METAGENOMIC AND METATRANSCRIPTOMIC ANALYSES OF CALCIFYING BIOFILMS

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie

der Georg-August University School of Science (GAUSS)

vorgelegt von

Dominik Schneider

aus Göttingen

Göttingen, 2013

Betreuungsausschuss

Prof. Dr. Rolf Daniel, Abteilung für Genomische und Angewandte Mikrobiologie, Institut für Mikrobiologie und Genetik

PD Dr. Michael Hoppert, Abteilung für Allgemeine Mikrobiologie, Institut für Mikrobiologie und Genetik

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Rolf Daniel, Abteilung für Genomische und Angewandte Mikrobiologie, Institut für Mikrobiologie und Genetik

Korreferent: PD Dr. Michael Hoppert, Abteilung für Allgemeine Mikrobiologie, Institut für Mikrobiologie und Genetik

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Stefanie Pöggeler, Abteilung für Genetik Eukaryotischer Mikroorganismen, Institut für Mikrobiologie und Genetik

PD Dr. Gernot Arp, Abteilung Geobiologie, Geoscience Centre

PD Dr. Stefan Irniger, Abteilung für Molekulare Mikrobiologie und Genetik, Institut für Mikrobiologie und Genetik

Jun.-Prof. Dr. Kai Heimel, Abteilung für Mikrobielle Zellbiologie, Institut für Mikrobiologie und Genetik

Tag der mündlichen Prüfung: 24.10.2013

Table of contents

List of publications

Blumenberg, M., Arp, G., Reitner, J., **Schneider, D.**, Daniel, R., Thiel, V. (2013). Bacteriohopanepolyols in a stratified cyanobacterial mat from Kiritimati (Christmas Island; Kiribati). Organic Geochemistry 55, 55-62.

Schneider, D., Arp, G., Reimer, A., Reitner, J., Daniel, R. (2013). Phylogenetic Analysis of a Microbialite-Forming Microbial Mat from a Hypersaline Lake of the Kiritimati Atoll, Central Pacific. PLoS ONE 8, e66662.

Schneider, D., Reimer, A., Hahlbrock, A., Arp, G., Reitner, J., Daniel, R. (2013). Metagenomic and metatranscriptomic analyses of bacterial commmunities derived from a calcifying karst water creek biofilm and tufa. Geomicrobiology Journal (submitted for publication).

Kachiuru, R. M., Makonde, H. M., **Schneider, D.**, Voget, S., Boga, H. I., Daniel, R. (2014). Prokaryotic microbial community composition of four Kenyan soda lakes as revealed by amplicon sequencing. (in preparation)

A INTRODUCTION

1 Microbiologist's "new" tools: Metagenomics and metatranscriptomics

At the end of the 20th century it has been estimated that less than 1% of earth's microorganisms are cultivable with standard protocols (Amann *et al.*, 1995). In addition, the vast majority of the to-date cultivated and genome-sequenced bacteria belong to only four phyla (*Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Bacteroidetes*) (Rinke *et al.*, 2013), whereas current data suggest the existence of at least 63 bacterial phyla and candidate groups (SILVA database version 115) (Pruesse *et al.*, 2007; Quast *et al.*, 2013). This gives us a fair idea of the largely undiscovered versatile genetic potential harbored by bacteria and archaea that evolved in billions of years.

In this light, a new discipline, metagenomics, evolved (Handelsman *et al.*, 1998). Metagenomics comprises the analysis of the entire genetic material of complex environmental samples that represents a certain biological niche or even an ecosystem. Metagenomic approaches aim to discover and genetically characterize the so far unknown microorganisms by employing culture-independent techniques (Daniel, 2005). For this purpose, extraction of DNA from an environmental sample is the first crucial step. Many factors influence the quantity and quality of isolated environmental DNA, turning isolation often into a challenging step. For example, humic acids present in soil and similar sample material can be co-extracted during DNA isolation and may hinder downstream applications such as PCR or ligation into a vector (Tebbe and Vahjen, 1993; Zipper *et al.*, 2003). However, approaches that remove these inhibitory compounds have been established successfully (Elshahed *et al.*, 2008; Shao *et al.*, 2011).

The isolation of DNA comprising the entire genetic material of an environmental sample circumvents the often complicated to impossible cultivation of bacteria (e.g., due to symbioses or dependency on other organisms). This enables the analysis of the metabolic potential of entire microbial communities within a certain environment such as different soil types (Nacke *et al.*, 2011a; Shah *et al.*, 2013; Will *et al.*, 2010), marine water (Hu *et al.*, 2010; Lloyd *et al.*, 2013; Venter *et al.*, 2004), glacial ice (Choudhari *et al.*, 2013; Edwards *et al.*, 2013; Simon *et al.*, 2009), human (Forslund *et al.*, 2013; Greenblum *et al.*, 2012; Hayashi *et al.*, 2002; Qin *et al.*, 2010) or animal (Turnbaugh *et al.*, 2009; Wang *et al.*, 2012; Xing *et al.*, 2013) guts, etc. Fur-

thermore, environmental DNA can be used to reveal the microbial community composition (e.g., by amplification of the ubiquitous rRNA genes) or for construction and screening of metagenomic libraries, allowing the identification of novel biocatalysts (Figure 1). Besides the isolation of the entire DNA from an environmental sample, the more challenging isolation of RNA, representing the metabolic *in situ* situation of a biological sample, has been established (Baldrian *et al.*, 2012; Wemheuer *et al.*, 2012). In this case, environmental RNA is reverse-transcribed to cDNA, which is then sequenced. This enables, for example, 16S rRNA gene analyses of the active prokaryotic community or determination of the functional profile of whole microbial communities (Stewart *et al.*, 2010).

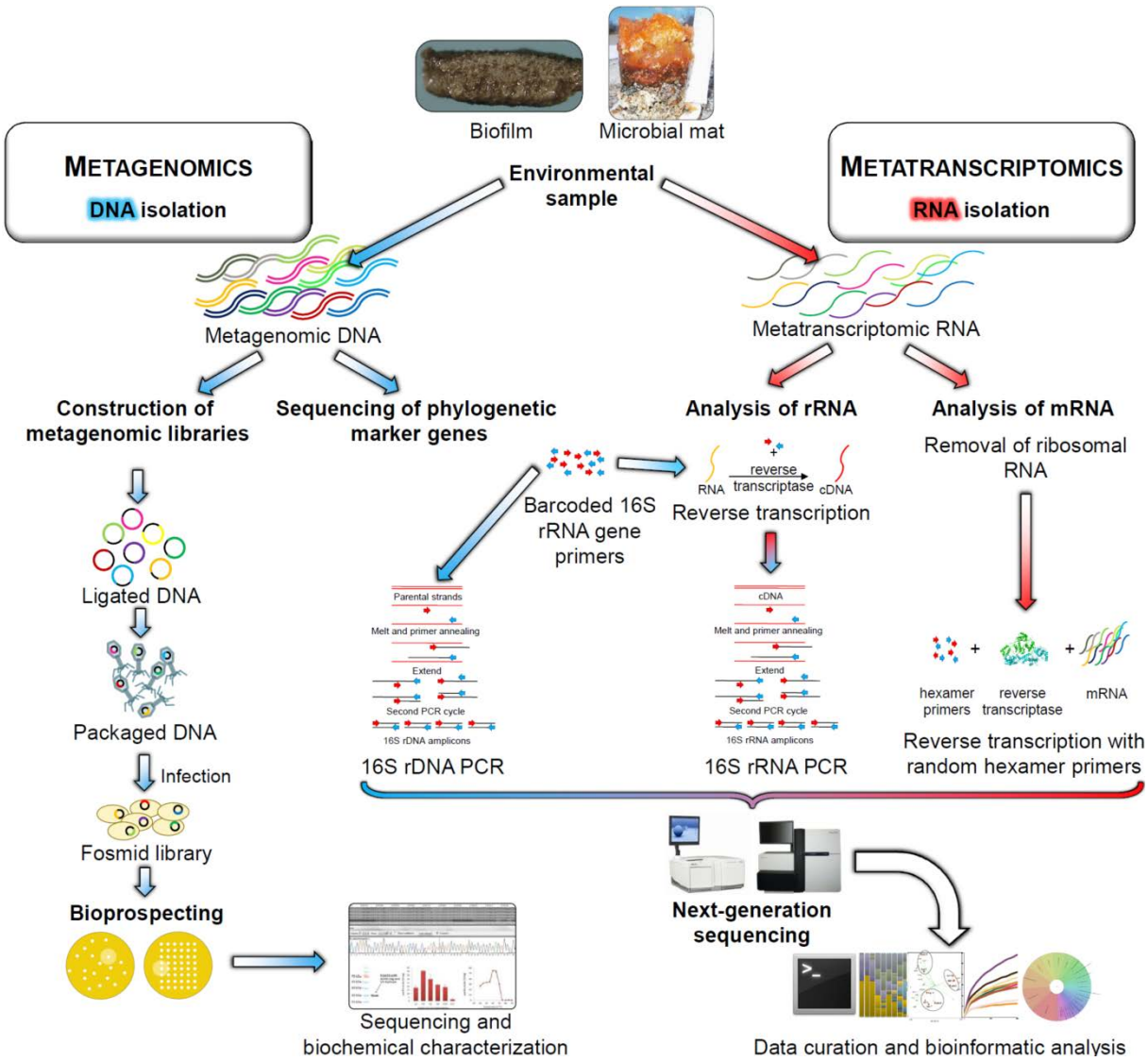


Figure 1. Schematic representation of metagenomic and metatranscriptomic approaches to analyze environmental microbial communities.

With the rapid improvement of high-throughput sequencing technologies (most significantly, Roche 454 and Illumina sequencing) the amount of sequences and sequence length raises drastically, which in turn increases the need for computational and bioinformatic solutions to tame the enormous information within these large data sets. However, by sequencing metagenomes derived from low-biodiversity ecosystems or by single-cell sequencing it was possible to reconstruct genomes from uncultured microorganisms (Herlemann *et al.*, 2013; Pelletier *et al.*, 2008; Rinke *et al.*, 2013; Strous *et al.*, 2006; Tyson *et al.*, 2004; Yelton *et al.*, 2013).

1.1 Microbial ecology based on phylogenetic marker genes – Revealing complex prokaryotic community structures in nature

With the development of a phylogenetic classification system based on the 16S rRNA gene, which is part of the small 30S subunit of the prokaryotic ribosome, Carl Woese and colleagues contributed to pave the way for the discipline of microbial ecology (Woese, 1987). Since then, the bacterial and archaeal phylogenetic tree grows and thereby continuously leads to expansions and renewal of existing phylogenetic tree branches. Owing to structure and function, the 16S rRNA gene possesses all aspects required for a phylogenetic marker: ubiquitary appearance in all prokaryotes, variable regions to infer phylogeny and conserved regions serving as primer sites. Today, 16S rRNA gene databases, for example, the well curated SILVA database (Pruesse *et al.*, 2007; Quast *et al.*, 2013), contain approximately 3.8 million 16S rRNA gene sequences (SSU version 115); most of these derived from uncultured prokaryotes of environmental studies. Furthermore, due to the detection of self-splicing introns within the 16S rRNA gene of giant sulfur bacteria, a recent study emphasizes that there might be still more to learn about this commonly used marker and the paradigms of bacteria (Salman *et al.*, 2012). Bacteria containing this special kind of 16S rRNA genes can only be identified by sequencing of RNA instead of DNA.

However, within the last two decades the analysis of particular habitats of bacterial and archaeal communities by sequencing the 16S rRNA gene has led to rising scientific interest, further pushed due to remarkable advances in next-generation sequencing techniques. An increasing number of studies aim to reveal microbial community composition of all imaginable kinds of ecosystems (Figure 2). Recently, the

sequencing of 16S rRNA gene transcripts instead of or additionally to DNA has become more frequent (Angel *et al.*, 2013; Baldrian *et al.*, 2012; Gosalbes *et al.*, 2011; Wemheuer *et al.*, 2012). This allows a potentially more precise approximation of the active community and, in case of the self-splicing introns, detection of further taxa, which could otherwise be missed (Salman *et al.*, 2012). Continuation of these research efforts will lead to a better understanding of the distribution and organization of prokaryotic communities in nature. Additionally, it will further sharpen our understanding of particular taxa's interactions and substrate cycling within habitats as well as ultimately lead to new insights into the evolution of life (Philippot *et al.*, 2010).

