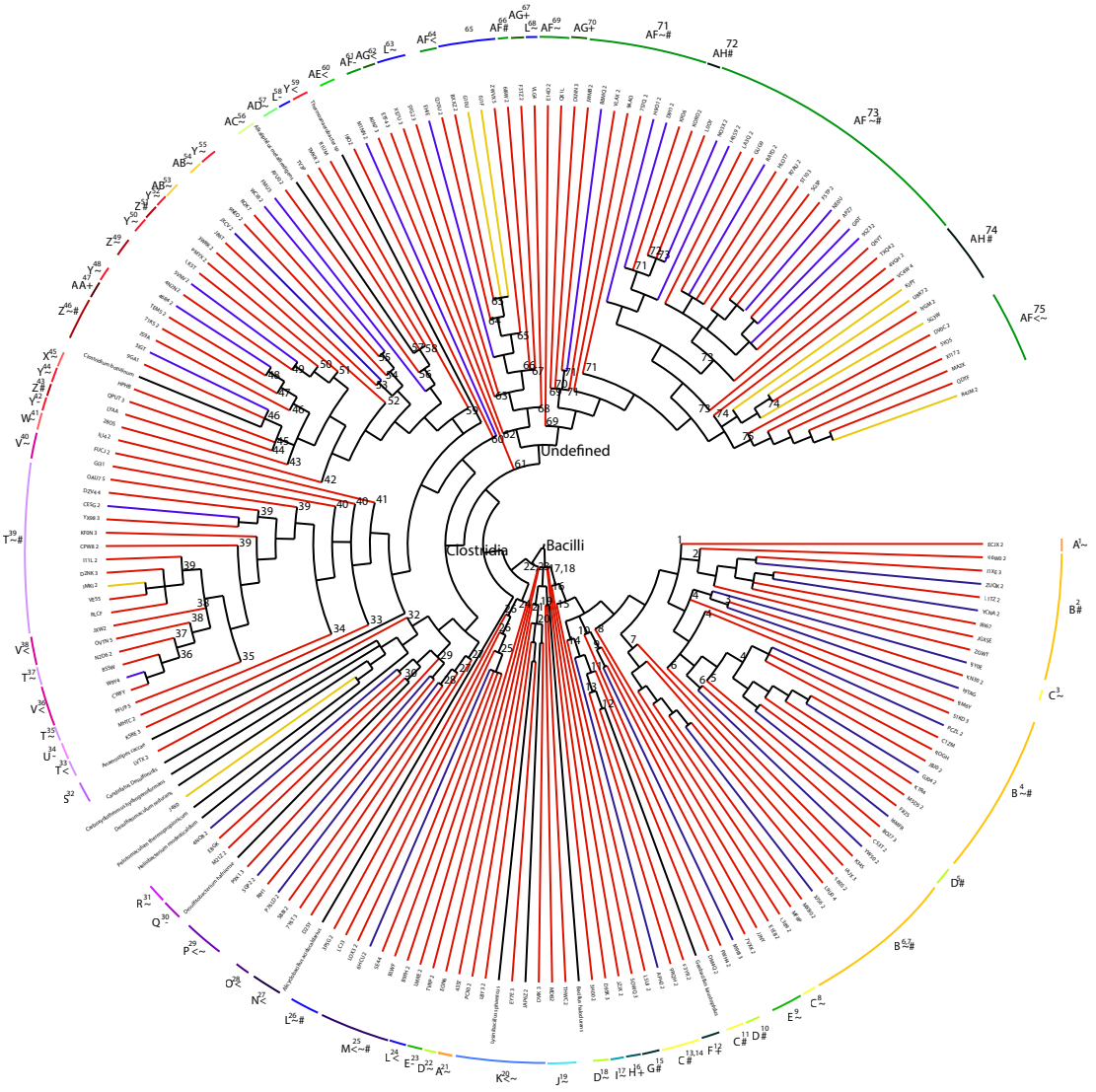


Figure 2.20: Cladogram of *spo0A* sequences from sediment of Lake Geneva extracted with protocols 1 (blue), 2 (yellow), and 3 (red). Classes of closest relative are shown in color with indication of the identity ranges (<65% identity (-), 65-74% (<), 75-84% (), 85-94%(#), >95% (+)). A *Bacillus amyloliquefaciens*, B *B. methanolicus*, C *Geobacillus* sp. (strain WCH70), D *B. cereus* subsp. *cytotoxis* (strain NVH 391-98), E *B. thuringiensis* , F *Geobacillus thermodenitrificans* (strain NG80-2), G *B. atrophaeus* (strain 1942), H *B. subtilis*, I *B. mycoides*, J *B. pseudofirmus* (strain OF4), K *Lysinibacillus sphaericus* (strain C3-41), L *Brevibacillus laterosporus*, M *Brevibacillus brevis* (strain 47), N *Thermincola potens* (strain JR), O *Desulfotomaculum acetoxidans* (strain ATCC 49208), P *Desulfosporosinus orientis* (strain ATCC 19365), Q *Thermosediminbacter oceani* (strain ATCC BAA-1034), R *Syntrophobotulus glycolicus* (strain DSM 8271), S *Heliobacterium medesticaldum* (strain ATCC 51547), T *Clostridium clariflavum* (strain DSM 19732), U *B. cereus* , V *C. thermocellum*, W *C. cellulovorans* (strain ATCC 35296), X *C. cellulolyticum* (strain ATCC 35319), Y *C. botulinum* , Z *C. lijungdahlii* (strain ATCC 55383), AA *C. perfringens*, AB *C. sporogenes* , AC *Alkaliphilus metalliredigens* (strain QYMF), AD *A. oremlandii* (strain OhILAs), AE *Desulfotomaculum kuznetsovii* (strain DSM 6115), AF *Geobacillus* sp. (strain Y412MC10), AG *Paenibacillus polymyxa* , AH *P. mucilaginosus* (strain KNP414).

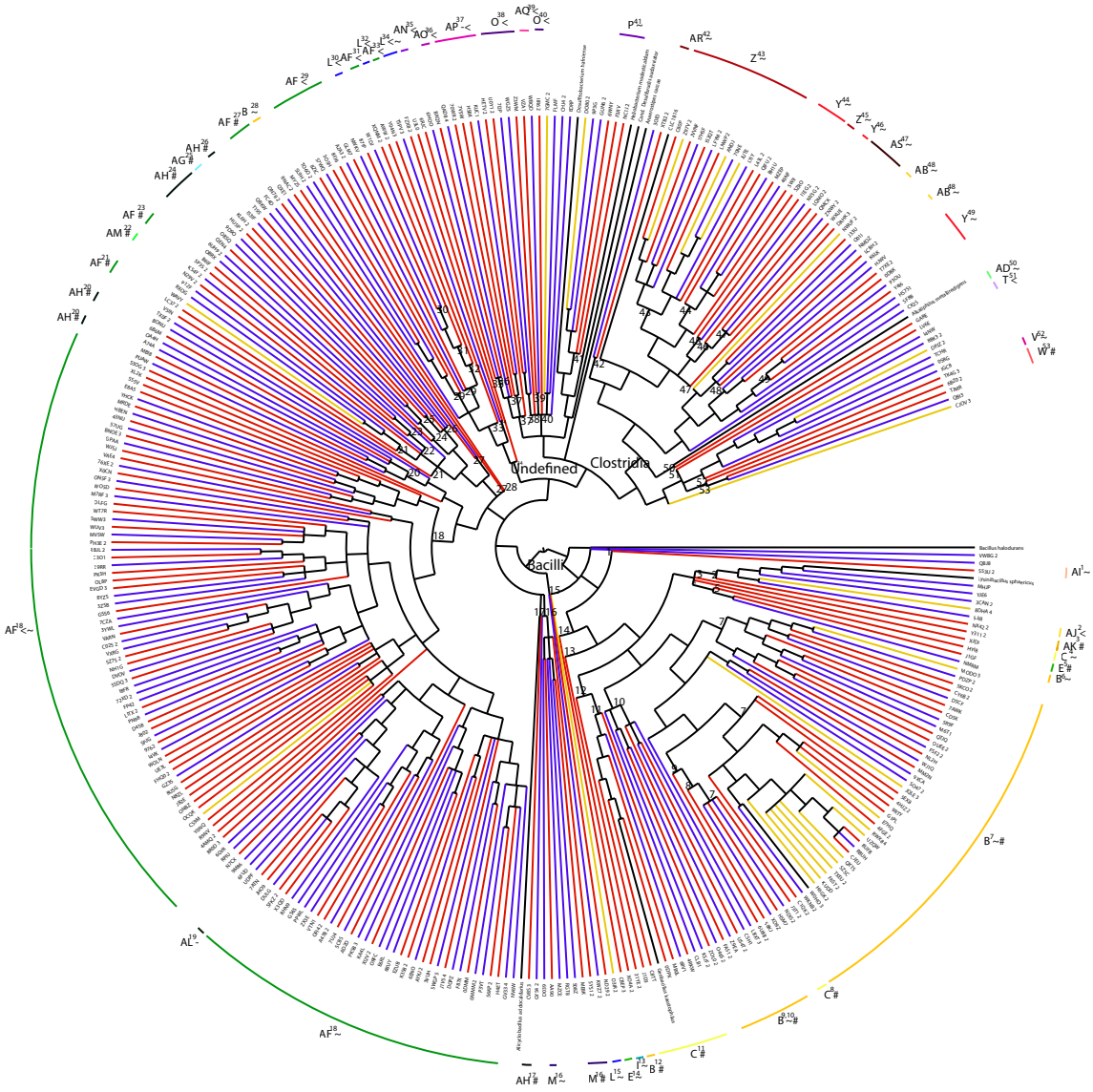


Figure 2.21: Cladogram of *spo0A* sequences from sediment of Lake Baikal extracted with protocols 1 (blue), 2 (yellow), and 3 (red). Closest relatives are shown in letters around the tree together with identity ranges (<65% identity (-), 65-74% (<), 75-84% (), 85-94%(#), >95% (+)). For classes see Table 3 and the following: AI *B. megaterium*, AJ *B. licheniformis* , AK *B. megaterium* (strain DSM 319) AL *C. haemolyticum*, AM *Paenibacillus* sp. (strain JDR-2), AN *B. cellulosilyticus* (strain ATCC 21833), AO *Sulfobacillus acidophilus* (strain TPY), AP *S. acidophilus* (strain ATCC 700253), AQ *Desulforudis audaxviator* (strain MP104C), AR *C. butyricum* , AS *C. kluyveri* (strain ATCC 8527).

2.5 References

Abecasis, A. B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J. B., and Henriques, A. O. (2013). A genomic signature and the identification of new sporulation genes. *J Bacteriol*, 195(9):2101–15.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389–402.

Arcuri, E. F., Wiedmann, M., and Boor, K. J. (2000). Phylogeny and functional conservation of sigma(E) in endospore-forming bacteria. *Microbiology*, 146(Pt 7):1593–1603.

Bartholomew, J. W. and Paik, G. (1966). Isolation and identification of obligate thermophilic sporeforming Bacilli from ocean basin cores. *J Bacteriol*, 92(3):635–8.

Bhandari, V., Ahmod, N. Z., Shah, H. N., and Gupta, R. S. (2013). Molecular signatures for *Bacillus species*: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. *Int J Syst Evol Microbiol*, 63(Pt 7):2712–2726.

Brill, J. A. and Wiegel, J. (1997). Differentiation between spore-forming and asporogenic bacteria using a PCR and southern hybridization based method. *J Microbiol Methods*, 31(1–2):29–36.

Bueche, M., Wunderlin, T., Roussel-Delif, L., Junier, T., Sauvain, L., Jeanneret, N., and Junier, P. (2013). Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene *spo0A*. *Appl Eviron Microbiol*, 79(17):5302–12.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7(5):335–6.

Challacombe, J. F., Altherr, M. R., Xie, G., Bhotika, S. S., Brown, N., Bruce, D., et al. (2007). The complete genome sequence of *Bacillus thuringiensis* Al Hakam. *J Bacteriol*, 189(9):3680–3681.

Chen, X. H., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., et al. (2007). Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nature Biotech*, 25(9):1007–1014.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science*, 322(5899):275–278.

de Rezende, J. R., Kjeldsen, K. U., Hubert, C. R., Finster, K., Loy, A., and Jorgensen, B. B. (2013). Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J*, 7(1):72–84.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., et al. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol*, 77(4):1315–24.

Dineen, S. M., Aranda, R. t., Anders, D. L., and Robertson, J. M. (2010). An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *J Appl Microbiol*, 109(6):1886–96.

Driks, A. (2002). Overview: development in bacteria: spore formation in *Bacillus subtilis*. *Cell Mol Life Sci*, 59(3):389–391.

Dubey, G. P., Narayan, A., Mattoo, A. R., Singh, G. P., Kurupati, R. K., Zaman, M. S., et al. (2009). Comparative genomic study of *spo0E* family genes and elucidation of the role of Spo0E in *Bacillus anthracis*. *Arch Microbiol*, 191(3):241–253.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*, 8(3):186–94.

Feinstein, L. M., Sul, W. J., and Blackwood, C. B. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl Environ Microbiol*, 75(16):5428–33.

Feng, L., Wang, W., Cheng, J. S., Ren, Y., Zhao, G., Gao, C. X., et al. (2007). Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Pro Natl Acad Sci USA*, 104(13):5602–5607.

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., and Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol*, 14(11):2870–90.

Gioia, J., Yerrapragada, S., Qin, X., Jiang, H. Y., Igboeli, O. C., Muzny, D., et al. (2007). Paradoxical DNA repair and peroxide resistance gene conservation in *Bacillus pumilus* SAFR-032. *PLoS One*, 2(9):e928.

Gontang, E. A., Fenical, W., and Jensen, P. R. (2007). Phylogenetic diversity of Gram-positive bacteria cultured from marine sediments. *Appl Environ Microbiol*, 73(10):3272–82.

Han, C. S., Xie, G., Challacombe, J. F., Altherr, M. R., Bhotika, S. S., Bruce, D., et al. (2006). Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. *J Bacteriol*, 188(9):3382–3390.

Hu, X. M., Fan, W., Han, B., Liu, H. Z., Zheng, D., Li, Q., et al. (2008). Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related *Bacillus* species. *J Bacteriol*, 190(8):2892–2902.

Hubert, C., Arnosti, C., Bruchert, V., Loy, A., Vandieken, V., and Jorgensen, B. B. (2010). Thermophilic anaerobes in arctic marine sediments induced to mineralize complex organic matter at high temperature. *Environ Microbiol*, 12(4):1089–104.

Junier, T. and Zdobnov, E. M. (2010). The newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, 26(13):1669–1670.

Klenk, H. P. and Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol*, 33(4):175–82.

Kriventseva, E. V., Rahman, N., Espinosa, O., and Zdobnov, E. M. (2008). OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res*, 36:D271–5. Database issue.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657):249–256.

Kuske, C. R., Banton, K. L., Adorada, D. L., Stark, P. C., Hill, K. K., and Jackson, P. J. (1998). Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl Environ Microbiol*, 64(7):2463–72.

Leff, L. G., Dana, J. R., McArthur, J. V., and Shimkets, L. J. (1995). Comparison of methods of DNA extraction from stream sediments. *Appl Environ Microbiol*, 61(3):1141–3.

Marchant, R., Banat, I. M., and Franzetti, A. (2011). Thermophilic bacteria in cool soils: metabolic activity and mechanisms of dispersal. In Fontaneto, D., editor, *Biogeography of Microscopic Organisms*, page 43–57. Cambridge University Press, Cambridge, UK.

More, M. I., Herrick, J. B., Silva, M. C., Ghiorse, W. C., and Madsen, E. L. (1994). Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl Environ Microbiol*, 60(5):1572–80.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Consortium, I., et al. (2002). InterPro: an integrated documentation resource for protein families, domains and functional sites. *Brief Bioinform*, 3(3):225–235.

Nazina, T. N., Tourova, T. P., Poltaraus, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., et al. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*. *Int J Syst Evol Microbiol*, 51(Pt 2):433–46.

Nicholson, W. L. (2002). Roles of *Bacillus* endospores in the environment. *Cell Mol Life Sci*, 59(3):410–416.

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., and Setlow, P. (2000). Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol Mol Biol Rev*, 64(3):548–572.

Onyenwoke, R. U., Brill, J. A., Farahi, K., and 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.

Ovreas, L., Forney, L., Daae, F. L., and 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.

Paredes, C. J., Alsaker, K. V., and Papoutsakis, E. T. (2005). A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol*, 3(12):969–78.

Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., et al. (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol*, 5(10):R77.

Sattley, W. M., Madigan, M. T., Swingley, W. D., Cheung, P. C., Clocksin, K. M., Conrad, A. L., et al. (2008). The genome of *Heliobacterium modesticaldum*, a phototrophic representative of the Firmicutes containing the simplest photosynthetic apparatus. *J Bacteriol*, 190(13):4687–4696.

Sauvain, L., Bueche, M., Junier, T., Masson, M., Wunderlin, T., Kohler-Milleret, R., et al. (2013). Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria. *Aquat Sci*, in press.

Schaeffer, P., Millet, J., and Aubert, J. P. (1965). Catabolic repression of bacterial sporulation. *Proc Natl Acad Sci USA*, 54(3):704–711.

Schleifer, K.-H. (2009). Phylum XIII. Firmicutes (Gibbons and Murray 1978), 5. In Vos, P., Garrity, G. M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., et al., editors, *Bergey's Manual of Systematic Bacteriology*, page 19–1317. Springer New York.

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., et al. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Gen*, 38(7):779–786.

Seedorf, H., Fricke, W. F., Veith, B., Bruggemann, H., Liesegang, H., Strittmatter, A., et al. (2008). The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *Pro Natl Acad Sci USA*, 105(6):2128–2133.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., et al. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Pro Natl Acad Sci USA*, 99(2):996–1001.

Staley, J. T. and Gosink, J. J. (1999). Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol*, 53:189–215.

Stragier, P. and Losick, R. (1996). Molecular genetics of sporulation in *Bacillus subtilis*. *Annu Rev Genet*, 30:297–341.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–80.

Traag, B. A., Pugliese, A., Eisen, J. A., and Losick, R. (2013). Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. *J Bacteriol*, 195(2):253–60.

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., et al. (2007). Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science*, 315(5815):1126–1130.

Webster, G., Newberry, C. J., Fry, J. C., and Weightman, A. J. (2003). Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. *J Microbiol Methods*, 55(1):155–64.

Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., et al. (2005). Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet*, 1(5):e65.

Yudkin, M. D. and Clarkson, J. (2005). Differential gene expression in genetically identical sister cells: the initiation of sporulation in *Bacillus subtilis*. *Mol Microbiol*, 56(3):578–589.

Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol*, 62(2):316–22.

3 Targeted metagenomics reveals unprecedented diversity of spore-forming bacteria in lake sediments

Abstract

In targeted metagenomics the focus on specific fractions of the environmental genome pool reduces the complexity of the genetic information obtained. Here we present a method for targeted metagenomics that singles out the fraction corresponding to bacterial endospores in environmental samples. We applied the method to sediment samples from Lakes Geneva and Baikal. An enrichment of endospore-formers was demonstrated by the quantification of *spo0A* (endospore-specific) relative to bacterial genes (*rpoB* and 16S rRNA gene) and an increase in the fraction of the community corresponding to Firmicutes (over 90% in the treated samples) according to 16S rRNA gene amplicon sequencing. An 1.7 and 2.4-fold increase in the number of *de novo* OTUs was obtained as a result of the targeted approach. This led to the detection of 34 genera unique to the targeted metagenome, including some supposedly asporogenic groups (e.g. *Ethanoligenens* and *Trichococcus*). A complementary analysis using amplicon sequencing of *spo0A* gene allowed the identification of additional groups (e.g. *Brevibacillus* and *Desmospora*) in the targeted metagenome. This approach constitutes a leap forward in the understanding of the role of endospore-forming bacteria in environmental samples. Moreover, the unexpected diversity in the two investigated sediments suggests that current general methods can only assess up to 10% of the diversity of all bacteria in environmental samples.

3.1 Introduction

Evaluating microbial diversity in the environment and understanding the way in which microorganisms interact with each other and how they reflect changes in environmental conditions, are of great interest but also among the greatest challenges in microbial ecology (Suenaga 2012). The large majority of microorganisms is not amenable to culturing, so diversity studies using genetic material directly extracted from environmental samples emerged as a promising approach to study microbial communities (Suenaga 2012; Amann et al. 1995; Warnecke and Hugenholtz 2007; Xu 2006; Zarraonaindia et al. 2013; Olsen et al. 1986). Pioneering studies amplifying and sequencing the 16S rRNA gene from environmental DNA, in a process dubbed "metagenomics" by Handelsman and colleagues (Handelsman et al. 1998), have started to unveil the vast phylogenetic diversity of the microbial world (Sogin et al. 2006; Huber et al. 2007; Harris et al. 2013). For example, bacterial abundance and diversity in soil have been estimated to 2 x 10^9 cells and 8 x 10^6 species per gram (Gans et al. 2005), respectively. Metagenomic analyses have expanded to include genomes of entire communities (Venter et al. 2004) and have been successfully applied to a variety of environments, helping to reveal new clades and metabolisms, while circumventing the limitations of culturing (Béjà et al. 2000, 2001; Ram et al. 2005; Voget et al. 2003). However in diverse environments such as ocean water, the current sequencing efforts can only cover <0.00001% of the DNA in a sample (Gilbert and Dupont 2011). Only in low diversity environments has it been possible to reconstruct complete genomes from

a metagenomics dataset (Tyson et al. 2004; Chivian et al. 2008). In addition, recent sequencing technologies generate such large quantities of data as to bring along a new set of challenges in data analysis, the so-called bioinformatics bottleneck (Scholz et al. 2012), and although un-assembled sequences can be mapped against databases of known genomes, the required curated databases still do not reflect actual environmental diversity. Furthermore, microbial communities are often too complex to draw meaningful biological conclusions from the assembled metagenomes alone (Sogin et al. 2006; Quince et al. 2008).

An approach to improve coverage and increase resolution of metagenomics is the targeted metagenomics approach. For these methods, the sequencing effort is focused on specific genes or on a subset of metagenomes (Suenaga 2012). Several techniques have been proposed to sequence a reduced fraction of the community (Summerer 2009). Those include amplicon sequencing (e.g. the 16S rRNA gene or functional genes; Iwai et al. 2010; Farnelid et al. 2011) or functional screening based on specific traits (e.g. incorporation of labelled substrates into DNA; Neufeld et al. 2008). Recently approaches including single-cell genomics have focused on sub-communities, opening the black box of the "rare biosphere" or "microbial dark matter" (Rinke et al. 2013). Taking advantage of the large morphological diversity of microbes can be yet another approach to single out specific groups. This has been successfully achieved with magnetotactic bacteria (Lin et al. 2011) or through the filtering of specific size fractions (Baker et al. 2006).

In this study, a targeted metagenomics approach for spore-forming bacteria is proposed. This group of bacteria is able to produce resistance structures when under environmental stress. Spore-like structures have been reported in Actinobacteria (Gram-positive filamentous bacteria; Ensign 1978; Chater and Chandra 2006), Myxobacteria (Gram-negative Deltaproteobacteria; Strauch and Hoch 1992; Thomas et al. 2008), Cyanobacteria (aerobic phototrophic bacteria; Adams and Duggan 1999), and Firmicutes (Gram-positive low G+C content bacteria), although not all members of this phylum sporulate (Galperin et al. 2012). The building of spores in Firmicutes is initiated through asymmetric cell division and the spore maturation happens within the mother cell. For this reason they are called "endo" spores (Green and Cutting 1999). Spores have been overlooked in metagenomic studies, due to their hardy outer cortex, rendering it difficult to extract DNA with traditional methods (von Mering et al. 2007; Delmont et al. 2011). More rigorous DNA extraction methods have been developed that are tailored to endospores (described in chapter 2). This resistant property of endospores is used for the targeted metagenomics approach described here. Using specific treatments the more fragile vegetative cells of a sample can be destroyed, while endospores remain intact. The isolated endospore-fraction can then be subjected to DNA extraction and metagenomic sequencing. This method was tested on strains and sediment samples and resulted in the detection of endospore communities at high resolution.

3.2 Materials and Methods

Strains and culture conditions

To test different cell disruption treatments, we selected the strains *Escherichia coli* (Gram-negative), grown in Nutrient Broth (NB); *Lactococcus lactis* subsp. *lactis* (Gram-positive non-endospore-forming Firmicute), grown in DSMZ 92 medium (`www.dsmz.de`); *Paenibacillus alvei*, *Bacillus subtilis*, and *Bacillus megaterium* (Gram-positive endospore-forming Firmicutes), all grown in NB. All strains were incubated at 30 °C. Endospores were obtained by culturing at 37 °C and vigorous shaking in Schaeffer sporulation (SG) medium (Schaeffer et al. 1965) for *B. megaterium*, and modified versions of this medium (2 x SG for *B. subtilis*; 0.5 x SG for *P. alvei*). Numbers of cells and endospores were determined microscopically using a Neubauer counting chamber. All endospore preparations consisted of >95% endospores. Prior to the treatments, cells or endospores were recovered by centrifugation at 7,500 rpm for 5 min and resuspended in 500 µl of physiological solution. Densities of 10^8 cells/ml or endospores were achieved, which corresponded to an absorbance of 0.3 A at 600 nm.

Development of the treatment with pure cultures

Various physical and chemical treatments were assayed for the disruption of vegetative cells. UV (treatment 1) was applied to open centrifuge tubes under a laminar flow hood at 20 cm distance from the UV-lamp emitting 30 W/cm. Sonication (2) was done one ice with a sterilized probe at 40 W output. Wet heat treatments were done on 500 µl sample put into a water bath heated to 65 °C (3) or 85 °C (4, 5). The following chemical or enzymatic agents were used: EDTA (1 mM) (treatment 6) (Sigma-Aldrich, Buchs, CH), NaOH (0.5 N or 1 N) (7, 8), sodium dodecyl sulphate (SDS) (1%) (9) (Sigma-Aldrich) and lysozyme in Tris-EDTA (5 mg/ml) (10) (Sigma-Aldrich). Stock solutions of the different chemical agents or lysozyme were added to the 500 µl sample as to reach the final concentrations. Finally a cocktail was tested consisting of successive treatment with lysozyme and then NaOH (0.5 N) and SDS (1%) (treatment 11), as well as a combined treatment of heat at 65 °C and the cocktail treatment afterwards (12). In the combined treatment the chemicals were added without the removal of the previous treatment and the volume increased over each consecutive step. After the treatment, the samples were centrifuged at 7,500 rpm for 5 min, supernatant was removed, and samples were resuspended in 500 µl of physiological solution.

Treated and control (untreated) samples were diluted 1/10 in NB. Re-growth of treated cells or endospores was assessed by the most probable number (MPN) counting method (Rowe et al. 1977) in triplicates at dilutions of 10^{-1} to 10^{-7} in 200 µl volume in 96-well microplates. Plates were sealed and incubated at 37 °C and measured with a spectrophotometer at 600 nm wavelength (ASYS Hightech, Cambridge, UK). After blank subtraction (sterile NB medium), wells with a value above 0.05 were considered as positive. The scale of growth was obtained by comparing the re-growth of the treated samples with an untreated control sample.

Treatment evaluation with pure cultures

A combined treatment consisting of wet heat at 65 ℃ for 20 min, followed by lysozyme digestion for 60 min and digestion with 0.5 N NaOH and 1% SDS for another 60 min was applied to endospore preparations of *B. subtilis*, *B. megaterium*, and *P. alvei* and to vegetative cells of *E. coli*. Growth of control cultures and re-growth of treated cultures was measured as described above. Measurements were taken every 60 min during 15 hours.

Environmental samples

Sediment samples were collected during a research campaign with the MIR manned submersibles in Lake Baikal (Russia)(52°53.00 N, 107°10.00 E, at 1598.5 m depth) in August 2010 and in Lake Geneva (Switzerland-France)(46°27.03 N, 6°42.52 E, at 284 m depth) in June 2011. Sediment cores of approximately 30 cm length were retrieved using a push-corer. Upon return to the surface, the fraction two-to-seven cm of the cores was immediately sub-sampled in the centre of the core as to avoid contamination using sterile cut-open syringes. Samples were then stored at -20 ℃ until further processing.

Sediment treatment

The procedure to extract spores from sediment is schematically shown in Figure 3.1. In a first step, 3 g of sediment was used, and the biomass was separated from the sediment particles as described in chapter 2. The slurry was then centrifuged at 20 x g for 1 min to remove coarse particles. The supernatant (containing the cells) was collected on a nitrocellulose membrane of 0.2 µm pore size and 47 mm diameter (Whatman, Dassel, Germany). The separation of biomass from sediment was repeated for a second time on the same sample. The filter membrane was cut in half and immediately frozen in liquid nitrogen and stored at -20 ℃. One half of the filter membrane was kept as a control (untreated) and the other was subjected to the treatment for destruction of vegetative cells. The filter half was transferred into a centrifuge tube with 900 µl tris-EDTA buffer and gently dispersed by vortexing. The mix was then incubated in a water bath at 65 ℃ for 20 min and cooled at 4 ℃ for 5 min. Lysozyme was added to the sample in a final concentration of 2 mg/ml and incubated at 37 ℃ for 60 min and 120 rpm shaking. Then 250 µl of NaOH and 250 µl of SDS were added to obtain final concentrations of 0.5 N and 1%, respectively. The mix was then incubated at room temperature for 60 min with slow agitation (100 rpm). The sample was afterwards filtered through 0.2 µm pore size nitrocellulose membrane of 17 mm diameter with a vacuum pump and washed with sterile physiological solution to remove residual detergent. The vacuum pump was turned off for subsequent DNAse treatment directly on the filter membrane. The mix of 400 µl containing water, 1 x reaction buffer and 2 µl DNase (New England Biolabs, Ipswich, MA, USA) was added directly onto the membrane and left standing for 15 min to digest the extracellular DNA in the sample. Then the sample was washed with physiological solution by applying the vacuum again.

Once dried, the filter membrane was stored at -20 ℃ until DNA extraction. The treated and un-treated membranes (half filter used as control) were both subjected to a repetitive DNA extraction protocol described in detail in chapter 2. Briefly, DNA was extracted directly from the membranes using the MP FastDNA SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA, USA) but with a modified protocol that consisted of three repetitive extraction cycles. In the end, the three extracts were pooled and DNA precipitated with 0.3 M sodium-acetate and ethanol (99%) and washed with ethanol (70%) before being re-suspended in sterile water (40 µl for control samples, 20 µl for treated samples). DNA yield was measured with the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using the Quant-iT dsDNA BR assay kit, following the manufacturer's instructions. DNA quality was verified by agarose gel electrophoresis.

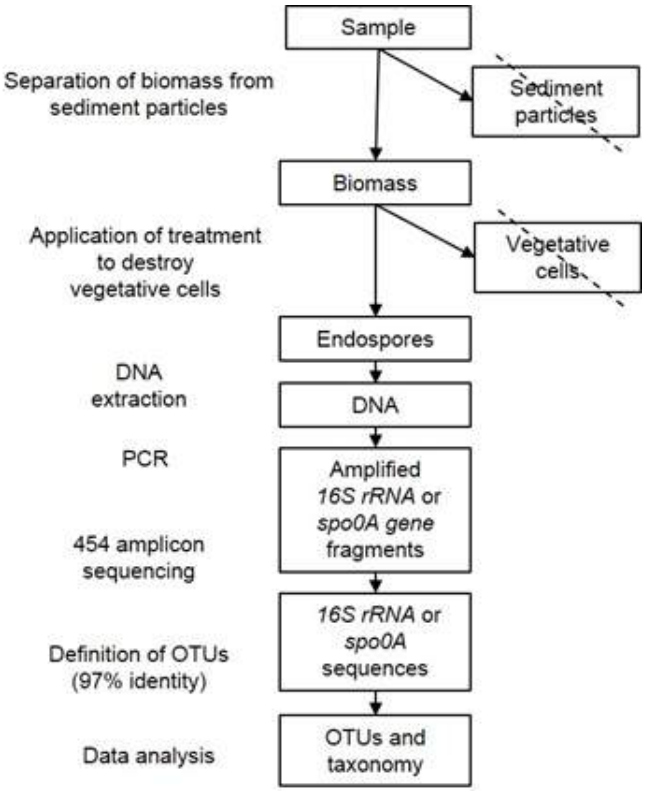


Figure 3.1: Schematic overview of the experimental procedure used to determine diversity of endospore-forming bacteria in sediments. For the global bacterial diversity the step of treatment to remove vegetative cells was omitted.

Quantitative PCR (qPCR)

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 the primers 338f and 520r (Ovreas et al. 1997). The detailed protocol is described in chapter 2.

The qPCR mix contained 0.5 ng DNA template. All extracts were analyzed in triplicates. For quantification three independent plasmid standards series with 300 to 3,000,000 gene copies/µL of the 16S rRNA gene of an environmental clone were included. Quantification of *spo0A* gene was done as mentioned elsewhere (Bueche et al. 2013) but with TaKaRa SYBR Premix Ex Taq II (TaKaRa, Shiga, Japan) with 1.4 or 0.5 ng DNA as template. The reaction mix was composed of 1 x TaKaRa Premix, 0.75 µM of forward and 0.45 µM of reverse primer and water to total volume of 20 µl. Three independent plasmid standards series with 30 to 300,000 gene copies/µL of *spo0A* gene of *B. subtilis* were included. Quantification of the *rpoB* gene was done with 0.5 ng DNA template, 0.3 µM of primers rpoB-f-4 and rpoB-r-2 (Dahllöf et al. 2000) and 1 x TaKaRa Premix in a total reaction volume of 20 µl. All qPCR reactions were run with a Rotor-GeneTM 6000 instrument (QIAGEN, Hilden, Germany) with the program: enzyme activation at 95 ℃ for 5 min, 40 cycles of denaturation at 95 ℃ for 5 sec, annealing at 55 ℃ for 20 sec and extension at 72 ℃ for 30 sec. Thresholds (Th), Ct values, and derivatives of melting curves were determined using Rotor-Gene 6 software (QIAGEN).