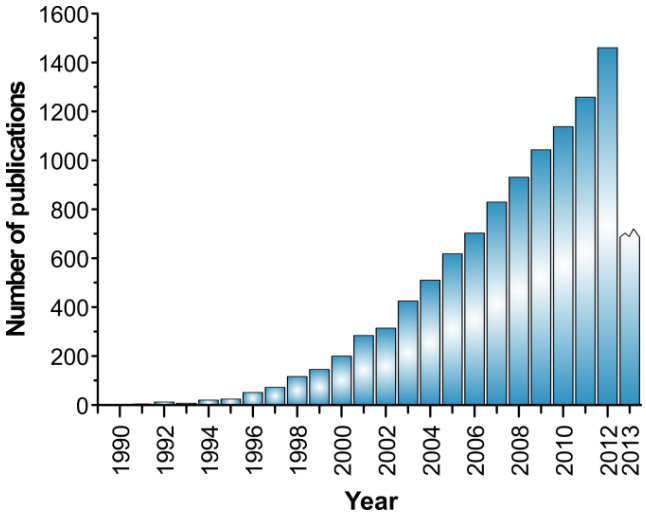


Figure 2. Bibliographic analysis on 16S rRNA gene studies of bacterial and archaeal communities. Search term used was "16S r*NA communit*" employing the Web of knowledge (http://apps.webofknowledge.com/). Time of data collection was 11-07-2013.

1.2 Mining metagenomes for novel biocatalysts

During an estimated 3.5 giga-annum (Ga) lasting evolution of bacteria and archaea (Allwood *et al.*, 2006; Bontognali *et al.*, 2012; Van Kranendonk *et al.*, 2008), a vast number of enzymes arose (Schomburg *et al.*, 2002). These enzymes differ in many aspects, first of all their amino acid sequence. This finally results in ingenious three-dimensional architectures, leading to specific reaction mechanisms. Furthermore, this specifies the enzymes properties such as stability against harmful agents, optimal operating temperatures, and requirement of specific ions. As stated above, only a small amount of this enormous metabolic potential is directly available by culture-based techniques. By the construction of metagenomic libraries (see Figure 1) from isolated environmental DNA, various novel enzymes could already be identified. Some examples are amylases (Lammle *et al.*, 2007; Sharma *et al.*, 2010; Yun *et al.*, 2004), proteases (Gupta *et al.*, 2002b; Lee *et al.*, 2007; Waschkowitz *et al.*, 2009), cellulases (Kim *et al.*, 2008; Nacke *et al.*, 2012; Voget *et al.*, 2006; Wang *et al.*,

2009), xylanases (Hu *et al.*, 2008; Nacke *et al.*, 2012), agarases (Voget *et al.*, 2003), lipases and esterases (Elend *et al.*, 2007; Elend *et al.*, 2006; Mayumi *et al.*, 2008; Nacke *et al.*, 2011b; Ouyang *et al.*, 2013; Wei *et al.*, 2009; Yu *et al.*, 2011). Consequently, established industrial processes can be further improved due to application of optimal enzymes. Additionally, as more metagenomes become available, sequence-based screenings (Pathak *et al.*, 2012) may be an attractive alternative to function-based screenings, with the disadvantage that the hereby detected enzymes might not be properly expressed by the foreign host (e.g., in commonly used *Escherichia coli*).

1.2.1 Proteases

Proteases are enzymes that catalyze the hydrolysis of peptide bonds of proteins. This is an essential process in all living cells and, therefore, their appearance is ubiquitary. The hydrolysis of peptide bonds represents an immensely important process in general protein catabolism, cleavage of signal peptides, and activation of zymogens inside or outside of cells (Gupta *et al.*, 2002a). Based on the mode of action, proteases can be classified as endo- or exoproteases (Rawlings *et al.*, 2012). Endoproteases cleave within a polypeptide chain at specific amino acid motifs. For example, chymotrypsin is a gastric enzyme that preferentially cleaves amino acids with hydrophobic side chains (i.e., tyrosine, tryptophan, and phenylalanine). In contrast, exoproteases release amino acids from the N-terminus (aminoproteases), C-terminus (carboxyproteases), or of both terminal regions of polypeptide chains. Given the nature of their catalytic centers, proteases are further classified into the families of aspartic-, cysteine-, glutamic-, metallo-, asparagine-, serine-, and threonine proteases (also some mixed and unknown catalytic types exist) (Rawlings *et al.*, 2012). Table 1 gives an overview and examples of typical representatives of each family.

Due to their variety in specificity, proteases are important industrial enzymes. Bacterial proteases are used as additive in washing and cleaning agents, pharmacology manufacture, waste treatment, modification of food proteins, or in the production of cheese (Ruttloff *et al.*, 1995). Thus, metagenomic screening for novel proteases represents a promising way to increase the available palette of enzymes for industrial demands.

Table 1. Overview of known protease families.

Family	Functional group	Number of subfamilies	Examples
Aspartic-proteases	Aspartate	25	Pepsin, renin, signalpeptidase II
Cysteine-proteases	Cysteine	76	Papain, caspase
Glutamic-proteases	Glutamate	2	Scytalidoglutamic peptidase
Metalloproteases	Metal complexes, Zn^{2+}, Ca^{2+}	67	Thermolysin, vibriolysin
Asparagine-proteases	Asparagine	10	Nodavirus peptide lyase, poliovirus capsid VP0-type self-cleaving protein
Serine-proteases	Serine	53	Trypsin, chymotrypsin, elastase, subtilisin Carlsberg
Threonine-proteases	Threonine	6	Archaean proteasome (beta component), gamma-glutamyltransferase 1
Mixed proteases	Cysteine, serine, threonine	1	DmpA aminopeptidase
Unknown catalytic type proteases	Unknown	8	Collagenase, Lit peptidase

1.2.2 Cellulases

Cellulose [$(C_6H_{10}O_5)_n$] is produced by plants, algae, oomycetes, and some bacteria (Lasa, 2006). It is the most abundant organic compound (1.5×10^{12} tons per annum) on earth (Klemm *et al.*, 2005). The chemical nature of cellulose is characterized by long β-1,4-linked D-glucose molecules, which are insoluble in water, build crystalline microfibrils, and are highly resistant to enzymatic hydrolysis. Therefore, cellulose breakdown requires several synergistically acting enzymes (Beguin and Aubert, 1994). These include exoglucanases (1,4-β-D-glucan glucanohydrolases), endoglucanases (1,4-β-D-glucan-4-glucanohydrolases and 1,4-β-D-glucan cellobiohydrolases), and β-glucosidases often associated with carbohydrate-binding modules (CBMs) (Lynd *et al.*, 2002). An up-to-date list of enzymes involved in cellulose degradation and associated modules is available at http://www.cazy.org/. Known organizations of cellulolytic systems are divided in non-complexed or complexed systems (i.e., cellulosomes). Fungi and some bacteria (e.g., *Actinomycetes*) can penetrate cellulosic substrates by hyphae and possess the non-complexed system, which secretes the cellulolytic enzymes in close proximity of the substrate. In bacteria synthesizing complexed cellulase systems, enzymes are mounted to a scaffoldin, which in turn is attached to the cell wall (Lynd *et al.*, 2002).

Cellulolytic enzymes have several applications in textile, detergent, and food industry. Additionally, cellulases have potential in generation of biofuel, as the released sugar glucose can be easily fermented to e.g., ethanol (Rao *et al.*, 2013). Unfortunately, the enzymatic breakdown of cellulosic biomass within industrial procedures is still inefficient (Hess *et al.*, 2011). However, besides protein engineering, the application of metagenomic approaches might be another promising way to discover proper enzymes.

2 Bacterial biofilms

A bacterial biofilm is generally defined as a polymicrobial (e.g., bacteria, archaea, fungi, protozoa) assemblage attached to a natural or artificial surface (also called substratum) embedded and protected by self-produced mucilage, commonly known as extracellular polymeric substance (EPS). The EPS is composed of numerous different compounds, which in total are difficult to analyze (for detailed information see (Flemming *et al.*, 2007; Flemming and Wingender, 2010)). The EPS composition is mainly dependent on the biofilm community and habitat properties. Generally, EPS is composed of polysaccharides (e.g., alginate), proteins (including extracellular enzymes), lipids, extracellular DNA (eDNA) (Tang *et al.*, 2013), humic substances, and sometimes microbialites (Arp *et al.*, 2010; Schneider *et al.*, 2013), and is highly hydrated (water is the main component of EPS) (Flemming and Wingender, 2010). In contrast to a planktonic lifestyle, bacteria living in biofilms enjoy several benefits. These advantages comprise higher resistance towards harmful substances (e.g., antibiotics (Stewart and Costerton, 2001)), and mechanical stability, as well as protection against salinity, grazers, desiccation, starvation, UV-radiation, and fluctuations in pH or oxygen. Additional positive effects are a higher possibility or even necessity, due to changing environmental factors, for horizontal gene transfer (Madsen *et al.*, 2012), nutrition cycling and waste disposal, establishment of micro-domains, etc. In summary, living in biofilms seems to be the ultimate perfection of survival and persistence in nearly all naturally and artificially occurring habitats (Flemming and Wingender, 2010).

2.1 Development of a (poly)microbial biofilm

Biofilm formation has mainly been studied in laboratory-grown individual cultures (e.g., *Pseudomonas aeruginosa* (Davies *et al.*, 1998; Ma *et al.*, 2009; Yang *et al.*, 2011), *Bacillus subtilis* (Vlamakis *et al.*, 2013)), which has led to several insights regarding development, attachment to surfaces, composition of EPS, and cell differentiation. In nature, biofilms are usually far more complex, meaning that they are composed of many species from different phyla and research in this area is still in its infancy. Nevertheless, microbial biofilm development can roughly be divided into four stages (Figure 3). The first step in biofilm formation is the adhesion of microbes to a suitable surface, the substratum (usually solid, but also pellicles are possible at water-air interfaces, see (Vlamakis *et al.*, 2013)). This is achieved by several effects, for example, initially by van-der-Waals forces and later by production of adhesins. More precisely, polysaccharides, proteins (e.g., flagella, pili) (Li *et al.*, 2012)), eDNA (Das *et al.*, 2010) or amphiphilic molecules (Flemming and Wingender, 2010) serve as adhesins. Possible initiators of biofilm formation can be starvation, quorum signaling molecules (e.g., peptides), or external stress factors like antibiotics (Lopez *et al.*, 2010). After attachment, cells start to multiply and form a monolayer, followed by further division into a multilayered biofilm (Lemon *et al.*, 2008). Furthermore, cells start to produce an outer matrix, i.e., EPS. It has also been reported that dramatic cellular differentiations occur during biofilm development. For example, motile cells involved in the initiation of the biofilm evolve to matrix-producing cells (Vlamakis *et al.*, 2013). During the ongoing maturation of the biofilm, a complex three-dimensional structure develops, which can be highly diverse. The mature biofilm offers several benefits like diffusion/transport of nutrients, physical stability, and protection against desiccation or antibiotics. Also the occurrence of micro-domains within a biofilm that allow the performance of sensitive reactions (e.g., reduction of sulfate) is likely. Due to the close proximity of cells, exchange of metabolites, cell-to-cell signals (quorum sensing), and also exchange of genetic information (Madsen *et al.*, 2012) are common in biofilms (Flemming and Wingender, 2010). Finally, after a biofilm has formed, parts of the biofilm can be separated, e.g., by physical forces or genetically mediated encapsulation, which can lead to distribution of the biofilm.

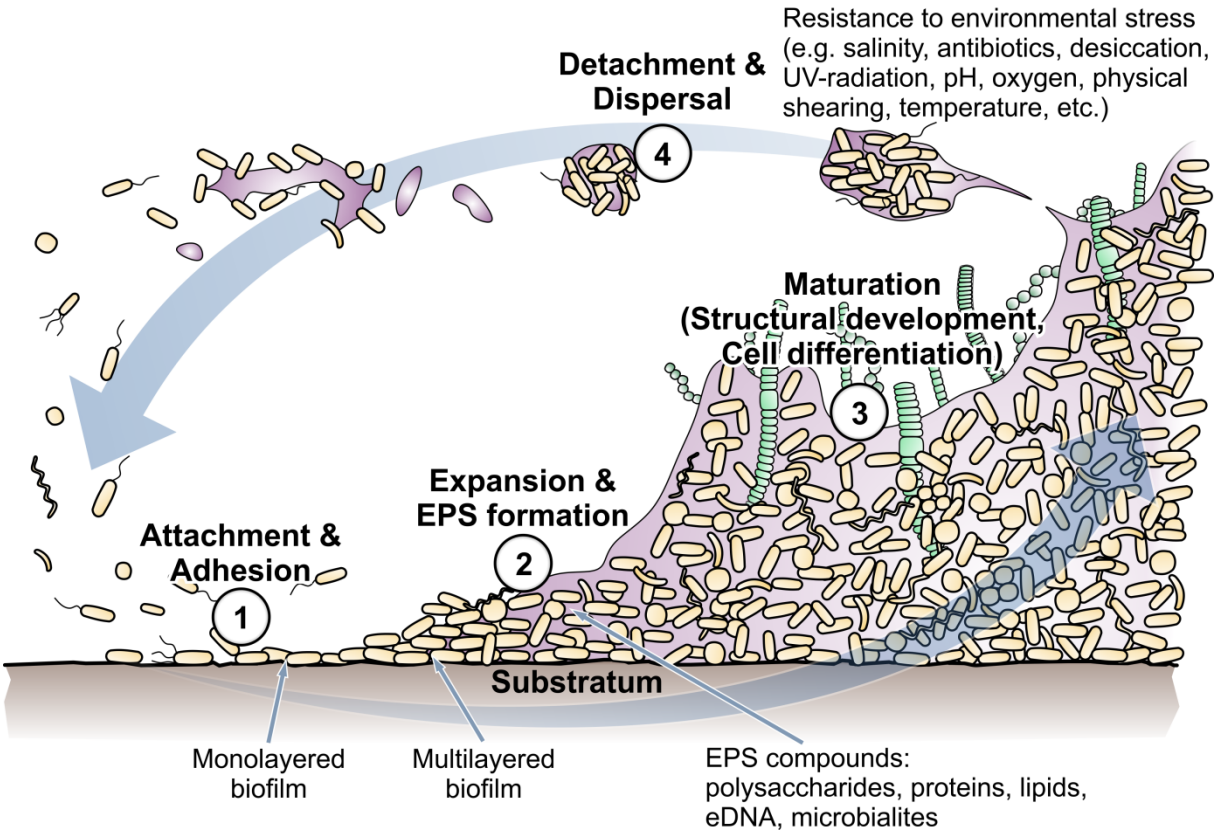


Figure 3. General scheme of bacterial biofilm attachment, development, and (re-)distribution. ① Cells adhere to the substratum and start building a (biofilm) monolayer. ② The cells form a multilayered structure and produce EPS. ③ The biofilm further maturates and cells specify. ④ Parts of the biofilm can distribute by shear force or active dispersion. Figure inspired by Monroe (2007) and citations within chapter 2.1.

2.2 Ancient and present biofilms

Geological findings (e.g., stromatolites) and results of current research indicate that biofilms are approved and widespread life forms of microbial communities long through earth history due to their protective, gene exchanging, and nutrition cycling benefits. They are found nearly everywhere where life is feasible. Current data suggest that first life forms occurred approximately 3.5 Ga ago (Allwood *et al.*, 2006; Bontognali *et al.*, 2012; Van Kranendonk *et al.*, 2008) or even earlier (Figure 4).

Under the harsh (from a human point of view) and “unbalanced” environmental conditions of the premature earth, the first molecules of life have been synthesized (Miller, 1953; Miller and Urey, 1959). Although exact mechanisms are still unclear, preferred hypotheses suggest RNAs (Gilbert, 1986) and/or proteins (Longo *et al.*, 2013) as the first molecules of life, which further evolved to replicating units, i.e.,

"cells". This resulted in the first life forms from which all today known life has evolved, beginning with the so called last universal common ancestor (LUCA). In the still harsh environmental settings, the first prokaryotic life forms may soon have "learned" to gain energy from sulfur compounds and light (anoxygenic photosynthesis). Then, after approximately 0.5 Ga (Madigan *et al.*, 2010), evolution led to the creation of the first cyanobacteria, which were able to produce O_2 by the use of more complex photosystems. O_2 is toxic to all obligate anaerobic life forms, which might have led to the first multi-layered biological systems such as microbial mats, because a separation of the different metabolic lifestyles became necessary (although not solely forced by oxygen).

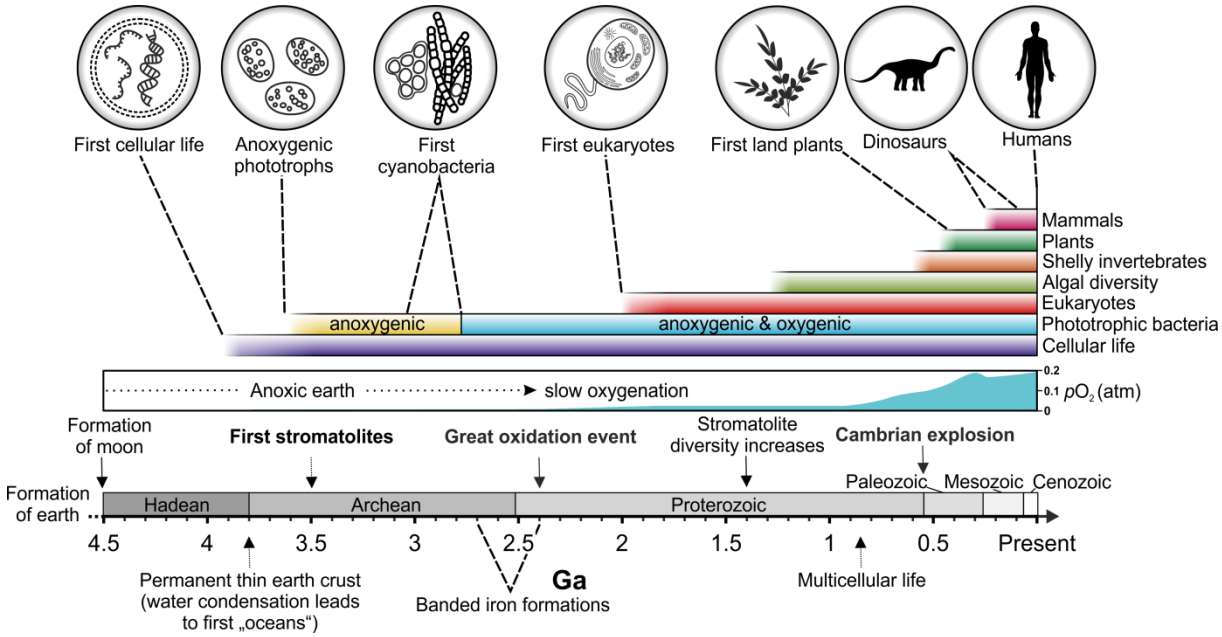


Figure 4. Draft of planet earth's history with focus on development of life, atmosphere, and connected events. Events and occurrence of life forms were adapted from (Madigan *et al.*, 2010; Resag, 2012), pO_2 of atmosphere was adapted from (Holland, 2006). Geological time scales were adapted from http://www.stratigraphy.org/ICSchart/ChronostratChart2013-01.pdf.

Today's stromatolites are the existing evidence of these systems, which are still actively formed at several "extreme" places on earth (see (Schneider *et al.*, 2013) and references therein). However, for a long time the bulk of organisms toxic oxygen was captured by organic matter (e.g., dead bacteria) and dissolved iron ions, which are believed to have resulted in precipitated iron oxides (banded iron formations) (Posth *et al.*, 2011). This kept the O_2 concentrations low and might have allowed O_2 sensitive microorganisms to adapt to lethal O_2 concentrations. After the main part of

oxidizable iron was removed, earth “slowly” got oxygenated and the ozone layer developed. Subsequently, more complex life forms appeared, possibly culminating in the Cambrian explosion (Marshall, 2006). However, until now (and most likely in the future), microbes are the most abundant and versatile life forms on earth (Dinsdale *et al.*, 2008), outlasting every known natural disaster.

3 Samples analyzed in this study

Main focus of this study was the analysis of calcifying biofilms with metagenomic and metatranscriptomic techniques. Mineralization by microorganisms is still not fully understood. Two distinct calcifying biofilm systems, with respect to geochemical properties, have therefore been selected: a freshwater creek biofilm located in Germany and a multilayered biofilm, better known as microbial mat, located on the atoll Kiritimati (also known as Christmas Island) (Figure 5). Additionally, environmental samples of four African soda lakes and artificial biofilms from microbial fuel cells (MFC) and their corresponding plankton have been analyzed with respect to their prokaryotic community compositions.

The Westerhöfer Bach is located to the west of the Harz Mountains. The stream is less than 2 m wide and receives its water from a spring area discharging from the Middle Triassic Muschelkalk-Group aquifer. The catchment area has a size of approximately 3.5 km^2 and is covered exclusively by deciduous and coniferous forest (Arp *et al.*, 2010). The creek shows karst water’s typical high partial pressure of carbon dioxide and enrichments in calcium, magnesium, and sulfate ions. The investigated section of the stream is approximately 75 m long and covered by a thin (approximately 3 mm) calcified biofilm, which covers a laminated tufa. The three sampling sites were located 262 (WB3), 287 (WB4), and 313 m (WB5) downstream of the creek spring area. The bacterial community composition of the biofilm and underlying tufa has been analyzed.

Kiritimati (Gilbertese for “Christmas”, formerly known as Christmas Island) is a coral atoll in the central Pacific Ocean, belonging to the northern Line Islands, and is part of the Republic of Kiribati. With a land area of approximately 321 km^2 it is the world’s largest atoll (Woodroffe and McLean, 1998). In the late 1950s and early

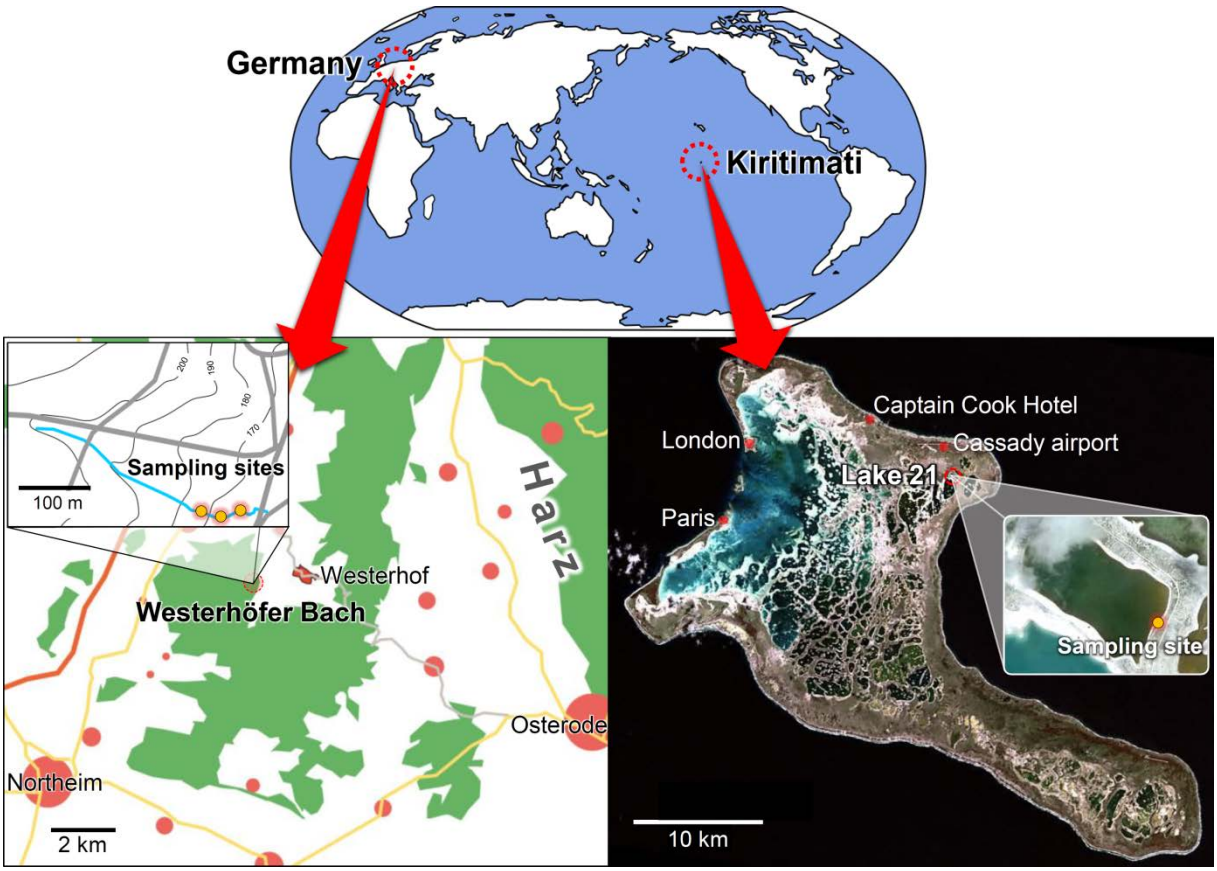


Figure 5. Geographic situation of Westerhöfer Bach, Germany (left) and Lake 21, Kiritimati (right) with detail on the sampling sites. Map from Westerhöfer Bach was adapted from Google Maps, Kiritimati satellite pictures were obtained from Landsat 7 (NASA Landsat Program, scene L706105919990916) and Google Maps.

1960s the United Kingdom and the United States performed several nuclear tests on and around the atoll (as parts of operation Grapple and operation Dominic). The island harbors approximately 500 lakes with different salinities ranging from nearly freshwater to hypersaline conditions. The lakes are largely occupied by lithifying and non-lithifiying microbial mats (Saenger *et al.*, 2006; Trichet *et al.*, 2001). The sampling site was located to the southeast shore line of the hypersaline Lake 21, which exhibited an average lake water salinity of 170‰ and generally an up to six-fold increase of total dissolved solids (TDS) compared to seawater, mainly due to evaporation. The microbial mat showed a colored lamination into nine layers, which were sampled and analyzed with respect to their prokaryotic community composition and hydrochemistry (Schneider *et al.*, 2013).