Preparation for 16S rRNA gene amplicon sequencing

All samples were amplified with the primers Eub9-27f and Eub1542r to produce a 1,533 bp fragment of the 16 s rRNA gene (Liesack et al. 1991). PCR reactions were performed with 0.5 ng DNA template in 1 x reaction buffer (TaKaRa), 2 mM MgCl2, 10 µg bovine serum albumin (BSA; New England Biolabs, Ipswich, MA, USA), 1 U of Ex Taq Polymerase (Proof-reading polymerase; TaKaRa), 200 µM of each dNTP and 200 nM 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 ℃ for 2 min; then 10 cycles of denaturation at 94 ℃ for 30 sec, touchdown annealing starting at 60 ℃ with decrease of 0.5 ℃ per cycle for 45 sec and elongation at 72 ℃ for 1 min; followed by 30 cycles of denaturation at 94 ℃ for 30 sec, annealing at 55 ℃ for 45 sec and elongation at 72 ℃ for 1 min; and a final extension at 72 ℃ for 5 min. Duplicate reactions (total volume of 100 µl) were then pooled and purified with MultiScreen PCRµ96 plate (Merck Millipore, Darmstadt, Germany). Re-eluted sample in 20 µl were then loaded onto a 1% agarose electrophoresis gel and run for 40 min at 80 V. Gel bands of 1,533 bp size were excised and purified with QiaQuick Gel extraction kit (QIAGEN). A minimum of 500 ng of amplified DNA was then sent for barcode amplicon sequencing with Roche GS FLX+ (Eurofins MWG Operon, Ebersberg, Germany).

Sequence analysis of 16S rRNA gene data

Sequences were analysed with the QIIME pipeline for high-throughput 16S rRNA sequences (Caporaso et al. 2010). After quality filtering (minimum quality score of 25), the sequences were clustered *de novo* into putative OTUs (identity of >97%)

with the pick-otus.py program using the Uclust method (Edgar 2010). Taxonomy was assigned with the RDP classifier (Wang et al. 2007). Alpha diversity was calculated using chao1-index (Chao 1984), and Beta diversity was calculated via an Euclidean distance matrix. Rarefaction curves were calculated using chao1 and plotted with the QIIME tools. All metagenomic sequences were submitted to Sequence Read Archive (SRA) under BioProject number PRJNA214154 (BioSamples SRS431098 (Lake Geneva) and SRS431097 (Lake Baikal)) and accession numbers SRR1011310, SRR1011311, SRR1011312, SRR1011313, SRR1011314, SRR1011315, SRR1011316, SRR1011317.

Preparation for *spo0A* gene amplicon sequencing

A 602 bp sequence of the *spo0A* gene was amplified with the primers spoA166f and spoA748r, described in chapter 2. PCR reactions were performed with 1 ng DNA template, 1 x reaction buffer (TaKaRa), 3 mM MgCl2, 10 µg BSA (New England Biolabs), 1 U of 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 run on an Arktik Thermo Cycler (ThermoFisher Scientific) with the following temperature program: initial denaturation at 94 ℃ for 5 min; then 10 cycles of denaturation at 94 ℃ for 30 sec, touchdown annealing starting at 55 ℃ with decrease of 0.3 ℃ per cycle for 30 sec and elongation at 72 ℃ for 1 min; followed by 30 cycles of denaturation at 94 ℃ for 30 sec, annealing at 52 ℃ for 30 sec and elongation at 72 ℃ for 1 min; and a final extension at 72 ℃ for 5 min. Reactions were done in quintuplets that were pooled together and purified with MultiScreen PCRµ96 plate (Merck Millipore) using a vacuum pump and afterwards eluted in 20 µl molecular grade sterile water. The purified samples were loaded onto a 1% agarose gel and electrophoresis run for 40 min at 80 V. The bands of the correct size (602 bp) were excised and purified with QiaQuick Gel extraction kit (QIAGEN). Purified amplicons were then sent to Eurofins MWG Operon for barcode amplicon sequencing with Roche GS FLX+.

Sequence analysis of *spo0A* gene data

Sequences were first binned according to their barcode and the corresponding sample. For quality filtration, the nucleotide sequences were translated to their amino acid sequences, based on ORF detection. The amino acid sequences were then aligned and compared to a Gribskov-style protein profile of *spo0A* sequence (Gribskov et al. 1987) that was built based on 27 known *spo0A* sequences described in chapter 2 (Supplementary Table 2.5). Filtration was applied as function of the profile score and profile alignment length, which separates noise or negatives hits from true postives *spo0A* sequences. The nucleotide sequences were clustered into operational taxonomic units (OTU), at 97% sequence identity using uclust (Edgar 2010) in the same way as done for the 16S rRNA gene. The centroid of each OTU

cluster was then retrieved and the closest relative available in GenBank database was searched by BLAST of the translated protein sequence.

Numerical analysis

Diversity indices (Shannon-Wiever, Pielou's Evenness and Simpson index) were calculated with the software R and the package *vegan* (R Core Team (2012); Oksanen et al. (2007)). Principal component analysis was done using R (R Core Team (2012)) with the community matrix of all OTUs identified as members of the phylum Firmicutes.

3.3 Results

Development of a method to target endospores

Based on the fact that endospores are resistant structures that tolerate more vigorous disruption conditions than vegetative cells, we devised a method to separate them from vegetative cells. We tested 13 cell disruption treatments consisting of chemical (EDTA, NaOH, SDS), physical (heat, UV, sonication), or enzymatic (lysozyme) agents, as well as a combination of these. The treatments were tested on vegetative cells and endospores from pure cultures. We selected three types of vegetative cells with different cell wall compositions: *E. coli* (Gram-negative), *L. lactis* (Gram-positive non-spore-former) and *P. alvei* (Gram-positive endospore former). An endospore preparation composed of 95% endospores of *P. alvei* (sporulation verified by optical microscopy) was also included. The efficiencies of the treatments were evaluated by re-growth of the sample after treatment and estimation of cell numbers by the most probable number (MPN) counting method. The results were scored relative to untreated controls that were grown and quantified in parallel (Figure 3.2).

Vegetative cells of *P. alvei* were in general easily removed with any of the treatments, although a minor re-growth was detected after treatments 2 (sonication), 4 and 5 (heat 85 ℃ for 20 min and 60 min), 6 (EDTA), 7 (NaOH 0.5 N), 10 (lysozyme), 11 (combined methods). Vegetative cells of *E. coli* and *L. lactis* were much more resistant and substantial re-growth was observed after treatments 2, 6, 9 (SDS) and 10. Finally, endospores of *P. alvei* were, as expected, resistant to most of the treatments, except to treatment 8 (NaOH 1 N). Complete re-growth was only measured in treatments 3 (heat 65 ℃ for 20 min), 5, 11, and 12 (heat followed by the cocktail). In summary, three treatments killed the vegetative cells of the tested strains (Gram-negative and Gram-positive) without harming the endospores of *P. alvei*. These treatments were heat (65 ℃) (3), a combination of lysozyme digestion for 60 min and digestion with a mix of NaOH and SDS (11), as well as a combination of heat (65 ℃), lysozyme and the mix of NaOH and SDS (12). The limited number of strains that were used for the screening does not account for morphological variation of cells and endospores in environmental samples. To ensure disruption of cells of different morphologies, the combined treatment consisting of heat, lysozyme and a mix of NaOH and SDS was chosen as final treatment (treatment 12, framed in figure 3.2A).

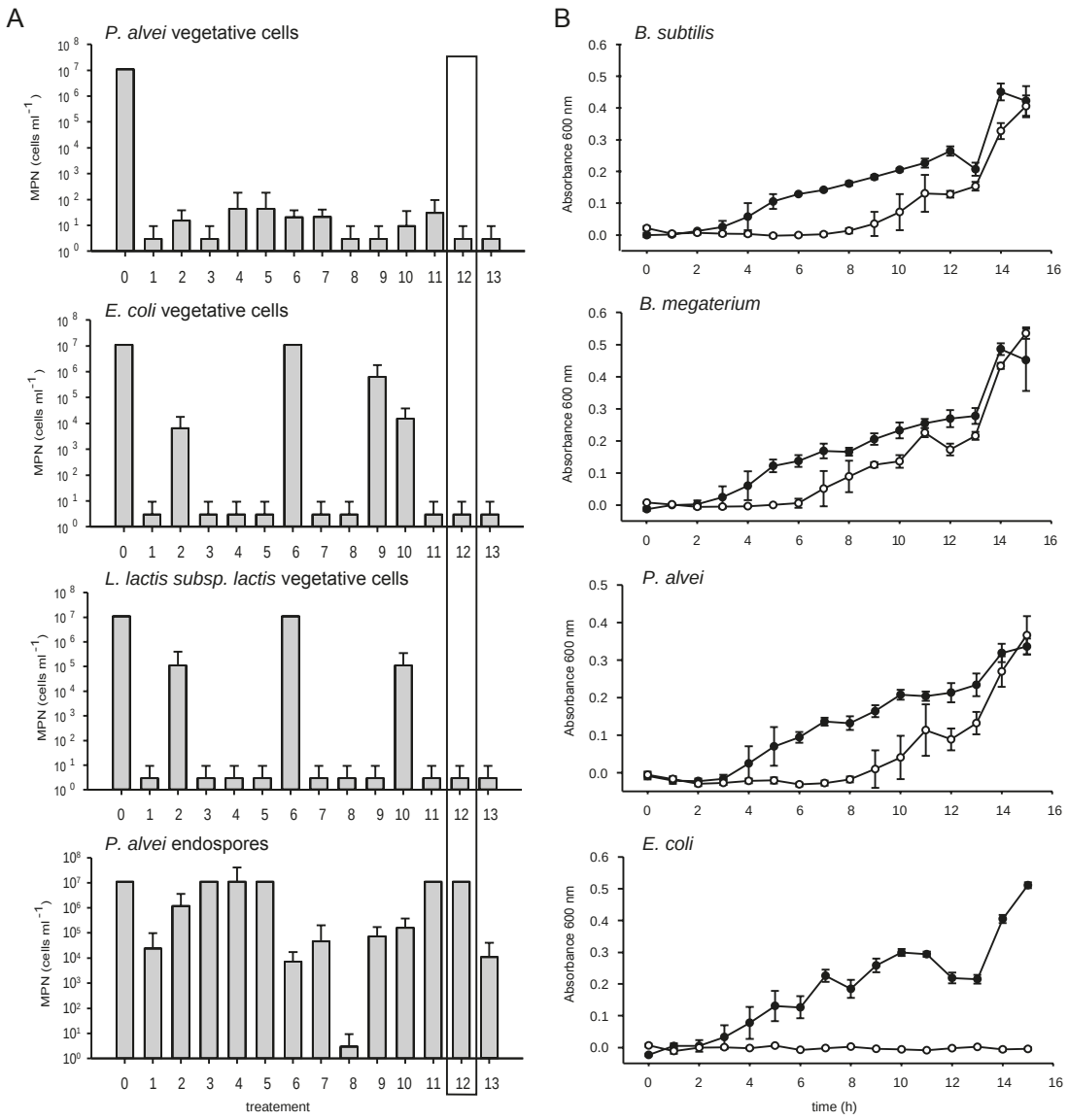


Figure 3.2: Development and evaluation of cell disruption methods.
A. Assays of re-growth after a series of disruption treatments were applied to preparations of cell and endospores. The quantification of growth was carried out using the most probable number (MPN) counting method. Counting above 10^1 cells/ml was considered positive growth. Treatments: (0) untreated, (1) UV for 2 min, (2) sonication for 30 sec, (3) heat at 65 ℃ for 20 min, (4) heat at 85 ℃ for 20 min, (5) heat at 85 ℃ for 60 min, (6) EDTA 1 mM for 60 min, (7) NaOH 0.5 N for 60 min, (8) NaOH 1 N for 60 min, (9) SDS 1% for 60 min, (10) lysozyme 5 mg/ml for 60 min, (11) cocktail (lysozyme 10 mg/ml for 60 min then NaOH 0.5 N and SDS 1% for 60 min, (12) mix of heat 65 ℃ for 20 min and cocktail (see 11), (13) heat 85 ℃ for 10 min then cocktail (see 11). Treatment 12 selected for further experiments is framed. Error bars from three independent MPN counts. **B**. Growth curves verifying the effectiveness of the disruption method applied to different bacterial species. Growth curves of treated and control endospore preparations from *Bacillus subtilis*, *Bacillus megaterium*, and *Paenibacillus alvei*, and a cell culture of *Escherichia coli*. Treated = white circles. Untreated = black circles. Error bars from three independent cultures.

The efficiency of this treatment was then verified using endospore preparations (>95% endospores) of *P. alvei* and two additional endospore-forming species (*B. subtilis* and *B. megaterium*). These endospore preparations were compared to a cell culture of *E. coli*. A treated and an untreated culture of each strain was incubated in NB and growth was measured using optical density at 600 nm. The growth curves (Figure 3.2B) showed growth in treated endospore preparations, demonstrating that the endospores withstand the treatment (albeit with an extended lag phase before germination and multiplication). In comparison no growth was observed in the vegetative cell sample from *E. coli* even after 15 hours of incubation.

Targeted sequencing of endospore-formers in sediment samples

The disruption treatment was then applied to two sediment samples from Lake Geneva (LB) and Lake Baikal (LB). Cells and spores were initially separated from the sediment particles because this has been shown to improve coverage when studying the diversity of endospore-formers, as described in chapter 2. The disruption treatment was applied to half of the recovered biomass (targeted approach) and the results were compared to untreated samples (global approach). To avoid contamination with exogenous DNA released during the disruption step in the treated sample, a DNAse treatment was carried out prior to DNA extraction. An overview of the results is presented in Table 3.1. Although DNA yields from the treated samples were substantially lower than from the untreated samples, the total number of 16S rRNA, *rpoB*, and *spo0A* gene copies was higher in the treated samples. The enrichment of the endospore-forming community with the treatment is demonstrated by the ratio in copy numbers of the *spo0A* gene (present in endospore-forming bacteria, Bueche et al. 2013) and the single copy gene *rpoB* (present in all bacteria). The ratio of *spo0A* to *rpoB* in LG samples increased from 0.9 (untreated) to 16.9 (treated) and from 3.1 to 13.8 in LB samples (Table 3.1).

Community composition based on 16S rRNA gene sequencing

To compare the composition of the bacterial communities obtained from the global approach (untreated samples) with the targeted approach (treated samples), amplicon sequencing of the 16S rRNA gene was done. Amplicon sequencing of the four samples (LG and LB, both treated and untreated) resulted in a total of 86,985 sequence reads. Between 45.1 and 48.9% of the sequences were retained after quality filtering. The sequences were clustered into operational taxonomic units (OTUs) based on 97% sequence identity.

The richness detected in LG samples was 3,814 OTUs with the global approach and 6,385 OTUs with the targeted approach, corresponding to a 1.7-fold increase. For LB samples, the increase in richness was even larger (2.4-fold) from global approach (3,551 OTUs) to targeted approach (8,612 OTUs). These results demonstrate a detection of higher richness when applying the targeted approach.

The community composition of phyla representing >1% of the global community is shown in Table 3.1 (entire list in Table 3.4). With the global approach,

Table 3.1: Overview of molecular data for untreated and treated sediment samples in Lake Geneva (LG) and Lake Baikal (LB). Gene numbers were measured by quantitative PCR. Values are normalized to 1 ng DNA. Standard deviations correspond to triplicate measurements. Ratios were calculated using the average values. Community composition with sequence counts and percentage values (in brackets) are shown for phyla with percentages of >1% in the untreated samples (detailed composition in supplementary Table 3.4). Abundances for Firmicutes are highlighted in bold. QC = quality control.

	Lake Geneva		Lake Baikal	
	Untreated	Treated	Untreated	Treated
DNA yield (ng/g of sediment)	3360	28	924	25.2
16S rRNA gene numbers (x 10^5)	2.57±0.4	11.0±1.5	2.39±0.2	22.0±1.4
rpoB gene numbers (x 10^3)	4.48±0.6	4.99±0.3	3.72±0.1	8.32±0.2
spo0A gene numbers	40.74±3.3	842.86±50	116.9±11.5	1148.67±101
Ratio *spo0A*/16S rRNA genes ratio	0.02	0.08	0.05	0.05
Ratio *spo0A*/*rpoB* genes ratio	0.91	16.88	3.14	13.81
Amplicon sequencing for of the 16S rRNA gene				
Sequence counts after QC	8647	11702	7137	14327
Nr of OTUs after QC	3814	6385	3551	8612
Ratio OTUs/sequence counts	2.27	1.83	2.01	1.66
Sequence counts per phylum (%)				
Acidobacteria	466 (5.4)	6 (0.05)	286 (4)	34 (0.2)
Actinobacteria	239 (2.8)	111 (0.9)	336 (4.7)	213 (1.5)
Bacteria unclassified	792 (9.2)	298 (2.5)	762 (10.7)	399 (2.8)
Chloroflexi	258 (3)	54 (0.05)	226 (3.2)	44 (0.3)
Firmicutes	863 (10)	10806 (92.3)	2194 (30.7)	12972 (90.5)
Nitrospirae	121 (1.4)	1 (0.01)	127 (1.8)	36 (0.3)
Planctomycetes	414 (4.8)	7 (0.06)	147 (2.1)	29 (0.2)
Proteobacteria	4938 (57.1)	395 (3.4)	2590 (36.3)	545 (3.8)
Alpha-proteobacteria	1147 (13.3)	104 (0.9)	578 (8.1)	244 (1.7)
Beta-proteobacteria	1453 (16.8)	44 (0.4)	1500 (21)	223 (1.6)
Delta-proteobacteria	781 (9)	185 (1.6)	266 (3.7)	49 (0.3)
Epsilon-proteobacteria	256 (3)	3 (0.03)	78 (1.1)	6 (0.04)
Gamma-proteobacteria	861 (10)	40 (0.3)	-	-
Verrucomicrobia	111 (1.3)	6 (0.05)	-	-

the most abundant phyla that were detected were Proteobacteria (Alpha- and Beta-proteobacteria), Firmicutes, Acidobacteria, Planctomycetes, Chloroflexi, and Actinobacteria. In contrast, with the targeted approach the phylum Firmicutes (only phylum known to comprise endospore-formers), was by far the most abundant phylum with 92.3% (LG) and 90.5% (LB) of the total community. These results confirm the successful removal of most vegetative cells when applying the treatment. Taxa other than Firmicutes were detected in minor abundance with the targeted approach (Table 3.5). The most abundant non-Firmicute taxa were affiliated to Proteobacteria (Delta-, Alpha-, and Beta-proteobacteria), Actinobacteria, and Chloroflexi. However, none of the phyla represented more than 1.5% of the community (896 (LG) and 1,355 (LB) sequences from a total of 11,702 and 14,327 sequences, respectively).

Community composition of the phylum Firmicutes

Firmicutes clearly dominated the targeted community (treated samples), their diversity was therefore analysed in detail (Figure 3.3). A comparison of the Firmicutes community of all samples showed that the community structure did not vary considerably between the global and targeted approaches (Figure 3.3A). This finding was confirmed by the grouping (by lake rather than by treatment) of the samples in the principal components analysis (PCA) of the Firmicutes community (Figure 3.3B).

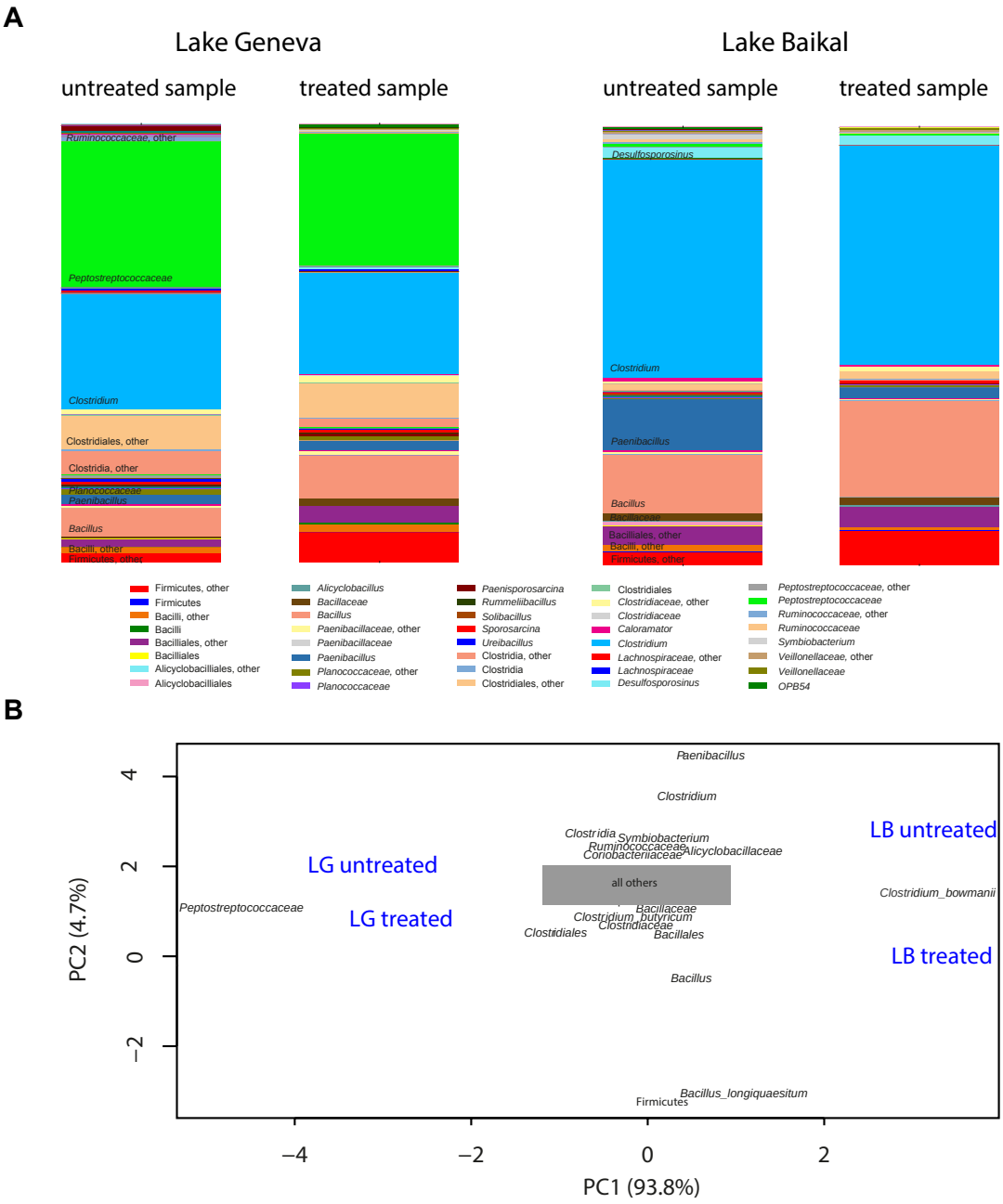


Figure 3.3: Comparison of diverstiy of Firmicutes. **A**. Taxonomic distribution of the OTUs defined from the 16S rRNA gene affiliated to the phylum Firmicutes revealed with the global (untreated) and the targeted approach (treated samples). **B**. Principal component analysis (PCA) based on the community structure of Firmicutes in the treated and untreated samples of Lake Geneva (LG) and Lake Baikal (LB). The majority of the sequences clustered in the middle, delineated by the gray box of "all others".

In the PCA the treated and untreated samples from LG or LB grouped together based on PC1 (explaining 93.8% of the variance) and were slightly differentiated in PC2 (explaining an additional 4.7% of the variance). Although each lake is clearly separated, surprisingly only a few endospore-forming groups appear to be responsible for this, namely Peptostreptococcaceae for LG samples, and *Clostridium bowmanii* for LB samples (Figure 3.3B). The majority of the members grouped closely in the middle.

Specific sequences disappeared or newly appeared after applying the treatment (Table 3.6). Those disappearing corresponded to a small fraction (14 OTUs in LG and 8 in LB samples). More remarkably, a total of 6,462 OTUs were newly detected (3,395 in LG and 3,067 in LB) with the targeted approach. In LG samples 34 genera within Firmicutes were newly detected when applying the targeted approach. The largest numbers of sequences retrieved were affiliated to the families Peptostreptococcaceae and Paenibacillaceae and genera *Bacillus* and *Clostridium*. A large fraction of the sequences could not be resolved to a finer taxonomic level, for example those clustering with Peptostreptococcaceae. However in some cases the sequences could be assigned to a species level. Eleven *Bacillus*, eight *Clostridium*, and seven *Paenibacillus* species could be assigned (Table 3.6). In the LB sample, the most abundant members were affiliated to *Bacillus*, *Clostridium*, *Desulfosporosinus*, *Sporosarcina*, and *Veillonellaceae*. In this sample, 1,572 sequences were identified as *Bacillus longiquaesitum* and 2,892 sequences to *C. bowmanii* (Table 3.7).

Structure of endospore community based on the *spo0A* gene

The diversity of endospore-forming bacteria in the treated samples of LG and LB was also analysed by sequencing the *spo0A* gene. This molecular marker has been demonstrated to be specific to endospore-forming bacteria in chapter 2. The *spo0A* amplicon sequencing of the treated LG and LB samples resulted in 1,174 and 2,460 sequence reads, respectively, of which 81 and 85% were retained after quality filtering (Table 3.2). OTU picking was done using the same criteria as for the 16S rRNA gene (97% identity level).

The community composition was defined to genus level, where possible (Table 3.2). In the LG sample, members affiliated to *Paenibacillus* (over 50% of the community), *Bacillus*, *Clostridium*, *Desmospora*, and *Brevibacillus* dominated the communities. Composition in LB was more evenly distributed with 26% of OTUs affiliated to *Paenibacillus* or *Bacillus*. Other minor groups such as *Sporomusa*, *Desmospora*, and *Geobacillus* were also detected.

The *spo0A*-based community composition was compared to the community composition based on the 16S rRNA gene (Figure 3.4). The first obvious difference between the two datasets is the fraction of unclassified groups, which was more than 61% LG and 21% (LB) in the 16S rRNA gene data, and only 3% in the *spo0A* data. Among the identified groups, the genera *Paenibacillus*, *Bacillus*, *Clostridium*, *Geobacillus*, *Alicyclobacillus*, *Anoxybacillus*, *Desulfotomaculum*, and *Sporosarcina* were found in both datasets, although with important differences in their relative abundances (with the exception of *Bacillus*). Several groups were only found in

Table 3.2: Community composition based on *spo0A* amplicon sequencing in the treated samples from Lake Geneva and Lake Baikal. OTUs were defined and classified to genus level. Sequence counts are given and the percentage abundance in parenthesis. QC = quality control.

	Lake Geneva	Lake Baikal
Sequence counts after QC	950	2084
Nr of OTUs after QC	406	734
Ratio OTUs/sequence counts	2.3	2.8
Sequence counts per genus (%)		
Alicyclobacillus	2 (0.002)	-
Amphibacillus	-	6 (0.3)
Anoxybacillus	1 (0.1)	20 (1.0)
Bacillaceae	6 (0.6)	4 (0.2)
Bacillus	98 (10.3)	532 (25.5)
Bacteria	24 (2.5)	54 (2.6)
Bhargavaea	-	2 (0.1)
Brevibacillus	65 (6.8)	104 (5.0)
Candidatus *Desulforudis*	-	3 (0.1)
Carboxydothermus	-	3 (0.1)
Clostridiales	-	1 (0.05)
Clostridium	85 (8.9)	73 (3.5)
Desmospora	82 (8.6)	122 (5.9)
Desulfotomaculum	10 (1.1)	37 (1.8)
Geobacillus	33 (3.5)	96 (4.6)
Heliobacterium	7 (0.7)	17 (0.8)
Moorella	3 (0.3)	28 (1.3)
Oceanobacillus	-	2 (0.1)
Paenibacillus	493 (51.9)	551 (26.4)
Pasteuria	8 (0.8)	12 (0.6)
Sporomusa	20 (2.1)	380 (18.2)
Sporosarcina	7 (0.7)	23 (1.1)
Sulfobacillus	4 (0.4)	10 (0.5)
Thermacetogenium	2 (0.2)	4 (0.2)

either one of the datasets (Table 3.7). Around 3% of the genera found in the 16S rRNA gene dataset could not be detected in the *spo0A* data, while 23% (LG) and 36% (LB) of genera were only detected with the *spo0A* gene, including some of the highly abundant genera such as *Sporomusa*, *Desmospora*, and *Brevibacillus*.

Increased coverage with targeted approach

The comparison between diversity measures from the global and the targeted approach suggested substantially better detection of diversity with the targeted approach (3.3). The richness increased 10-fold in the LG sample and 4-fold in the LB sample when targeted metagenomics is applied. Similar values were obtained based on Chao's diversity measure, with an increase of 7.5 x (LG) and 3.5 x (LB) with the targeted approach. To estimate the sequencing coverage, the number of genomes in each sediment extract was estimated. The calculation was based on the DNA yields of the different extracts and an average genome size of 2 x 10^6

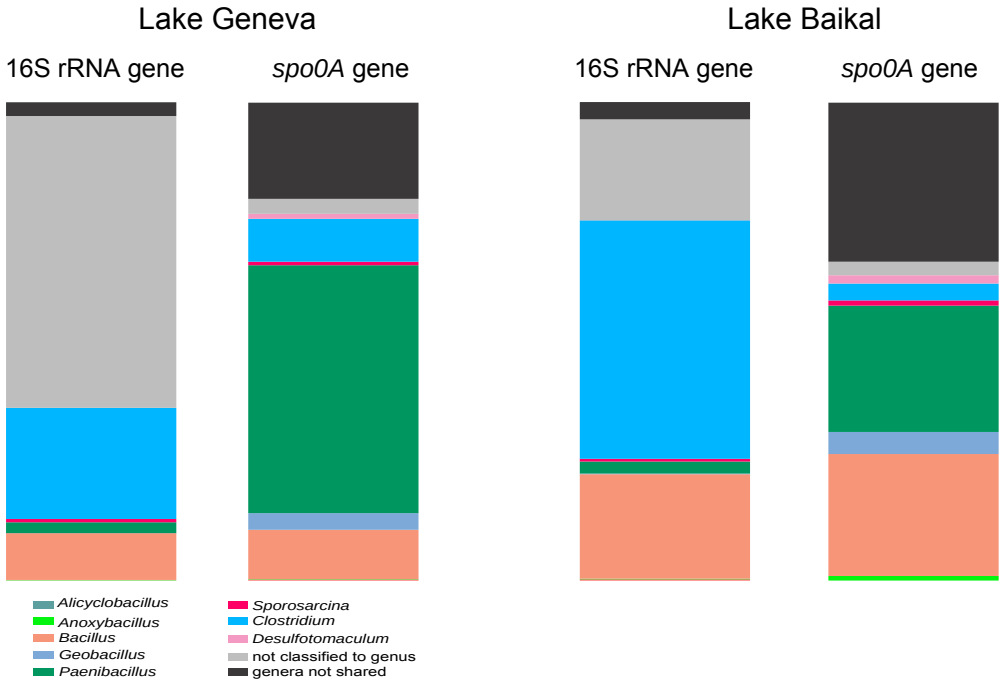


Figure 3.4: Comparison of the taxonomic distribution of genera belonging to Firmicutes derived from 16S RNA gene and *spo0A* datasets for the treated samples from Lakes Geneva and Baikal. (Detailed information is provided in supplementary Table 4.)

base pairs (at average weight of 650 dalton per base pair). Using the number of genomes in the sample and the number of sequence reads that were obtained from the 16S rRNA amplicon sequencing, the coverage can be estimated (Table 3.3). Coverage was extremely low but increased by two orders of magnitude with the targeted approach.

Detection of rare species

The extent to which targeted metagenomics improved the detection of the rare biosphere, i.e. bacteria in low abundance, was determined by the species abundance curves (Figure 3.5). The curves are based on the detection of sequences classified in the phylum Firmicutes based on 16S rRNA sequencing. The curves for the treated samples (targeted approach) indicate substantial extensions of the tail on the x-axis, corresponding to OTUs in low abundance (mostly singletons). The number of OTUs corresponding to single sequence counts is substantially extended (10- and 4- fold) with the targeted approach (Table 3.3). The vast majority of the singletons corresponded to OTUs only detected as a result of the targeted approach. Only a minor fraction of the OTUs in the 16S rRNA gene data are shared between the untreated and treated samples (69 shared OTUs in LG and 85 shared OTUs LB samples), which correspond to the most abundant OTUs. The ratios of singletons to total OTUs remained relatively constant (close to 68%) in both, the 16S rRNA gene and the *spo0A* gene data (Table 3.3). The same ratio

Table 3.3: Diversity metrics for the community affiliated to Firmicutes based on the 16S rRNA gene dataset and on the *spo0A* dataset. NA = not applicable.

		Lake Geneva		Lake Baikal	
		untreated	treated	untreated	treated
Firmicutes (16S rRNA gene)	Richness (R)	367	3751	949	4010
	Singletons (S)	252	2574	657	2682
	Ratio S/R	0.69	0.69	0.69	0.67
	Chao1	988.3	7435.5	2123.2	7563.8
	Shannon-Wiever	4.97	6.48	5.97	6.6
	Pielou's evenness	0.84	0.79	0.87	0.8
	Simpson (1-D)	0.97	0.97	0.99	0.98
	Estimated genomes	1.6×10^9	1.3×10^7	4.3×10^8	1.2×10^7
	Sequence reads	17,660	24,083	15,818	29,424
	Coverage (x)	0.000012	0.0019	0.000037	0.0025
spo0A gene data	Richness	NA	406	NA	734
	Singletons	NA	261	NA	419
	Ratio S/R	NA	0.64	NA	0.57
	Shannon-Wiever	NA	5.36	NA	5.86
	Pielou's evenness	NA	0.89	NA	0.89
	Simpson (1-D)	NA	0.99	NA	0.99

was calculated when 16S rRNA gene sequences were analysed based on clustering with a lower identity level (90%; Table 3.8).

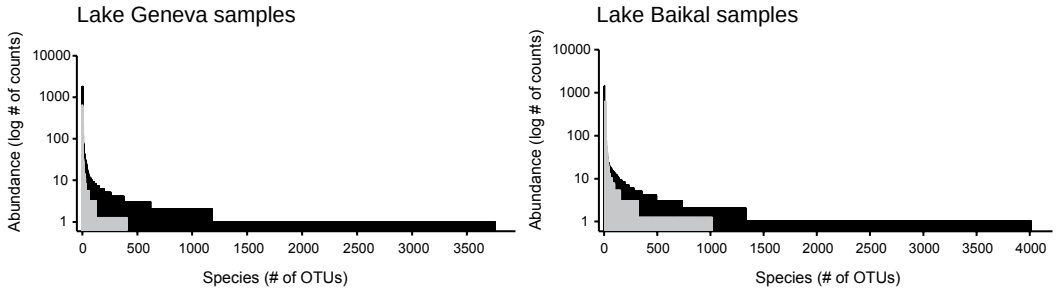


Figure 3.5: Species abundance curves for the community affiliated to the phylum Firmicutes (based on 16S rRNA gene data). The curves show the number of operational taxonomic units (OTUs) at their respective abundance (number of sequence counts), determined with the global approach (untreated, light gray bars) and the targeted approach (treated, dark gray bars).

3.4 Discussion

In this chapter a new approach of targeted metagenomics is described based on sporulation as a physiological and functional trait. A treatment is described to

destroy vegetative cells and isolate the endospore fraction of the community before subjecting this fraction to metagenomic sequencing. This method provides a diversity analysis of endospore-forming bacteria in high resolution.

In the past, pasteurization and density gradient centrifugation have been proposed as methods to enrich endospore-formers (Nicholson and Law 1999). Our preliminary tests however, showed that after density gradient centrifugation endospores were still detectable in the bulk sediment samples. Also, due to physiological differences and different densities, this approach is not appropriate as a general method to isolate endospores from environmental samples. In contrast, with the cell disruption treatment described in this chapter, the endospore-forming community was successfully targeted and the community structure revealed in detail.

Endospore-forming bacteria are easily dispersed (Driks 2002; Onyenwoke et al. 2004; Yudkin and Clarkson 2005) and have therefore been found in a wide range of environments (Nicholson 2002; Nicholson et al. 2000). Endospore-formers are also relevant in many domains related to human life in positive (e.g. biocontrol; McSpadden Gardener 2004, industrial production; Logan 2012) or negative ways (e.g. pathogens; Logan 2012 and food-spoilage; Brown 2000). Some endospore-forming species have extraordinary metabolisms such as CO oxidation (Wu et al. 2005) and metal reduction or oxidation (Junier et al. 2009; Francis and Tebo 1999; Lee and Tebo 1994). Endospore-formers have been shown to reflect past lake conditions (Renberg and Nilsson 1992), and they have been detected as the dominant bacterial life forms in deep sediments (Rothfuss et al. 1997; Lomstein et al. 2012). To date, Firmicutes (the phylum housing all known endospore-forming bacteria) is the second most represented bacterial phylum in culture collections (Klenk and Goker 2010) but to determine the environmental diversity of endospore-forming bacteria has remained challenging.

The results of the targeted metagenomics approach described here contribute significant knowledge about endospore-forming bacterial diversity in lake sediments. Although the diversity of endospore-forming bacteria detected with the global approach was substantial (Simpson indices were close to 1 in all cases), the richness was significantly increased with the targeted approach. With the targeted metagenomics approach described here, more than 90% of the community members were affiliated to the phylum Firmicutes. The improved detection of species with the targeted method was in the range of 1.4 to 15%, and the sequencing coverage was increased by two orders of magnitude. Due to higher sequencing coverage, the resolution of diversity detection was improved. This is a considerable advantage for the endospore-forming community, which has a large metabolic diversity and a versatile ecology. Representatives from all three known classes of Firmicutes were detected, and in some cases, detection down to a species level was possible, providing information about the metabolic capabilities of some bacteria in the sediment. Examples of abundant members detected here are *Symbiobacterium thermophilum*, the first known high GC-content endospore-forming bacterium (Ueda et al. 2004); *Paenibacillus chitinolyticus*, a chitin-degrader (Kuroshima et al. 1996); *Clostridium stercorarium*, a cellulolytic and thermophilic anaerobe (Madden 1983); or *Clostridium ljungdahlii*, an acetogenic bacterium previously isolated from chicken

yard waste (Tanner et al. 1993). Also, we detected members from *Alicyclobacillaceae*, an endospore-forming family with a poorly resolved taxonomy (Klenk et al. 2011).

Interestingly, a large number of OTUs were affiliated to *C. bowmanii* particularly in Lake Baikal but also in Lake Geneva, although to a lesser extent. The type strain of this psychrophilic species was isolated from a microbial mat in Lake Fryxell in Antarctica (Spring et al. 2003). Likewise abundant species affiliated to Peptostreptococcaceae, were detected in Lake Geneva. This family has been associated with human microbiota (Ezaki et al. 2006) and possibly reflects human impact on the microbial communities of Lake Geneva. In Laka Baikal Peptospreptococcaceae are not abundant, in accordance with lower anthropogenic impact on this lake.

The targeted approach based on 16S rRNA sequencing revealed a series of supposedly asporogenic species, including *Ethanoligenens harbinense*, a non-spore-forming member of obligate anaerobic Clostridia (Xing et al. 2006). Also members of the family Coriabacteriaceae (Clavel et al. 2010) were detected, which comprises a dozen genera, four of which have been described as non-spore-formers (Maruo et al. 2008; Minamida et al. 2008; Clavel et al. 2009), as well as *Trichococcus*, also reported as a non-spore-former (Liu et al. 2002). In past studies, some supposedly asporogenic Firmicutes have been discovered to form spores only after their full genome was sequenced (e.g.*Carboxydothermus*; Wu et al. 2005). The detection of these groups after the treatment suggest that these are essentially endospore-formers, but sporulation has not been confirmed in culture.

Well-known problems in sequencing and sequencing analysis were also observed in the results here. For example, there are significant differences in the community structure derived from the 16S rRNA gene and the functional *spo0A* gene. These differences could originate from amplification and sequencing biases, or due to different sequence curation (quality filtering). Most likely the method for assigning phylogenetic affiliation has the largest influence. The method of classification can be an issue as well as the strong dependency on the reference datasets.

Constantly novel and improved sequencing technologies are available and sequencing prices have dropped considerably. In consequence large sequencing projects are becoming very common (e.g. TerraGenome (Vogel et al. 2009) or the human microbiome project (Turnbaugh et al. 2007)). However, questions that remain open are: how well can diversity be assessed? And what conclusions can be drawn from diverstiy inventories? The results here, in particular the comparison between global and targeted approach suggest that one can still expect a progression in the number of rare species detected when sequence coverage is increased. This means that if another 10-fold increase in coverage is considered (close to 40,000 sequences), an expected 26,000 additional OTUs could be obtained. Overall if the data for Firmicutes are extrapolated to the whole bacterial diversity, it results in an estimated number of 30,000 OTUs in Lake Geneva sediment and 13,000 OTUs in Lake Baikal without a full community coverage. Therefore, microbiologists are still facing a daunting task when assessing diversity and probably the focus is better placed on strategic groups as is suggested here for endospore-forming bacteria.

3.5 Supplementary Information

Detection of growth by MPN method in microplates

The effectiveness of the cell disruption treatment was measured by re-growth of the treated cultures. This growth was detected using absorbance measurements (600 nm wavelength) in microplates, a simple and efficient way to measure bacterial growth. This method is described in Schrader et al. (1997).

DNase treatment tests

On the sediment samples, the free DNA was digested with DNase (New England Biolabs) after the treatment and before the DNA extraction. This DNase treatment was first tested on a DNA sample. After digestion, the DNA was amplified using the 16S rRNA primers described in materials and methods section of this chapter. The image of the gel electrophoresis shows samples after the PCR on lane 1 (original DNA sample) and on lane 2 the PCR of digested DNA (Figure 3.5). The results confirm that DNase successfully breaks down the DNA into tri-nucleotide pieces, which cannot be amplified.

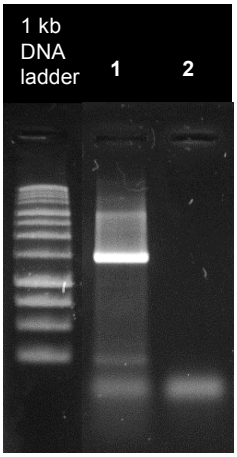


Figure 3.6: Gel-electrophoresis image of PCR amplified 16SrRNA gene of an original DNA sample (1) and a DNase digested sample (2).

rpoB quantitative PCR

For the quantification of a universal bacterial single-copy gene the *rpoB* gene, coding the RNA polymerase beta subunit. The rpoB primer sequences were obtained from Dahllöf et al. 2000. They were then tested on strains of endospore-forming bacteria (*Desulfotomaculum reducens*, *Paenibacillus alvei*, *Bacillus tusciae*, *Bacillus subtilis*, *Clostridium* sp.), two non spore-forming bacteria (*Escherichia coli*, *L. lactis* subsp. *lactis*) and the two sediment samples from Lake Geneva and Lake Baikal. The primers amplify a fragment of 360 bp length. PCR reactions of 50

µl volume contained 1 u Taq Polymerase (New England Biolabs), 1 x buffer, 0.2 mM dNTPs, 0.2 µM of each primer, PCR-grade water and a 10 x dilution of DNA sample (all previously confirmed to amplify with using the 16S rRNA gene primers). The PCR reaction was composed of 30 cycles of denaturation at 94 ℃ for 30 sec, annealing at 55 ℃ for 30 sec and elongation at 68 ℃ for 30 sec. Successful amplification of the correct fragment size was verified by gel electrophoresis (3.7). The image confirms successful amplification of all strains, even though at different identities and for some strains, there were unspecific bands (*P. alvei* (lane 2), *B. tusciae* (3) and *L. lactis* subsp. *lactis* (8). Also there is a low amplification of the blank control reaction, albeit at very low intensity. No unspecific bands are visible in the environmental samples (lane 4 and 5).

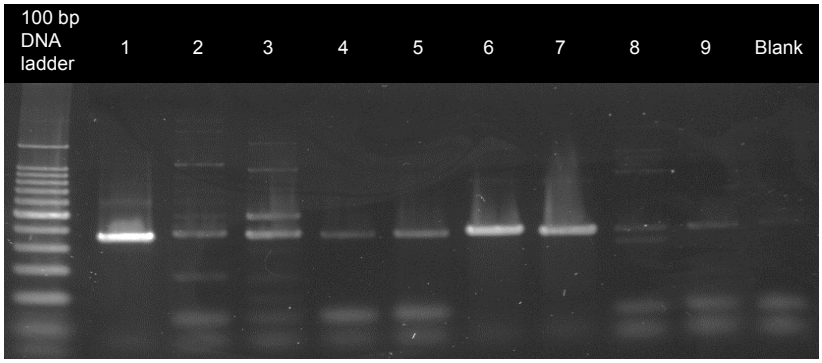


Figure 3.7: Gel-electrophoresis image of amplification test of *rpoB* primers. Samples were 1: *Desulfotomaculum reducens*, 2: *Paenibacillus alvei*, 3: *Bacillus tusciae*, 4: Lake Geneva sediment, 5: Lake Baikal sediment, 6: *Bacillus subtilis*, 7: *Escherichia coli*, 8: *Lacotcoccus lactis* subps. *lactis*, 9: *Clostridium* sp. and 10: Blank (no template).

For quantification, a *rpoB* standard series was procuded. To do this, the *rpoB* amplification products of *B. subtilis* and *E. coli* were purified with MultiScreen Millipore 96-well plate and then inserted into plasmids and cloned into competent cells with Zero Blunt TOPO PCR cloning kit (Invitrogen). Cloned plasmids were extracted using Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI, USA) and linearized with *Ncol* restriction enzyme (Promega). Linearized plasmids were loaded onto 1% agarose gel and run at 80 V for 30 min to verify the correct size of 4300 bp. Plasmid concentration was then measured and copy numbers calculated. Two individual standard series of dilutions from 3,000 to 30,000,000 copies/µl were produced with plasmids housing the *rpoB* gene fragment from *B. subtilis* and from *E. coli*. Quantitative PCR was then run with the standard series from both strains (3.8). The standard series of the plasmid with *rpoB* from *B. subtilis* is more accurate, with a better R^2 value.

Growth curve of treated samples

After the treatment, a sub-sample of Lake Geneva and Lake Baikal was incubated in nutrient broth (NB) at 30 ℃ and growth was measured by optical density (600

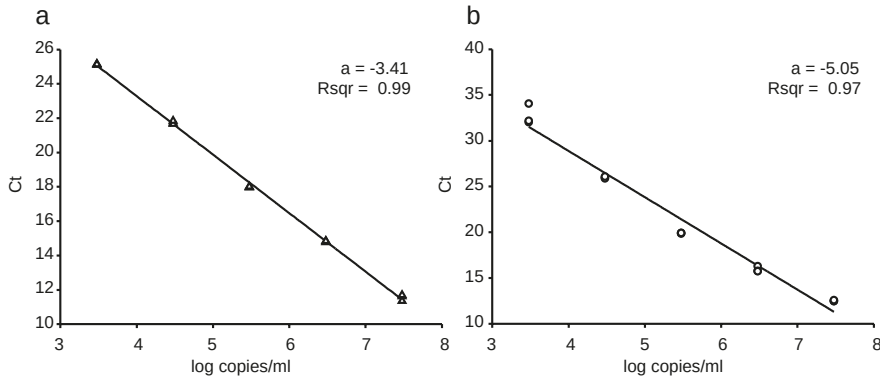


Figure 3.8: Quantitative PCR standard series of *rpoB* copies from *Bacillus subtilis* (a) and *Escherichia coli* (b).

nm)(Figure 3.9). The sample of Lake Geneva showed growth after 20 h of incubation, while samples from Lake Baikal did not grow until 25 h of incubation.

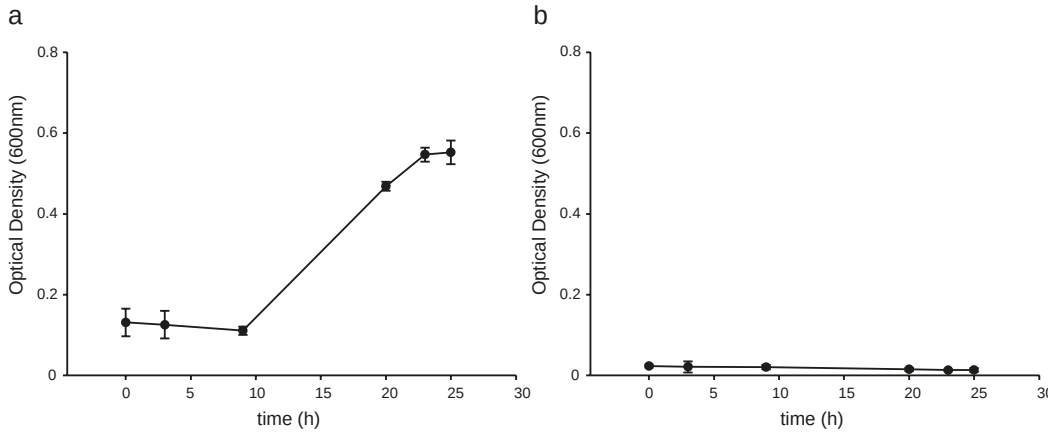


Figure 3.9: Growth curve of sediment samples after treatment for destruction of vegetative cells. Growth measured by optical density at 600 nm. (a) Lake Geneva and (b) Lake Baikal.

Isolation

The sediments from Lake Geneva and Lake Baikal were enriched in liquid Schaeffer sporulation medium (Schaeffer et al. 1965) for 14 days at room temperature to enrich for endospore-forming bacteria. The enriched samples were then plated onto NB agar plates and strains isolated by colony picking and dilution streaking. Once the strains were isolated, the agar plates were left standing for 20 days at room temperature and afterwards checked for endospores (Table 3.4). The isolates, where sporulation was confirmed were grown in liquid medium and their DNA was extracted using the MP FastDNA Spin Kit (MP Biomedicals) for soil,

following the manufacturer's instruction. The 16S rRNA gene and the *spo0A* gene were amplified and sent to Microsynth (Balgach, Switzerland) for Sanger sequencing. The closest relative was then determined using BLAST or BLASTx search (`http://blast.ncbi.nlm.nih.gov/Blast.cgi`).

Table 3.4: Summary of isolation of strains in sediment samples from Lake Geneva and Lake Baikal. Endospore presence or absence was determined with phase-contrast microscopy.

Sample	Sediment depth (cm)	Nr of isolates with endospores	Nr of isolates without endospores
GeC1-7	12-14	1	8
GeC1-8	14-16	2	2
GeC13-1	0-1	1	5
GeC1-28	54-56	0	3
GeC1-52	102-103	1	4
BkC1-2	12-17	8	0
BkC1-4	17-22	4	0

Table 3.5: Description of endospores and identification of isolates by Sanger sequencing and BLAST determination of closest relative based on partial sequences of 16S rRNA and the *spo0A* gene. E (endospores), L (length of sequence in nucleotides), I (identity %). ND: not determined.

			spo0A gene			16S rRNA gene		
Isolate	E	Description	Closest relative	L (nt)	I (%)	Closest relative	L (nt)	I (%)
1/4	y	central, oval shaped	*Bacillus cereus*	502	96	*Bacillus weihenstephanensis*	1039	99
7/1	n		ND			*Stenotrophomonas* sp.	72	94
7/5	y	terminal, free spores	*Paenibacillus* sp.	559	96	*Paenibacillus lautus*	893	98
7/6	n		ND			*Comamonas* sp.	965	99
7/9	n		ND			*Comamonas testosteroni*	1082	99
8/1	y	central	ND			*Bacillus cereus*	1083	99
8/2	y	terminal, round spore	*Lysinibacillus sphaericus*	158	87	*Lysinibacillus fusiformus*	1055	99
28/2	n		ND			*Pseudomonas aeruginosa* PAO581 genome	89	94
28/3	n		ND			*Delftia* sp.	1008	99
52/1	n		ND			*Bacillus weihenstephanensis*	1059	99
52/2	n		ND			*Stenotrophomonas maltophilia*	977	99
52/4	n		ND			*Pseudomonas* sp.	751	99
52/5	y	terminal, round spore	*Lysinibacillus fusiformus*	370	96	*Lysinibacillus* sp.	961	98
2/1	y	central, bacillus shaped	ND			ND		
2/2	y	sub-terminal, bacillus shaped	ND			ND		
2/3	y	sub-terminal, round	ND			ND		
2/4	y	sub-terminal, round	ND			ND		
2/5	y	sub-terminal, bacillus shaped	ND			ND		
2/6	y	bacillus shaped	ND			ND		
2/7	y	bacillus shaped	ND			ND		
2/8	y	terminal, bacillus shaped	ND			ND		
4/1	y	sub-terminal, bacillus shaped	ND			ND		
4/2	y	small, bacillus shaped	ND			ND		
4/3	y	central, bacillus shaped	ND			ND		
4/4	m	maybe spores, not round, dark	ND			ND		

Sequence data

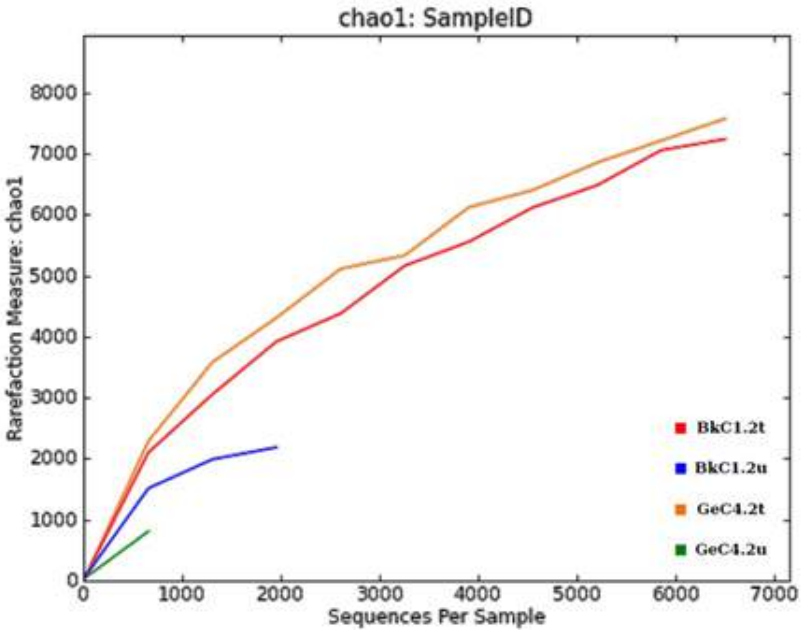


Figure 3.10: Chao rarefaction curves of species richness for sequences clustered in the phylum Firmicutes. Green line (Lake Geneva, global approach), orange line (Lake Geneva targeted approach), blue line (Lake Baikal global approach) and red line (Lake Baikal targeted approach).

Table 3.6: Detailed diversity of all bacteria, based on 16S rRNA sequencing. (u) are the untreated samples, and (t) are the treated samples. Values are given in sequence reads and percentage (in brackets). Firmicutes are in bold.

Phylum	Total Nr of OTUs	Lake Geneva		Lake Baikal	
		untreated	treated	untreated	treated
AC1	15	2 (0.02)	-	20 (0.3)	1 (0.01)
Acidobacteria	284	466 (5.4)	6 (0.05)	286 (4)	34 (0.2)
Actinobacteria	356	239 (2.8)	111 (0.9)	336 (4.7)	213 (1.5)
Armatimonadetes	65	54 (0.6)	2 (0.02)	24 (0.3)	7 (0.05)
Bacteria unclassified	1685	792 (9.2)	298 (2.5)	762 (10.7)	399 (2.8)
Bacteroideteds	62	63 (0.7)	5 (0.04)	37 (0.5)	1 (0.01)
BRC1	2	2 (0.02)	-	-	-
Caldiserica	5	5 (0.06)	-	-	-
Chlorobi	40	46 (0.5)	-	29 (0.4)	-
Chloroflexi	255	258 (3)	54 (0.05)	226 (3.2)	44 (0.3)
Cyanobacteria	32	57 (0.7)	9 (0.08)	11 (0.2)	4 (0.03)
Elusimicrobia	3	-	-	5 (0.07)	-
Fibrobacteres	6	3 (0.03)	-	7 (0.1)	-
Firmicutes	**8502**	**863 (10)**	**10806 (92.3)**	**2194 (30.7)**	**12972 (90.5)**
GAL15	7	-	-	-	1 (0.01)
Gemmatimonadetes	25	45 (0.5)	-	66 (0.9)	16 (0.1)
GN02	2	1 (0.01)	-	1 (0.01)	-
GN04	3	3 (0.03)	-	-	-
GOUTA4	2	-	-	9 (0.13)	-
Hyd24-12	2	5 (0.06)	-	-	-
LCP-89	2	1 (0.01)	-	1 (0.01)	-
Lentisphaerae	3	1 (0.01)	-	2 (0.03)	-
MVP-21	1	-	-	2 (0.03)	-
NC10	23	2 (0.02)	-	35 (0.5)	6 (0.04)
Nitrosphirae	72	121 (1.4)	1 (0.01)	127 (1.8)	36 (0.3)
NKB19	5	2 (0.02)	-	3 (0.04)	-
OC31	1	-	-	1 (0.01)	-
OD1	9	14 (0.2)	-	13 (0.2)	-
OP1	1	2 (0.02)	-	-	-
OP11	9	6 (0.07)	-	8 (0.1)	-
OP3	27	-	-	67 (0.9)	3 (0.02)
OP8	30	21 (0.2)	-	28 (0.4)	4 (0.03)
OP9	1	2 (0.02)	-	-	-
Planctomycetes	250	414 (4.8)	7 (0.06)	147 (2.1)	29 (0.2)
Proteobacteria	3230	4938 (57.1)	395 (3.4)	2590 (36.3)	545 (3.8)
Alph-proteobacteria	632	1147 (13.3)	104 (0.9)	578 (8.1)	244 (1.7)
Beta-proteobacteria	1216	1453 (16.8)	44 (0.4)	1500 (21)	223 (1.6)
Delta-proteobacteria	435	781 (9)	185 (1.6)	266 (3.7)	49 (0.3)
Epsilon-proteobacteria	67	256 (3)	3 (0.03)	78 (1.1)	6 (0.04)
Gamma-proteobacteria	424	861 (10)	40 (0.3)	48 (0.7)	5 (0.03)
SBR1093	7	-	-	8 (0.1)	1 (0.01)
Spirochaetes	10	8 (0.09)	-	4 (0.06)	3 (0.02)
TM6	54	69 (0.8)	2 (0.02)	23 (0.3)	3 (0.02)
TM7	1	1 (0.01)	-	-	-
Verrucomicrobia	81	111 (1.3)	6 (0.05)	41 (0.6)	3 (0.02)
WS1	9	6 (0.07)	-	8 (0.11)	1 (0.01)

Table 3.6: continued

Phylum	Total Nr of OTUs	Lake Geneva		Lake Baikal	
WS2	8	13 (0.2)	-	5 (0.07)	-
WS3	10	6 (0.07)	-	9 (0.1)	1 (0.01)
WS4	1	2 (0.02)	-	-	-
WYO	2	3 (0.03)	-	-	-
ZB3	2	-	-	2 (0.03)	-
Total Nr of sequences		8647	11702	7137	14327

Table 3.7: Detailed diversity of Firmicutes determined by 16S rRNA sequencing. (global) are the untreated samples, (targeted) are the treated samples. Values are given in sequence reads and percentage (in brackets).

	Genus	Lake Geneva		Lake Baikal	
		global	targeted	global	targeted
Firmicutes		19 (2.2)	738 (6.8)	70 (3.2)	1029 (7.9)
Bacilli		11 (1.3)	235 (2.2)	31 (1.4)	111 (0.9)
Bacillales		18 (2.0)	422 (3.9)	96 (4.4)	585 (4.5)
Alicyclobacillaceae		-	2 (0.02)	21 (1.0)	15 (0.1)
	Alicyclobacillus	-	5 (0.05)	3 (0.1)	38 (0.3)
Bacillaceae		4 (0.5)	164 (1.5)	38 (1.7)	238 (1.8)
	Anaerobacillus	-	-	-	2 (0.02)
	Anoxybacillus	-	12 (0.1)	1 (0.05)	13 (0.1)
	Bacillus	55 (6.4)	1056 (9.8)	294 (13.4)	2840 (21.9)
	Geobacillus	-	1 (0.01)	1 (0.05)	15 (0.1)
	Marinibacillus	-	-	3 (0.1)	2 (0.02)
	Terribacillus	1 (0.1)	-	-	1 (0.01)
	Virgibacillus	-	1 (0.01)	-	-
Paenibacillaceae		3 (0.4)	111 (1.0)	8 (0.4)	32 (0.3)
	Cohnella	4 (0.5)	10 (0.09)	8 (0.4)	16 (0.1)
	Paenibacillus	19 (2.2)	247 (2.3)	257 (11.7)	324 (2.5)
	Thermobacillus	-	4 (0.04)	-	-
Planococcaceae		13 (1.5)	99 (0.9)	12 (0.6)	86 (0.7)
	Kurthia	-	1 (0.01)	-	-
	Lysinibacillus	2 (0.2)	7 (0.06)	5 (0.2)	18 (0.1)
	Paenisporosarcina	3 (0.4)	48 (0.4)	3 (0.1)	12 (0.1)
	Rummeliibacillus	1 (0.1)	19 (0.2)	2 (0.09)	1 (0.01)
	Solibacillus	1 (0.1)	17 (0.2)	5 (0.2)	8 (0.06)
	Sporosarcina	4 (0.5)	72 (0.7)	8 (0.4)	74 (0.6)
	Ureibacillus	6 (0.7)	25 (0.2)	-	1 (0.01)
	Pullulanibacillus	-	-	-	2 (0.02)
	Sporolactobacillus	-	1 (0.01)	-	-
Thermoactinomycetaceae		-	2 (0.02)	2 (0.09)	1 (0.01)
	Laceyella	-	0	2 (0.09)	-
	Planifilum	-	1 (0.01)	1 (0.05)	-
	Shimazuella	-	1 (0.01)	-	-
	Thermoactinomyces	-	2 (0.02)	-	-
Lactobacillales		1 (0.1)	-	-	-
	Trichococcus	7 (0.8)	3 (0.03)	-	-
	Turicibacte	2 (0.2)	11 (0.1)	-	-
Clostridia		50 (5.8)	249 (2.3)	10 (0.5)	45 (0.4)

Table 3.7: continued

	Genus	Lake Geneva		Lake Baikal	
		global	targeted	global	targeted
Clostridiales		67 (7.8)	869 (8.0)	28 (1.3)	235 (1.8)
Catabacteriaceae		1 (0.1)	-	-	-
Clostridiaceae		10 (1.2)	181 (1.7)	12 (0.6)	141 (1.1)
	Caloramator	0	13 (0.1)	18 (0.8)	55 (0.4)
	Clostridium	226 (26.2)	2514 (23.3)	1094 (49.9)	6474 (49.9)
	Sarcina	-	9 (0.08)	-	-
	Sedimentibacter	-	2 (0.02)	-	-
	Tepidimicrobium	-	13 (0.1)	-	-
	D*ehalobacterium*	3 (0.4)	-	-	-
EtOH8		-	-	-	2 (0.02)
Eubacteriaceae		-	1 (0.01)	-	-
	Acetobacterium	1 (0.1)	-	-	-
Heliobacteriaceae		-	-	1 (0.05)	-
Lachnospiraceae		7 (0.8)	59 (0.6)	4 (0.2)	14 (0.1)
	Roseburia	-	1 (0.01)	-	-
Peptococcaceae		-	6 (0.06)	1 (0.05)	3 (0.02)
	Desulfitobacterium	-	1 (0.01)	1 (0.05)	-
	Desulfosporosinus	1 (0.1)	46 (0.4)	53 (2.4)	286 (2.2)
	Desulfotomaculum	-	-	-	2 (0.02)
	Pelotomaculum	2 (0.2)	9 (0.08)	-	-
	WCHB1-84	-	3 (0.03)	-	-
Peptostreptococcaceae		287 (33.3)	3280 (30.35)	14 (0.6)	53 (0.4)
	Clostridium	-	13 (0.1)	4 (0.2)	4 (0.03)
Ruminococcaceae		9 (1.0)	41 (0.4)	20 (0.9)	18 (0.1)
	Ethanoligenens	-	9 (0.1)	3 (0.1)	8 (0.06)
	Ruminococcus	4 (0.5)	12 (0.1)	-	-
Symbiobacteriaceae		-	-	-	1 (0.01)
	Symbiobacterium	1 (0.1)	6 (0.06)	24 (1.1)	16 (0.1)
Syntrophomonadaceae		1 (0.1)	2 (0.02)	-	-
Veillonellaceae		2 (0.2)	50 (0.5)	17 (0.8)	111 (0.9)
	Desulfosporomusa	-	5 (0.05)	1 (0.05)	6 (0.05)
	Thermosinus	-	3 (0.03)	-	-
	vadinHB04	3 (0.3)	-	-	-
Coriobacteriaceae		10 (1.2)	-	9 (0.4)	3 (0.02)
Halanaerobiaceae		-	4 (0.04)	-	-
MBA08		-	8 (0.07)	-	-
Anaerobrancaceae		-	4 (0.04)	-	-
	A55-D21	-	1 (0.01)	-	-
	Dethiobacter	-	1 (0.01)	-	1 (0.01)
OPB54		1 (0.1)	65 (0.6)	8 (0.4)	10 (0.08)
SHA-98		2 (0.2)	-	-	-
Thermoanaerobacteraceae		-	4 (0.04)	-	7 (0.05)
Erysipelotrichi	PSB-M-3	1 (0.1)	-	-	-

Table 3.7: continued

	Genus	Lake Geneva		Lake Baikal	
		global	targeted	global	targeted
Negativicutes	BSV43	-	5 (0.05)	1 (0.05)	13 (0.1)
Total		863	10,806	2,194	12,972

Table 3.8: List of groups that newly appeared when applying the treatment. These results are based on the 16S rRNA sequencing data. Values correspond to sequence reads. (u) untreated, (t) treated, (diff.) difference.