Soda lakes are characterized by high pH (9-12), high salinities (20‰ up to saturation), and high productivity by photosynthesis. Soil, water, wet and dry sedi-

ments, and a microbial mat of four Kenyan soda lakes were sampled. The lakes Bogoria, Sonachi (Crater Lake), Elmenteita, and Magadi are located within the Great East African Rift Valley. The samples have been analyzed with respect to their bacterial and archaeal community composition.

Microbial fuel cells are promising renewable energy sources, able to convert waste (e.g., wastewater) into electric current and in the case of microbial electrolysis cells (MEC) additionally into useful fuels. In cooperation with the TU Braunschweig the effect of different substrates (acetate, citric acid, glycerol, mannitol, and oleic acid) on small scale microbial fuel cells (MFC) inoculated with wastewater has been analyzed. Therefore, several parameters for the estimation of the efficiency of conversion of test substrate into electric current have been measured at the TU Braunschweig (Baudler, 2012). After incubation the planktonic and biofilm communities were collected and their bacterial composition was analyzed.

4 Aim of this thesis

Focus of this study was the analysis of the prokaryotic community composition and the inherent functional repertoire of different biofilms by metagenomic and metatranscriptomic approaches. To achieve this goal, the microbial, i.e., bacterial and partially archaeal community compositions of two very distinct calcifying biofilm ecosystems, different soda lake samples, and ten microbial fuel cells have been analyzed by large-scale analyses of 16S rRNA genes. In addition, the biofilm of the Westerhöfer Bach was analyzed by a metatranscriptomic approach to reveal key metabolic functions of the inherent microbial community. Furthermore, large-insert metagenomic fosmid libraries were constructed and subjected to a function-based screening for novel proteolytic and cellulolytic biocatalysts.

5 References

Allwood, A.C., Walter M.R., Kamber B.S., Marshall C.P., Burch I.W. (2006). Stromatolite reef from the Early Archaean era of Australia. Nature **441**: 714-718.

Amann, R.I., Ludwig W., Schleifer K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev **59**: 143-169.

Angel, R., Pasternak Z., Soares M.I., Conrad R., Gillor O. (2013). Active and total prokaryotic communities in dryland soils. FEMS Microbiol Ecol.

Arp, G., Bissett A., Brinkmann N., Cousin S., de Beer D.*, et al.* (2010). Tufa-forming biofilms of German karstwater streams: microorganisms, exopolymers, hydrochemistry and calcification. The Geological Society of London **336**: 83-118.

Baldrian, P., Kolarik M., Stursova M., Kopecky J., Valaskova V.*, et al.* (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. Isme J **6**: 248-258.

Baudler, A. (2012). Elektrochemisch aktive Biofilme auf Basis mikrobieller Konsortien: Grundlagenuntersuchung zur Verwendung von Mono - und Mischsubstraten (Master thesis). Braunschweig: Technische Universität Braunschweig. 106 p.

Beguin, P., Aubert J.P. (1994). The biological degradation of cellulose. FEMS Microbiol Rev **13**: 25-58.

Bontognali, T.R., Sessions A.L., Allwood A.C., Fischer W.W., Grotzinger J.P.*, et al.* (2012). Sulfur isotopes of organic matter preserved in 3.45-billion-year-old stromatolites reveal microbial metabolism. Proc Natl Acad Sci U S A **109**: 15146-15151.

Choudhari, S., Smith S., Owens S., Gilbert J.A., Shain D.H.*, et al.* (2013). Metagenome sequencing of prokaryotic microbiota collected from Byron Glacier, Alaska. Genome Announc **1**: e0009913.

Daniel, R. (2005). The metagenomics of soil. Nat Rev Microbiol **3**: 470-478.

Das, T., Sharma P.K., Busscher H.J., van der Mei H.C., Krom B.P. (2010). Role of extracellular DNA in initial bacterial adhesion and surface aggregation. Appl Environ Microbiol **76**: 3405-3408.

Davies, D.G., Parsek M.R., Pearson J.P., Iglewski B.H., Costerton J.W.*, et al.* (1998). The involvement of cell-to-cell signals in the development of a bacterial biofilm. Science **280**: 295-298.

Dinsdale, E.A., Edwards R.A., Hall D., Angly F., Breitbart M.*, et al.* (2008). Functional metagenomic profiling of nine biomes. Nature **452**: 629-632.

Edwards, A., Pachebat J.A., Swain M., Hegarty M., Hodson A.J.*, et al.* (2013). A metagenomic snapshot of taxonomic and functional diversity in an alpine glacier cryoconite ecosystem. Environmental Research Letters **8**.

Elend, C., Schmeisser C., Hoebenreich H., Steele H.L., Streit W.R. (2007). Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (R)-ibuprofen esters. J Biotechnol **130**: 370-377.

Elend, C., Schmeisser C., Leggewie C., Babiak P., Carballeira J.D.*, et al.* (2006). Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl Environ Microbiol **72**: 3637-3645.

Elshahed, M.S., Youssef N.H., Spain A.M., Sheik C., Najar F.Z.*, et al.* (2008). Novelty and uniqueness patterns of rare members of the soil biosphere. Appl Environ Microbiol **74**: 5422-5428.

Flemming, H.C., Neu T.R., Wozniak D.J. (2007). The EPS matrix: the "house of biofilm cells". J Bacteriol **189**: 7945-7947.

Flemming, H.C., Wingender J. (2010). The biofilm matrix. Nat Rev Microbiol **8**: 623-633.

Forslund, K., Sunagawa S., Kultima J.R., Mende D., Arumugam M.*, et al.* (2013). Country-specific antibiotic use practices impact the human gut resistome. Genome Research.

Gilbert, W. (1986). Origin of life: The RNA world. Nature **319**: 1.

Gosalbes, M.J., Durban A., Pignatelli M., Abellan J.J., Jimenez-Hernandez N.*, et al.* (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS One **6**: e17447.

Greenblum, S., Turnbaugh P.J., Borenstein E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci U S A **109**: 594-599.

Gupta, R., Beg Q.K., Khan S., Chauhan B. (2002a). An overview on fermentation, downstream processing and properties of microbial alkaline proteases. Appl Microbiol Biotechnol **60**: 381-395.

Gupta, R., Beg Q.K., Lorenz P. (2002b). Bacterial alkaline proteases: molecular approaches and industrial applications. Appl Microbiol Biotechnol **59**: 15-32.

Handelsman, J., Rondon M.R., Brady S.F., Clardy J., Goodman R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol **5**: R245-249.

Hayashi, H., Sakamoto M., Benno Y. (2002). Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. Microbiol Immunol **46**: 535-548.

Herlemann, D.P., Lundin D., Labrenz M., Jurgens K., Zheng Z.*, et al.* (2013). Metagenomic de novo assembly of an aquatic representative of the verrucomicrobial class *Spartobacteria*. MBio **4**: e00569-00512.

Hess, M., Sczyrba A., Egan R., Kim T.W., Chokhawala H.*, et al.* (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science **331**: 463-467.

Holland, H.D. (2006). The oxygenation of the atmosphere and oceans. Philos Trans R Soc Lond B Biol Sci **361**: 903-915.

Hu, Y., Fu C., Huang Y., Yin Y., Cheng G.*, et al.* (2010). Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. FEMS Microbiol Ecol **72**: 228-237.

Hu, Y., Zhang G., Li A., Chen J., Ma L. (2008). Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. Appl Microbiol Biotechnol **80**: 823-830.

Kim, S.J., Lee C.M., Han B.R., Kim M.Y., Yeo Y.S.*, et al.* (2008). Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett **282**: 44-51.

Klemm, D., Heublein B., Fink H.P., Bohn A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. Angew Chem Int Ed Engl **44**: 3358-3393.

Lammle, K., Zipper H., Breuer M., Hauer B., Buta C.*, et al.* (2007). Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol **127**: 575-592.

Lasa, I. (2006). Towards the identification of the common features of bacterial biofilm development. Int Microbiol **9**: 21-28.

Lee, D.G., Jeon J.H., Jang M.K., Kim N.Y., Lee J.H.*, et al.* (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. Biotechnol Lett **29**: 465-472.

Lemon, K.P., Earl A.M., Vlamakis H.C., Aguilar C., Kolter R. (2008). Biofilm Development with an Emphasis on *Bacillus subtilis*. In: Romeo T, editor. Bacterial biofilms. Berlin Heidelberg: Springer-Verlag.

Li, G., Brown P.J., Tang J.X., Xu J., Quardokus E.M.*, et al.* (2012). Surface contact stimulates the just-in-time deployment of bacterial adhesins. Mol Microbiol **83**: 41-51.

Lloyd, K.G., Schreiber L., Petersen D.G., Kjeldsen K.U., Lever M.A.*, et al.* (2013). Predominant archaea in marine sediments degrade detrital proteins. Nature **496**: 215-218.

Longo, L.M., Lee J., Blaber M. (2013). Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. Proc Natl Acad Sci U S A **110**: 2135-2139.

Lopez, D., Vlamakis H., Kolter R. (2010). Biofilms. Cold Spring Harb Perspect Biol **2**: a000398.

Lynd, L.R., Weimer P.J., van Zyl W.H., Pretorius I.S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev **66**: 506-577, table of contents.

Ma, L., Conover M., Lu H., Parsek M.R., Bayles K.*, et al.* (2009). Assembly and development of the *Pseudomonas aeruginosa* biofilm matrix. PLoS Pathog **5**: e1000354.

Madigan, M.T., Martinko J.M., Stahl D.A., Clark D.P. (2010). Brock Biology of Microorganisms. Englewood Cliffs, N.J.: Prentice-Hall. 1152 p.

Madsen, J.S., Burmølle M., Hansen L.H., Sørensen S.J. (2012). The interconnection between biofilm formation and horizontal gene transfer. FEMS Immunology & Medical Microbiology **65**.

Marshall, C.R. (2006). Explaining the Cambrian “Explosion” of Animals. Annual Review of Earth and Planetary Sciences **34**: 355-384.

Mayumi, D., Akutsu-Shigeno Y., Uchiyama H., Nomura N., Nakajima-Kambe T. (2008). Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. Appl Microbiol Biotechnol **79**: 743-750.

Miller, S.L. (1953). A production of amino acids under possible primitive earth conditions. Science **117**: 528-529.

Miller, S.L., Urey H.C. (1959). Organic compound synthesis on the primitive earth. Science **130**: 245-251.

Monroe, D. (2007). Looking for chinks in the armor of bacterial biofilms. PLoS Biol **5**: e307.

Nacke, H., Engelhaupt M., Brady S., Fischer C., Tautzt J.*, et al.* (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett **34**: 663-675.

Nacke, H., Thurmer A., Wollherr A., Will C., Hodač L.*, et al.* (2011a). Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. PLoS One **6**: e17000.

Nacke, H., Will C., Herzog S., Nowka B., Engelhaupt M.*, et al.* (2011b). Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. FEMS Microbiol Ecol **78**: 188-201.

Ouyang, L.M., Liu J.Y., Qiao M., Xu J.H. (2013). Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl Biochem Biotechnol **169**: 15-28.

Pathak, G.P., Losi A., Gartner W. (2012). Metagenome-based screening reveals worldwide distribution of LOV-domain proteins. Photochem Photobiol **88**: 107-118.

Pelletier, E., Kreimeyer A., Bocs S., Rouy Z., Gyapay G.*, et al.* (2008). "*Candidatus* Cloacamonas Acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. J Bacteriol **190**: 2572-2579.

Philippot, L., Andersson S.G., Battin T.J., Prosser J.I., Schimel J.P.*, et al.* (2010). The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol **8**: 523-529.

Posth, N.R., Konhauser K.O., Kapper A. (2011). Banded Iron Formations. In: Reitner J, Thiel V, editors. Encyclopedia of Geobiology. pp. 92-103.

Pruesse, E., Quast C., Knittel K., Fuchs B.M., Ludwig W.*, et al.* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35**: 7188-7196.

Qin, J., Li R., Raes J., Arumugam M., Burgdorf K.S.*, et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464**: 59-65.

Quast, C., Pruesse E., Yilmaz P., Gerken J., Schweer T.*, et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **41**: D590-596.

Rao, L.V., Chandel A.K., Chandrasekhar G., Rodhe A.V., Sridevi J. (2013). Cellulases of Thermophilic Microbes. In: Satyanarayana T, Littlechild J, Kawarabayasi Y, editors. Thermophilic Microbes in Environmental and Industrial Biotechnology. Dordrecht: Springer Science+Business Media. pp. 771-793.

Rawlings, N.D., Barrett A.J., Bateman A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res **40**: D343-350.

Resag, J. (2012). Zeitpfad. Berlin-Heidelberg: Spektrum Akademischer Verlag.

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

Ruttloff, H., Huber J., Zickler F., Mangold K.H. (1995). Industrielle Enzyme. Hamburg: Springer-Verlag GmbH. 1038 p.

Saenger, C., Miller M., Smittenberg R.H., Sachs J.P. (2006). A physico-chemical survey of inland lakes and saline ponds: Christmas Island (Kiritimati) and Washington (Teraina) Islands, Republic of Kiribati. Saline Systems **2**: 8.

Salman, V., Amann R., Shub D.A., Schulz-Vogt H.N. (2012). Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. Proc Natl Acad Sci U S A **109**: 4203-4208.

Schneider, D., Arp G., Reimer A., Reitner J., Daniel R. (2013). Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the kiritimati atoll, central pacific. PLoS One **8**: e66662.

Schomburg, I., Chang A., Hofmann O., Ebeling C., Ehrentreich F.*, et al.* (2002). BRENDA: a resource for enzyme data and metabolic information. Trends Biochem Sci **27**: 54-56.

Shah, V., Zakrzewski M., Wibberg D., Eikmeyer F., Schluter A.*, et al.* (2013). Taxonomic Profiling and Metagenome Analysis of a Microbial Community from a Habitat Contaminated with Industrial Discharges. Microb Ecol.

Shao, K., Gao G., Qin B., Tang X., Wang Y.*, et al.* (2011). Comparing sediment bacterial communities in the macrophyte-dominated and algae-dominated areas of eutrophic Lake Taihu, China. Can J Microbiol **57**: 263-272.

Sharma, S., Khan F.G., Qazi G.N. (2010). Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. Appl Microbiol Biotechnol **86**: 1821-1828.

Simon, C., Wiezer A., Strittmatter A.W., Daniel R. (2009). Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. Appl Environ Microbiol **75**: 7519-7526.

Stewart, F.J., Ottesen E.A., DeLong E.F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. Isme J **4**: 896-907.

Stewart, P.S., Costerton J.W. (2001). Antibiotic resistance of bacteria in biofilms. Lancet **358**: 135-138.

Strous, M., Pelletier E., Mangenot S., Rattei T., Lehner A.*, et al.* (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature **440**: 790-794.

Tang, L., Schramm A., Neu T.R., Revsbech N.P., Meyer R.L. (2013). Extracellular DNA in adhesion and biofilm formation of four environmental isolates: a quantitative study. FEMS Microbiol Ecol.

Tebbe, C.C., Vahjen W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl Environ Microbiol **59**: 2657-2665.

Trichet, J., Défarge C., Tribble J., Tribble G.W., Sansone F.J. (2001). Christmas Island lagoonal lakes, models for the deposition of carbonate-evaporite-organic laminated sediments. Sedimentary Geology **140**: 177–189.

Turnbaugh, P.J., Ridaura V.K., Faith J.J., Rey F.E., Knight R.*, et al.* (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med **1**: 6ra14.

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

Van Kranendonk, M.J., Philippot P., Lepot K., Bodorkos S., Pirajno F. (2008). Geological setting of Earth's oldest fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. Precambrian Research **167**: 93-124.

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

Vlamakis, H., Chai Y., Beauregard P., Losick R., Kolter R. (2013). Sticking together: building a biofilm the *Bacillus subtilis* way. Nat Rev Microbiol **11**: 157-168.

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

Voget, S., Steele H.L., Streit W.R. (2006). Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol **126**: 26-36.

Wang, F., Li F., Chen G., Liu W. (2009). Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol Res **164**: 650-657.

Wang, W., Archbold T., Kimber M.S., Li J., Lam J.S.*, et al.* (2012). The porcine gut microbial metagenomic library for mining novel cellulases established from growing pigs fed cellulose-supplemented high-fat diets. J Anim Sci **90 Suppl 4**: 400-402.

Waschkowitz, T., Rockstroh S., Daniel R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. Appl Environ Microbiol **75**: 2506-2516.

Wei, P., Bai L., Song W., Hao G. (2009). Characterization of two soil metagenome-derived lipases with high specificity for p-nitrophenyl palmitate. Arch Microbiol **191**: 233-240.

Wemheuer, B., Wemheuer F., Daniel R. (2012). RNA-based assessment of diversity and composition of active archaeal communities in the German Bight. Archaea **2012**: 695826.

Will, C., Thurmer A., Wollherr A., Nacke H., Herold N.*, et al.* (2010). Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol **76**: 6751-6759.

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

Woodroffe, C.D., McLean R.F. (1998). Pleistocene morphology and Holocene emergence of Christmas (Kiritimati) Island, Pacific Ocean. Coral Reefs **17**: 235-248.

Xing, M., Hou Z., Yuan J., Liu Y., Qu Y.*, et al.* (2013). Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (*Scophthalmus maximus*). FEMS Microbiol Ecol.

Yang, L., Hu Y., Liu Y., Zhang J., Ulstrup J.*, et al.* (2011). Distinct roles of extracellular polymeric substances in *Pseudomonas aeruginosa* biofilm development. Environ Microbiol **13**: 1705-1717.

Yelton, A.P., Comolli L.R., Justice N.B., Castelle C., Denef V.J.*, et al.* (2013). Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. BMC Genomics **14**: 485.

Yu, E.Y., Kwon M.A., Lee M., Oh J.Y., Choi J.E.*, et al.* (2011). Isolation and characterization of cold-active family VIII esterases from an arctic soil metagenome. Appl Microbiol Biotechnol **90**: 573-581.

Yun, J., Kang S., Park S., Yoon H., Kim M.J.*, et al.* (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Appl Environ Microbiol **70**: 7229-7235.

Zipper, H., Buta C., Lammle K., Brunner H., Bernhagen J.*, et al.* (2003). Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments. Nucleic Acids Res **31**: e39.

B PUBLICATIONS AND RESULTS

1 Bacteriohopanepolyols in a stratified cyanobacterial mat from Kiritimati (Christmas Island, Kiribati)

Martin Blumenberg[1,2], Gernot Arp[1], Joachim Reitner[1,2], Dominik Schneider[3], Rolf Daniel[3], Volker Thiel[1,2]

Organic Geochemistry (2013), Volume 55, p. 55-62

[1]Geoscience Center, Geobiology Group, Georg-August-University Göttingen, Goldschmidtstr. 3, 37077 Göttingen, Germany

[2]Courant Center Geobiology, Georg-August-University Göttingen, Goldschmidtstr. 3, 37077 Göttingen, Germany

[3]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany

---

Author contributions:

Conceived and designed the experiments: MB GA JR DS RD VT.

Performed the experiments: MB GA DS.

Analyzed the data: MB GA DS RD VT.

Wrote the paper: MB DS RD VT.

Organic Geochemistry 55 (2013) 55–62

Contents lists available at SciVerse ScienceDirect

Organic Geochemistry

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

Bacteriohopanepolyols in a stratified cyanobacterial mat from Kiritimati (Christmas Island, Kiribati)

Martin Blumenberg [a,b,*], Gernot Arp [a], Joachim Reitner [a,b], Dominik Schneider [c], Rolf Daniel [c], Volker Thiel [a,b]

[a] Geoscience Center, Geobiology Group, Georg-August-University Göttingen, Goldschmidtstr. 3, 37077 Göttingen, Germany
[b] Courant Center Geobiology, Georg-August-University Göttingen, Goldschmidtstr. 3, 37077 Göttingen, Germany
[c] Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 25 September 2012
Received in revised form 5 November 2012
Accepted 7 November 2012
Available online 15 November 2012

ABSTRACT

Bacteriohopanepolyols (BHPs) are lipids of distinct bacterial groups and are, along with geohopanoids as their diagenetic products, ubiquitous in microbial mats, sediments, soils and oil. Among BHP-producing bacteria, Cyanobacteria are of special interest since occurrences of C-2 methylated and other geohopanoids are interpreted as signals of oxygenic photoautotrophs in the early oceans. However, many questions, with respect to the source and function of hopanoids, remain open. Cyanobacterial mats are complex systems in terms of biogeochemical zones and hence, harbor phylogenetically and metabolically diverse photoautotrophic and heterotrophic microorganisms. Whereas Cyanobacteria are key microbial players for carbon fixation in these mat systems, most other bacterial and archaeal groups feed on their metabolic products and/or perform anoxygenic photosynthesis in deeper low-light regimes.

We have analyzed the abundance and distribution of BHPs in distinct layers of a stratified mat from a hypersaline lake on Kiritimati. Suites of BHPs were observed, with the majority probably originating from (proteo-)bacteria thriving in the deeper layers of the mat and not in upper layers where mat-forming Cyanobacteria are prevalent. Interestingly, the BHPs in the deepest layer include C-2 methylated structures, which likely do not originate from Cyanobacteria, but rather from α-Proteobacteria thriving at the redoxcline of this stratified microbial system.

1. Introduction

Cyanobacterial mats have been described from various environments and are often found in hypersaline settings (e.g. Des Marais, 2003; Wieland et al., 2008). They typically consist of well defined zones, shaped by microorganisms performing distinct metabolisms. Studies of such mats provide insight into the processes of burial and preservation of organic matter (OM) and are relevant for the interpretation of chemical signatures left in fossilized microbial structures such as stromatolites (Des Marais, 2003). Key organisms are different types of Cyanobacteria that photoautotrophically fix CO_2 and produce substances used by other bacterial and archaeal members of the community. These bacteria include phototrophs other than Cyanobacteria (e.g. anoxygenic phototrophs), heterotrophs (e.g. sulfate reducing bacteria and fermenting bacteria) and methanogenic archaea. Because of their analogy with stromatolites in the Earth's history and the occurrences of specific lipids in the geological record, biomarkers from cyanobacterial mats have been frequently studied (e.g. Boon et al., 1981; Dobson et al., 1988; Shiea et al., 1990; Grimalt et al., 1991; Kenig et al., 1995; Rontani and Volkman, 2005; Sachse and Sachs, 2008). Interestingly, high amounts of hopanoids, including ring A structures methylated at C-2, appear to be common in cyanobacterial mats and cultured Cyanobacteria (Summons et al., 1999). Findings of C-2 methyl hopanes in rocks as old as 2.7 billion years (Ga) have been taken as a strong argument for oxygenic photosynthesis, most likely driven by Cyanobacteria (Summons et al., 1999). However, the utility of C-2 methyl hopanes as biomarkers for *Cyanobacteria* has been compromised. First, the 2.7 Ga sample was reported to be likely contaminated with anthropogenic bitumen or oil from younger formations (Rasmussen et al., 2008). Second, C-2 methyl bacteriohopanepolyols (BHPs), the most likely biogenic precursors of C-2 methyl hopanes, have been found in anoxygenic phototrophic and methylotrophic bacteria (Renoux and Rohmer, 1985; Rashby et al., 2007) and a genetic survey also revealed the capability for C-2 methyl hopanoid production for Acidobacteria (Welander et al., 2010).

Although bulk cyanobacterial mats are frequently found to be rich in C-2 methyl BHPs, the distribution of these compounds in different mat layers is poorly known. Kiritimati in the Central Pacific offers a unique opportunity for studying well developed,

* Corresponding author at: Geoscience Center, Geobiology Group, Georg-August-University Göttingen, Goldschmidtstr. 3, 37077 Göttingen, Germany. Tel.: +49 0 551 3913756; fax: +49 0 551 397918.
E-mail address: martin.blumenberg@geo.uni-goettingen.de (M. Blumenberg).

0146-6380/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.orggeochem.2012.11.004

thick stratified mats, which grow in numerous lakes on this atoll (Trichet et al., 2001). In a recent study of layers of a cyanobacterial mat from Kiritimati (Lake 2a; numeration as in Bühring et al., 2009), biomarkers amenable to gas chromatography were reported, but only 17β(H),21β(H)-bacteriohopanetetrol (BHT), the most common BHP in nature, was found in traces in a deep layer of the mat (Bühring et al., 2009). The authors attributed it to Cyanobacteria of the genus *Aphanocapsa*, the predominant cyanobacterial morphotype in the layer. Highest cell numbers of *Aphanocapsa* were, however, found in another layer, where no BHT was observed (Bühring et al., 2009).

We have analyzed the spatial distributions of BHPs in a cyanobacterial mat from Lake 21 on Kiritimati, which has a similar salinity to Lake 2a. The microbial composition of the mat has been studied in detail (Arp et al., 2012). In the 20 mm thick top orange layer, coccoid and filamentous Cyanobacteria predominate [*Cyanothece* (i.e. the morphotype *Aphanocapsa* in Bühring et al., 2009) and *Leptolyngbia* morphotypes]. Various Cyanobacteria are also present in the sublacent green layer, where decaying colonies of *Cyanothece* and common *Chroococcus* colonies occur. In the underlying olive colored layer, and in the lowermost purple layer [most likely representing the daytime oxic–anoxic transition (redoxcline)], numbers of Cyanobacteria were relatively low or cyanobacterial sheaths empty, while various types of proteobacteria were found to be relatively important. Further, in the bottom layers, purple non-sulfur bacteria become increasingly abundant, and are likely responsible for the purple color of the layer.

Our study demonstrates depth-related distributions and abundances of BHPs in the different mat layers. The results indicate multiple sources of BHPs but only a minor, if any, contribution from Cyanobacteria in the mat.