	Lake Geneva			Lake Baikal		
	(u)	(t)	diff.	(u)	(t)	diff.
Firmicutes unclassified	19	738	719	70	1029	959
Bacilli	11	235	224	31	111	80
Bacillales	18	422	404	96	585	489
Alicyclobacillaceae	0	2	2	-	-	-
Alicyclobacillus	0	5	5	3	33	30
Alicyclobacillus pohliae	-	-	-	0	5	5
Bacillaceae	4	164	160	38	238	200
Anaerobacillus	-	-	-	0	2	2
Anoxybacillus	0	12	12	1	13	12
Bacillus	19	448	429	111	974	863
Bacillus acidicola	0	1	1	0	2	2
Bacillus asahii	2	9	7	6	11	5
Bacillus badius	0	3	3	1	14	13
Bacillus flexus	0	20	20	3	15	12
Bacillus foraminis	5	7	2	2	11	9
Bacillus horikoshii	0	1	1	-	-	-
Bacillus longiquaesitum	22	505	483	160	1732	1572
Bacillus muralis	2	38	36	10	64	54
Bacillus oleronius	-	-	-	0	1	1
Bacillus safensis	4	18	14	0	2	2
Bacillus thermoamylovorans	0	5	5	0	5	5
Bacillus trypoxylicola	-	-	-	1	9	8
Ureibacillus	6	25	19	0	1	1
Terribacillus	-	-	-	0	1	1
Virgibacillus	0	1	1	-	-	-
Geobacillus	-	-	-	1	15	14
Geobacillus thermodenitrificans	0	1	1	-	-	-
Paenibacillaceae	3	111	108	8	32	24
Thermobacillus	0	4	4	-	-	-
Cohnella	4	10	6	8	16	8
Paenibacillus	16	193	177	244	302	58
Paenibacillus alvei	-	-	-	0	1	1
Paenibacillus amylolyticus	0	3	3	1	5	4
Paenibacillus barengoltzii	0	1	1	-	-	-
Paenibacillus chondroitinus	2	38	36	2	11	9
Paenibacillus illinoisensis	0	3	3	-	-	-
Paenibacillus macerans	0	2	2	-	-	-
Paenibacillus mucilaginosus	0	3	3	0	1	1
Paenibacillus stellifer	0	4	4	0	2	2
Planococcaceae	13	99	86	12	86	74
Kurthia	0	1	1	-	-	-
Lysinibacillus	2	7	5	5	18	13
Paenisporosarcina	3	48	45	3	12	9
Rummeliibacillus	1	19	18	-	-	-
Solibacillus	1	17	16	5	8	3
Sporosarcina	4	68	64	7	66	59
Sporosarcina aquimarina	0	4	4	1	8	7
Pullulanibacillus	-	-	-	0	2	2
Sporolactobacillus	0	1	1	-	-	-

Table 3.8: continued

	Lake Geneva			Lake Baikal		
	(u)	(t)	diff.	(u)	(t)	diff.
Thermoactinomycetaceae	0	2	2	0	7	7
Planifilum	0	1	1	-	-	-
Shimazuella	0	1	1	-	-	-
Thermoactinomyces	0	2	2	-	-	-
Turicibacter	2	11	9	-	-	-
Clostridia	50	249	199	10	45	35
Clostridiales unclassified	67	869	802	28	235	207
Clostridiaceae	10	181	171	12	141	129
Caloramator	0	13	13	18	55	37
Clostridium	190	1977	1787	523	2976	2453
Clostridium	0	13	13	-	-	-
Clostridium acetobutylicum	-	-	-	0	1	1
Clostridium bowmanii	26	302	276	548	3440	2892
Clostridium butyricum	3	194	191	7	43	36
Clostridium cellulovorans	0	1	1	-	-	-
Clostridium intestinale	0	1	1	0	3	3
Clostridium ljungdahlii	1	16	15	-	-	-
Clostridium neonatale	0	1	1	-	-	-
Clostridium perfringens	0	21	21	-	-	-
Clostridium thermopalmarium	0	1	1	-	-	-
Sarcina	0	9	9	-	-	-
Sedimentibacter	0	2	2	-	-	-
Tepidimicrobium	0	13	13	-	-	-
EtOH8	-	-	-	0	2	2
Eubacteriaceae	0	1	1	-	-	-
Lachnospiraceae	7	59	52	4	14	10
Roseburia faecis	0	1	1	-	-	-
Peptococcaceae	0	6	6	1	3	2
Peptococcaceae WCHB1-84	0	3	3	-	-	-
Desulfotomaculum	-	-	-	0	1	1
Desulfotomaculum aeronauticum	-	-	-	0	1	1
Desulfosporosinus	0	20	20	20	144	124
Desulfosporosinus meridiei	1	26	25	33	142	109
Desulfitobacterium	0	1	1	-	-	-
Pelotomaculum	2	9	7	-	-	-
Peptostreptococcaceae	287	3280	2993	14	53	39
Ruminococcaceae	9	41	32	-	-	-
Ethanoligenens	0	7	7	0	8	8
Ethanoligenens harbinense	0	2	2	-	-	-
Ruminococcus	4	12	8	-	-	-
Symbiobacteriaceae	-	-	-	0	1	1
Symbiobacterium	0	1	1	-	-	-
Symbiobacterium thermophilum	1	5	4	1	5	4
Syntrophomonadaceae	1	2	1	-	-	-
Veillonellaceae	2	50	48	17	111	94
Desulfosporomusa	0	5	5	1	3	2
Desulfosporomusa polytropa	-	-	-	0	3	3
Thermosinus	0	3	3	-	-	-
Halanaerobiaceae	0	4	4	-	-	-
MBA08	0	8	8	-	-	-
Anaerobrancaceae	0	4	4	-	-	-

Table 3.8: continued

	Lake Geneva			Lake Baikal		
	(u)	(t)	diff.	(u)	(t)	diff.
Anaerobrancaceae A55-D21	0	1	1	-	-	-
Anaerobrancaceae	-	-	-	0	1	1
Dethiobacter	0	1	1	-	-	-
OPB54	1	65	64	8	10	2
Thermoanaerobacteraceae	0	4	4	-	-	-
BSV43	0	5	5	1	13	12
Total counts	825	10802	9977	2076	12904	10828

Table 3.9: List of groups that dissapearad when applying the treatment. These results are based on the 16S rRNA sequencing data. Values correspond to sequence reads. (u) untreated, (t) treated, (diff.) difference.

	Lake Geneva			Lake Baikal		
	(u)	(t)	diff.	(u)	(t)	diff.
Alicyclobacillaceae				21	15	-6
Marinibacillus				3	2	-1
Terribacillus	1	0	-1			
Paenibacillus barengoltzii				2	0	-2
Paenibacillus chitinolyticus	1	0	-1	7	2	-5
Paenibacillus illinoisensis				1	0	-1
Rummeliibacillus				2	1	-1
Thermoactinomycetaceae	-	-	-	2	1	-1
Laceyella				2	0	-2
Planifilum				1	0	-1
Lactobacillales	1	0	-1			
Trichococcus	7	3	-4			
Catabacteriaceae	1	0	-1			
Clostridium acetobutylicum	3	0	-3			
Clostridium cellulovorans				6	5	-1
Clostridium ljungdahlii				6	3	-3
Clostridium perfringens				4	3	-1
Clostridium stercorarium	3	0	-3			
Dehalobacterium	3	0	-3			
Acetobacterium	1	0	-1			
Heliobacteriaceae				1	0	-1
Desulfitobacterium				1	0	-1
Ruminococcaceae				20	18	-2
Ethanoligenens harbinense				3	0	-3
Symbiobacterium				23	11	-12
vadinHB04	3	0	-3			
Coriobacteriaceae	10	0	-10	9	3	-6
SHA-98	2	0	-2			
PSB-M-3	1	0	-1			
Sum of sequence counts			-34			-50

Table 3.10: List of genera that were found in both datasets from 16S rRNA and *spo0A* gene sequencing. Values correspond to number of sequence reads and percentage in brackets.

	16S rRNA dataset		*spo0A* dataset	
	Lake Geneva	Lake Baikal	Lake Geneva	Lake Baikal
Alicyclobacillus	5 (0.05)	38 (0.3)	2 (0.2)	0 (0)
Anoxybacillus	12 (0.1)	13 (0.1)	1 (0.1)	20 (1.0)
Bacillus	1056 (9.8)	2840 (21.9)	98 (10.3)	532 (25.5)
Geobacillus	1 (0.01)	15 (0.1)	33 (3.5)	96 (4.6)
Paenibacillus	247 (2.3)	324 (2.5)	493 (51.9)	551 (26.4)
Sporosarcina	72 (0.7)	74 (0.6)	7 (0.7)	23 (1.1)
Clostridium	2514 (23.3)	6474 (49.9)	85 (9.0)	73 (3.5)
Desulfotomaculum	0 (0)	2 (0.02)	10 (1.1)	37 (1.8)

Table 3.11: List of diversity determined by clustering at 90% sequence identity with 16S-rRNA gene data. (u) untreated, (t) treated. Values are given in number of sequence reads and percentage (in brackets).

	Lake Geneva		Lake Baikal	
	(u)	(t)	(u)	(t)
Bacteria	369 (24.3)	130 (12.0)	324 (26.3)	151 (12.69)
AC1	3 (0.2)	-	8 (0.7)	1 (0.1)
Acidobacteria	46 (3.0)	4 (0.4)	38 (3.1)	10 (0.9)
Actinobacteria	49 (3.2)	38 (3.5)	44 (3.6)	45 (3.9)
Armatimonadetes	19 (1.3)	2 (0.2)	13 (1.1)	6 (0.5)
Bacteroidetes	24 (1.6)	5 (0.5)	15 (1.2)	1 (0.1)
BRC1	2 (0.1)	-	-	-
Caldiserica	3 (0.2)	-	-	-
Chlorobi	10 (0.7)	-	12 (1.0)	-
Chloroflexi	59 (3.9)	17 (1.6)	48 (3.9)	17 (1.5)
Cyanobacteria	8 (0.5)	6 (0.6)	5 (0.4)	1 (0.1)
Elusimicrobia	-	-	3 (0.2)	-
Fibrobacteres	1 (0.07)	-	3 (0.2)	-
Firmicutes	164 (10.8)	776 (71.8)	304 (24.7)	775 (66.4)
GAL15	-	-	-	1 (0.1)
Gemmatimonadetes	4 (0.3)	-	3 (0.2)	3 (0.3)
GN02	1 (0.07)	-	1 (0.08)	-
GN04	2 (0.1)	-	-	-
GOUTA4	-	-	1 (0.08)	1 (0.1)
Hyd24-12	2 (0.1)	-	-	-
LCP-89	1 (0.07)	-	1 (0.08)	-
Lentisphaerae	1 (0.07)	-	1 (0.08)	-
MVP-21	-	-	1 (0.08)	-
NC10	1 (0.07)	-	5 (0.4)	3 (0.3)
Nitrospirae	9 (0.6)	1 (0.09)	14 (1.1)	11 (0.9)
NKB19	2 (0.1)	-	2 (0.2)	-
OC31	-	-	1 (0.08)	-
OD1	8 (0.5)	-	3 (0.2)	-
OP11	4 (0.3)	-	4 (0.3)	-
OP3	-	-	17 (1.4)	2 (0.2)
OP8	5 (0.3)	-	4 (0.3)	2 (0.2)
OP9	1 (0.07)	-	-	-
Planctomycetes	92 (6.1)	5 (0.5)	54 (4.4)	19 (1.6)
Proteobacteria	573 (37.7)	95 (8.9)	260 (21.1)	108 (9.3)
SBR1093	-	-	1 (0.08)	1 (0.1)
Spirochaetes	4 (0.3)	-	2 (0.2)	2 (0.2)
TM6	12 (0.8)	1 (0.09)	8 (0.7)	3 (0.3)
TM7	1 (0.07)	-	1 (0.08)	1 (0.1)
Verrucomicrobia	28 (1.8)	1 (0.09)	16 (1.3)	2 (0.2)
WS1	1 (0.07)	-	4 (0.3)	-
WS2	5 (0.3)	-	2 (0.2)	-
WS3	4 (0.3)	-	5 (0.4)	2 (0.2)
WS4	1 (0.07)	-	-	-
WYO	1 (0.07)	-	-	-
ZB3	1 (0.07)	-	2 (0.2)	-
Total	1521 (100)	1081 (100)	1230 (100)	1168 (100)

3.6 References

Adams, D. G. and Duggan, P. S. (1999). Tansley review no. 107. Heterocyst and akinete differentiation in cyanobacteria. *New Phytologist*, 144(1):3–33.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev*, 59(1):143–69.

Baker, B. J., Tyson, G. W., Webb, R. I., Flanagan, J., Hugenholtz, P., Allen, E. E., and Banfield, J. F. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. *Science*, 314(5807):1933–5.

Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., et al. (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, 289(5486):1902–1906.

Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., and DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, 411(6839):786–789.

Brown, K. L. (2000). Control of bacterial spores. *Br Med Bull*, 56(1):158–171.

Bueche, M., Wunderlin, T., Roussel-Delif, L., Junier, T., Sauvain, L., Jeanneret, N., and Junier, P. (2013). Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene *spo0A*. *Appl Eviron Microbiol*, 79(17):5302–12.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7(5):335–6.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand J Statistics*, 11(4):265–270.

Chater, K. F. and Chandra, G. (2006). The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev*, 30(5):651–72.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science*, 322(5899):275–278.

Clavel, T., Charrier, C., Braune, A., Wenning, M., Blaut, M., and Haller, D. (2009). Isolation of bacteria from the ileal mucosa of TNFdeltaARE mice and description of *Enterorhabdus mucosicola* gen. nov., sp. nov. *Int J Syst Evol Microbiol*, 59(Pt 7):1805–1812.

Clavel, T., Duck, W., Charrier, C., Wenning, M., Elson, C., and Haller, D. (2010). *Enterorhabdus caecimuris* sp. nov., a member of the family Coriobacteriaceae isolated from a mouse model of spontaneous colitis, and emended description of the genus *Enterorhabdus* (Clavel et al. 2009). *Int J Syst Evol Microbiol*, 60(Pt 7):1527–1531.

Dahllöf, I., Baillie, H., and Kjelleberg, S. (2000). rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol*, 66(8):3376–3380.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., et al. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol*, 77(4):1315–24.

Driks, A. (2002). Overview: development in bacteria: spore formation in *Bacillus subtilis*. *Cell Mol Life Sci*, 59(3):389–391.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–1.

Ensign, J. C. (1978). Formation, properties, and germination of actinomycete spores. *Annu Rev Microbiol*, 32(1):185–219.

Ezaki, T., Li, N., and Kawamura, Y. (2006). The anaerobic Gram-positive cocci. In Dr Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, pages 795–808. Springer US.

Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., et al. (2011). Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS ONE*, 6(4):e19223.

Francis, C. A. and Tebo, B. M. (1999). Marine *Bacillus* spores as catalysts for oxidative precipitation and sorption of metals. *J Mol Microbiol Biotechnol*, 1(1):71–78.

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., and Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol*, 14(11):2870–90.

Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, 309(5739):1387–1390.

Gilbert, J. A. and Dupont, C. L. (2011). Microbial metagenomics: beyond the genome. *Annu Rev Mar Sci*, 3:347–371.

Green, D. and Cutting, S. (1999). Sporulation in *Bacillus subtilis*. In Margaret CM Smith and R Elizabeth Sockett, editor, *Methods in Microbiology*, volume 29 of *Genetic Methods for Diverse Prokaryotes*, pages 467–486. Academic Press.

Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci USA*, 84(13):4355–4358.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*, 5(10):R245–9.

Harris, J. K., Caporaso, J. G., Walker, J. J., Spear, J. R., Gold, N. J., Robertson, C. E., et al. (2013). Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J*, 7(1):50–60.

Huber, J. A., Mark Welch, D. B., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., and Sogin, M. L. (2007). Microbial population structures in the deep marine biosphere. *Science*, 318(5847):97–100.

Iwai, S., Chai, B., Sul, W. J., Cole, J. R., Hashsham, S. A., and Tiedje, J. M. (2010). Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *ISME J*, 4(2):279–285.

Junier, P., Frutschi, M., Wigginton, N. S., Schofield, E. J., Bargar, J. R., and Bernier-Latmani, R. (2009). Metal reduction by spores of *Desulfotomaculum reducens*. *Environ Microbiol*, 11(12):3007–3017.

Klenk, H. P. and Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol*, 33(4):175–82.

Klenk, H.-P., Lapidus, A., Chertkov, O., Copeland, A., Del Rio, T. G., Nolan, M., et al. (2011). Complete genome sequence of the thermophilic, hydrogen-oxidizing *Bacillus tusciae* type strain (T2) and reclassification in the new genus, *Kyrpidia* gen. nov. as *Kyrpidia tusciae* comb. nov. and emendation of the family *Alicyclobacillaceae* (da Costa and Rainey, 2010). *Stand Genomic Sci*, 5(1):121–134.

Kuroshima, K.-I., Sakane, T., Takata, R., and Yokota, A. (1996). *Bacillus ehimensis* sp. nov. and *Bacillus chitinolyticus* sp. nov., new chitinolytic members of the genus *Bacillus*. *Int J Syst Bacteriol*, 46(1):76–80.

Lee, Y. and Tebo, B. M. (1994). Cobalt(II) oxidation by the marine manganese(II)-oxidizing *Bacillus* sp. strain SG-1. *Appl Environ Microbiol*, 60(8):2949–2957.

Liesack, W., Weyland, H., and Stackebrandt, E. (1991). Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microb Ecol*, 21(1):191–198.

Lin, W., Jogler, C., Schüler, D., and Pan, Y. (2011). Metagenomic analysis reveals unexpected subgenomic diversity of magnetotactic bacteria within the phylum Nitrospirae. *Appl Environ Microbiol*, 77(1):323–326.

Liu, J. R., Tanner, R. S., Schumann, P., Weiss, N., McKenzie, C. A., Janssen, P. H., et al. (2002). Emended description of the genus *Trichococcus*, description of *Trichococcus collinsii* sp. nov., and reclassification of *Lactosphaera pasteurii* as *Trichococcus pasteurii* comb. nov. and of *Ruminococcus palustris* as *Trichococcus palustris* comb. nov. in the low-G+C gram-positive bacteria. *Int J Syst Evol Microbiol*, 52(Pt 4):1113–26.

Logan, N. A. (2012). *Bacillus* and relatives in foodborne illness. *J Appl Microbiol*, 112(3):417–429.

Lomstein, B. A., Langerhuus, A. T., D'Hondt, S., Jorgensen, B. B., and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature*, 484(7392):101–4.

Madden, R. H. (1983). Isolation and characterization of *Clostridium stercorarium* sp. nov., cellulolytic thermophile. *Int J Syst Bacteriol*, 33(4):837–840.

Maruo, T., Sakamoto, M., Ito, C., Toda, T., and Benno, Y. (2008). *Adlercreutzia equolifaciens* gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended description of the genus *Eggerthella*. *Int J Syst Evol Microbiol*, 58(Pt 5):1221–1227.

McSpadden Gardener, B. B. (2004). Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. *Phytopathology*, 94(11):1252–1258.

Minamida, K., Ota, K., Nishimukai, M., Tanaka, M., Abe, A., Sone, T., et al. (2008). *Asaccharobacter celatus* gen. nov., sp. nov., isolated from rat caecum. *Int J Syst Evol Microbiol*, 58(Pt 5):1238–1240.

Neufeld, J. D., Chen, Y., Dumont, M. G., and Murrell, J. C. (2008). Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol*, 10(6):1526–1535.

Nicholson, W. L. (2002). Roles of *Bacillus* endospores in the environment. *Cell Mol Life Sci*, 59(3):410–416.

Nicholson, W. L. and Law, J. F. (1999). Method for purification of bacterial endospores from soils: UV resistance of natural Sonoran desert soil populations of *Bacillus* spp. with reference to *B. subtilis* strain 168. *J Microbiol Methods*, 35(1):13–21.

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., and 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, M., Oksanen, M., and Suggests, M. (2007). The vegan package.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., and Stahl, D. A. (1986). Microbial ecology and evolution: A ribosomal RNA approach. *Ann Rev Microbiol*, 40(1):337–365.

Onyenwoke, R. U., Brill, J. A., Farahi, K., and 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.

Ovreas, L., Forney, L., Daae, F. L., and 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.

Quince, C., Curtis, T. P., and Sloan, W. T. (2008). The rational exploration of microbial diversity. *ISME J*, 2(10):997–1006.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., et al. (2005). Community proteomics of a natural microbial biofilm. *Science*, 308(5730):1915–1920.

Renberg, I. and Nilsson, M. (1992). Dormant bacteria in lake sediments as palaeoecological indicators. *J Paleolimnol*, 7(2):127–135.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459):431–437.

Rothfuss, F., Bender, M., and Conrad, R. (1997). Survival and activity of bacteria in a deep, aged lake sediment (Lake Constance). *Microbiol Ecol*, 33(1):69–77.

Rowe, R., Todd, R., and Waide, J. (1977). Microtechnique for most-probable-number analysis. *Appl Environ Microbiol*, 33(3):675–680.

Schaeffer, P., Millet, J., and Aubert, J. P. (1965). Catabolic repression of bacterial sporulation. *Proc Natl Acad Sci USA*, 54(3):704–711.

Scholz, M. B., Lo, C. C., and Chain, P. S. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol*, 23(10):9–15.

Schrader, K. K., Regt, M. Q. D., Tucker, C. S., and Duke, S. O. (1997). A rapid bioassay for selective algicides. *Weed Technol*, 11(4):767–774.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA*, 103(32):12115–12120.

Spring, S., Merkhoffer, B., Weiss, N., Kroppenstedt, R. M., Hippe, H., and Stackebrandt, E. (2003). Characterization of novel psychrophilic clostridia from an antarctic microbial mat: description of *Clostridium frigoris* sp. nov., *Clostridium lacusfryxellense* sp. nov., *Clostridium bowmanii* sp. nov. and *Clostridium psychrophilum* sp. nov. and reclassification of *Clostridium laramiense* as *Clostridium estertheticum* subsp. *laramiense* subsp. nov. *Int J Syst Evol Microbiol*, 53(Pt 4):1019–1029.

Strauch, M. A. and Hoch, J. A. (1992). Sporulation in prokaryotes and lower eukaryotes. *Curr Opin Genet Dev*, 2(5):799–804.

Suenaga, H. (2012). Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol*, 14(1):13–22.

Summerer, D. (2009). Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics*, 94(6):363–368.

Tanner, R. S., Miller, L. M., and Yang, D. (1993). *Clostridium ljungdahlii* sp. nov., an acetogenic species in clostridial rRNA homology group I. *Int J Syst Bacteriol*, 43(2):232–236.

Thomas, S. H., Wagner, R. D., Arakaki, A. K., Skolnick, J., Kirby, J. R., Shimkets, L. J., et al. (2008). The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PloS One*, 3(5):e2103.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature*, 449(7164):804–810.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978):37–43.

Ueda, K., Yamashita, A., Ishikawa, J., Shimada, M., Watsuji, T.-o., Morimura, K., et al. (2004). Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res*, 32(16):4937–4944.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667):66–74.

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., van Elsas, J. D., et al. (2009). TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature*, 7(4):252–252.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E., and Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol*, 69(10):6235–6242.

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., et al. (2007). Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science*, 315(5815):1126–1130.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*, 73(16):5261–7.

Warnecke, F. and Hugenholtz, P. (2007). Building on basic metagenomics with complementary technologies. *Genome Biol*, 8(12):231.

Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., et al. (2005). Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet*, 1(5):e65.

Xing, D., Ren, N., Li, Q., Lin, M., Wang, A., and Zhao, L. (2006). *Ethanoligenens harbinense* gen. nov., sp. nov., isolated from molasses wastewater. *Int J Syst Evol Microbiol*, 56(Pt 4):755–760.

Xu, J. (2006). Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol*, 15(7):1713–1731.

Yudkin, M. D. and Clarkson, J. (2005). Differential gene expression in genetically identical sister cells: the initiation of sporulation in *Bacillus subtilis*. *Mol Microbiol*, 56(3):578–589.

Zarraonaindia, I., Smith, D. P., and Gilbert, J. A. (2013). Beyond the genome: community-level analysis of the microbial world. *Biol Philos*, 28(2):261–282.

4 Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva, (Switzerland-France)

This chapter is based on the accepted article:
Tina Wunderlin, Juan Pablo Corella, Thomas Junier, Matthieu Bueche, Jean-Luc Loizeau, Stéphanie Girardclos and Pilar Junier (*in press*). Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva (Switzerland-France). *Aquatic Sciences*.

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 analysed at a 0.2 cm resolution while DNA extracts, elemental geochemistry and grain size were obtained every four 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 and the phylogenetic affiliation of the endospore-formers. The results showed that despite the fact that the quantity of extracted DNA decreased with 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 the community composition was linked to the dominance of anaerobic Clostridia-like endospore-formers. This trend has reversed in the last ten 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 successfully introduce a new approach for paleoecology that could also be applied to ocean sediments and long sediment cores.

4.1 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 climatic 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 weathering. Many of the formers are direct consequences of human activities.

In paleolimnology, changes in community composition of organisms producing identifiable fossilized remnants can be analysed on different time-scales 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 stable 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 bi-

ological indicators for paleoecological studies (Gorham et al. 2001). In order 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 previously (Renberg and Nilsson 1992). The authors concluded that it is possible to isolate dormant bacteria (*Thermoactinomyces* 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 archaeological excavations containing plant debris deposited between 85-125 AD (Seaward et al. 1976). Spores are commonly found in sediments and isolation and enumeration 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 (Bartholomew 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 13000 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 exponentially 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; de 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 towards 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 has been quantified to account for endospores in sediment of the North Sea, where endospores have been found to make up to 3 % of the total community of Bacteria and Archaea (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 towards eutrophication in the late 1960s. Measures to reduce the phosphate input from early 70s on have been effective and have decreased the nutrient level by 2012 to nearly pre-1960 values (Lazzarotto et al. 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 the 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 are 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.

4.2 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 4.7) on the eastern side of the Rhone delta in Lake Geneva (Switzerland), which are inactive Rhone canyons since the river was channelled in 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 ℃.

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 in two lengthwise halves to proceed with the sedimentological description. Pictures were obtained using a digital camera and controlled light conditions. X-ray fluorescence was analysed in CAN02 core using an AVAATECH XRF core scanner (2000 A, 10kV and 30kV) every 2 mm at the University of Barcelona. CAN01 sediment core was sampled every 4 cm for total carbon (TC) and total nitrogen (TN) using an elemental analyser (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 (TOC) 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 analyser (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 fall-out 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 a cumulated sediment mass scale, because porosity didn't 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 *Bacillus subtilis*. Endospore preparations of 99% spores were subjected to successive DNA extractions following protocols for three commercially available kits. Between each extraction the freed DNA was separated from the remaining cell pellet by centrifugation for 5 min at 14,000 x g. Sediment DNA extractions were finally done 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/sec with the TissueLyser LT (QIAGEN, Hilden, Germany) for 10 min. The sample was then centrifuged for 5 min at 14,000 x 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-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. DNA quality was also verified by agarose gel electrophoresis.

Quantification of total bacteria

Quantification of bacterial DNA in sediment extracts was performed by real-time quantitative PCR of the V3 region of the 16S rRNA gene with primers 338f and 520r (Ovreas 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 program: enzyme activation at 95 ℃ for 5 min, 40 cycles of denaturation at 95 ℃ for 5 sec, annealing at 55 ℃ for 15 sec and extension at 72 ℃ for 20 sec. Thresholds (Th), Ct values, and derivatives of melting curves were determined using Rotor-Gene 6 software. All extracts were analysed in triplicate. For quantification three independent plasmid standards series with 300 to 3,000,000 gene copies/µL of the 16S rRNA gene of an environmental clone were included.

Quantification of endospore-forming bacteria

Quantification of *spo0A* gene was performed as mentioned above for the 16S rRNA gene but with the primers spo0A655f and spo0A923r (Bueche et al. 2013). The qPCR mix contained 5 ng DNA template, 0.76 µM of each primer and 1 x QuantiTect SYBR Green PCR Kit. Total reaction volume of 20 µL was reached with PCR-grade water. The program differed in an annealing at 52 ℃ for 30 sec and extension at 72 ℃ for 30 sec. All qPCR reactions were run in three technical replicates. For quantification three independent plasmid standards series with 30 to 300,000 gene copies/µL of *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 (described in chapter 2) were used for determination of the phylogenetic affiliation (Operational Taxonomic Units - OTUs) of endospore-forming bacteria. PCR reactions were performed with 0.5 ng DNA template, 1 x reaction buffer (TaKaRa Bio, Shiga, Japan), 3 mM MgCl2, 10 µg bovine serum albumin (BSA; New England Biolabs, Ipswich, MA, USA), 1 U of the proof-reading 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 ℃ for 5 min; then 10 cycles of denaturation at 94 ℃ for 30 min, touchdown annealing starting at 55 ℃ with decrease of 0.3 ℃ per cycle for 30 sec and elongation at 72°C for 1 min; followed by 30 cycles of denaturation at 94 ℃ for 30 min, annealing at 52 ℃ for 30 sec and elongation at 72 ℃ for 1 min; and a final extension at 72 ℃ 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, de-noising (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 Naïve 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 (EMBL 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 towards medically relevant species.

Statistical analysis and display

The distribution of the most abundant OTUs (over 40 sequences per OTU) per depth was analysed using the Heat-map function in R (R Core Team 2012). Correlations with depth were calculated using linear or exponential regression with the program SigmaPlot 12.0 (Systat Software, San Jose, CA). Significance of differences of parameters before and after 1960 was calculated using Mann-Whitney rank sum test or Student's t-test, when data was normally distributed. Correlations between community structure (only OTUs shared between at least two samples) and environmental variables 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 in the center of the lake at station SHL2) were retrieved from Lazzarotto et al. (2012), Annex 1, p.46. Phosphorus values before 1957 were assumed constant (average of values from 1957, 1958, and 1959).

4.3 Results

Sedimentary facies

The description of CAN01 and CAN02 sediment revealed unique sedimentary facies corresponding to hemipelagic sediments as previously shown (Loizeau 1991; Corella et al. 2011). This background sedimentation consisted of alternating triplets of; i) mm-thick organic-debris layers; ii) calcite-rich white laminae and; iii) mm- to cm-thick allochthonous detrital layers mostly transported within the river plume and dispersed as interflow. Grain size profile (Figure 4.1) revealed that sediments are fine silts with mean grain-size=17 µm, although an interval between (15-42 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 those 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 (Figure 4.1). The lack of ^{137}Cs in the sediment from 44 cm depth downwards suggests that deposition below this depth pre dates AD 1954. Two well-defined peaks were found at 37 and 19.5 cm depth (149 and 247 Bq/kg, 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± 0.05 cm/yr. As there is no ^{137}Cs signal below 44 cm, MS measurements were used to date the lower part of the core. MS show a peak at a depth of 65.5 cm, which can be correlated with peak # 8, dated to 1943 ± 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 about 1.83 cm/year, 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 1920 AD. Correlation between both sediment cores enables to estimate similar sedimentation rates for CAN02 (Figure 4.1).

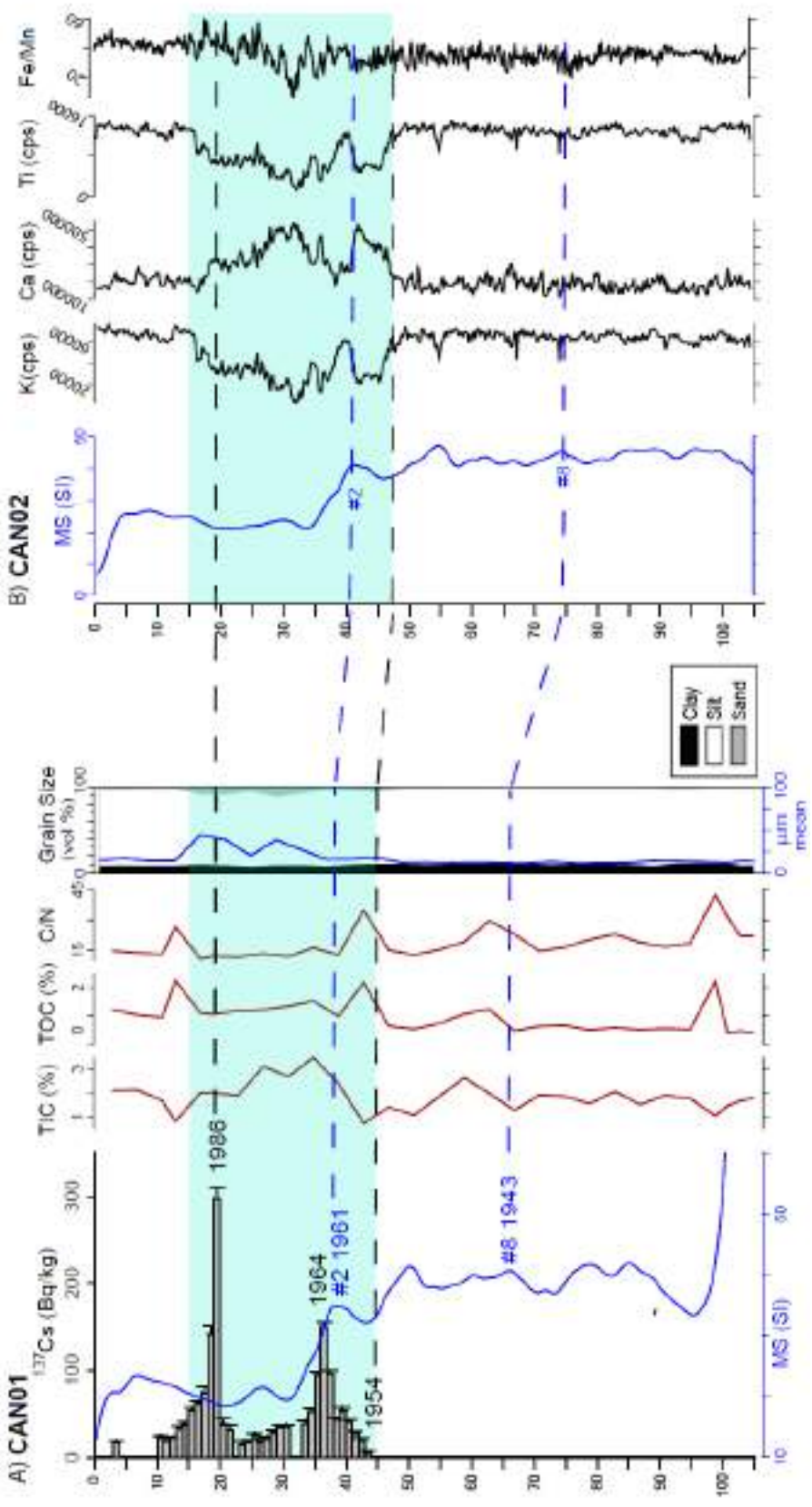


Figure 4.1: Sediment cores analysed in this study. **A**. From left to right: ^{137}Cs detection and magnetic susceptibility (MS) profiles and geochemical profiles of sediment core CAN01 (TIC: Total Inorganic Carbon; TOC: Total Organic Carbon; C/N: Organic Carbon/Nitrogen ratio), Grain-size (blue line indicate the mean grain-size). **B** From left to right: Magnetic susceptibility (MS) and XRF profiles (K, Ca, Ti, Fe/Mn ratio) in core CAN02. Dashed lines correspond to correlation horizons based on sedimentological and magnetic susceptibility profiles. Coloured area corresponds to a period of large limnological change.