2. Material and methods

2.1. Samples

We studied five layers from a 5 cm thick mat sampled in 2002 at a saline lake on Kiritimati (01°52′N; 157°20′W; Lake 21; Fig. 1). The site was chosen because of the extraordinary well developed microbial mats in the lake. Various layers were visually distinguishable: an upper orange layer [most likely corresponding to layer 1 in a mat studied by Bühring et al. (2009)], a subjacent green layer underlain by two olive layers below [similar to layer 2 in Bühring et al. (2009), a purple layer (similar to layer 3 in Bühring et al. (2009; Fig. 2)]. The layers were separated in the field and kept frozen until freeze drying and further processing in the home laboratory. A second sampling was carried out in 2011 and a similar mat in the same lake was sampled for metagenomic studies.

2.2. C/N bulk analysis

Homogenized aliquots of the dried samples were subjected to bulk total C/N analysis, performed using a Hekatech Euro EA CNS analyser. To determine the relative amount of organic carbon (C_{org}) and carbonate carbon (C_{carb}) the samples were analyzed with a temperature-programmable RC 412 carbon analyser (LECO, St. Josseph, MI, USA). The ground sample was placed in a quartz combustion boat and heated in a stream of O_2 (750 ml/min). The time- and temperature-resolved CO_2 release was detected by way of an infrared absorption cell. First, C_{org} was released and detected as CO_2, on heating from 150 to 520 °C at 90 °C/min. After a hold time of 120 s, the sample was further heated to 900 °C at 90 °C/min, resulting in the release of C_{carb} as CO_2.

2.3. Extraction

Aliquots (ca. 1 g) of freeze dried mat samples were extracted with dichloromethane (DCM), DCM/MeOH and MeOH using ultrasonication. All extracts were combined.

2.4. Workup and analysis of non-extended hopanoids using gas chromatography–mass spectrometry (GC–MS)

To extract and transesterify fatty acids (FAs) an aliquot of the combined extract was subjected to acid hydrolysis using trimethylchlorosilane/MeOH (1:8, v:v) after a method described by Blumenberg and Michaelis (2007). The reaction mixture was re-extracted with *n*-hexane and separated into hydrocarbons, FA methyl esters and alcohols/ketones using column chromatography (Blumenberg and Michaelis, 2007). The fractions were analyzed by way of GC–MS using a Varian CP-3800 gas chromatograph coupled to a Varian 1200L mass spectrometer. The system was equipped with a fused silica column (Phenomenex Zebron ZB-5MS; 30 m × 0.32 mm i.d., 0.25 μm film thickness). He was the carrier gas. Samples were injected 'on column' into a Varian PTV 1279 temperature-programmable injector. In the injector, the samples were heated from 80 °C [injection temperature, 0.2 min (isothermal) to 320 °C (held 5 min) at 150 °C/min]. The GC oven was

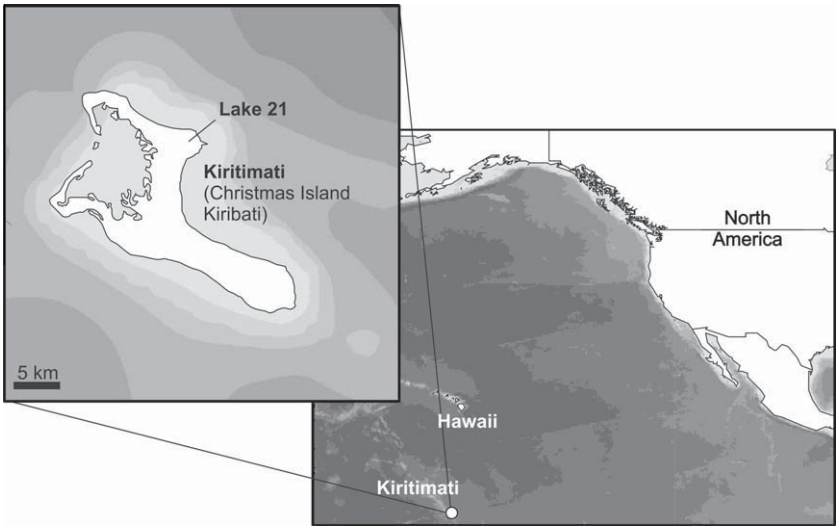


Fig. 1. Geographical position of Kiritimati Island (Republic of Kiribati) and sampling site (Lake 21).

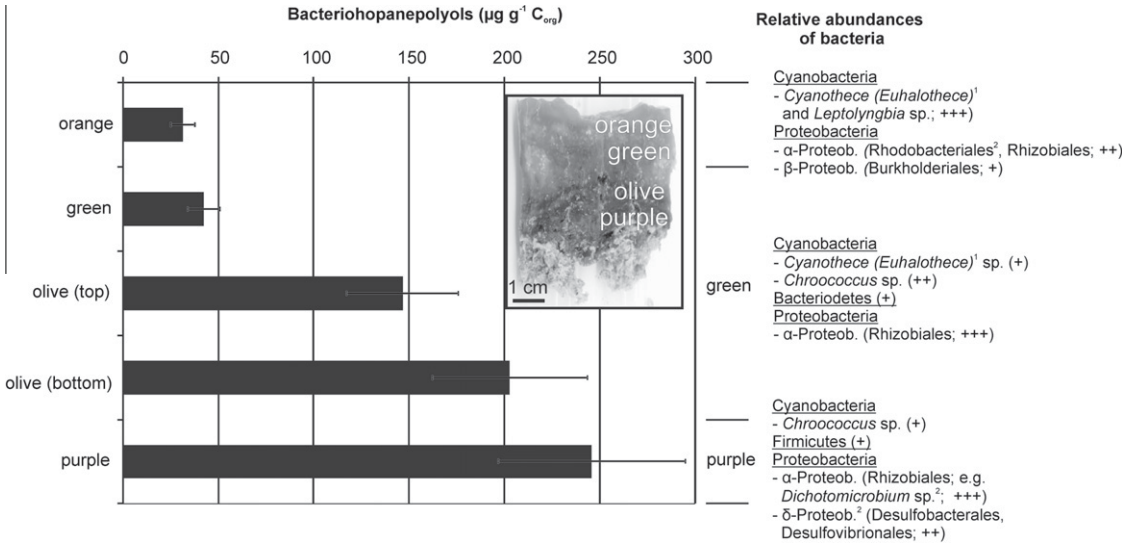


Fig. 2. BHP concentration in layers of cyanobacterial mat from Kiritimati and an image of the structured hypersaline mat. Most data were gained from the same mat from 2002 published by Arp et al. (2012) and are generally supported by metagenomic data from a similar mat sampled in 2011. [1,2]Specified results from the metagenomic approach. [1] *Euhalothece* sp. most likely corresponds to the morphologically identified *Cyanothece* sp. in Arp et al. (2012).

programmed from 80 °C (3 min) to 310 °C (held 25 min) at 4 °C/min. Biomarkers were assigned by comparing mass spectra and retention times with published data and/or reference compounds.

2.5. Workup and analysis of BHPs using liquid chromatography (LC)–MS

Another aliquot of the combined extract was acetylated using a mixture of Ac_2O and pyridine (1:1, v:v, 50 °C, 1 h then overnight at room temperature). Details are described by Blumenberg et al. (2009). Briefly, the pyridine/Ac_2O mixture was dried under vacuum and analyzed for BHPs via LC–MS, using a Prostar Dynamax high-performance liquid chromatography (HPLC) system interfaced with a 1200L triple quadrupole mass spectrometer (both Varian). HPLC separation was achieved using a Merck Lichrocart [Lichrosphere 100; RP C_{18e} column (250 × 4 mm) and a Merck Lichrosphere pre-column of the same material]. The flow rate was 0.5 ml min^{-1}. The solvent gradient profile was: 100% A (0–1 min) to 100% B at 35 min, then isocratic to 60 min (solvent A, MeOH/water 9/1; v/v; solvent B, MeOH/propan-2-ol 1/1; v/v; all Fisher Scientific HPLC grade). The MS instrument was equipped with an atmospheric pressure chemical ionization (APCI) source operated in positive ion mode (capillary temperature 150 °C, vaporizer temperature 400 °C, corona discharge current 8 µA, nebulizing gas flow 70 psi and auxiliary gas 17 psi). BHP concentration was determined by comparison of selected ions (single ion monitoring, SIM, mode) from acetylated BHP peaks with those of known concentrations of authentic BHP standards (acetylated BHT and 35-aminobacteriohopanetriol). BHP abundance was corrected for the individual response of amino- and non-amino-BHPs. The error in quantification was estimated to be ±20%. Assignment of individual BHPs was performed by comparison with elution times of previously identified compounds, as well as by MS–MS experiments. MS–MS was realized from ions with *m/z* 1002, 655, and 669. The occurrence of ring A methylated BHTs was tested via MS–MS experiments fragmenting *m/z* 669 (and via co-elution experiments with acetylated extracts of *Acetobacter xylinum* and *Beijerinckia indica*). C-2 methyl BHT elutes slightly before, while C-3 methyl BHT elutes after, desmethylated BHT (as reported by Pearson et al. (2009)). C-2 methyl BHT was assigned from the diagnostic ion at *m/z* 205, as well as from co-elution with an acetylated extract from *B. indica* containing C-2 methyl hopanoids. C-2 methyl BHT was not reported from *B. indica* (Vilcheze et al., 1994), but was found to be present in low abundance (own unpublished MS–MS and co-elution results). C-3 methyl BHT was assigned by the diagnostic ion at *m/z* 205, as well as from co-elution with an acetylated extract from *A. xylinum* containing C-3 methyl BHT (Peiseler and Rohmer, 1992).

2.6. Metagenomics

Additional information on the distribution of bacteria in the Kiritimati mats was obtained from metagenomic analyses of a mat sampled from the same lake in 2011. The mat contained identical macroscopic zones, but exhibited different thickness of the layers. Environmental DNA was isolated by employing the MoBio PowerBiofilm DNA isolation kit (MO BIO Laboratories, Carlsbad, USA). The isolation was performed using 100 mg microbial mat sample per layer as starting material. To identify taxonomic composition of bacterial communities, PCR-based 16S rRNA gene analysis was performed. The primers used covered the hypervariable regions V3 to V5 of the 16S rRNA gene. The resulting PCR products were purified by using the peqGold gel extraction kit (Peqlab Biotechnologie, Erlangen, Germany). All PCR reactions were performed 3× and were pooled in equal amounts. The Göttingen Genomics Laboratory determined the sequences of the partial 16S rRNA genes by using the Roche GS-FLX 454 pyrosequencer (Roche, Mannheim, Germany). Quality filtering, denoising and subsequent analyses of the recovered 16S rRNA gene sequences were performed using the QIIME 1.4 software package (Caporaso et al., 2010). To detect potential chimeric sequences UChime was employed (Edgar et al., 2011). Classification of the sequences was performed via similarity searches using BLAST (Altschul et al., 1990) against the SILVA database version 111 (Pruesse et al., 2007).

3. Results

3.1. Bulk data

Bulk geochemical data for individual mat layers are shown in Table 1. The highest amount of C_{org} (9.8%) was in the green mat. C_{carb} was observed in all mat layers, but was highest in the olive (top) sample (2.6%). Since the mats were not washed, NaCl comprised a significant and non-quantified portion of the dry wt.

3.2. Bacterial composition

Bacterial composition of the mat, based on morphotype-analysis and full and partial 16S rDNA sequencing, has been reported by Arp et al. (2012) and is briefly described above. In Fig. 2, the relative abundances of bacterial groups, showing a decrease in cyanobacterial and increase in proteobacterial numbers from top to bottom layers, are presented. Metagenomic data from a similar mat sampled in 2011 supported the general bacterial distributions, but were partly more specific. For instance, the metagenomic approach showed that *Cyanothece* morphotypes, described by Arp et al. (2012), most likely represent *Euhalothece*, which were the prominent Cyanobacteria in the mat. Moreover, in the purple layer, sulfate reducing δ-Proteobacteria and *Dichotomicrobium* spp. (α-Proteobacteria) were found to be important members of the bacterial community. Minor differences between both data sets (and/or mat samples) were also found. While *Leptolyngbia* morphotypes were reported to be relatively important (Arp et al., 2012), metagenomic data revealed only low numbers of this and other filamentous Cyanobacteria.

3.3. FA classes and non-extended hopanoids

FAs were found in all mat zones. Highest total concentration was in the lower olive layer (ca. 360 μg g^{-1} C_{org}; Table 1), which also had the highest relative abundance of polyunsaturated FAs (PUFAs). The distributions are not shown in detail, but were similar to those in a comparable mat (Bühring et al., 2009). Highest amounts of diplopterol, diploptene [detected as the acid-catalyzed transformation product hop-17(21)-ene (Ageta et al., 1987)], 17β(H),21β(H)-*bis*homohopanol and 17β(H),21β(H)-*bis*homohopanoic acid methyl ester were found in the lower olive part of the mat (Table 1).