Sediment geochemistry

The XRF data shown in Figure 4.1 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 Figure 4.1), 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 displays 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 as to target endospores as well as vegetative cells. Three different commercial DNA extraction kits were tested on an endospore preparation of *Bacillus subtilis* in order to decide which is 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, as described in chapter 2. Using the modified extraction method, the obtained DNA ranged from 0.8 to 16.4 µg/g sediment (Figure 4.2A). The highest DNA yield was not obtained at the top, but at 5 cm sediment depth (2004). Two other peaks could be identified at depths of 17 cm (1989) and 43 cm (1955). 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 (Figure 4.2B) ranged from 8.2×10^9 copies/g sediment in the upper part of the core (at depth of 5 cm (2004)) to 3.8×10^8 copies/g sediment at a depth of 99 cm (1923). 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, each showing a different trend. In the top section of the sediment core there is not a 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 sediment were measured, which was substantially higher than all other values. In the middle section (between 15 cm and 47 cm; 1955 to 1991), there was a trend towards an increasing abundance when moving upwards, although there was some variation between samples, especially in the upper half of this section. From 72 cm (1940) downwards the gene abundance was relatively constant, but one order of magnitude lower than those above.

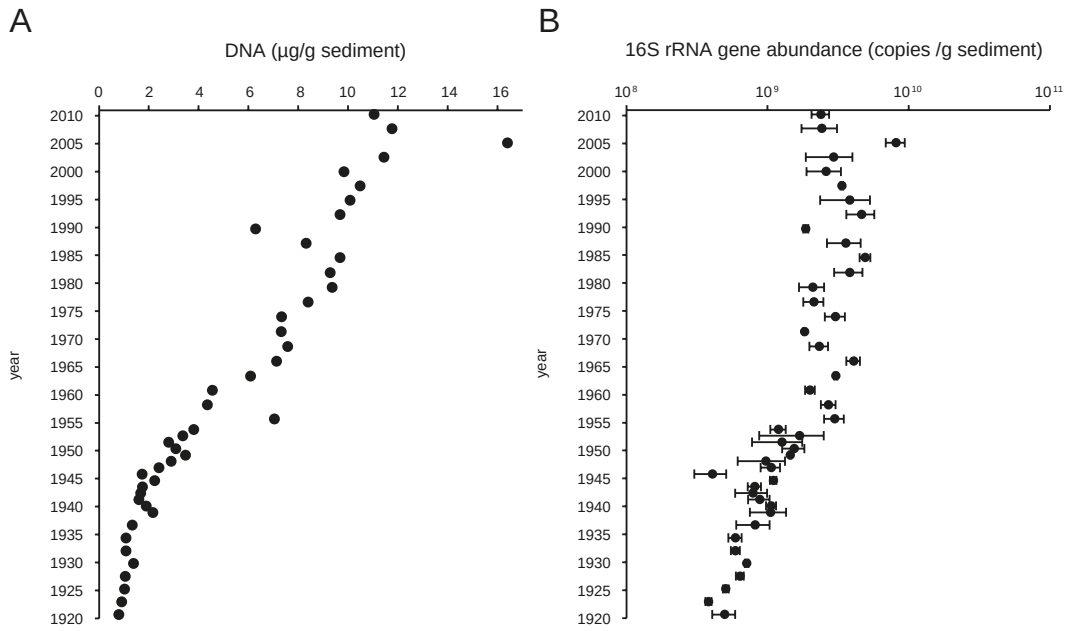


Figure 4.2: **A**. DNA content (µg/g sediment) with depth corresponding to calendar year timeline. **B**. Distribution of 16S rRNA gene abundance (copies/g sediment) with depth. Error bars correspond to independent technical qPCR replicates.

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 (Figure 4.3A) followed a different trend than the pattern from the 16S rRNA gene. Gene abundance ranged from 5.4 x 10^3 copies/g sediment at the bottom of the sediment core to 1.9 x 10^5 copies/g at the top. The highest value does not coincide with the sample that had 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 to 106 cm; 1921 to 1953), no overall trend could be seen. There were two samples with higher values than the others in this section (at 87 cm depth (1931) and in the area between 65 and 75 cm (1938 to 1944)). In the middle section of the core (15-50 cm; 1953 to 1991), there was a high variation between the samples and contrary to the 16S rRNA gene numbers, there was a slight decreasing trend when moving upwards in the core. Finally, in the top section of the core (0-15 cm; 1991 to 2011), the variation between samples was high, but overall, a trend could be seen shifting towards increasing values at the top of the core. The highest value of the *spo0A* gene abundance was found at depth of 3 cm (2007). The ratio between the 16S rRNA gene and the *spo0A* gene abundance were in the range of 0.7 x 10^{-3} % to 8 x 10^{-3} % and averaged 3.2 x 10^{-3} % (4.3B). The ratios were significantly higher in the bottom section of the core (below 50 cm; before 1953) than in the

upper half (Students t-test; p= <0.001).

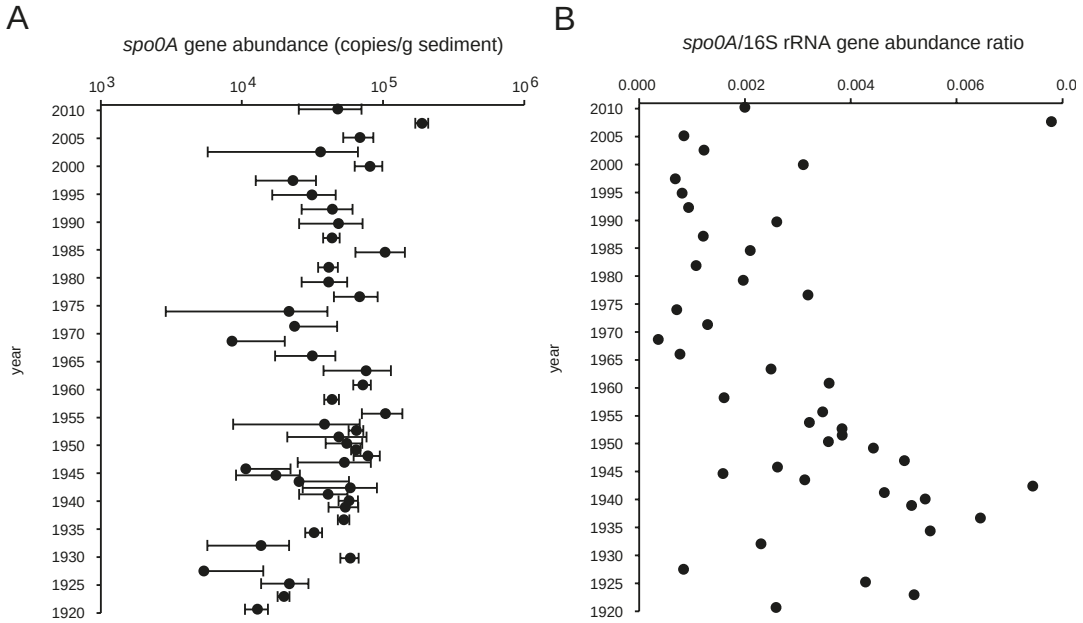


Figure 4.3: **A**. Depth distribution of *spo0A* gene abundance (copies/g sediment). Error bars correspond to independent technical qPCR replicates. **B**. Ratio of 16S rRNA/*spo0A* gene abundance.

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 i) as to represent depths spanning through the entire core and ii) 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 a total number of 5,144 sequences could be clustered into 552 OTUs. The results showed that despite the fact that the quantity of extracted DNA decreases with age of the sediment, the richness of endospore-forming bacteria does increase; richness is significantly smaller in the samples from depth (11, 19, 27, 39 cm) (t-test, p=0.007) with a mean of 31.5 OTUs per sample than in the other samples with mean of 65.4 OTUs per sample. Richness ranged from 15 (at 11 cm depth) to 96 OTUs at 103 cm depth (Figure 4.4). The majority (471) of the total identified phylotypes (554) only appeared as a single copy sequence. The heat-map (Figure 4.4) shows the relationship of the most abundant phylotypes, with representatives that have >40 sequences per OTU. The lighter the colour in the heat-map is, the more sequences per OTU were found. The samples from the top of the core (1961 to 2007) clustered separately from the older half of the core (1921 to 1948) and have distinct OTUs that are only found at one depth. On the contrary in the lower half of the core there are a

number of dominant phylotypes that were detected at several depths (OTUs Nr. 8, 9, 13, 20, 42) (Figure 4.4).

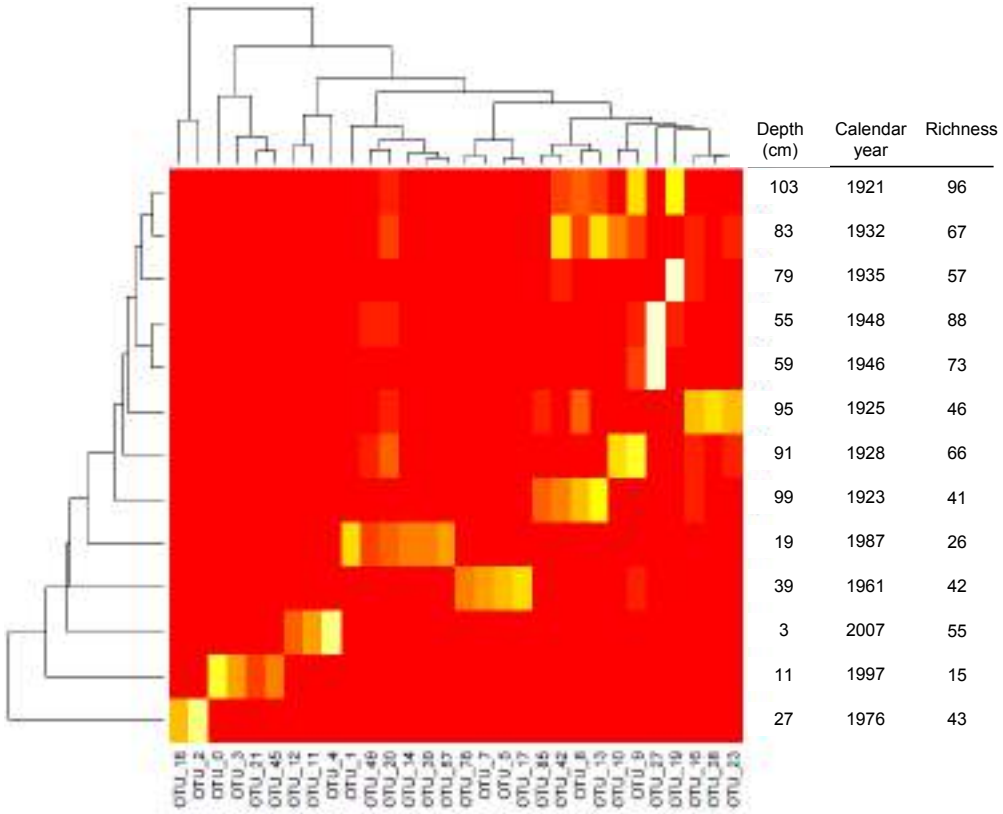


Figure 4.4: Heat-map of most dominant phylotypes obtained from *spo0A* sequences (>40 sequences per OTU) with clustering of samples from different depth according to similarities in community structure (OTU presence and abundance in the vertical axis). The frequency of OTUs is indicated by the colour 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 to the right.

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 Figure 4.4. 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 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 of 2.7, while other samples have on average 4.6 genera per sample. Between years 1961 to 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 of 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*.

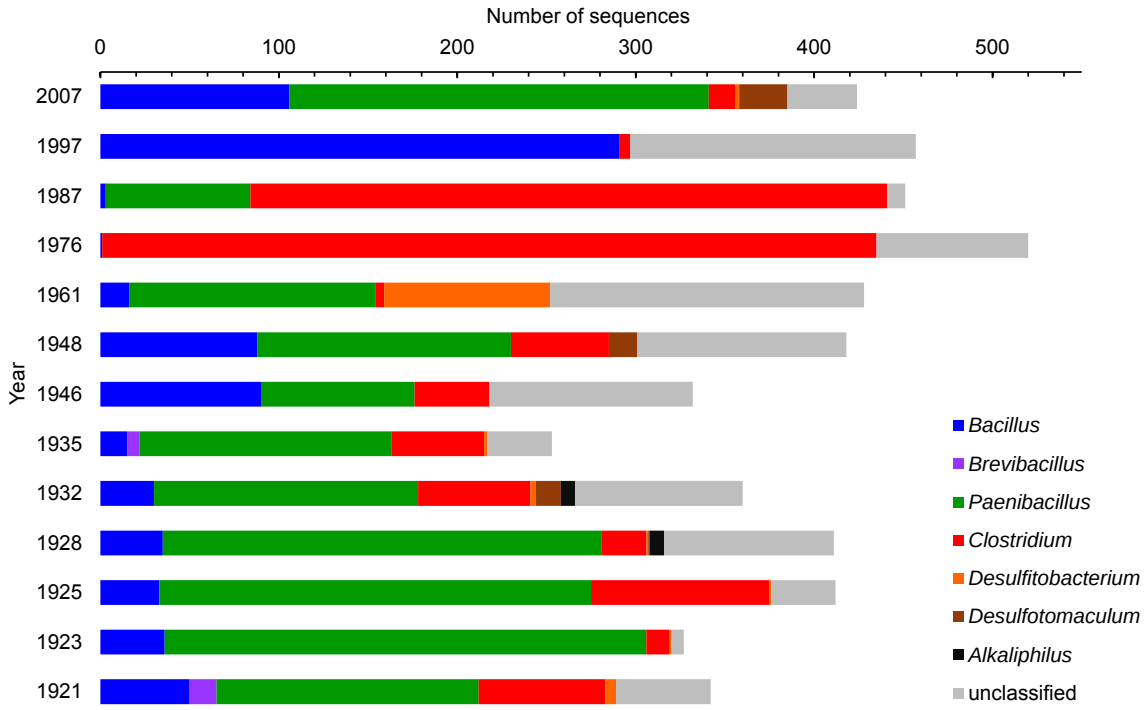


Figure 4.5: Endospore-forming community structure represented by *spo0A* sequences and their genus affiliation at different depths. Length of bar corresponds to total sequence abundance. Different genera are depicted in different colours. Grey bars are sequences that could not be classified to genus level (unclassified).

Correlation between physico-chemical and microbiological parameters

The main changes in biological parameters could be seen in the core section of 15 to 47 cm, where also the chemical and physical parameters showed greatest variations. The 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 5 sequences) of all samples and the parameters Ca, K, Ti, Fe, Mn, Fe/Mn, depth and total P (Figure 4.6). The community composition of endospore-forming bacteria from the samples 1976, 1987, 1997, and 2007 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 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.

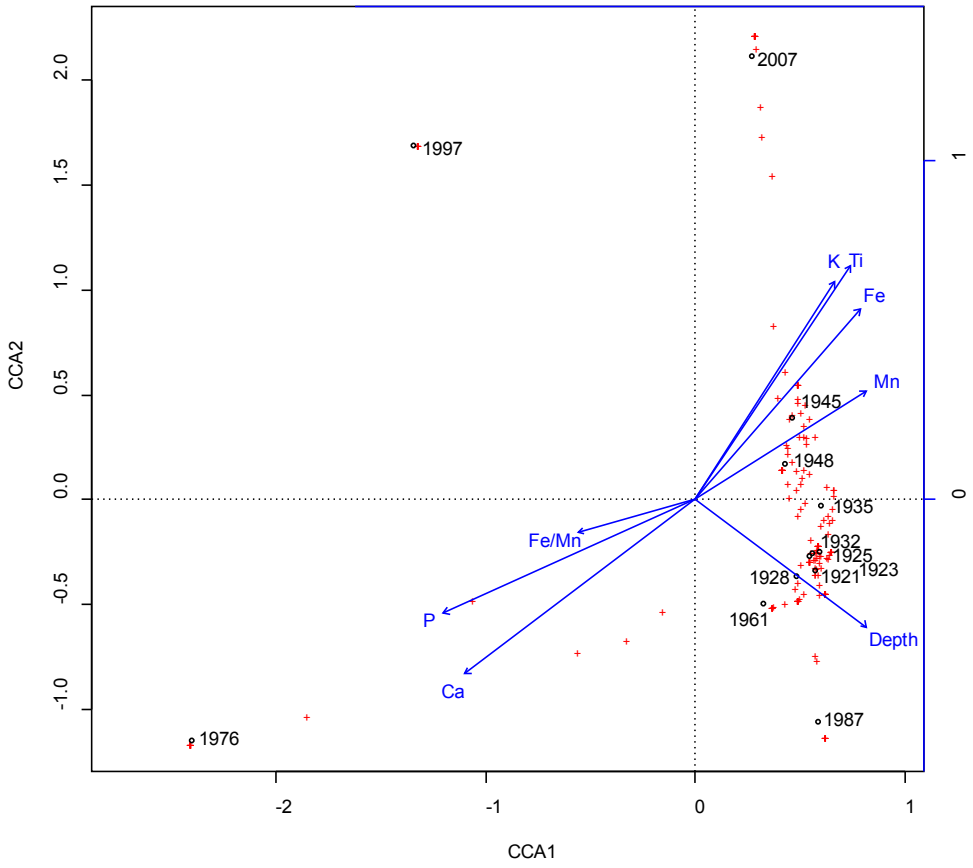


Figure 4.6: Canonical correspondence analysis (CCA) of the community matrix with all OTUs composed of minimum 5 sequences and the environmental matrix with values of depth, total P, Ca, K, Ti, Fe, Mn and Fe/Mn as constraining parameters. The total P (µg/L) values are average yearly values from the Grand Lac (Lazzarotto et al. 2012). X-axis explained 15.3% and Y-axis explained 15.0% of importance.

4.4 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 are decreasing exponentially with time. So is 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 decrease in 16S rRNA gene abundance, which can be explained by the contribution of eukaryote 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 x 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) that accounts for 25 x 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 analyse 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 to protect 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 can be seen by the ratio of *spo0A* and 16S rRNA gene increasing with depth (Figure 4.3B). 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 and 1987, the members from the genus *Clostridium* are most abundant and in the sample from 1997, it is members from *Bacillus* that 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, especially concerning aerobic or anaerobic growth conditions. The genera *Clostridium* (Schleifer 2009), *Desulfotomaculum* (Stackebrandt et al. 1997) and *Desulfitobacterium* (Utkin et al. 1994) are strict anaerobic heterotrophic bacteria. Bacteria from genera *Bacillus* (Schleifer 2009) and *Brevibacillus* (Shida et al. 1996) on the contrary, grow aerobically, while *Paenibacillus* (Shida et al. 1997) is composed of members that grow as facultative anaerobes. Therefore the 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 the 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 non-spore-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 non spore-forming strains with our *spo0A* primer have been negative (see Table 2.1). 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. However, in the present study this might not be the case. The high C/N ratio shown in Figure 4.1, 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 C/N ratio which means, that likely the majority of the endospore-forming community in the sediments is from an autochthonous source and less influenced by transport from rivers. An example for this could be members from *Paenibacillus*, facultative anaerobes known to hydrolyse 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 strongly decreasing in the top few cm of the sediment

to almost zero at 10 cm, suggesting little microbial activity below this depth (Thevenon et al. 2011). When conditions get oligothrophic and in particular if the carbon is low, endosporulation sets in due to starvation (Hageman et al. 1984). Based on that, 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 hyperpicnal 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 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 (Figure 4.6), it is difficult to disentangle endogenic Ca precipitation due to increased primary production with 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) favouring 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 interpreted as influenced by allochthonous inputs. Neither of the above single elements can therefore clearly be linked to changes solely due to the 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 reflects hypoxia at the sediment water interface due to preferential re-solubilization 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 (Figure 4.1) can be associated to 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 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 to 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 run-off have had an important impact on the ecosystem. The lake has been closely monitored since 1957 and data is made publicly available by the "Commission Internationale de la Protection des Eaux du Léman" (CIPEL) in the form of yearly published reports (`http://www.cipel.org`) the most recent one released in 2012 (Lazzarotto et al. 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 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 the 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 <4mg O_2/L) have been registered in the deepest part of the lake (300 m) in the years from 1973-78 and 1986-1998, by CIPEL (Lazzarotto et al. 2012). Important changes have been seen in the composition of pelagic primary and secondary producers as 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 towards the recovery of endospore-forming communities, however, 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 8 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 is breaking 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 sequence 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.

Acknowledgements

The work was supported by Swiss National Science Foundation grant No. 31003A-132358/1 and the eLEMO Grants 301 and 302. The authors acknowledge Philippe Arpagaus and Angel Arantegui for their support during the fieldwork and laboratory analyses. The authors thank two anonymous reviewers for the valuable comments to improve this manuscript.

4.5 Supplementary Information

The two sediment cores were taken in the upper part of Lake Geneva, close to the Rhone delta. The exact locations are indicated in Figure 4.7.

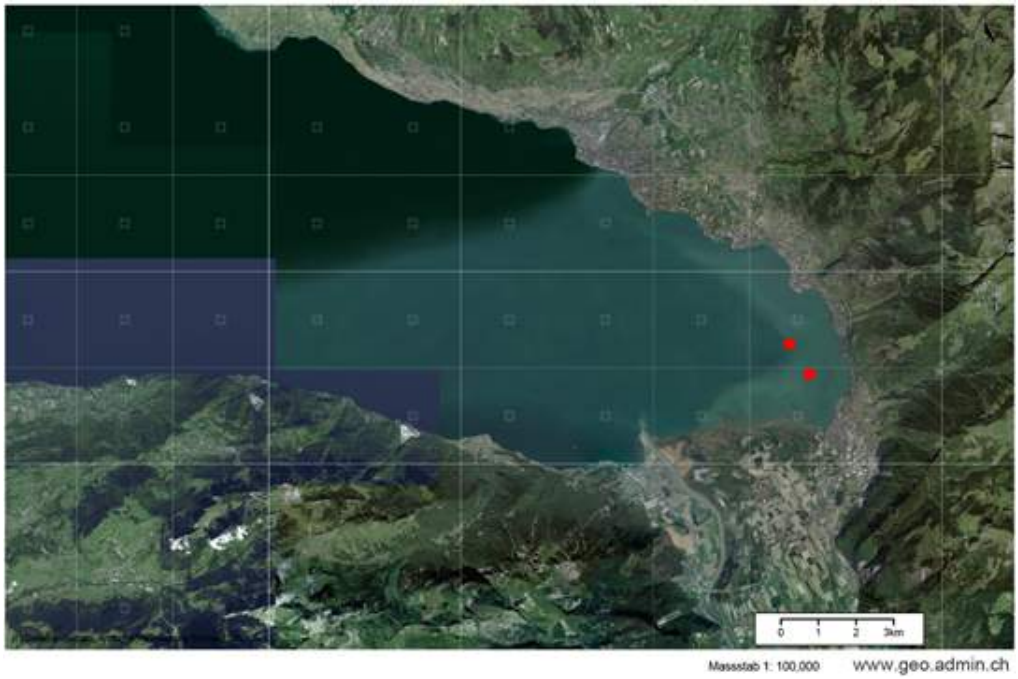


Figure 4.7: Satellite image of the eastern part of Lake Geneva (`www.geo.admin.ch`; retrieved april 2013). The locations of sampling are indicated by a red star.

4.6 References

Abecasis, A. B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J. B., and Henriques, A. O. (2013). A genomic signature and the identification of new sporulation genes. *J Bacteriol*, 195(9):2101–15.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev*, 59(1):143–69.

Anneville, O. and Pelletier, J. P. (2000). Recovery of Lake Geneva from eutrophication: quantitative response of phytoplankton. *Arch Hydrobiol*, 148(4):607–624.

Bartholomew, J. W. and Paik, G. (1966). Isolation and identification of obligate thermophilic sporeforming Bacilli from ocean basin cores. *J Bacteriol*, 92(3):635–8.

Bueche, M., Wunderlin, T., Roussel-Delif, L., Junier, T., Sauvain, L., Jeanneret, N., and Junier, P. (2013). Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene *spo0A*. *Appl Eviron Microbiol*, 79(17):5302–12.

Corella, J. P., Brauer, A., Mangili, C., Rull, V., VegasVilarrúbia, T., Morellón, M., and ValeroGarcés, B. L. (2012). The 1.5-ka varved record of Lake Montcortès (southern Pyrenees, NE Spain). *Quaternary Res*, 78(2):323–332.

Corella, J. P., El Amrani, A., Sigró, J., Morellón, M., Rico, E., and Valero-Garcés, B. (2011). Recent evolution of Lake Arreo, northern Spain: influences of land use change and climate. *J Paleolimnol*, 46(3):469–485.

de Rezende, J. R., Kjeldsen, K. U., Hubert, C. R., Finster, K., Loy, A., and Jorgensen, B. B. (2013). Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J*, 7(1):72–84.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–1.

EMBL (2005). The EMBL nucleotide sequence database.

Fichtel, J., Koster, J., Rullkotter, J., and Sass, H. (2007). Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol Ecol*, 61(3):522–32.

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., and Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol*, 14(11):2870–90.

Gerdeaux, D. and Perga, M.-E. (2006). Changes in whitefish scales 13C during eutrophication and reoligotrophication of subalpine lakes. *Limnol Oceanogr*, 51(1 part 2):772–780.

Gorham, E., Brush, G. S., Graumlich, L. J., Rosenzweig, M. L., and Johnson, A. H. (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, J. H., Shankweiler, G. W., Wall, P. R., Franich, K., McCowan, G. W., Cauble, S. M., Grajeda, J., and Quinones, C. (1984). Single, chemically defined sporulation medium for *Bacillus subtilis*: growth, sporulation, and extracellular protease production. *J Bacteriol*, 160(1):438–41.

Hobbs, W. O., Lalonde, S. V., Vinebrooke, R. D., Konhauser, K. O., Weidman, P. R., Graham, M. D., and Wolfe, A. P. (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., and Jorgensen, B. B. (2010). Thermophilic anaerobes in arctic marine sediments induced to mineralize complex organic matter at high temperature. *Environ Microbiol*, 12(4):1089–104.

Kindt, R. and 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, J. A., Dunbar, J. M., and Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol*, 66(4):1328–33.

Koinig, K., Shotyk, W., Lotter, A., Ohlendorf, C., and Sturm, M. (2003). 9000 years of geochemical evolution of lithogenic major and trace elements in the sediment of an alpine lake – the role of climate, vegetation, and land-use history. *J Paleolimnol*, 30(3):307–320.

Kroos, L. (2007). The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. *Ann Rev Genetics*, 41:13–39.

Lazzarotto, J., Klein, A., and Commission pour la Protection des Eaux du Léman (CIPEL) (2012). Physical-chemical changes in the waters of Lake Geneva (major elements). Technical report, Lausanne, Switzerland.

Loizeau, J.-L. (1991). *La sédimentation récente dans le delta du Rhône, Léman: processus et évolution*. PhD thesis, Université de Genève, Geneva, Switzerland.

Loizeau, J.-L., Dominik, J., Luzzi, T., and 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, B. A., Langerhuus, A. T., D'Hondt, S., Jorgensen, B. B., and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature*, 484(7392):101–4.

Meyers, P. A. (2003). Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. *Org Geochem*, 34(2):261–289.

Molinero, J. C., Anneville, O., Souissi, S., Balvay, G., and Gerdeaux, D. (2006). Anthropogenic and climate forcing on the long-term changes of planktonic rotifers in Lake Geneva, Europe. *J Plankton Res*, 28(3):287–296.

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., and 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, M., Oksanen, M., and Suggests, M. (2007). The vegan package.

Onyenwoke, R. U., Brill, J. A., Farahi, K., and 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.

Ovreas, L., Forney, L., Daae, F. L., and 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, M. R., Rollion-Bard, C., Farah, R., and Vasconcelos, C. (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., and 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 Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rellstab, C., Keller, B., Girardclos, S., Anselmetti, F. S., and Spaak, P. (2011). Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing daphnia in unproductive lakes. *Limnol Oceanogr*, 56(1):292.

Renberg, I. and Nilsson, M. (1992). Dormant bacteria in lake sediments as palaeoecological indicators. *J Paleolimnol*, 7(2):127–135.

Robles, S., Rodriguez, J. M., Granados, I., and Guerrero, M. C. (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):187–91.

Rothfuss, F., Bender, M., and Conrad, R. (1997). Survival and activity of bacteria in a deep, aged lake sediment (Lake Constance). *Microbiol Ecol*, 33(1):69–77.

Ryves, D. B., Battarbee, R. W., Juggins, S., Fritz, S. C., and Anderson, N. J. (2006). Physical and chemical predictors of diatom dissolution in freshwater and saline lake sediments in North America and West Greenland. *Papers Geosci*, 53(3):1355–1368.

Sastre, V., Loizeau, J.-L., Greinert, J., Naudts, L., Arpagaus, P., Anselmetti, F., and Wildi, W. (2010). Morphology and recent history of the Rhone river delta in Lake Geneva (Switzerland). *Swiss J Geosci*, 103(1):33–42.

Schaller, T. and Wehrli, B. (1996). Geochemical-focusing of manganese in lake sediments — an indicator of deep-water oxygen conditions. *Aquat Geochem*, 2(4):359–378.

Schleifer, K.-H. (2009). Phylum XIII. Firmicutes (Gibbons and Murray 1978), 5. In Vos, P., Garrity, G. M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., et al., editors, *Bergey's Manual of Systematic Bacteriology*, page 19–1317. Springer New York.

Seaward, M. R., Cross, T., and Unsworth, B. A. (1976). Viable bacterial spores recovered from an archaeological excavation. *Nature*, 261(5559):407–8.

Shida, O., Takagi, H., Kadowaki, K., and Komagata, K. (1996). Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. *Int J Syst Evol Microbiol*, 46(4):939–46.

Shida, O., Takagi, H., Kadowaki, K., Nakamura, L. K., and 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):289–98.

Sneath, P. H. (1962). Longevity of micro-organisms. *Nature*, 195:643–6.

Stackebrandt, E., Sproer, C., Rainey, F. A., Burghardt, J., Päuker, O., and 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, J. T. and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*, 39:321–46.

Thevenon, F., Adatte, T., Wildi, W., and Pote, J. (2012). Antibiotic resistant bacteria/genes dissemination in lacustrine sediments highly increased following cultural eutrophication of Lake Geneva (Switzerland). *Chemosphere*, 86(5):468–76.

Thevenon, F., Graham, N. D., Herbez, A., Wildi, W., and 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):609–17.

Traag, B. A., Pugliese, A., Eisen, J. A., and Losick, R. (2013). Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. *J Bacteriol*, 195(2):253–60.

Utkin, I., Woese, C., and 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):612–9.

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., et al. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482(7385):357–62.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*, 73(16):5261–7.

Willard, D. A. and Cronin, T. M. (2007). Paleoecology and ecosystem restoration: Case studies from Chesapeake Bay and the Florida Everglades. *Front Ecol Environ*, 5(9):491–498.

5 Endospores in sediment as a proxy for 20th century environmental lake history

Abstract

Lake Geneva (Switzerland-France) is a prime example of the impact of human activities on freshwater ecosystems. This lake has experienced an eutrophication period in the early 1960s, followed by measures to decrease the nutrient input in 1980s leading to partial restoration of its trophic status. The example of Lake Geneva can be used to establish important baseline information for policy makers, and the public in general, to implement measures for ecosystem management for freshwater bodies. Reference conditions such as local biodiversity and ecosystem responses to environmental perturbation are needed, but they are challenging to establish. Lake sediments provide an ideal environmental archive of historical and reference conditions. In this study we use bacterial endospores, natural biological time capsules, to unveil the lake history and effect of recent events on the sediment microbiota of Lake Geneva. We provide evidence that variations in community structure can be linked to environmental factors such as eutrophication, sulfate metabolism, input of terrestrial organic matter, and specific climate events. A treatment allowing the separation of endospores from vegetative cells and sequencing of the endospore fraction provides additional information on the dormant bacterial remains in sediment. It also gives information on the "seed bank" component of the microbial pool. This study reveals the community composition of dormant bacterial endospores in sediment and uses this group as a paleoecological proxy for the reconstruction of the lake history.

5.1 Introduction

Most freshwater lakes are heavily subjected to anthropogenic pressure due to population growth, agriculture and industrial pollution, and modification of land use, leading in particular to eutrophication due to high nutrient load (Smith et al. 1999; Smith 2003). Eutrophication threatens water quality and health of aquatic ecosystems, which are of important social, ecological, and economic value (Wilson and Carpenter 1999; Dudgeon et al. 2006).

Lake Geneva, bordering Switzerland and France, is the largest lake in central Europe (580 km^2, maximum depth 310 m and mean depth 153 m) and provides about 80,000 m^3 of drinking water per year for a population of over half a million (Lazzarotto et al. 2012). Lake Geneva sustains a fishery industry with an output of about 1000 tons of fish per year. In addition, the lake has an important recreational value for residents and tourists. The International Commission for Protection of Lake Geneva (CIPEL) has been monitoring the status of the lake since 1963. Yearly measures of water quality have shown an eutrophication of the lake due to increased input of phosphorus (P) and nitrogen (N) since the 1950s. This shift in nutrient status has led to the extinction of whitefish (Vonlanthen et al. 2012) and to changes in the zoo- and phytoplankton communities (Anneville et al. 2002; Tadonléké et al. 2009). In 1972, measures to reduce the P input have been taken by adding precipitation steps during wastewater treatment and in 1986 a legal ban of polyphosphates in detergents was added (Anneville et al. 2002). Today, the

P values are back to pre-1960 values, describing the lake status as mesotrophic (Lazzarotto et al. 2012). Future goals in ecosystem restoration are to further decrease the P values below 20 g/l, to maintain low levels of micropollutants, and to ameliorate ecosystem quality (International Commission of the Protection of Lake Geneva (CIPEL) 2010).

The problem of eutrophication has been observed in many other lakes and has been acknowledged by policy makers. The European Water Framework Directive has set an objective to reach "natural ecological status" for all European lakes by 2015 (European Union (EU) 2000). However, to reach and particularly to monitor this goal, parameters for ecological quality need to be defined. First the reference conditions (i.e. the pre-disturbance status, natural fluctuations) have to be known. Then the magnitude of change and the link between ecological pressure and ecosystem response have to be established. Only with this background knowledge future predictions on ecosystem health can be made. To date, however, most of these parameters are poorly known.

An archive of past environmental conditions and reference conditions can be found in the lake sediments (Willard and Cronin 2007). Sediment cores offer a record of chemical and biological components and their change over time and can therefore be used to understand the relationship between ecological disturbance and its impact on the ecosystem. Biological components typically correspond to organisms producing identifiable fossilized structures (e.g. pollen grains or siliceous microfossils) that remain unaltered in the sediments and can be analysed on different time-scales (Gorham et al. 2001). Such paleolimnological studies provide the range of natural fluctuations and the impact of anthropogenic pressure on biological indicators (Thevenon et al. 2012). Furthermore, it provides data about the trend of recovery after an impact (resilience). In Lake Geneva, paleolimnological studies conducted on diatoms and fish scales have described the eutrophication and re-mesotrophication process (Gerdeaux and Perga 2006).

Paleolimnological studies focusing on bacteria have been done so far mostly using fossil pigments (Gorham et al. 2001; Dressler et al. 2007). However, the potential of bacteria as indicators of paleoenvironmental conditions is high considering their large cumulative mass in lakes (water column and sediment), their phylogenetic diversity and the wide range of different metabolisms they represent (Nealson 1997). Nevertheless, methods for the comprehensive detection of bacteria in sediments are challenging, especially because the biomass decreases strongly with sediment depth and because of DNA degradation with age.

An alternative to this is the use of bacterial resting states, a topic discussed by Renberg and Nilsson (1992). The authors conclude that it is possible to isolate spores from sediment and, by combining the analyses of biochemical and genetic data, to infer past environmental conditions. 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 (Onyenwoke et al. 2004). Revival and isolation of endospores has been used for paleoecological reconstructions (Bartholomew and Paik 1966; Renberg and Nilsson 1992; Rothfuss et al. 1997) or to demonstrate the dispersal of metabolic inactive thermophiles

into cold sediments (Bartholomew and Paik 1966; Hubert et al. 2010; de Rezende et al. 2013). However, such culturing studies are biased towards a small cultivable fraction of the community (Staley and Konopka 1985; Amann et al. 1995). In a culture-independent approach, the use of an endospore-specific biomarker (dipicolinic acid) suggested that endospores in sediments of the North Sea represent 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). In chapter 2 of this thesis, an alternative culture-independent molecular approach based on the *spo0A* gene is described. Using this method the applicability of endospore-forming bacteria as proxies for eutrophication in Lake Geneva was shown in chapter 4.

In the study presented here, we have extended the analysis to higher depth resolution and to include other environmental factors in a sediment core spanning the past 100 years of history in Lake Geneva. Also, the active (vegetative cells) and the dormant fraction (endospores) were analysed individually and compared to each other. The results of this study show that endospore-forming bacteria can be used to track specific events in the history of Lake Geneva and that the differentiation between the active and inactive communities help to establish the prevailing environmental conditions at particular time points.

5.2 Materials and Methods

Collection of environmental samples

The sediment samples are described in detail in chapter 4. Briefly, two sediment cores were retrieved with a gravity corer (UWITEC, Mondstein, Au) in two canyons (C1 and C2) on the eastern side of the Rhone delta in Lake Geneva (Switzerland) (CAN01, coordinates 559901-139859, 79 m depth, 105 cm; and CAN02, coordinates 559405-140504, 96 m depth, 107 cm) The cores were stored at 4 ℃ immediately after retrieval.

Sedimentological description

Detailed description of measurements of the physical and geochemical properties is given in chapter 4. Pictures were obtained using a digital camera and controlled light conditions (Supplementray Figure 5.5). 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 fall-out 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 a cumulated sediment

mass scale, because porosity didn't change drastically with depth; therefore no significant bias was introduced by porosity variations.

CAN01 sediment core was sampled every 4 cm for total carbon (TC) and total nitrogen (TN) using an elemental analyser (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 (TOC) 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 analyser (Marlvern instruments Ltd, USA) at the Pyrenean Institute of Ecology (Zaragoza, Spain).

X-ray fluorescence was analysed in CAN02 core using an AVAATECH XRF core scanner (2000 A, 10kV and 30kV) every 2 mm at the University of Barcelona.

The phoshorus (P) concentrations of the lake water were retrieved from the yearly monitoring reports of Lake Geneva (Lazzarotto et al. 2012).

The core CAN01 was sub-sampled aseptically every 2 cm for biological analysis and samples were stored at -20 ℃ until processing.

DNA extraction

DNA was extracted using a modified protocol of the MP FastDNA SPIN kit for soil (MP Biomedicals, Santa Ana, CA, USA) described in detail in Chapter 4. The modification consisted of an addition of three repetitive extraction cycles on 0.5 g of sediment as starting material. The three final DNA extracts were pooled and DNA was precipitated with 0.3 M sodium-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.

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 in chapter 3 of this thesis. Briefly, 0.5 g sediment was dispersed in 900 µl tris-EDTA buffer and incubated in a waterbath at 65 ℃ for 20 min. After cooling the samples for 5 min at 4 ℃ 100 µl of lysozyme was added to reach a final concentration of 2 mg/ml, and the mix was incubated at 37 ℃ for 60 min and 120 rpm shaking. Afterwards, 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 60 min with low agitation (100 rpm). Samples were then filtered onto 17 mm cross section and 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, USA) 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 (with sediment particles and endospores) was then removed from the apparatus and stored at -20 ℃ until DNA extraction as described above.

Quantification of total bacteria

Quantification of bacterial DNA in extracts was carried out by real-time quantitative PCR (qPCR) of the V3 region of the 16S rRNA gene with primers 338f and 520r (Ovreas 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 mix (QIAGEN, Hilden, Germany). 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: initial activation at 95 ℃ for 5 min, then 40 cycles of denaturation at 95 ℃ for 5 sec, annealing at 55 ℃ for 15 sec and extension at 72 ℃ for 20 sec. Thresholds (Th), Ct values, and derivatives of melting curves were determined using Rotor-Gene 6 software (QIAGEN). All the qPCR reactions were run in triplicates. For quantification three independent plasmid standard series with 300 to 3,000,000 copies/µL of the 16S rRNA gene of an environmental clone were included.

Quantification of endospore-forming bacteria

Quantification of the spo0A gene was done by qPCR with primers spo0A655f and spo0A923r (Bueche et al. 2013). The qPCR mix contained 5 ng DNA template, 0.76 µM of each primer and 1 x QuantiTect SYBR Green PCR mix (QIAGEN). Total reaction volume of 20 µL was reached with PCR-grade water. The program contained an initial activation at 95 ℃ for 5 min, then 45 cycles of denaturation at 95 ℃ for 5 sec, annealing at 52 ℃ for 15 sec and extension at 72 ℃ for 30 sec. All qPCR reactions were run in triplicates. Three independent plasmid standards series with 30 to 300,000 copies/µL of *spo0A* gene of *Bacillus subtilis* were included.

spo0A amplicon sequencing

Degenerate primers (spo0A166Af and spo0A748Ar) (described in chapter 2) amplifying a 602 nucleotide fragment of the *spo0A* gene 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), 3 mM MgCl2, 10 µg 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 program: initial activation at 94 ℃ for 5 min, then 10 cycles of denaturation (94 ℃; 30 sec), touchdown annealing from 55 ℃ to 52 ℃ (-0.3 ℃/cycle) for 30 sec and elongation (72 ℃; 1 min); followed by 30 cycles of denaturation (94 ℃; 30 sec), annealing (55 ℃; 45 sec) and elongation

(72℃; 1 min) and final extension (72℃; 5 min). PCR products were then purified with MultiScreen PCRμ96 plate (Merck Millipore, Billerica, USA) using a vacuum pump and eluted in 40 µl water. The purified samples were sent to Eurofins MWG Operon for barcode amplicon sequencing with Roche GS FLX+. All metagenomic sequences were submitted to GenBank Sequene Read Archive under BioProject (PRJNA206067) and accession numbers SRR1011318 to SRR1011346.

Determination of OTUs of endospore-forming bacteria

Sequences were 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 (Gribskov et al. 1987) that was built based on 27 known Spo0A sequences listed in supplementary information of chapter 2 (Table 2.5). Filtration was applied as function of the profile score and profile alignment length, which separates true Spo0A hits (5.9). For determination of operational taxonomic units (OTU), the nucleotide sequences were clustered at 97% sequence identity using uclust (Edgar 2010) in the same way as commonly done for the 16S rRNA gene. The centroid of each OTU cluster was then used in BLASTx to find the closest relative available from the GenBank database (`http://www.ncbi.nlm.nih.gov/genbank/`).

Numerical analysis

Simpson and beta diversity indices were calculated with the software R and the package *vegan* (R Core Team (2012); Oksanen et al. (2007)). Canonical correspondence analysis (CCA) and redundancy analysis (RDA) were done using R (R Core Team (2012) with the community matrix including all depths and their community composition on the level of genus distribution. Regressions were calculated using Sigmaplot software 12.0 (Systat Software, San Jose, CA).

5.3 Results

Sedimentological description

The ^{137}Cs activity in core CAN01 ranged from 7 to 297 Bq/kg (Figure 4.1). The lack of ^{137}Cs in the sediment from 44 cm depth downwards suggests that deposition below this depth pre dates AD 1954. Two well-defined peaks were found at 37 and 19.5 cm depth (149 and 247 Bq/kg, 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± 0.05 cm/yr. As there is no ^{137}Cs signal below 44 cm, MS measurements were used to date the lower part of the core. MS show a peak at a depth of 65.5 cm, which can be correlated with peak # 8, dated to 1943 ± 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 about 1.83 cm/year, 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 1920 AD.

Succession of endospore-forming bacteria in the sediment core

Sediment samples were taken every four cm at a resolution of about four years spanning the period from 1921 to 2010. Endospore-forming bacterial communities were analysed using the *spo0A* gene in all samples. An overview of the results is shown in Table 5.1.

The *spo0A* abundance per gram sediment over depth shows a linear decrease of about 3.8% at each depth (R^2 0.55, p = <0.0001; Supplementary Figure 5.6 A). The oldest sample (1921) has 14.7% of the *spo0A* gene abundance compared to the most recent sediment (2010). After amplicon sequencing, *spo0A* sequences were clustered *de novo*. The numbers of OTUs detected at the different sediment ages ranged from 22 (2010) to 475 OTUs (1930) with a significant increase with age (R^2 0.66, p = <0.0001; Supplementary Figure 5.6B). Also the number of singletons (OTUs with a single sequence) detected in each sample increased with sediment age (R^2 0.64, p = <0.0001; Supplementary Figure 5.6C). The highest abundance (most dominant OTU) is significantly higher (RankSumTest, p = 0.001) in the period of 1950 to 2007 than the time before.

The OTUs were affiliated to 18 different genera. In addition, some OTUs could only be classified to Bacteria, Bacillaceae, or Clostridiales. The relative abundance of genera over time is shown in Figure 5.1 (total and percentage values are provided in Supplementary Table 5.3, 5.4 and 5.5).

Based on the community structure, the record could be divided into three main sections (Figure 5.1). Section I is the deeper (older) half of the core (1921 to 1950) with a constant distribution of genera *Bacillus*, *Paenibacillus*, *Clostridium* and some minor groups such as *Desmospora* and *Moorella*. The Simpson index in these samples ranged from 0.27 (1923) to 0.74 (1948) with a mean of 0.56. The exception to this relatively constant pattern was the sample from 1930, where the relative

Table 5.1: Summary of the molecular data obtained for samples from the period from 1921 to 2010 in the sediment record of Lake Geneva. Singletons are OTUs with a single sequence. Highest abundance is the percentage at which the most common OTU is present.

	spo0A/g sediment (x10^3)	Sequences (Number)	OTUs (Number)	Singletons (Number)	Highest abundance (%)
2010	83.6 ± 63.1	37	22	16	14
2007	157.1 ± 99.5	3823	156	45	26
2002	67.3 ± 62.2	845	129	44	7
1997	55.3 ± 53.7	2941	73	27	23
1986	46.5 ± 26.1	4205	141	45	16
1976	68.3 ± 42.4	4288	169	48	9
1970	97.7 ± 100.2	1040	72	25	17
1965	34.6 ± 20.7	1184	63	45	28
1961	48.6 ± 27.8	3755	290	112	12
1957	64.4 ± 34.8	1011	68	31	22
1955	52.3 ± 33.8	1426	104	50	19
1952	38.3 ± 24.7	2515	183	62	6
1950	62.3 ± 42.2	3175	311	107	7
1948	63.1 ± 64.7	2400	239	75	5
1945	58.9 ± 45.4	2451	163	55	8
1943	50.9 ± 39.5	2717	218	72	11
1940	43.2 ± 29.2	2526	322	108	5
1938	33.4 ± 22.0	2368	263	96	10
1935	25.9 ± 17.5	2057	214	75	10
1932	9.6 ± 7.0	2976	339	134	8
1930	42.6 ± 29.2	4633	475	204	10
1928	5.9 ± 5.8	3634	390	143	4
1925	28.2 ± 21.5	3421	275	85	6
1923	16.6 ± 11.0	3248	339	120	5
1921	12.3 ± 8.2	3121	461	193	3

abundance of OTUs affiliated to *Bacillus* was reduced, and instead other genera such as *Sporomusa* and *Moorella* increased. The beta-diversity values between samples from two successive sediment ages was low with an average of 0.81 in section I.

A clear drop in diversity between 1950 to 1952 (from 0.72 to 0.31) and a shift in community composition marked the beginning of the second section (II). This shift is also seen in the beta-diversity index that is at 0.99 between the sample from 1950 and 1952.

In 1952, *Paenibacillus* was by far the most dominant genus with 82% abundance, while all representatives from *Bacillus* disappeared. The communities in sediments from 1955, 1957, and 1961 fluctuated importantly as seen in the variation of the Simpson index, and were marked by the alternated dominance of *Clostridium*, *Paenibacillus*, *Moorella* or *Desulfotomaculum*. The dominance of *Moorella* or *Desulfotomaculum* in these three samples is remarkable considering that those genera are only a minor fraction of the community in other samples. In sediments from 1965 to 1986 the diversity was generally low (Simpson index of 0.36, 0.66, 0.009, and 0.31) and *Paenibacillus* and *Clostridium* dominated the

communities. The extreme case was the year 1976 in which *Clostridium* made up 99.5 % of the community. Finally, the community in 1997 was peculiar because of the dominance of *Brevibacillus*, which was not common in other depths. Overall in this section (section II) the diversity is highly fluctuating as seen in the high beta-diversity values (average of 0.96).

The communities from 2002 and 2010 (section III) were more diverse (0.66, 0.56, and 0.76) with the percentage of OTUs affiliated to *Bacillus* increasing in 2002 and 2007 (25.1 and 21.0%) compared to the majority of the samples from 1952 to 1997. The results from the most recent sediment (2010) should be interpreted with care because a substantially fewer number of sequences was obtained. Nonetheless, this sample had high diversity and evenness and was dominated by *Heliobacterium*.

Correlation of endospore-forming bacterial diversity and environmental parameters

The redundancy analysis (RDA) plot relating the distribution of genera with sediment age and environmental parameters is shown in Figure 5.2. The sample of 2010 was removed from this analysis because of the lower number of sequences obtained. The environmental constraints corresponded to the carbon to nitrogen ratio (C/N ratio), total organic carbon (TOC), and the concentration levels of P in the water (medium, low and high) derived from data by Lazzarotto et al. (2012).

The samples from older sediment ages (1921 to 1952) as well as the sample from 1965 and 1970 built a cluster in the center of the plot together with most

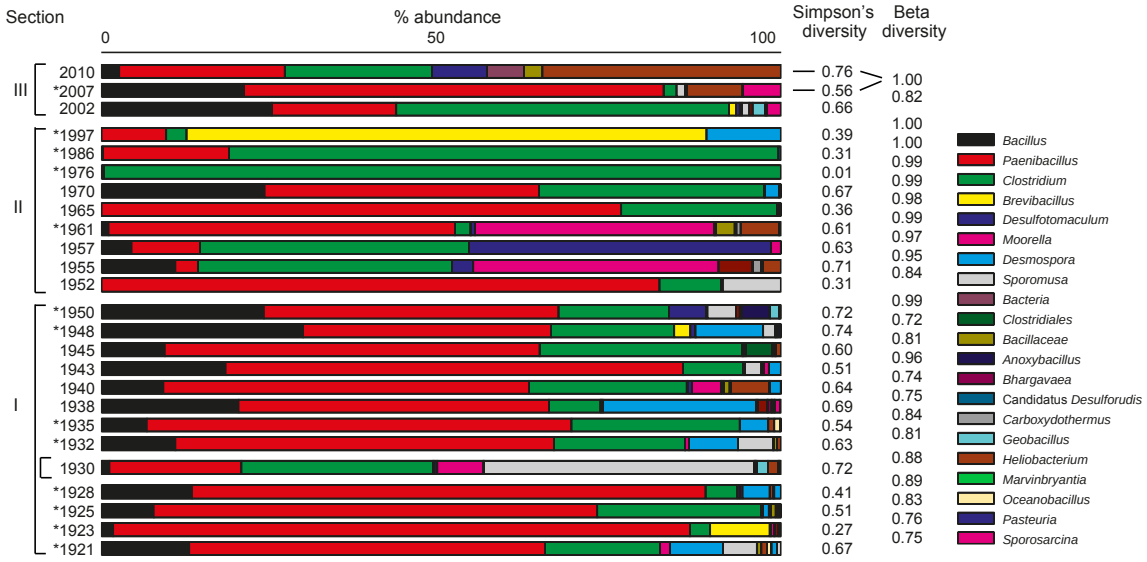


Figure 5.1: Distribution of endospore-forming genera over time. The Simpson diversity index calculated for each sample and the beta diversity index between two successive ages are indicated on the right of the bars.

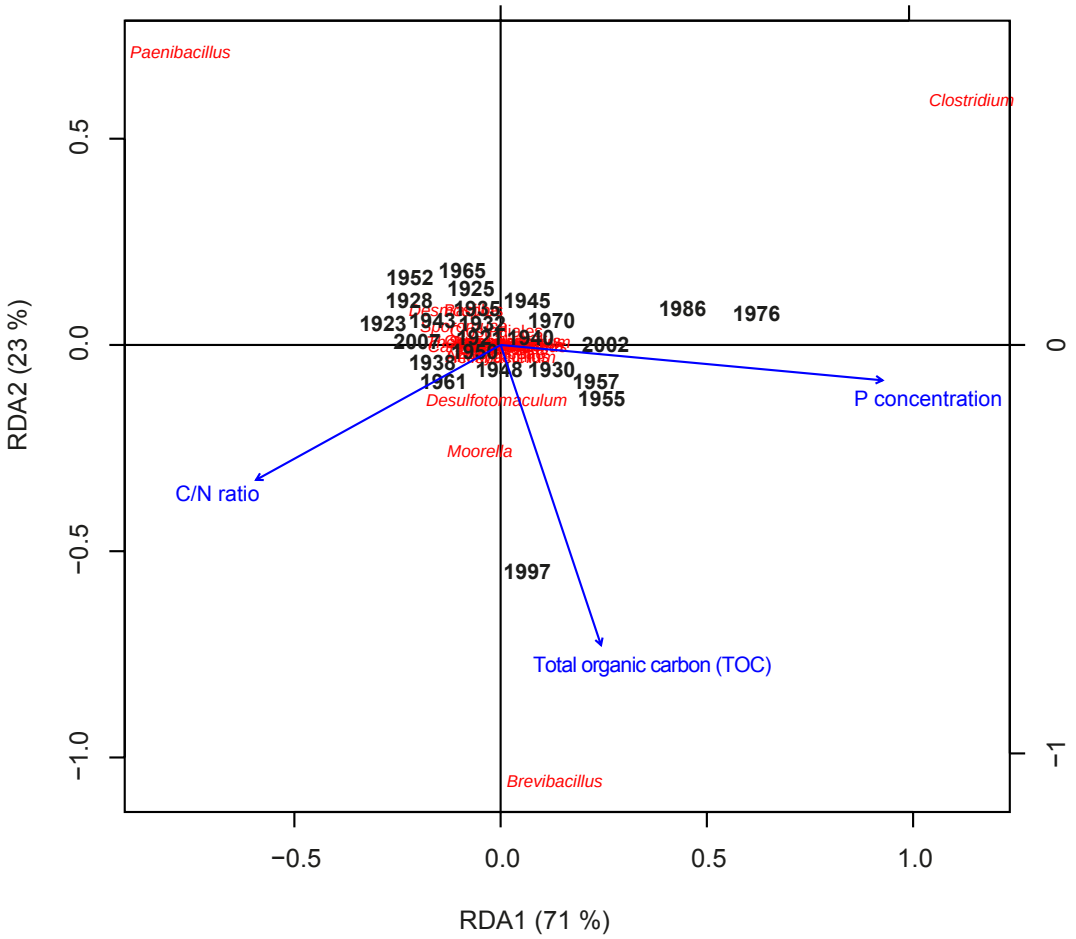


Figure 5.2: Redundancy analysis (RDA) plot of the community structure (genera composition per year) and the environmental parameters consisting of C/N ratio, total organic carbon (TOC), and a measure of phosphorous in the water (medium, low and high) according to Lazzarotto and Klein (Lazzarotto et al. 2012). The year 2010 was omitted in this analysis.

of the genera. The communities from 1955, 1957, and 2002 are further away from this cluster. The samples from 1976 and 1986 were removed from the center on the first axis (RDA1), and in the direction of the largest influence of the P concentration. This offset was linked to the dominance of *Clostridium*. The sample from 1997 is then placed apart from the other samples on the second axis into the direction of large influence of the total organic carbon (TOC) values of the sediment and influenced by the genus *Brevibacillus*. The combined analysis of the community structure and the ordination plots indicated a clear differentiation of a set of samples for which a more detailed analysis was carried out.

In the case of the samples from 1955 and 1957, the increase in the dominance of sulfate-reducing endospore-formers such as *Desulfotomaculum* or *Moorella* suggested a change in the content of sulfate either in the water column or entering the sediment. Since sulfate-reduction can lead to the production of insoluble metal

sulfides, this hypothesis was verified by comparing the content of sulfur (S) in the sediment core to the prevalence of specific genera (Figure 5.3A). Other non-sulfate-reducing genera were included as well in the analysis. The results showed that while the variation in the prevalence of groups such as *Bacillus*, *Paenibacillus* or *Clostridium* is independent of S, the peak of prevalence of *Desulfotomaculum* and *Moorella* is correlated with an increase in S registered in the sediment from 1952 to 1957. However, a second peak in S content around the year 1965 does not correlate with these bacterial groups.

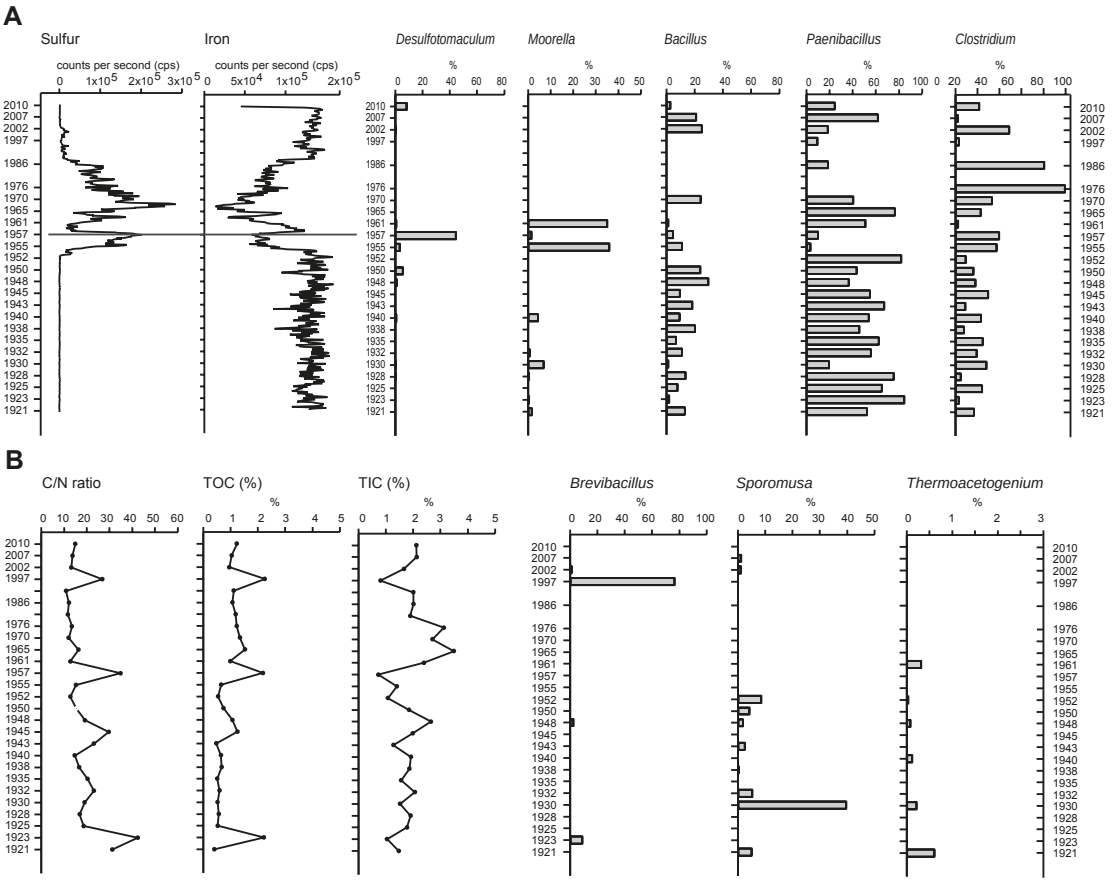


Figure 5.3: Correlation of specific environmental parameters and selected genera of endospore-forming bacteria. **A**. Correlation of sulfur and iron profiles in the sediment and the percentage contribution of *Desulfotomaculum*, *Moorella*, *Bacillus*, *Paenibacillus*, and *Clostridium* to the total endospore-forming community. A line showing the correlation of a sulfur peak in 1957 and the precipitation of iron in the top sediment is indicated. **B**. Correlation of carbon and nitrogen ratios (C/N ratio), percentage of total organic carbon (TOC) and total inorganic carbon (TIC) and the percentage contribution of *Brevibacillus*, *Sporomusa*, and *Thermoacetogenium* to the total endospore-forming community.

A second parameter analysed in detail was the carbon content in the sediments, in particular in relationship with the community changes occurring in the samples

from 1997 and 1930. Indeed, the prevalence of *Brevibacillus*, which explained the differentiation of the sample from 1997, was correlated with peaks in TOC and C/N ratio registered in the sediment from 1923, 1945, 1957, and 1997. In comparison this was not the case of other genera such as *Sporomusa* and *Thermoacetogenium* (Figure 5.3B). The latter two genera were included in the analysis because they explained the differentiation of the sample from 1930. However, none of the previous environmental variables (sulfur, TOC, C/N ratio, and total inorganic carbon (TIC)) could explain the changes in the endospore-forming community registered at this period.

Activity versus dormancy of endospore-forming bacteria at different sediment depths

The results presented until here were based on the total fraction of sporulation genes, corresponding to a mix of vegetative cells and inactive endospores. In order to focus on the inactive fraction, a treatment to remove all vegetative cells and thus isolate endospores was applied in selected samples. The treatment was applied to four samples corresponding to different years, precisely 1921, 1930, 1976, and 2010. The samples from 1921 and 2010 corresponded to the oldest and most recent periods, expected to give the most distinct images of the active and dormant communities. In addition, the samples from 1930 and 1976 were selected because they corresponded to significant events in the sediment record that could not be entirely explained by the factors analysed previously. The summary of the results of amplicon sequencing in the treated samples is shown in Table 5.2. The Simpson index (based on OTUs) for all four treated samples is exceptionally high in the inactive fraction from 1930 and 1921. The number of OTUs and singletons was significantly higher in the treated samples than in the untreated (t-test, p = <0.001).

Table 5.2: Summary of molecular data for samples corresponding to the endospore fraction in the sediment. The *spo0A* gene numbers is given in x 10^3.

	2010	1976	1930	1921
DNA (ng/g sediment)	108.4	117.6	55.6	32.4
% DNA in treated	1	1.4	4	4.1
spo0A (gene number/g sediment)	0.4± 0.2	83.9± 1.9	111.1± 4.3	9.1± 1.4
% *spo0A* in treated	0.5	123	261	74
Nr of sequences	1529	2786	1970	2480
Nr of OTUs	349	566	643	739
Nr of singletons	161	287	308	362
Highest abundance	10	16	3	2
Simpson index (OTU-level)	0.983	0.97	0.99	0.99

The community composition in the treated sediments (inactive endospores) was then compared to the untreated (cells and endospores) samples mentioned before (Figure 5.4). Some genera could only be detected in the endospore fraction. These were *Marvinbryantia*, *Anaerofustis*, and *Alicyclobacillus*. Other genera, such as *Turicibacter* and Candidatus *Desulforudis* were not detected after the treatment.

In the oldest sediment sample (1921), the untreated and inactive communities were very similar in both cases showing the dominance of *Paenibacillus* and in lesser extent of *Bacillus*, *Clostridium*, and *Desmospora*. A similar community was found in the endospore fraction of 1930 sample, with the exception of a higher abundance of *Desulfotomaculum* and the presence of *Sporomusa* and *Moorella*. In contrast, the total *spo0A* community from 1930 was entirely different. The abundances of *Clostridium*, *Moorella*, *Heliobacterium*, and particularly *Sporomusa* abundances were substantially higher, while *Paenibacillus*, *Desulfotomaculum*, and *Bacillus* were highly reduced.