3.4. BHPs

The most abundant structures in the cyanobacterial mat included BHT, 35-aminobacteriohopanetriol and a BHT cyclitol ether. We also observed isomers of 35-aminobacteriohopanetriol and, most likely, of a BHT cyclitol ether. The latter was evidenced by the MS–MS fragmentation of *m/z* 1002, typical for BHT cyclitol ether, using a green mat extract, which gave ions at *m/z* 330 for both peaks, and at *m/z* 348 only for the first. Elution characteristics and MS–MS spectra were similar to BHPs with *m/z* 1002 described by Blumenberg et al. (2010); interpretations here are based upon fragmentations reported by Talbot et al. (2003). Fragments diagnostic for BHT glucosamine, for example ions at *m/z* 270 and 210 (Cooke, 2011), were not observed, indicating the second peak to be an isomer of BHT cyclitol ether (Table 2). C-2 and C-3 methyl BHPs were below detection limit in this layer.

The lowest concentration of total BHPs was in the top orange layer (total 34 μg g^{-1} C_{org}; Table 1, Fig. 2). Notably, total BHP concentration showed a continuous increase with depth (Table 1; Fig. 2). Indeed, it was nearly an order of magnitude more abundant in the lowermost purple layer (246 μg g^{-1} C_{org}) than at the top of the mat. Maximum individual concentration in the deeper mat layers was again observed for BHT and 35-aminobacteriohopanetriol. Within the deeper parts of the mat, a low amount of ring A methylated BHTs was found in the lower olive (C-3 methyl BHT) and the purple layer (C-2 and C-3 methyl BHT; Fig. 3).

In the deeper mat layers, a relatively high amount of a BHT isomer, eluting slightly later than the common BHT, was observed. From its elution and MS characteristics [see discussion by Blumenberg et al. (2010)], it was tentatively assigned to have the uncommon 22*S*-configuration. Isomerisation at C-17 (β to α) could

Table 1
Bulk geochemistry, FA class distribution (% total FAs) and hopanoid biomarker data from top (orange) to bottom (purple) layers of the Kiritimati mat.

Layer	Bulk data				Biomarkers							
	C_{org} (%)	C_{carb} (%)	N (%)	C_{org}/N	BHPs (μg g^{-1} C_{org})	Diploptene (μg g^{-1} C_{org})[a]	Diplopterol (μg g^{-1} C_{org})	β,β-C_{32}-BH[b]	β,β-C_{32}-BHA[c]	Sum FAs (μg g^{-1} C_{org})	PUFAs (%)[d]	MUFAs (%)[e]	Sat. FAs (%)[f]
1 Orange	6.4	1.7	0.6	10.7	33.7	4.1	–	–	–	257.0	7.2	23.1	52.8
2 Green	9.8	0.9	0.8	12.3	61.2	7.8	–	–	–	232.7	4.5	17.0	60.0
3 Olive (top)	6.1	2.6	0.7	8.7	146.7	17.4	38.0	3.5	4.6	174.7	1.6	21.0	2.5
4 Olive (bottom)	7.0	0.4	0.6	11.7	202.8	26.4	3.6	19.4	20.0	365.2	1.8	16.4	54.4
5 Purple	6.4	1.9	0.9	7.1	245.8	13.4	1.3	4.5	–	126.0	–	45.2	54.8

[a] Analyzed as hop-17(21)-ene (H^+ treatment of extract).
[b] 17β(H),21β(H)-*bis*homohopanol.
[c] 17β(H),21β(H)-*bis*homohopanoic acid methyl ester.
[d] Polyunsaturated FAs.
[e] Monounsaturated FAs.
[f] Saturated FAs. Remaining percentages of sum of FAs, PUFAs and MUFAs from 100 comprise *iso*- and *anteiso*- and cyclic FAs.

Table 2
Assignment of BHPs in individual mat layers of the Kiritimati mat to tentative source bacteria.[a,b] Ions from MS–MS experiments, where abundances allowed respective studies, are also given.

	Diagnostic ions[c]	High abundance in layer(s)	Tentative source in the mat[d,e]
BHT cyclitol ether	SIM 1002 (MS–MS of 1002: ions 330, 348)	1 (orange)	Various, Cyanobacteria?[d]
BHT cyclitol ether (2nd isomer)?	SIM 1002 (MS–MS of 1002: ion 330)	2 (green)	Cyanobacteria?[d,1]
35-Aminobacteriohopanetriol (2nd isomer)		1–2 (orange and green) 3–5 (olive to purple; minor)	Cyanobacteria?[d,2]
BHT isomer	SIM 655 (MS–MS of 655: ions 475, 191)	3–5 (olive top to purple)	α-Proteobacteria; e.g. Rhizobiales? (*Rhodopseudomonas* relatives?)[d,3] or acetic acid bacteria[4]
C-3 methyl BHT	SIM 669 (MS–MS of 669: strong ion at 205)	4–5 (olive bottom to purple)	α-Proteobacteria (acetic acid or methylotrophic bacteria[5,6])
C-2 methyl BHT	SIM 669 (MS–MS of 669: strong ion at 205)	5 (purple)	α-Proteobacteria (*Rhodopseudomonas* relative[7], redoxcline specific unknown bacterium?[d,8])

[a] Note that identification of compounds is also based on coelution experiments with previously identified compounds (see Section 2).
[b] The most abundant but non-specific BHPs (BHT and 35-aminobacteriohopanetriol) are not included.
[c] SIM = Selected ion monitoring, ions from MS–MS experiments, where abundance allowed, are also given.
[d] Based on bacterial distribution in the mat layers.
[e] References are as follows: [1] Hermann et al., (1996), [2] no report from cultures, but high in a Cyanobacteria-rich setting (Talbot et al., 2008b)), [3] Neunlist et al. (1988), [4] Peiseler and Rohmer (1992), [5] Zundel and Rohmer (1985), [6] Renoux and Rohmer (1985), [7] Rashby et al. (2007), [8] Sáenz et al. (2011).

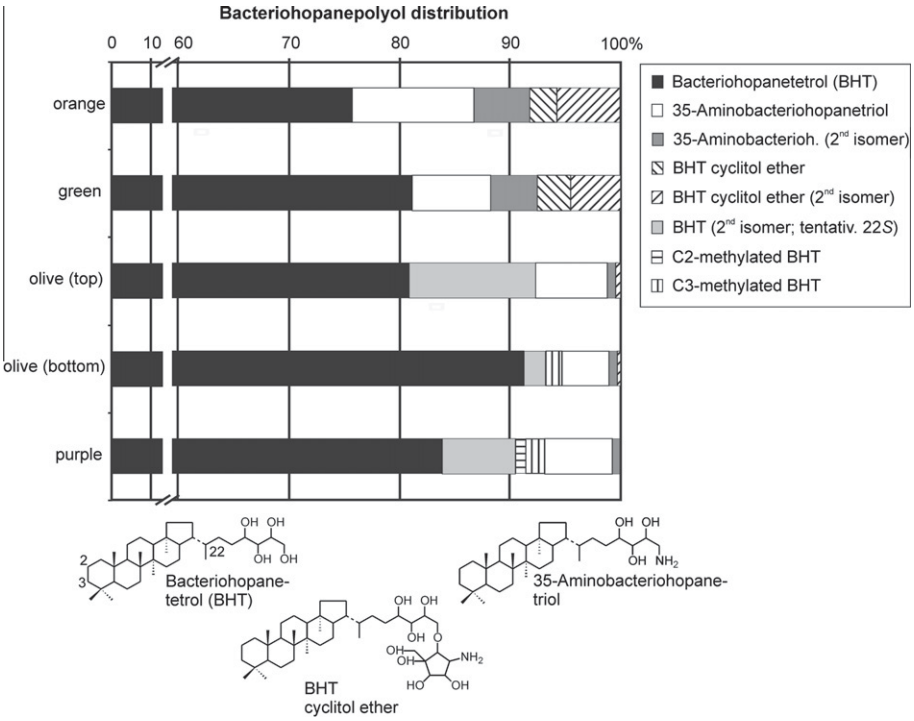


Fig. 3. Structures and distributions of BHPs in individual layers of the cyanobacterial mat from Kiritimati.

be excluded because of the different elution characteristics of 17α(H),21β(H)-BHT (Talbot et al., 2008a) vs. the isomer in the mat.

4. Discussion

Our study was aimed at an understanding of sources of bacteriohopanepolyols in a stratified microbial mat from a hypersaline lake in Kiritimati (Lake 21). The salinity of the lake at the surface was 124.6. The proportion of C_{org} and proportion of N appear not to be related to mat depth. The highest amount of C_{org} in the green layer possibly indicates the most active mat layer with respect to carbon fixation. A similar C_{org} distribution were reported for a microbial mat from a different, brackish Lake on Kiritimati (Lake 30; Trichet et al., 2001). In the mat of our study, C_{carb} showed no increase with depth and, except for scattered aragonite spherulites, major carbonate precipitation took place in the deeper parts of the green layer and the purple layer and proceeded at a low rate in anoxic zones below [see Arp et al. (2012) for details].

Cyanobacterial mats are abundant in the various lakes and ponds on Kiritimati and have been studied for the deposition of carbonate and organic laminated sediments (Trichet et al., 2001), the influence of salinity on the fractionation of D/H ratio values in selected lipids (Sachse and Sachs, 2008) and phototrophic carbon fixation (Bühring et al., 2009). In the latter study, the carbon uptake into individual lipids after stable isotope probing with ^{13}C labeled substrates was investigated. It was shown that the food chain is based on photoautotrophic carbon fixation by Cyanobacteria.

The microbial mat investigated here for hopanoid distributions strongly resembles a hypersaline microbial mat from a nearby location that has been described in detail with respect to its microbial inventory and biogeochemistry (Arp et al., 2012). Briefly, in the top part coccoid *Cyanothece*- and filamentous *Leptolyngbya*-like

morphotypes predominated, while in the deeper layers chroococcoid Cyanobacteria were prominent (Fig. 2 for general bacteria distribution). The relative abundance of Cyanobacteria decreased with depth, while Proteobacteria (mainly members of the α-Proteobacteria, such as prominent Rhizobiales, minor β-Proteobacteria, and δ-Proteobacteria such as Desulfobacterales and Desulfovibrionales) increased. Unfortunately, neither Arp et al. (2012) nor Bühring et al. (2009) further subdivided the green-colored layers in their mat profiles [layer 2 of Bühring et al. (2009)] and thus, specific information on the sub-samples used in this study (i.e. green, top olive and bottom olive layers) is lacking. Nevertheless, the study of Arp et al. (2012 – Fig. 17) showed that the deepest, olive parts of the green-colored layer differ from higher parts by a collapse of the exopolymer framework and onset of microcrystalline aragonite precipitation, abundant non-phototrophic bacteria (mainly heterotrophs) and degradation of cyanobacterial colonies.

The distribution of microbial groups is in principle mirrored in the major classes of FAs. The highest relative amount of polyunsaturated FAs (PUFAs) was found in the top orange layer (Table 1). PUFAs are often attributed to eukaryotes, particularly diatoms, which have, although in low abundance, been reported from the top part of the mat (Arp et al., 2012). A high abundance of PUFAs was also found in numerous Cyanobacteria (Kenyon, 1972), which may thus represent a considerable source of PUFAs in the Kiritimati mat. Consequently, the relative decrease in PUFAs with depth can be interpreted not only in terms of their susceptibility to diagenetic alteration, but also by a decreasing abundance of diatoms and/or cyanobacterial biomass with depth. Other more specific cyanobacterial biomarkers – e.g. methylated heptadecanes (Shiea et al., 1990) – were found only in traces.

Non-extended hopanoids were found in all layers of the mat. Diploptene was highest in the lower olive layer (ca. 26 $\mu g\ g^{-1}\ C_{org}$) and lowest in the orange top layer (ca. 4 $\mu g\ g^{-1}\ C_{org}$), while the other layers had intermediate concentration (Table 2). Diploptene, diplopterol and BHPs, which often occur together in bacteria (Rohmer et al., 1984), demonstrated comparable, but not identical, increasing trends with depth. A similar high proportion of hopanoids in the deeper layers of cyanobacterial mats has occasionally been reported (Grimalt et al., 1992; Scherf, 2008; Scherf and Rullkötter, 2009), but further details on the distributions of BHPs in these systems are lacking.

Distinct suites of BHPs in the different layers of the Kiritimati mat indicate multiple BHP sources. In accord with the finding of BHT in layer 3 in a similar Kiritimati mat (Bühring et al., 2009), we observed the highest concentration of BHPs in this (purple) layer (Figs. 2 and 3). In addition to production in this layer, BHP accumulation with depth is a plausible explanation. But, the relevance of this and other artificial processes is most likely negligible, as indicated by the lack of accumulation of other lipid classes with comparable stability (e.g. summed FAs; Table 1). Likewise, dilution of the lipid signal by exopolymeric substances (EPSs) has to be considered as a factor influencing the abundance of hopanoids as a fraction of C_{org}. In the upper part of cyanobacterial mats, considerable portions of the OM consist of carbohydrate-like substances (Wieland et al., 2008). On the other hand, a detailed study revealed that EPSs make up only a few percent of C_{org} in a similar cyanobacterial mat (Klock et al., 2007). Therefore, dilution of the lipid signal by EPSs is probably not a major control on the hopanoid distributions.