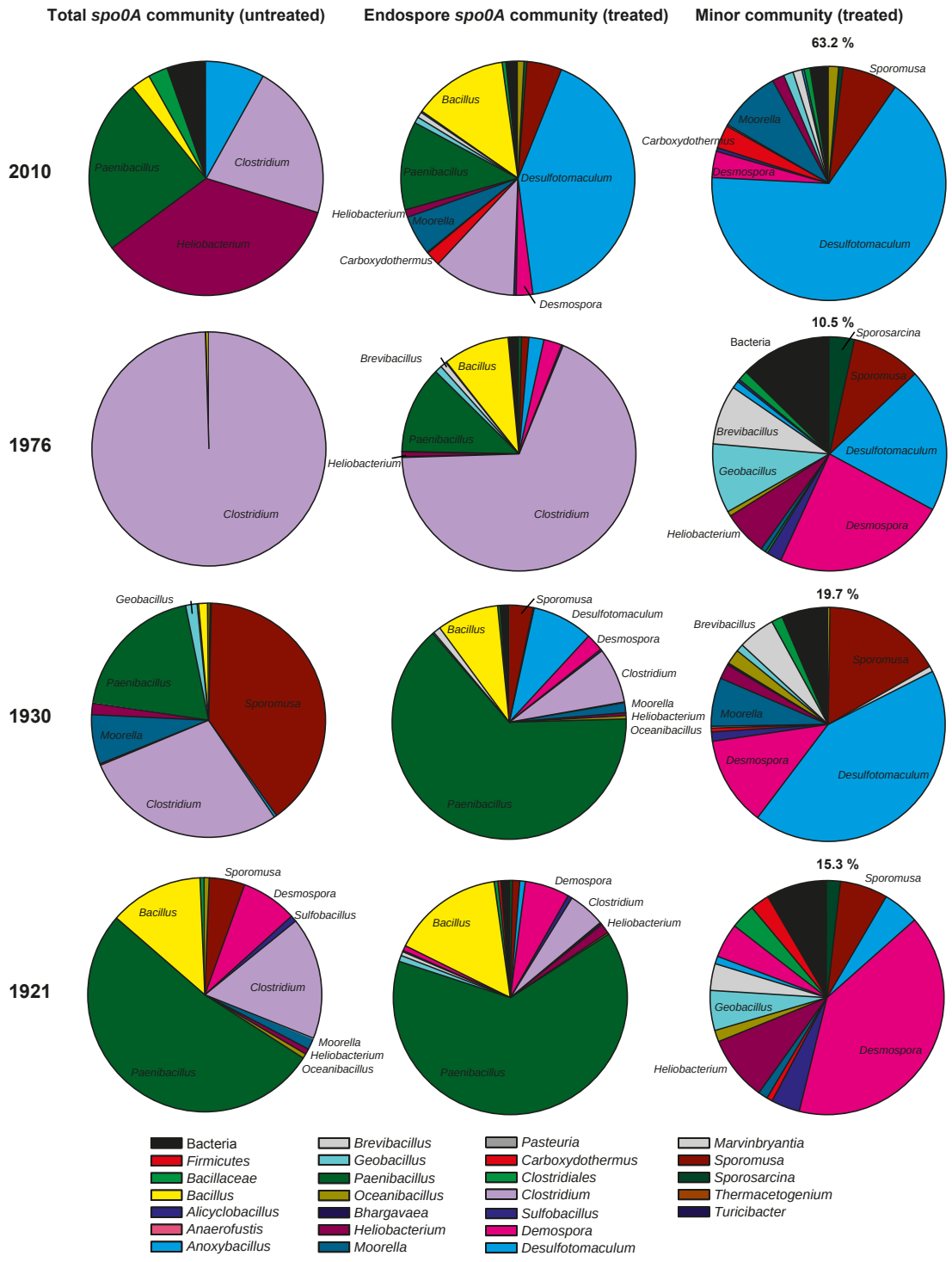


Figure 5.4: Pie charts displaying the community composition of the total (untreated samples; left column), endospore (treated sample; middle column), and seed bank fractions (minor community treated; left column) of endospore-forming bacteria in 2010, 1976, 1930, and 1921. The minor fraction corresponds to all members of the community except the most abundant (*Bacillus, Paenibacillus*, and *Clostridium*). The cumulative percentage of the minor fraction is shown on top of the respective pie chart.

These results once more point to a change around the year 1930 that fundamentally affected the active endospore-forming community. To support this specific effect, we compared the active communities of the samples before (1928) and after 1930 (1932) (Supplementary Figure 5.7). In both samples, the community was similar to the active community from 1921 and the endospore communities in 1921 and 1930, and thus one might expect a similar pattern for the endospore fraction surrounding the event in 1930.

The active community from 1976 was very different from any other in Figure 5.4 as this sample was almost entirely dominated by *Clostridium*. In the endospore fraction of this sample, *Clostridium* was still dominant but to a lesser extent. A larger diversity was found (Simpson index changing from 0.01 in the active versus 0.97 in the endospore fraction), with community members of genera *Bacillus*, *Brevibacillus*, *Geobacillus*, *Paenibacillus*, *Heliobacterium*, *Desmospora*, *Desulfotomaculum*, and *Sporomusa* being present.

In the most recent sediment sample (2010) the communities composed of active cells and endospores and of only endospores were highly different. In the active cell community, *Heliobacterium*, *Clostridium*, and *Paenibacillus* 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*. Also *Bacillus*, *Paenibacillus*, and a small abundance of *Heliobacterium*, *Desmospora*, and *Sporomusa* were observed. In addition, this was the only sample harboring *Carboxydothermus*. The constrained ordination plot (Supplementary Figure 5.8) of the active cell and endospore communities at the four time points showed closest distances for the two samples from each year, with the exception of 1930 (treated), that clustered more closely to the two samples from 1921. Here, the environmental constraints used were the P concentrations, which pointed into the same orientation as the 1976 samples. The number of *spo0A* genes was inversely correlated with the 2010 samples. The highest variation of C/N ratio in the sediment was placed in the same quartile as the oldest sediment samples.

In all the samples a small fraction of the community (10-19% except for 2010) corresponded to a series of minor groups that were constantly found in the endospore fraction. Those genera corresponded to the "seed bank", a series of endospores that get deposited in the sediment with minor to no *in situ* activity. The seed bank was composed of members from 19 different genera (Figure 5.4), such as *Desulfotomaculum*, *Desmospora*, *Geobacillus*, *Sporomusa*, and *Brevibacillus*. The seed bank is of ecological importance as a source of specialized groups whose abundances vary depending on the environmental conditions. This was clearly the case for *Desulfotomaculum* in 1957 and 2010, *Brevibacillus* in 1997, or *Sporomusa* and *Moorella* in 1930.

5.4 Discussion

Isolation of inactive endospores from lake sediments has been used to infer past and present environmental conditions (Renberg and Nilsson 1992; Hubert et al. 2010; de Rezende et al. 2013). In contrast, sequencing studies to detect the diver-

sity of endospore-forming bacteria in sediments have only been developed in the research for this thesis. With this method, a shift in the endospore-forming community could be related to eutrophic conditions of Lake Geneva between 1960 and 1990 (see Chapter 4). In the study presented here, diversity of endospore-forming bacteria from a sediment core spanning 1921 to 2010 was revealed by targeted sequencing of the sporulation gene *spo0A*. This gene is specific to endospore-forming bacteria and a few asporogenic bacteria from the phylum Firmicutes (Galperin et al. 2012; Abecasis et al. 2013). In addition, a treatment to destroy vegetative cells and single out the endospores (detailed description of the method in Chapter 3) was applied to four samples from the core (corresponding to the years 1921, 1930, 1976 and 2010). This targeted approach revealed the diversity of the dormant endospore fraction in the sediment. Both communities (the total and the inactive fraction) were compared with chemical sediment parameters or historical records of lake conditions.

The diversity of endospores and cells from endospore-forming bacteria was analysed in an approach analogous to the common approach used to study bacterial communities based on the 16S rRNA gene sequencing (Pace 2009). However, the advantage of selecting this specific bacterial group is the amount of information regarding metabolic capabilities that can be obtained thanks to the existence of a large collection of culturable representatives (Klenk and Goker 2010). The definition of OTUs from the *spo0A* sequences and their classification into meaningful units is nevertheless a challenge since there are well-known conflicts in the phylogenetic relationships of this group (Collins et al. 1994; Yutin and Galperin 2013) and the analysis of environmental strains using this functional gene is still limited (Brill and Wiegel 1997; Rueckert et al. 2006, and this thesis). Nonetheless, prominent shifts in the community of endospore-forming bacteria in the sediment were seen in certain years. The years prior to eutrophication in Lake Geneva were marked by a constant community distribution with the exception of the community in 1930. In this year a remarkably different community was detected, with a shift towards high abundance of acetogenic *Sporomusa* and fermentative *Clostridium*, both anaerobic bacteria previously isolated from lake sediments (Breznak 2006; Wiegel et al. 2006). *Sporomusa* are homoacetogens that require CO_2 and H_2 for growth and thus are frequently reported as growing in syntrophic associations (Cord-Ruwisch and Ollivier 1986). 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 (Kotsyurbenko et al. 1995; Simankova et al. 2000). Therefore, a possible explanation for the increase in *Sporomusa* in 1930 could be favourable environmental conditions of anoxia at the sediment-water interface and cold temperatures due to low temperatures registered in the winter 1929 that generate little mixing in the water column and partial lake freezing (Maurer 1929; Hendricks-Franssen and Scherrer 2008). Clearly, the effect was pronounced on the active fraction of the community, as the analysis of the endospore-fraction did not reveal an increase in *Sporomusa*.

During the period from 1952 to 1961, the community was less stable than before, with important fluctuations of the dominant members. In 1952, *Paeni-*

bacillus dominates the community. Contrary to this, in 1955 and 1961, there were almost no OTUs affiliated to *Paenibacillus*, but instead *Moorella* (1955 and 1961) became abundant. Members from the genus *Moorella* are described as thermophilic, homoacetogenic bacteria (Collins et al. 1994), most likely outcompeting methanogens in these layers. In 1957 the most abundant group was the sulfate-reducing *Desulfotomaculum* (Rabus et al. 2013). Acetate utilizing sulfate reducers such as *Desulfotomaculum* have been shown to outcompete methanogens in lake sediments, when sulfur concentrations and organic matter load increase (Jones and Simon 1985). Also, in such circumstances the acetogenic bacteria, able to co-exist with sulfate reducers and methanogens, may play a more important role in the carbon cycle (Lever 2012). In the year 1957, the *Desulfotomaculum* peak coincides with high sulfur values in the sediment and increased iron values, likely due to co-precipitation of iron with reduced sulfide in the form of pyrite (FeS) (Berner 1970). Even though organic matter input during the eutrophication period was larger, no activity of sulfate-reducing endospore-forming bacteria can be seen during this period.

In the year 1976, when the P values were at their maximum in the lake, the community was largely dominated by members from the anaerobic genus *Clostridium*. There dominance suggests anoxia at the sediment-water interface during that time. The dominance of *Clostridium* was also observed in the endospore fraction of the community in 1976, even though the remaining community was largely different from the active fraction. This data supports that the community corresponding to year 1976 does indeed reflect the conditions at the time of sediment burial and not the conditions of today at 27 cm sediment depths. In addition, the endospore community in 1976 was also very different from the endospore-fraction in 2010 and 1930, alluding to a variation in the communities at sediment surface before burial started and sporulation had set in. Apart from high P values, the year 1976 also marked extremely low annual discharge (Loizeau and Dominik 2000).

The year 1997 was again a particular year, where the community had completely shifted towards high dominance of *Brevibacillus*. This is a common metal resistant bacterium isolated from soils (Vivas et al. 2003; de Oliveira et al. 2004). The dominance of *Brevibacillus* is potentially related to allochthonous input of organic material as was confirmed by the positive correlation between *Brevibacillus* and the TOC values in the sediment. Also, in 1997 there is a peak in the C/N value, which indicated influence of terrestrial organic matter as opposed to aquatic organic matter (Meyers 2003).

In the recent years (2002 and 2007) the community shifted again towards evenly distributed compositions with *Paenibacillus* and *Clostridium*. The community of the top sediment sample (2010) had high content of *Heliobacterium*, a group of anoxygenic phototrophic bacteria. This genus is interesting because its members have mainly been associated to terrestrial environments (e.g. rice paddy soils), rather than aquatic habitats (Asao and Madigan 2010). In our samples however, members of *Heliobacterium* were present in the vegetative cells fraction and not in the endospore fraction, suggesting that they are active at the lake sediment surface studied here.

An interesting concept has been revealed by the diversity analysis of the in-

active endospores throughout the core. The community could be divided into two parts, the dominant community members (*Bacillus* at overall 9.7%, *Clostridium* at 25.7% and *Paenibacillus* at overall 44.9%). The second part is the "rare biosphere" composed of 23 different genera, representing an overall abundance of 19.7%. Even though the community composition heavily fluctuated in the last 100 years, particularly in respect to the dominant members in each community, the rare members of the inactive endospore community were rather stable. This fraction of the community corresponds to the seed bank that has been described in other studies (Lennon and Jones 2011; Caporaso et al. 2012). The sudden shift of a rare member to become a dominant member at a specific timepoint (for example *Desulfotomaculum* in 1957) confirmed the potential of the rare biosphere as a seed reservoir, where species can suddenly emerge upon environmental change. Additionally the seed bank concept was asserted by the resilience of the system (rapid re-establishment of the community in 1932 after the event in 1930). Endospores as seed banks have been previously studied in cold marine sediments, where the presence of thermophilic endospores was explained by an allochthonous source (de Rezende et al. 2013).

The origin of the endospores (allochthonous or autochthonous) in the sediment studied here cannot be defined conclusively. Evolutionary, Firmicutes are considered as a terrestrial clade of bacteria (Battistuzzi and Hedges 2009) and only a few abundant members of endospore-forming freshwater pelagic bacteria are known (Newton et al. 2011). In this study however, we have detected substantial abundance and diversity of endospore-forming bacteria in Lake Geneva sediments. Furthermore, by comparing the diversity of vegetative cells and endospores, the communities differed and some groups (for example *Heliobacterium, Sporomusa* and *Clostridium*) appeared only in the active fraction of the sediment. Also a BLAST search of *Bacillus thuringiensis* in all sequences had negative results. *B. thuringiensis* endospores are commercially used as insecticides in agriculture. High abundance of this bacterium in the sediment would therefore suggest influence of run-off, which was not the case in our analysis.

The data presented here promote the use of endospores in sediment as promising paleolimnological proxies. By combining the diversity analysis of the vegetative cell fraction together with the dormant endospore fraction, the past environmental history of Lake Geneva could be reconstructed. Endospores are promising new indicators for paleoecology that could also be applied to other freshwater systems, marine sediments, ice cores and deep sediments spanning longer timescales.

5.5 Supplementary information

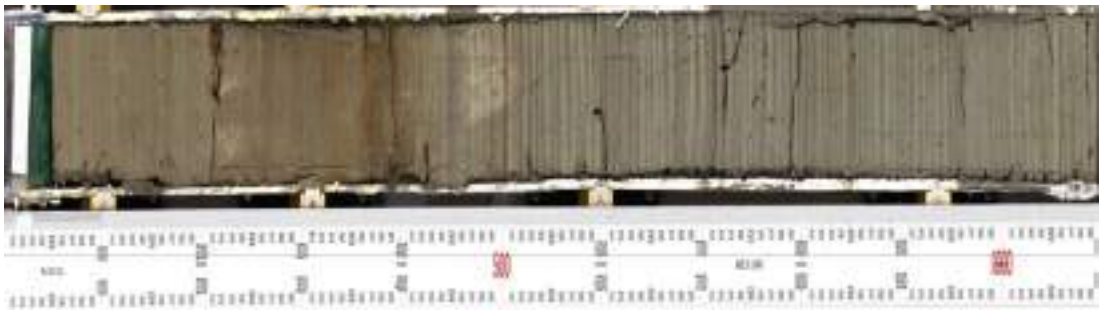


Figure 5.5: Image of sediment core transect from top (left) to bottom (right) showing varved sediment layers without turbidities.

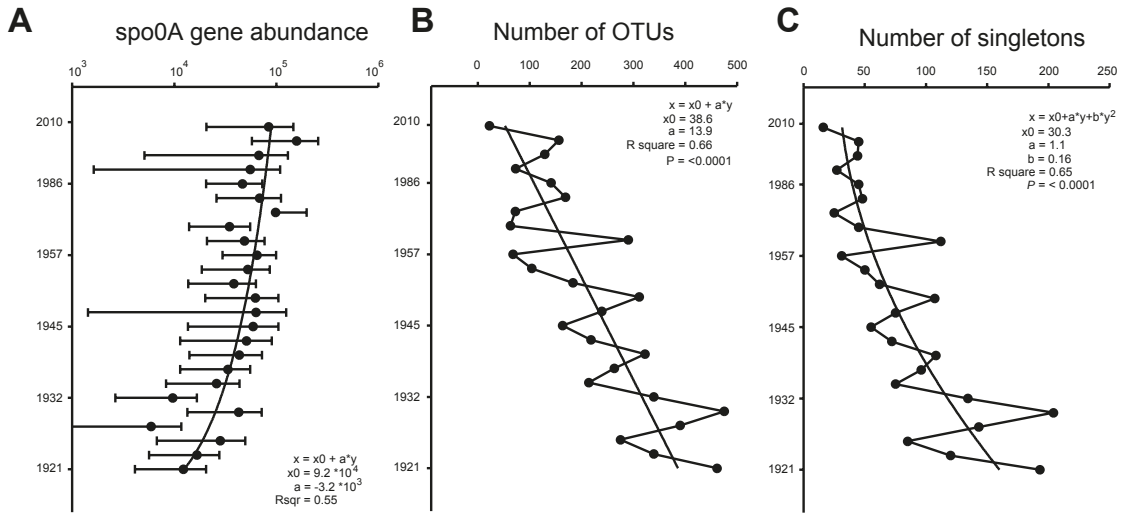


Figure 5.6: Regression lines for *spo0A* gene abundance, the number of OTUs and the number of singletons with sediment depth.

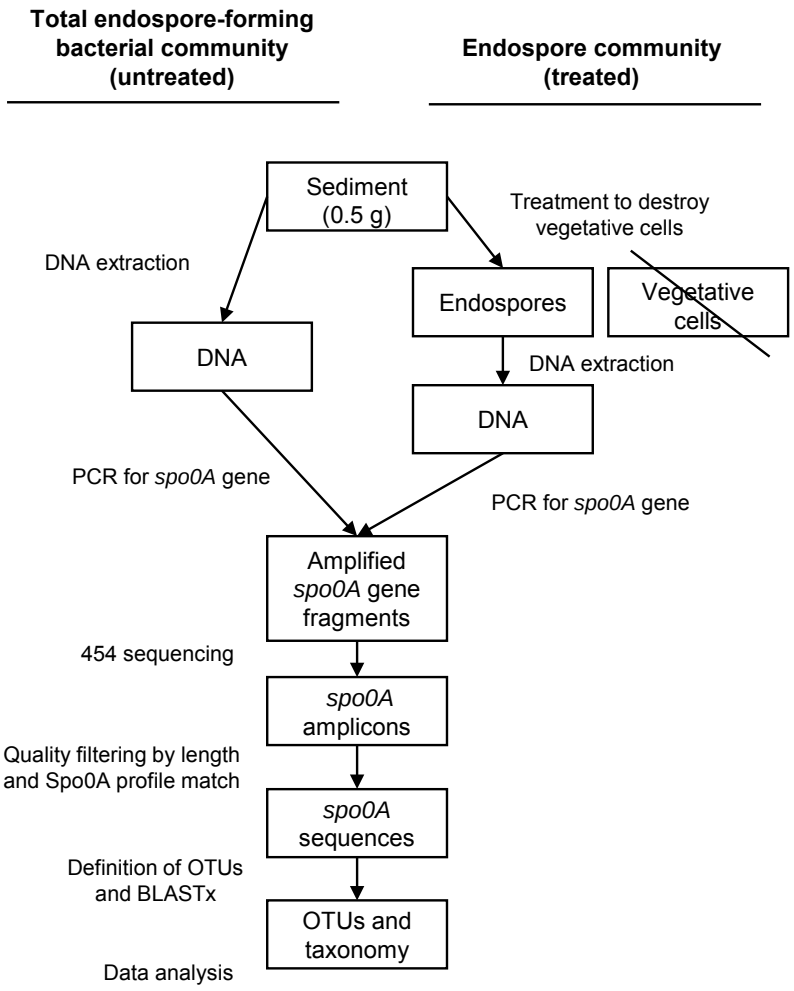


Figure 5.7: Experimental procedure to determine the diversity of the endospore-forming bacteria and endospores in sediment.

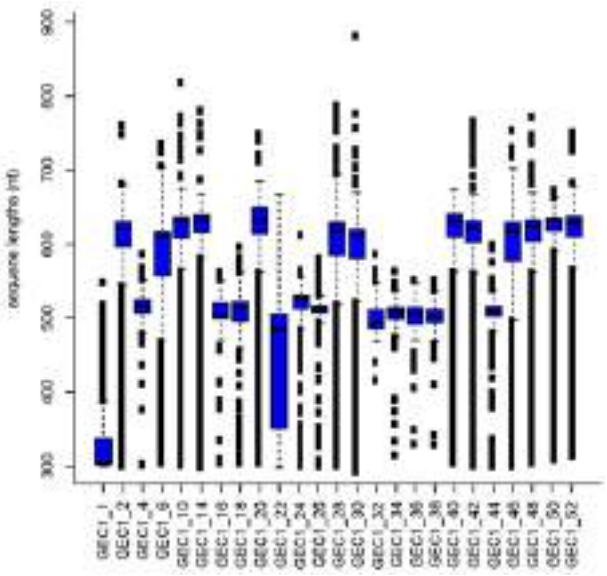


Figure 5.8: Length distribution of *spo0A* sequence reads of all samples of the sediment core.

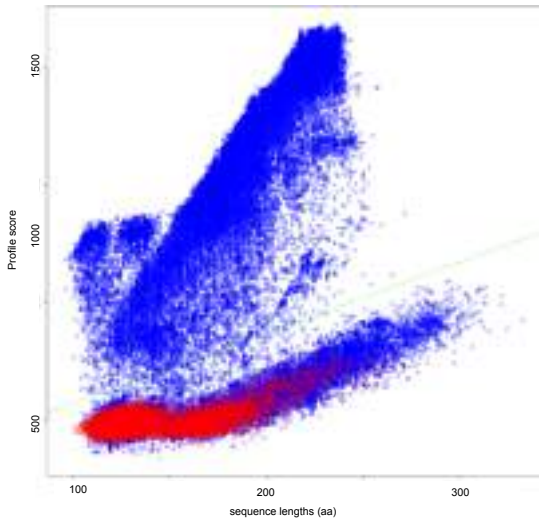


Figure 5.9: Dot-plot of lengths versus profile score for quality filtering. Blue dots are real sequences and red dots are randomly shuffled sequences. All sequences below the green line were removed from the dataset.

Table 5.3: Distribution of genera by number of sequences and percent abundance in brackets for untreated samples (2010 to 1961).

	2010	2007	2002	1997	1986	1976	1970	1965	1961
Bacteria	2 (5.4)	2 (0.05)	4 (0.5)	0	0	3 (0.07)	0	0	9 (0.2)
Bacillaceae	1 (2.7)	0	0	0	0	0	0	1 (0.08)	102 (2.7)
Bacillus	1 (2.7)	801 (21.0)	212 (25.1)	1 (0.03)	13 (0.3)	15 (0.4)	250 (24.0)	0	42 (1.1)
Brevibacillus	0	0	9 (1.1)	2248 (76.4)	0	0	1 (0.1)	1 (0.08)	1 (0.03)
Desmospora	0	0	3 (0.4)	322 (10.9)	0	0	21 (2.0)	3 (0.3)	1 (0.03)
Geobacillus	0	0	16 (1.9)	1 (0.03)	0	0	3 (0.3)	0	0
Heliobacterium	13 (35.1)	312 (8.2)	1 (0.1)	0	0	0	0	0	211 (5.6)
Oceanobacillus	0	0	0	0	0	0	1 (0.1)	0	1 (0.03)
Paenibacillus	9 (24.3)	2360 (61.7)	155 (18.3)	280 (9.5)	779 (18.5)	3 (0.07)	420 (40.4)	905 (76.4)	1911 (50.9)
Pasteuria	0	2 (0.05)	1 (0.1)	0	0	0	0	0	0
Sporosarcina	0	215 (5.6)	18 (2.1)	0	0	0	0	0	0
Clostridium	8 (21.6)	74 (1.9)	413 (48.9)	89 (3.0)	3394 (80.7)	4267 (99.5)	344 (33.1)	272 (23.0)	87 (2.3)
Anoxybacillus	0	12 (0.3)	0	0	0	0	0	0	11 (0.3)
Candidatus *Desulforudis*	0	0	0	0	19 (0.5)	0	0	0	0
Carboxydothermus	0	0	0	0	0	0	0	0	23 (0.6)
Desulfotomaculum	3 (8.1)	0	4 (0.5)	0	0	0	0	2 (0.2)	25 (0.7)
Marvinbryantia	0	0	0	0	0	0	0	0	0
Moorella	0	0	0	0	0	0	0	0	1319 (35.1)
Sporomusa	0	45 (1.2)	9 (1.1)	0	0	0	0	0	0
Sulfobacillus	0	0	0	0	0	0	0	0	0
Thermacetogenium	0	0	0	0	0	0	0	0	12 (0.3)
Total Nr of sequences	37	3823	845	2941	4205	4288	1040	1184	3755

Table 5.4: Distribution of genera by number of sequences (continued) (1957 to 1940).

	1957	1955	1952	1950	1948	1945	1943	1940
Bacteria	0	71 (5.0)	0	18 (0.6)	3 (0.1)	6 (0.2)	3 (0.1)	7 (0.3)
Bacillaceae	1 (0.1)	2 (0.1)	0	9 (0.3)	5 (0.2)	2 (0.08=	0	19 (0.8)
Bacillus	46 (4.6)	156 (10.9)	2 (0.08)	760 (23.9)	713 (29.7)	231 (9.4)	498 (18.3)	231 (9.1)
Bhargavaea	0	0	0	0	0	0	0	3 (0.1)
Brevibacillus	0	0	0	0	57 (2.4)	0	0	3 (0.1)
Desmospora	0	0	5 (0.2)	6 (0.2)	238 (9.9)	3 (0.1)	0	1 (0.04)
Geobacillus	0	4 (0.3)	0	42 (1.3)	0	5 (0.2)	8 (0.3)	1 (0.04)
Heliobacterium	0	38 (2.7)	0	2 (0.06)	0	18 (0.7)	0	142 (5.6)
Oceanobacillus	0	0	0	5 (0.2)	0	0	0	1 (0.04)
Paenibacillus	101 (10.0)	48 (3.4)	2062 (82.0)	1376 (43.3)	875 (36.5)	1349 (55.0)	1826 (67.2)	1358 (53.8)
Sporosarcina	0	0	0	0	0	0	19 (0.7)	0
Clostridiales	0	0	0	0	0	96 (3.9)	0	0
Clostridium	400 (39.6)	532 (37.3)	229 (9.1)	516 (16.3)	434 (18.1)	729 (29.7)	240 (8.8)	585 (23.2)
Anoxybacillus	0	0	1 (0.04)	130 (4.1)	10 (0.4)	7 (0.3)	0	0
Candidatus *Desulforudis*	0	0	0	0	0	0	0	0
Carboxydothermus	0	16 (1.1)	0	1 (0.03)	0	1 (0.04)	0	2 (0.08)
Desulfotomaculum	449 (44.4)	44 (3.1)	0	172 (5.4)	18 (0.8)	0	7 (0.3)	16 (0.6)
Marvinbryantia	0	0	0	0	0	0	0	0
Moorella	14 (1.4)	514 (36.0)	0	0	1 (0.04)	4 (0.2)	0	108 (4.3)
Sporomusa	0	0	215 (8.6)	133 (4.2)	43 (1.8)	0	66 (2.4)	5 (0.2)
Sulfobacillus	0	0	0	5 (0.2)	1 (0.04)	0	50 (1.8)	41 (1.6)
Thermacetogenium	0	0	1 (0.04)	0	2 (0.08)	0	0	3 (0.1)
Total Nr of sequences	1011	1426	2515	3175	2400	2451	2717	2526

Table 5.5: Distribution of genera by number of sequences (continued) (1938 to 1921).

	1938	1935	1932	1930	1928	1925	1923	1921
Bacteria	34 (1.4)	0	4 (0.1)	1 (0.02)	0	11 (0.3)	16 (0.5)	3 (0.1)
Bacillaceae	0	1 (0.05)	14 (0.5)	5 (0.1)	0	24 (0.7)	0	17 (0.5)
Bacillus	479 (20.2)							
	138 (6.7)	326 (11.0)	57 (1.2)	487 (13.4)	265 (7.8)	59 (1.8)	405 (13.0)	
Bhargavaea	13 (0.6)	0	0	2 (0.04)	0	10 (0.3)	0	0
Brevibacillus	2 (0.08)	0	2 (0.07)	9 (0.2)	0	4 (0.1)	285 (8.8)	0
Desmospora	533 (22.5)	85 (4.1)	212 (7.1)	2 (0.04)	145 (4.0)	30 (0.9)	0	242 (7.8)
Geobacillus	1 (0.04)	1 (0.05)	0	74 (1.6)	0	13 (0.4)	10 (0.3)	0
Heliobacterium	9 (0.4)	18 (0.9)	17 (0.6)	73 (1.6)	17 (0.5)	2 (0.06)	0	24 (0.8)
Oceanobacillus	0	18 (0.9)	1 (0.03)	2 (0.04)	5 (0.1)	0	0	20 (0.6)
Paenibacillus	1080 (45.6)	1284 (62.4)	1656 (55.7)	900 (19.4)	2742 (75.5)	2231 (65.2)	2753 (84.8)	1632 (52.3)
Pasteuria	1 (0.04)	1 (0.05)	0	0	0	0	0	1 (0.03)
Sporosarcina	17 (0.7)	2 (0.1)	0	8 (0.2)	0	3 (0.1)	0	0
Clostridiales	0	0	0	0	0	0	0	1 (0.03)
Clostridium	180 (7.6)	509 (24.7)	571 (19.2)	1306 (28.2)	170 (4.7)	823 (24.1)	96 (3.0)	527 (16.9)
Candidatus *Desulforudis*	0	0	0	0	0	0	0	0
Carboxydothermus	0	0	0	9 (0.2)	0	0	0	2 (0.06)
Desulfotomaculum	6 (0.3)	0	0	17 (0.4)	14 (0.4)	0	6 (0.2)	0
Moorella	0	0	18 (0.6)	317 (6.8)	14 (0.4)	5 (0.2)	14 (0.4)	47 (1.5)
Sporomusa	6 (0.3)	0	154 (5.2)	1841 (39.7)	3 (0.08)	0	4 (0.1)	154 (4.9)
Sulfobacillus	6 (0.3)	0	1 (0.03)	0	37 (1.0)	0	5 (0.2)	27 (0.9)
Thermacetogenium	0	0	0	10 (0.2)	0	0	0	19 (0.6)
Total Nr of sequences	2368	2057	2976	4633	3634	3421	3248	3121

Table 5.6: Distribution of genera in treated samples by number of sequences and percent abundance in brackets.

	2010	1976	1930	1921
Bacteria	25 (1.6)	37 (1.3)	25 (1.3)	32 (1.3)
Firmicutes	0	0	0	10 (0.4)
Bacillaceae	8 (0.5)	4 (0.1)	6 (0.3)	13 (0.5)
Bacillus	201 (13.2)	254 (9.1)	168 (8.5)	384 (15.5)
Alicyclobacillus	0	1 (0.04)	0	0
Bhargavaea	0	0	1 (0.05)	0
Brevibacillus	12 (0.8)	24 (0.9)	20 (1.0)	14 (0.6)
Desmospora	35 (2.3)	70 (2.5)	48 (2.4)	153 (6.2)
Geobacillus	13 (0.9)	28 (1.0)	4 (0.2)	21 (0.9)
Heliobacterium	16 (1.1)	18 (0.7)	8 (0.4)	34 (1.4)
Oceanobacillus	0	2 (0.07)	8 (0.4)	6 (0.2)
Paenibacillus	187 (12.2)	336 (12.1)	1263 (64.1)	1590 (64.1)
Pasteuria	2 (0.1)	0	1 (0.05)	0
Sporosarcina	6 (0.4)	10 (0.4)	0	7 (0.3)
Clostridiales	0	1 (0.04)	0	0
Clostridium	174 (11.4)	1904 (68.3)	151 (7.7)	127 (5.1)
Anaerofustis	0	0	0	18 (1.0)
Anoxybacillus	3 (0.2)	3 (0.1)	0	4 (0.2)
Candidatus *Desulforudis*	0	0	0	0
Carboxydothermus	31 (2.0)	0	2 (0.1)	3 (0.1)
Desulfotomaculum	640 (41.9)	58 (2.1)	166 (8.4)	19 (0.8)
Marvinbryantia	0	0	3 (0.2)	0
Moorella	84 (5.5)	2 (0.07)	26 (1.3)	5 (0.2)
Sporomusa	74 (4.8)	28 (1.0)	64 (3.3)	25 (1.0)
Sulfobacillus	5 (0.3)	6 (0.2)	5 (0.3)	15 (0.6)
Thermacetogenium	13 (0.9)	0	1 (0.05)	0
Total Nr of sequences	1529	2786	1970	2480

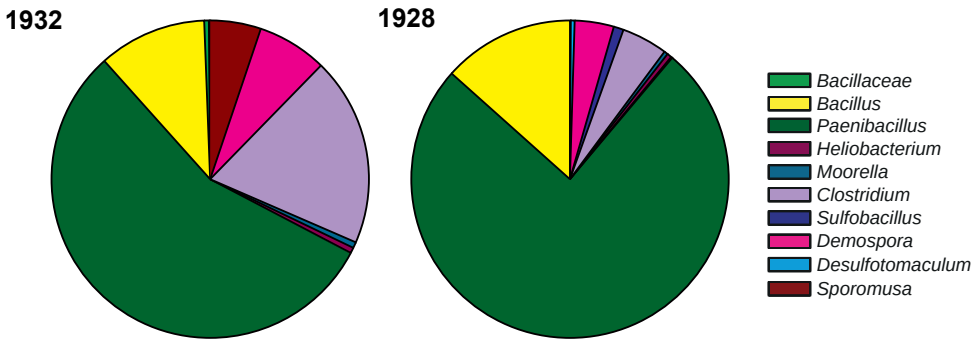


Figure 5.10: Pie chart of the active cell fraction of samples from 1928 and 1932.

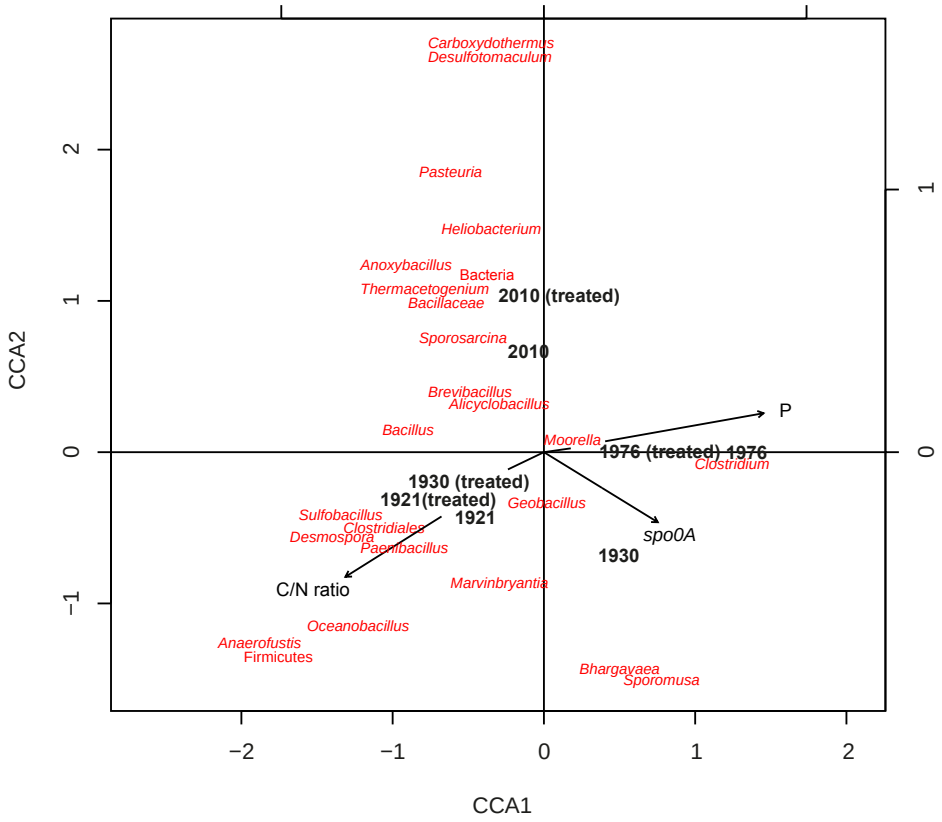


Figure 5.11: Canonical correspondence analysis of samples from 1921, 1930, 1976 and 2010 untreated (total endospore-forming bacterial community) and treated (dormant endospore community). Constraints are done with phosphorous (P) values obtained from Lazzarotto and Klein (2012), *spo0A* gene quantities, and ratio of carbon to nitrogen in the sediment (C/N ratio).

5.6 References

Abecasis, A. B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J. B., and Henriques, A. O. (2013). A genomic signature and the identification of new sporulation genes. *J Bacteriol*, 195(9):2101–15.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev*, 59(1):143–69.

Anneville, O., Ginot, V., Druart, J.-C., and Angeli, N. (2002). Long-term study (1974–1998) of seasonal changes in the phytoplankton in Lake Geneva: a multi-table approach. *J Plankton*, 24(10):993–1008.

Asao, M. and Madigan, M. T. (2010). Taxonomy, phylogeny, and ecology of the *Heliobacteria*. *Photosynth Res*, 104(2-3):103–111.

Bartholomew, J. W. and Paik, G. (1966). Isolation and identification of obligate thermophilic sporeforming Bacilli from ocean basin cores. *J Bacteriol*, 92(3):635–8.

Battistuzzi, F. U. and Hedges, S. B. (2009). A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol*, 26(2):335–43.

Berner, R. A. (1970). Sedimentary pyrite formation. *Am J Sci*, 268(1):1–23.

Breznak, J. A. (2006). The genus *Sporomusa*. In Dworking, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, page 991–1001. Springer.

Brill, J. A. and Wiegel, J. (1997). Differentiation between spore-forming and asporogenic bacteria using a PCR and southern hybridization based method. *J Microbiol Methods*, 31(1–2):29–36.

Bueche, M., Wunderlin, T., Roussel-Delif, L., Junier, T., Sauvain, L., Jeanneret, N., and Junier, P. (2013). Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene *spo0A*. *Appl Eviron Microbiol*, 79(17):5302–12.

Caporaso, J. G., Paszkiewicz, K., Field, D., Knight, R., and Gilbert, J. A. (2012). The western english channel contains a persistent microbial seed bank. *ISME J*, 6(6):1089–1093.

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., et al. (1994). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol*, 44(4):812–826.

Cord-Ruwisch, R. and Ollivier, B. (1986). Interspecific hydrogen transfer during methanol degradation by *Sporomusa acidovorans* and hydrogenophilic anaerobes. *Arch Microbiol*, 144(2):163–165.

de Rezende, J. R., Kjeldsen, K. U., Hubert, C. R., Finster, K., Loy, A., and Jorgensen, B. B. (2013). Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J*, 7(1):72–84.

Dressler, M., Hübener, T., Görs, S., Werner, P., and Selig, U. (2007). 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(2):205–219.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., et al. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev*, 81(2):163–182.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–1.

European Union (EU) (2000). *EUR-Lex - 32000 L0060 - EN*. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.

Fichtel, J., Koster, J., Rullkotter, J., and Sass, H. (2007). Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol Ecol*, 61(3):522–32.

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., and Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol*, 14(11):2870–90.

Gerdeaux, D. and Perga, M.-E. (2006). Changes in whitefish scales 13C during eutrophication and reoligotrophication of subalpine lakes. *Limnol Oceanogr*, 51(1 part 2):772–780.

Gorham, E., Brush, G. S., Graumlich, L. J., Rosenzweig, M. L., and Johnson, A. H. (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.

Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci USA*, 84(13):4355–4358.

Hendricks-Franssen, H. J. and Scherrer, S. C. (2008). Freezing of lakes on the Swiss plateau in the period 1901–2006. *Int J Climatol*, 28(4):421–433.

Hubert, C., Arnosti, C., Bruchert, V., Loy, A., Vandieken, V., and Jorgensen, B. B. (2010). Thermophilic anaerobes in arctic marine sediments induced to mineralize complex organic matter at high temperature. *Environ Microbiol*, 12(4):1089–104.

International Commission of the Protection of Lake Geneva (CIPEL) (2010). *Plan d'action 2011-2010 en faveur du Léman, du Rhône et de leurs affluents*. CIPEL.

Jones, J. G. and Simon, B. M. (1985). Interaction of acetogens and methanogens in anaerobic freshwater sediments. *Appl Environ Microbiol*, 49(4):944–948.

Klenk, H. P. and Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol*, 33(4):175–82.

Kotsyurbenko, O. R., Simankova, M. V., Nozhevnikova, A. N., Zhilina, T. N., Bolotina, N. P., Lysenko, A. M., and Osipov, G. A. (1995). New species of psychrophilic acetogens: *Acetobacterium bakii* sp. nov., *A. paludosum* sp. nov., *A. fimetarium* sp. nov. *Arch Microbiol*, 163(1):29–34.

Lazzarotto, J., Klein, A., and Commission pour la Protection des Eaux du Léman (CIPEL) (2012). Physical-chemical changes in the waters of Lake Geneva (major elements). Technical report, Lausanne, Switzerland.

Lennon, J. T. and Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol*, 9(2):119–130.

Lever, M. A. (2012). Acetogenesis in the energy-starved deep biosphere - a paradox? *Front Microbiol*, 2.

Loizeau, J.-L. and Dominik, J. (2000). Evolution of the upper rhone river discharge and suspended sediment load during the last 80 years and some implications for lake geneva. *Aquat Sci*, 62(1):54–67.

Loizeau, J.-L., Dominik, J., Luzzi, T., and 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, B. A., Langerhuus, A. T., D'Hondt, S., Jorgensen, B. B., and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature*, 484(7392):101–4.

Maurer, J. (1929). Der strenge Winter und die große Seegefrörne Februar-März 1929 in der Schweiz. *Das Wetter*, 46.

Meyers, P. A. (2003). Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. *Org Geochem*, 34(2):261–289.

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

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., and Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbio Mol Biol Rev*, 75(1):14–49.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M., Oksanen, M., and Suggests, M. (2007). The vegan package.

Onyenwoke, R. U., Brill, J. A., Farahi, K., and 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.

Ovreas, L., Forney, L., Daae, F. L., and 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.

Pace, N. R. (2009). Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev*, 73(4):565–76.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabus, R., Hansen, T. A., and Widdel, F. (2013). Dissimilatory sulfate- and sulfur-reducing Prokaryotes. In Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E., and Thompson, F., editors, *The Prokaryotes*, page 309–404. Springer Berlin Heidelberg.

Renberg, I. and Nilsson, M. (1992). Dormant bacteria in lake sediments as palaeoecological indicators. *J Paleolimnol*, 7(2):127–135.

Rothfuss, F., Bender, M., and Conrad, R. (1997). Survival and activity of bacteria in a deep, aged lake sediment (Lake Constance). *Microbiol Ecol*, 33(1):69–77.

Rueckert, A., Ronimus, R. S., and Morgan, H. W. (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(3):220–230.

Simankova, M. V., Kotsyurbenko, O. R., Stackebrandt, E., Kostrikina, N. A., Lysenko, A. M., Osipov, G. A., and Nozhevnikova, A. N. (2000). *Acetobacterium tundrae* sp nov, a new psychrophilic acetogenic bacterium from tundra soil. *Arch Microbiol*, 174(6):440–447.

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems: a global problem. *Environ Sci Pollut Res Int*, 10(2):126–139.

Smith, V. H., Tilman, G. D., and Nekola, J. C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ Pollut*, 100(1-3):179–196.

Staley, J. T. and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*, 39:321–46.

Tadonléké, R. D., Lazzarotto, J., Anneville, O., and Druart, J.-C. (2009). Phytoplankton productivity increased in Lake Geneva despite phosphorus loading reduction. *J Plankton Res*, 31(10):1179–1194.

Thevenon, F., Adatte, T., Wildi, W., and Pote, J. (2012). Antibiotic resistant bacteria/genes dissemination in lacustrine sediments highly increased following cultural eutrophication of Lake Geneva (Switzerland). *Chemosphere*, 86(5):468–76.

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., et al. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482(7385):357–62.

Wiegel, J., Tanner, R., and Rainey, F. A. (2006). An introduction to the family *Clostridiaceae*. In Dr Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, page 654–678. Springer US.

Willard, D. A. and Cronin, T. M. (2007). Paleoecology and ecosystem restoration: Case studies from Chesapeake Bay and the Florida Everglades. *Front Ecol Environ*, 5(9):491–498.

Wilson, M. A. and Carpenter, S. R. (1999). Economic valuation of freshwater ecosystem services in the United States: 1971-1997. *Ecol App*, 9(3):772–783.

Yutin, N. and Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced Clostridia. *Environ Microbiol*, 15(10):2631–2641.

6 Synthesis

6.1 Main findings

The main objective of this work was to develop methods for targeted metagenomics to reveal the diversity of endospore-forming bacteria in sediments.

Chapter 2 – The gene *spo0A* as molecular marker

In the first chapter, primers were designed for the molecular diversity study of endospore-forming bacteria in sediment samples.

Individual findings were:

- The gene *spo0A* was confirmed as a molecular marker for sporulation.
- Amplicon sequencing of a 602 bp *spo0A* gene fragment revealed diversity of endospore-forming bacteria in sediments.
- A DNA extraction method based on indirect and repetitive lysis cycles improved the detection (in quantity and diversity) of endospore-forming bacteria.
- Previously infrequently detected genera such as *Geobacillus* or *Paenibacillus* were revealed in high abundance.
- Sequences of *spo0A* were detected that were not closely related to any *spo0A* sequences available in databases, suggesting possible new clades of endospore-forming bacteria.

Chapter 3 – Targeted metagenomics reveals unprecedented diversity of endospores

In this chapter, a new method for targeted metagenomics was promoted based on a physical selection of endospores in sediment.

Detailed findings were:

- A treatment of heat, chemical and enzymatic agents, successfully destroyed vegetative cells, while resistant endospores were unaffected.
- Application of this treatment to sediment samples, successfully targeted endospores for subsequent metagenomics analysis.
- The targeted community was composed of more than 90% Firmicutes, the phylum comprising all known endospore-forming bacteria.
- The detection of endospore-forming bacteria was substantially ameliorated with the targeted metagenomics approach. A large number of sequences were defined at high resolution and in some cases down to species level.
- With the targeted approach the sequence coverage was substantially improved and a large number of members from the rare biosphere was detected.

Chapter 4 – Endospore-forming bacteria as new proxies to assess impact of lake eutrophication

In this chapter the use of endospore-forming bacteria from sediments as paleoecological proxies for eutrophication was established.

Specific findings were:

- Analysis of a sediment core from Lake Geneva revealed dating, chemical and biological parameters to infer the environmental history from 1921 to 2010, covering the eutrophication period from 1960 to 1990.

- The extracted DNA and the total bacterial abundance in the sediment core significantly decreased with age of the sediment.

- On the contrary, the quantification of endospore-forming bacteria with age of the sediment was markedly stable, proving high detection of endospores at older sediment ages.

- The community of endospore-forming bacteria from different sediment ages reflected nutrient load and anoxic conditions, corresponding to the eutrophic lake period.

Chapter 5 – Endospores in sediment as a proxy for environmental lake history

This research confirmed the use of endospores as paleoecological indicators for the last 100 years of environmental history of Lake Geneva.

Individual findings were:

- Endospore-forming bacterial abundance from sediment samples spanning the period between 1921 and 2010 ranged from 22 to 475 taxonomic units.

- Abundance and community composition of endospores reflected the environmental conditions or specific events at the time of sedimentation.

- The community composition in the sediment corresponding to year 1930, had high abundance of *Sporomusa* and *Moorella*, likely in response to an exceptionally cold previous winter.

- The taxonomic richness decreased in the years between 1950 and 1990, when the lake was in eutrophic state. During this period, high abundance of *Clostridum* species was detected.

- In the year 1997 a strong shift in the community from *Paenibacillus* and *Clostridium* to high dominance of *Brevibacillus* was seen, likely related to high input of terrestrial material at this time.

- In the surface sample of the sediment, high abundance of vegetative cells of *Heliobacterium* was detected, an anoxygenic photosynthetic bacteria, previously undocumented to be active in Lake Geneva.

- The targeted metagenomics approach to reveal diversity of the dormant endospore fraction was applied to four samples and revealed differences between the active and dormant community.

- A seed bank component of the endospore-community was seen, reflecting high diversity at low abundance.

These findings confirm the successful development of targeted metagenomics methods for endospore-forming bacteria. The results also established an important diversity of this group in lake sediments and suggested that endospores in sediment can be used as indicators of past environmental lake history.

6.2 Challenges and Perspectives

Even though a targeted metagenomics study already improved some issues of molecular studies such as low coverage or large complexity of the communities, there are still challenges to be solved for the methods used in the research presented in this thesis.

The gene *spo0A*

One important question is the suitability of *spo0A* as a molecular marker for endospore-forming bacteria. Results have shown amplification of a large number of different endospore-forming taxa with the *spo0A* primers. They have however also shown groups that are not or less efficiently targeted using these primers, such as strains from *Acidophilus* or *Sulfobacillus*. Also, it is likely that sequences from the genus *Clostridum* are more difficult to target than sequences from *Bacillus*. This difference is obvious from the comparison of the endospore community based on 16S rRNA and *spo0A* gene sequencing. The fact that *Clostridium* were more difficult to amplify is not surprising, since the primers have initially been designed based on alignments of *Bacillus spo0A* sequences.

Another issue with the primers for *spo0A* is the targeting of asporogenic bacteria. We have amplified and sequenced bacteria from genera, where sporulation has not been confirmed experimentally (asporogenic bacteria). These bacteria have a putative *spo0A* gene but it is not known whether they can sporulate. It is possible for bacteria to have sporulation genes, but by having lost some genes in the course of evolution, they can no longer form endospores. Such strains would be detected by targeting the *spo0A* gene, even though the trait of sporulation has been lost. For this reason, it was interesting to apply the second targeting method (physically separate endospore from vegetative cells) in order to ensure that the fraction of the community studied was indeed the endospore-forming fraction. In fact, using

the physical targeting method for endospores, sequences from supposedly asporogenic bacteria were retrieved. However, further analyses will be needed to analyse potential new spore-forming candidates in more detail.

DNA extraction methods

A repetitive DNA extraction protocol and particularly an indirect method is substantially more laborious and costly than a simple direct extraction. Depending on the number of samples to be processed this can be a significant challenge. On the other hand, seeing how much time and budget is invested in downstream processes such as sequencing and data analysis, it might be worth investing more in this initial step. A possible disadvantage of an indirect lysis protocol is the omission of extraction of cells that are tightly associated with the sediment particles and are therefore not released into the solution. The indirect extraction with a dispersing agent is repeated twice on the same sample as to maximize the release of cells and endospores from the sediment matrix, but residual cells cannot be excluded. In a future study, both fractions could be extracted and sequenced to compare the two communities.

For the research in this thesis, the indirect extraction protocol was only applied to the samples from Lake Geneva and Lake Baikal presented in chapters 2 and 3. The sediment samples from the long core of Lake Geneva (chapters 4 and 5) have been extracted using direct *in situ* lysis and three repetitive extraction cycles. This was a choice of time, cost and experience at that time. As is shown in chapter 2, this DNA extraction method (dubbed protocol 2) resulted in substantial quenching of quantitative PCR amplification and in reduced detection of taxonomic units compared to the *ex situ* lysis (indirect extraction protocol 3). It is therefore possible that some members of the endospore-forming community from the long sediment core of Lake Geneva were overlooked because of a less adapted extraction protocol. Nonetheless a substantial amount of information on the diversity and ecology of endospore-forming bacteria was retrieved from these samples.

A single extraction protocol will never be adjusted to every type of sample and every physiology or every morphological state of bacteria. An ideal approach would be to use several different DNA extraction protocols for the same samples to increase the targeting of all types of cells.

Sequencing studies

Another issue is that none of the herein presented samples have been sequenced in replicates. Due to the expensive nature of sequencing technologies, it is uncommon to do replicates. Apart from lacking repeatability, it is challenging and in some cases impossible to use statistical analysis, as one result per sample provides low statistical power. Due to lacking statistical methodology and single sequencing per sample, the analyses are in most cases descriptive. The exception are the analysis of taxa-matrix and environmental constrains for the results of the sediment core samples. Descriptive studies are of high value and inevitable in a research field that is comparably novel, such as microbial molecular diversity. However, as

sequencing studies become more frequent and millions of bases of sequencing data are generated each year, it will be increasingly important to have hypothesis driven research approaches. It soon will not be enough to provide diversity information. What becomes more and more interesting and important is to know the ecological role of bacteria and the circumstances that influence the presence or absence of taxa in a spatial or temporal scale.

Phylogeny and taxonomy

When comparing the endospore-forming communities between the four chapters in this thesis, we can see that the community was not always the same, even though the majority of the samples stemmed from the same lake (Lake Geneva), albeit at different sampling sites. The analysis from chapters 2 and 4 are even based on the identical samples. The difference in community composition can be due to environmental differences (depth of sediment, location) but more importantly due to methodological differences. As mentioned above, the DNA extraction method used in the different studies was not identical. Also, the amplification and purification methods before sequencing varied between samples and studies. The samples were also sequenced in different sequencing projects (different reaction pools), which can lead to variation in the data. Furthermore, the method for sequence analysis, notably quality filtering and curation, has been adapted to samples and current experience at the time of analysis.

But most likely, variations in community composition could be seen because of different methods to define genus affiliation or methods to find the closest relative for each sequence. There were two different methods to identify taxa (based on the *spo0A* sequencing). One of those was based on BLASTx identification (basic local alignment search tool for translated nucleotide sequences) of the closest relative found in the database from the National Center of Biotechnology Information (NCBI). This method was applied to the data in chapter 2, 3, and 5. In contrast, the diversity described in chapter 4 was revealed by a classifier developed specifically for the *spo0A* gene. Essentially, upon a database compiled with known protein sequences of Spo0A retrieved from public databases, a classification was done to define the genus affiliation of each previously defined operational taxonomic unit of the sequence data. Both methods have individual problems, such as for BLAST, the definition of a "closest" relative, even though the sequence identity could be very poor. For the classifier, the method is highly dependent on the compiled database (no errors in annotated Spo0A proteins) and on the phylogenetic relationship of the *spo0A* genes. The phylogeny assessment issue is very important since the phylogeny of endospore-forming bacteria is subject to frequent updates and re-classification and many groups have not been placed with high confidence. This study has revealed substantial problems in the taxonomy. The genera of *Geobacillus* and *Paenibacillus* for example are not conclusively placed in the phylogeny. The cladogram of the *spo0A* sequences has placed these two genera in clades closer to the *Clostridium* than the *Bacillus*. It is possible, that part of the sequences that we have classified as *Geobacillus* and *Paenibacillus* are in reality

belonging to unknown groups of endospore-forming bacteria as mentioned in the synthesis.

The fact that there are differences in community composition depending on above mentioned methodological or environmental parameters supports the claim made herein that choosing an adapted method tailored to the bacteria of interest is highly important for metagenomics community studies. Furthermore, it addresses the current lack of knowledge about the phylogeny of endospore-forming bacteria as well as limited information in publicly available databases.

6.3 Outlook

There are several questions that have been raised with this research that would be interesting to answer. One question is what metabolic activity and genetic potential of metabolism lie in endospore-forming bacteria. Some answers to these questions could be found by shotgun metagenomics sequencing. In some cases, where DNA yield is low (deep sediments) it might be necessary to amplify the metagenomic DNA for example by multiple displacement amplification (MDA) (Raghunathan et al. 2005). Since endospore-forming bacteria are common bacteria for the production of industrial products, such a shotgun metagenomic analysis could also reveal novel and interesting biomolecules or pathways. In sediments, there are still many open questions about the *in situ* activity, and their changes in activity over the course of seasonal variations. *In situ* studies, time-line analysis or growth studies in the laboratory could shed light on some of these questions.

Also, the origin of endospores in the sediment for some members of the community remains obscure. It has been shown that thermophilic endospore-forming bacteria are abundant in cold sediments and that they are locally inactive (de Rezende et al. 2013). In this thesis, several members of thermophilic (or supposedly) thermophilic endospore-formers have been found (*Desulfotomaculum*, *Geobacillus*, and others). However, if there is *in situ* activity of these groups has not been studied and it is difficult to make reliable predictions on this. One simple way to investigate this is to study the endospore-forming bacteria in the water column and in run-off from land or lake tributaries. In order to ensure a causal connection between variation in community compositions, for example, increase in *Clostridium* or sudden dominance of *Sporomusa*, an analogue lake system, where eutrophication has occurred should be studied to verify the same pattern. Also incubation experiments of sediment with specific carbon sources or other nutrients or at specific temperatures would be interesting.

Last but not least, the biogeographic distribution of these bacteria are of high ecological interest. Endospores have been shown to be globally distributed but little is known on diversity across continents apart from a few studies on isolates. Questions to answer could be: Are there specific endospore-forming bacteria as ecotypes (freshwater strains)? Is it possible to find the same species in different sediment? What factors explain the differences (soil as seed bank, terrestrial organic matter, Cyanobacteria or other pelagic bacteria)?

In a biogeographic context, the bacterial communities from sediments of Lake Geneva and Lake Baikal studied for this thesis, are surprisingly similar, even though the properties of the sites are very different. Lake Geneva is heavily influenced by anthropogenic pressure and has high nutrient loads (Thevenon et al. 2011). Also, the sediment rates are much higher than Lake Baikal (Vologina and Sturm 2009). The latter is oligotrophic, extremely deep and entirely frozen in winter months (Chernitsina et al. 2007). This similarity of sediment communities points towards ecological coherence between freshwater sediments (Philippot et al. 2010; Newton et al. 2011), meaning that sediment communities of two lakes, even when thousands of kilometres apart, are more closely related than other habitats such as marine environments or soil.

There is still a substantial information gap to describe freshwater biodiversity and make predictions on vulnerability to anthropogenic disturbance. No global inventory of freshwater bacteria and their function is known. Also, knowledge about extinction, resilience and functional redundancy is missing. Furthermore, most of the bacteria are members of the rare biosphere, making them difficult to reveal and study. Due to this rare biosphere, accurate estimations of species richness are almost impossible, in addition to the difficulty of actually defining species or taxa. In recent years, new threats have emerged for freshwater lakes, related to the release of antibiotics (Czekalski et al. 2012), micropollutants and endocrine active substances (Morasch et al. 2010). This poses new challenges for freshwater ecosystems and conservation management.

6.4 Conclusion

The research presented in this thesis successfully demonstrated the development of methods for targeted metagenomics on endospore-forming bacteria from sediment. It is the first study of endospore-forming bacteria in the environment based on metagenomics sequencing. Prior to this research, endospore-forming bacteria could be quantified in bulk via microscopy or detection of specific biomolecules such as dipicholynic acid (DPA)(Hill and Gray 1967; Fell et al. 2001; Fichtel et al. 2007). Alternatively, endospore-formers were studied by isolation and culturing (Bartholomew and Paik 1966; Ettenauer et al. 2010; Logan et al. 2000; Alawi et al. 2011). Based on molecular studies, the data on endospore-forming bacteria is limited, since they have never been targeted specifically. In the research described here, the endospore-forming bacteria were targeted by either using primers for the sporulation specific gene *spo0A*, or by a physical separation of endospores from vegetative cell prior to DNA extraction and metagenomic sequencing. Using these methods resulted in the detection of high endospore-forming bacterial diversities in lake sediments.

An unprecedented number of endospore-forming taxa have been revealed from sediment samples. A large culture collection of endospore-forming bacteria and an increasing number of fully sequenced genomes (i.e. Junier et al. 2009; Chivian et al. 2008; Li and Xia 2012) have added substantial knowledge on different metabolic properties and genetic potentials of endospore-forming bacteria.

Additionally, and independent of the focus on endospore-forming bacteria of this study, the data generated here also provide substantial knowledge about sediment bacteria in general. Sediment bacteria, in particular from freshwaters are not frequently studied and only a few reviews exist (Nealson 1997; Newton et al. 2011). Few metagenomics studies are available on freshwater sediment, for example, on methane oxidizing bacteria (Beck et al. 2013). To the best of our knowledge, a full metagenomics study on lake sediment bacteria and their functions has not been done. The datasets produced for this thesis contain substantial information on sediment bacteria. The global approach based on 16S rRNA gene describes a high diversity of bacteria in sediments. There were for example sequences detected that classified in previously uncultured candidate phyla such as OP3 and OP8 from the Obsidian Pool, Yellow Stone National Park (Rohini Kumar and Saravanan 2010) or TM6 named after "Torf, mittlere Schicht" (Rheims et al. 1996), or WS1 from the Wurtsmith aquifer (Dojka et al. 1998).

Functional and metabolic properties of bacteria can only be inferred when the sequence data provides high resolution taxonomic results. Bacteria in general are highly diverse and the same metabolisms are found in different groups. For example, just knowing that members from phyla Firmicutes or Alphaproteobacteria are present in a sample does not provide meaningful functional information. On the other hand detecting a species in high abundance, for example *Bacillus methanolicus* (as shown in chapter 2) provides evidence that methanol utilizing bacteria play a role in sediment biogeochemistry (Heggeset et al. 2012). The latter obviously provides more detailed information about the environmental functioning of bacteria. It is thus important to link taxa to environmental functions in order to provide an ecosystemic view on diversity and distribution of bacteria.

As shown in the individual chapters, the resolution of endospore-forming bacterial diversity was substantially increased with the targeted metagenomics approaches described in this thesis and in a number of cases, functional information could be inferred from detailed genus or species-level taxonomic identifications. Using these methods, substantial knowledge about endospore-forming bacteria in sediments was generated and an unprecedented diversity was revealed.

6.5 References

Alawi, M., Lerm, S., Vetter, A., Wolfgramm, D. M., Seibt, D. A., and Würdemann, H. (2011). Diversity of sulfate-reducing bacteria in a plant using deep geothermal energy. *Grundwasser*, 16(2):105–112.

Bartholomew, J. W. and Paik, G. (1966). Isolation and identification of obligate thermophilic sporeforming Bacilli from ocean basin cores. *J Bacteriol*, 92(3):635–8.

Beck, D. A., Kalyuzhnaya, M. G., Malfatti, S., Tringe, S. G., Glavina del Rio, T., Ivanova, N., Lidstrom, M. E., and Chistoserdova, L. (2013). A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the Methylococcaceae and the Methylophilaceae. *PeerJ*, 1:e23.

Chernitsina, S. M., Zemskaya, T. I., Vorob'eva, S. S., Shubenkova, O. V., Khlystov, O. M., and Kostornova, T. Y. (2007). Comparative molecular biological analysis of the microbial community of the Holocene and Pleistocene deposits of Posol'skaya Shoal, Lake Baikal. *Microbiology*, 76(1):102–111.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science*, 322(5899):275–278.

Czekalski, N., Berthold, T., Caucci, S., Egli, A., and Bürgmann, H. (2012). Increased levels of multiresistant bacteria and resistance genes after wastewater treatment and their dissemination into Lake Geneva, Switzerland. *Front Microbiol*, 3:106.

de Rezende, J. R., Kjeldsen, K. U., Hubert, C. R., Finster, K., Loy, A., and Jorgensen, B. B. (2013). Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J*, 7(1):72–84.

Dojka, M. A., Hugenholtz, P., Haack, S. K., and Pace, N. R. (1998). Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl Environ Microbiol*, 64(10):3869–3877.

Ettenauer, J., Sterflinger, K., and Piñar, G. (2010). Cultivation and molecular monitoring of halophilic microorganisms inhabiting an extreme environment presented by a salt-attacked monument. *Int J Astrobiol*, 9(01):59–72.

Fell, N. F., Pellegrino, P. M., and Gillespie, J. B. (2001). Mitigating phosphate interference in bacterial endospore detection by Tb dipicolinate photoluminescence. *Anal Chim Acta*, 426(1):43–50.

Fichtel, J., Koster, J., Rullkotter, J., and Sass, H. (2007). Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol Ecol*, 61(3):522–32.

Heggeset, T. M. B., Krog, A., Balzer, S., Wentzel, A., Ellingsen, T. E., and Brautaset, T. (2012). Genome sequence of thermotolerant *Bacillus methanolicus*: features and regulation related to methylotrophy and production of L-lysine and L-glutamate from methanol. *Appl Environ Microbiol*, 78(15):5170–5181.

Hill, I. R. and Gray, T. R. G. (1967). Application of the fluorescent-antibody technique to an ecological study of bacteria in soil. *J Bacteriol*, 93(6):1888–1896.

Junier, P., Frutschi, M., Wigginton, N. S., Schofield, E. J., Bargar, J. R., and Bernier-Latmani, R. (2009). Metal reduction by spores of *Desulfotomaculum reducens*. *Environ Microbiol*, 11(12):3007–3017.

Li, N.-Z. and Xia, T. (2012). Genome sequence of *Paenibacillus* sp. strain Aloe-11, an endophytic bacterium with broad antimicrobial activity and intestinal colonization ability. *J Bacteriol*, 194(8):2117–8.

Logan, N. A., Lebbe, L., Hoste, B., Goris, J., Forsyth, G., Heyndrickx, M., et al. (2000). Aerobic endospore-forming bacteria from geothermal environments in northern Victoria Land, Antarctica, and Candlemas Island, South Sandwich archipelago, with the proposal of *Bacillus fumarioli* sp. nov. *Int J Syst Evol Microbiol*, 50(5):1741–1753.

Morasch, B., Bonvin, F., Reiser, H., Grandjean, D., de Alencastro, L. F., Perazzolo, C., Chèvre, N., and Kohn, T. (2010). Occurrence and fate of micropollutants in the Vidy Bay of Lake Geneva, Switzerland. part II: Micropollutant removal between wastewater and raw drinking water. *Environ Toxicol Chem*, 29(8):1658–1668.

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

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., and Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbio Mol Biol Rev*, 75(1):14–49.

Philippot, L., Andersson, S. G. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., and Hallin, S. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol*, 8(7):523–529.

Raghunathan, A., Ferguson, H. R., Bornarth, C. J., Song, W., Driscoll, M., and Lasken, R. S. (2005). Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol*, 71(6):3342–3347.

Rheims, H., Rainey, F. A., and Stackebrandt, E. (1996). A molecular approach to search for diversity among bacteria in the environment. *J Indust Microbiol*, 17(3-4):159–169.

Rohini Kumar, M. and Saravanan, V. S. (2010). Candidate OP phyla: Importance, ecology and cultivation prospects. *Indian J Microbiol*, 50(4):474–477.

Thevenon, F., Graham, N. D., Herbez, A., Wildi, W., and 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):609–17.

Vologina, E. and Sturm, M. (2009). Types of Holocene deposits and regional pattern of sedimentation in Lake Baikal. *Russ Geol Geophys*, 50(8):722–727.