In concert with the low concentrations of BHPs in the top orange and green layer of the mat, the bacteria thriving in this level produce rather non-specific structures, mostly BHT and 35-aminobacteriohopanetriol. These BHPs are produced by many bacterial groups and are therefore widespread in the geosphere. It can be argued that they originate at least partly from the filamentous (*Leptolyngbya*) or coccoid (*Cyanothece*) Cyanobacteria predominating in this interval. A cyanobacterial source is also likely for a second isomer of 35-aminobacteriohopanetriol, which has never been reported from a bacterial isolate, but from a cyanobacterial-rich lake setting (Talbot et al., 2008b). Similarly, the second isomer of a BHT cyclitol ether (Fig. 3) has only been reported from *Zymomonas mobilis* and the cyanobacterium *Anacystis montana* (Chroococcales; Herrmann et al., 1996), so that an origin from *Anacystis*-related Chroococcales like *Aphanothece* appears plausible. C-2 methylated BHPs were not observed in the top layers, excluding the predominating Cyanobacteria in the mat as source for this class of BHPs.

The low amount of BHPs in the orange top and in the underlying green layer suggest that the predominating Cyanobacteria contribute only a minor fraction of the total BHPs in the Kiritimati mat. Our data are in line with the scattered occurrence of hopanoid production in Cyanobacteria in general (Talbot et al., 2008b) and studies revealing the lack of hopanoids in specific strains [*Leptolyngbya* (related to *Microcoleus*; own unpublished results); Summons et al., 1999; Bauersachs et al., 2010]. However, from other related strains, hopanoid production has been frequently reported, including C-2 methylated structures (Jahnke et al., 2004).

In the deeper, olive layers, the abundance of BHPs clearly increased, accompanied by the appearance of more specific BHP structures. In this depth interval, distinct non-*Leptolyngbya* morphotypes of Cyanobacteria and empty cyanobacterial sheaths were still present in low abundance, whereas other (proteo-)bacteria became increasingly important, including particularly members of the order Rhizobiales (Arp et al., 2012). Relatives of these Proteobacteria are reported as being a rich source of BHPs in soils and other environments (Kannenberg et al., 1995; Cvejic et al., 2000). Likewise, the tentatively assigned 22*S*-BHT isomer, an abundant compound in the deeper mat layers, has only been reported from proteobacteria. Known sources of 22*S* BHT are acetic acid bacteria (Peiseler and Rohmer, 1992; e.g. *Acetobacter* spp.) and the purple non-sulfur bacterium *Rhodopseudomonas acidophila* (Neunlist et al., 1988). In the mat, the abundances of both the tentative 22*S*-BHT isomer and Proteobacteria increased with depth, so we consider bacteria of this group as a main source of this and other BHPs in the olive and purple layers. Support for this comes from the presence of C-3 methyl BHT in the lower olive and purple mat layers, a compound that has only been described from the Proteobacteria *Gluconacetobacter xylinus* (an acetic acid bacterium, Zundel and Rohmer, 1985) and *Methylobacterium organophilum* (Renoux and Rohmer, 1985). C-2 methyl BHPs were in relatively low concentration and occurred only in the deepest (purple) layer of the mat, where processes other than oxygenic photosynthesis prevailed. Plausible candidate microbes are purple non-sulfur bacteria performing anoxygenic photosynthesis, an important process in stratified microbial mats in general (Fourçans et al., 2004). Whereas *Rhodopseudomonas palustris* (order Rhizobiales), a member of the purple non-sulfur bacteria and a known C-2 methyl BHP producer, was not found in the Kiritimati mat (Arp et al., 2012), other members of the Rhizobiales may be responsible for the occurrence of C-2 methyl BHPs as well as for the intense color (e.g. *Dichotomicrobium* relatives, which were prominent in the deeper mat layers, although these bacteria have not been tested for BHP production). Contributions from anaerobic bacteria can also not be excluded as further feasible BHP sources, since Planctomycetes, *Geobacter* spp. and *Desulfovibrio* spp. are capable of hopanoid production (Sinninghe Damsté et al., 2004; Fischer et al., 2005; Härtner et al., 2005; Blumenberg et al., 2006, 2012), as well as bacteria related to *Desulfovibrio*, the latter of which were found to be present in the purple layer (and below; not shown). Nevertheless, the high amount of BHPs in the purple layer remains somewhat enigmatic, also because bacterial lipid contributions appeared to be generally lower in this layer, as indicated by the relatively low

amount of FAs. Second, purple sulfur bacteria were probably important components of this layer (e.g. Bühring et al., 2009), but have not been reported to contain BHPs.

Even if the abundance of BHPs is overestimated because of the relatively large error in the LC–MS quantification, the major BHP production obviously took place in the deeper layers of the mat. Comparable biogeochemical zones in other settings are also often enriched in BHPs (e.g. at redoxcline of stratified marine water columns; Wakeham et al., 2007; Sáenz et al., 2011). Moreover, although not further characterized, an isomer of BHT similar to that we found near the redoxcline of the hypersaline microbial mat was previously reported from marine redoxclines (Sáenz et al., 2011), and may be considered specific for these conditions. In the corresponding depth interval we also observed a C-2 methyl BHT, and its occurrence was also related to the oxic–anoxic transition zone in the Black Sea water column (Wakeham et al., 2007). Consequently, we may speculate that the respective bacterial sources of these particular hopanoids appear to be related to a specific biogeochemical zone and not a definite environment. Future studies should aim to identify the phylogenetic affiliation and/or the conditions fostering growth of these bacteria or the enhanced production of BHPs at redoxclines.

5. Conclusions

Our study indicated that the majority of BHPs in a hypersaline cyanobacterial mat from Kiritimati was produced by Proteobacteria thriving below the carbon-fixing Cyanobacteria. In the topmost layers of the mat, filamentous Cyanobacteria were the most likely producers of low amounts of mostly non-specific BHPs, together with a more specific isomer of BHT cyclitol ether and an isomer of 35-aminobacteriohopanetriol. Bacteria in the deeper mat layers were found to produce much higher amounts of BHPs, including a tentatively assigned 22*S*-BHT and C-3- and C-2 methylated structures. Respective depth distributions in the mat excluded Cyanobacteria as a source for the latter BHPs. Our study of a microbial mat and analogies with BHP abundances in stratified marine water columns suggest that BHP production is enhanced at redoxclines. Moreover, the production of a BHT isomer appeared to be also related to bacteria, most likely of the α-proteobacterial sub-group, thriving in these zones.

Acknowledgements

We thank K. Nödler and T. Licha for support with LC–MS, J. Germer for analytical support, A. Reimer for helpful discussions and W. Dröse for preparation of the original sample. We thank three anonymous reviewers, H.M. Talbot and S. Bühring for helpful comments on this and on an earlier version of the manuscript. The DFG (Grants BL 971/1-3, RE665/18, AR335/5, and through the Courant Center Geobiology), is kindly acknowledged for financial support. This is Publication No. 108 of the Courant Center Geobiology and FOR 571 "Geobiology of Organo- and Biofilms".

Associate Editor—**B. van Dongen**

References

Ageta, H., Shiojima, K., Arai, Y., 1987. Acid-induced rearrangement of triterpenoid hydrocarbons belonging to the hopane and migrated hopane series. Chemical and Pharmaceutical Bulletin 35, 2705–2716.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Arp, G., Helms, G., Karlinska, K., Schumann, G., Reimer, A., Reitner, J., Trichet, J., 2012. Photosynthesis versus exopolymer degradation in the formation of microbialites on the atoll of Kiritimati, Republic of Kiribati, Central Pacific. Geomicrobiology Journal 29, 29–65.

Bauersachs, T., Schouten, S., Compaoré, J., Stal, L.J., Sinninghe Damsté, J.S., 2010. Occurrence of C_{35}–C_{45} polyprenols in filamentous and unicellular cyanobacteria. Organic Geochemistry 41, 867–870.

Blumenberg, M., Krüger, M., Nauhaus, K., Talbot, H.M., Oppermann, B., Seifert, R., Pape, T., Michaelis, W., 2006. Biosynthesis of hopanoids by sulfate-reducing bacteria (genus *Desulfovibrio*). Environmental Microbiology 8, 1220–1227.

Blumenberg, M., Michaelis, W., 2007. High occurrences of brominated lipid fatty acids in boreal sponges of the order Halichondrida. Marine Biology 150, 1153–1160.

Blumenberg, M., Seifert, R., Kasten, S., Bahlmann, E., Michaelis, W., 2009. Euphotic zone bacterioplankton sources major sedimentary bacteriohopanepolyols in the Holocene Black Sea. Geochimica et Cosmochimica Acta 73, 750–766.

Blumenberg, M., Mollenhauer, G., Zabel, M., Reimer, A., Thiel, V., 2010. Decoupling of bio- and geohopanoids in sediments of the Benguela Upwelling System (BUS). Organic Geochemistry 41, 1119–1129.

Blumenberg, M., Hoppert, M., Krüger, M., Dreier, A., Thiel, V., 2012. Novel findings on hopanoid occurrences among sulfate reducing bacteria: is there a direct link to nitrogen fixation? Organic Geochemistry 49, 1–5.

Boon, J.J., Hines, H., Burlingame, A.L., Klok, J., Rijpstra, I.C., De Leeuw, J.W., Edmunds, K.E., Eglinton, G., 1981. Organic geochemical studies of Solar Lake laminated cyanobacterial mats. In: Bjorøy, M. et al. (Eds.), Advances in Organic Geochemistry. John Wiley, Chichester, pp. 207–227.

Bühring, S.I., Smittenberg, R.H., Sachse, D., Lipp, J.S., Golubic, S., Sachs, J.P., Hinrichs, K.-U., Summons, R.E., 2009. A hypersaline microbial mat from the Pacific Atoll Kiritimati: insights into composition and carbon fixation using biomarker analyses and a ^{13}C-labeling approach. Geobiology 7, 308–323.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7, 335–336.

Cooke, M., 2011. The Role of Bacteriohopanepolyols as Biomarkers for Soil Bacterial Communities and Soil Derived Organic Matter. PhD Thesis, Newcastle University, UK.

Cvejic, J.H., Putra, S.R., El-Beltagy, A., Hattori, R., Hattori, T., Rohmer, M., 2000. Bacterial triterpenoids of the hopane series as biomarkers for the chemotaxonomy of *Burkholderia*, *Pseudomonas* and *Ralstonia* spp. FEMS Microbiology Letters 183, 295–299.

Des Marais, D., 2003. Biogeochemistry of hypersaline microbial mats illustrates the dynamics of modern microbial ecosystems and the early evolution of the biosphere. Biological Bulletin 204, 160–167.

Dobson, G., Ward, D.M., Robinson, N., Eglinton, G., 1988. Biogeochemistry of hot spring environments: extractable lipids of a cyanobacterial mat. Chemical Geology 68, 155–179.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.

Fischer, W.W., Summons, R.E., Pearson, A., 2005. Targeted genomic detection of biosynthetic pathways: anaerobic production of hopanoid biomarkers by a common sedimentary microbe. Geobiology 3, 33–40.

Fourçans, A., de Oteyza, T.G., Wieland, A., Solé, A., Diestra, E., van Bleijswijk, J., Grimalt, J.O., Kühl, M., Esteve, I., Muyzer, G., Caumette, P., Duran, R., 2004. Characterization of functional bacterial groups in a hypersaline microbial mat community (Salins-de-Giraud, Camargue, France). FEMS Microbiology Ecology 51, 55–70.

Grimalt, J.O., De Wit, R., Teixidor, P., Albaigés, J., 1992. Lipid biogeochemistry of *Phormidium* and *Microcoleus* mats. Organic Geochemistry 19, 509–530.

Grimalt, J.O., Yruela, I., Saiz-Jimenez, C., Toja, J., de Leeuw, J.W., Albaiges, J., 1991. Sedimentary lipid biogeochemistry of an hypereutrophic alkaline lagoon. Geochimica et Cosmochimica Acta 55, 2555–2577.

Härtner, T., Straub, K.L., Kannenberg, E., 2005. Occurrence of hopanoid lipids in anaerobic *Geobacter* species. FEMS Microbiology Letters 243, 59–64.

Herrmann, D., Bisseret, P., Connan, J., Rohmer, M., 1996. Relative configurations of carbapseudopentose moieties of hopanoids of the bacterium *Zymomonas mobilis* and the cyanobacterium '*Anacystis montana*'. Tetrahedron Letters 37, 1791–1794.

Jahnke, L.L., Embaye, T., Hope, J., Turk, K.A., van Zuilen, M., Des Marais, D.J., Farmer, J.D., Summons, R., 2004. Lipid biomarker and carbon isotopic signatures from stromatolite-forming, microbial mat communities and *Phormidium* cultures from Yellowstone National Park. Geobiology 2, 31–47.

Kannenberg, E.L., Perzl, M., Härtner, T., 1995. The occurrence of hopanoid lipids in *Bradyrhizobium* bacteria. FEMS Microbiology Letters 127, 255–262.

Kenig, F., Sinninghe Damsté, J.S., Kock-van Dalen, A.C., Rijpstra, I.C., Huc, A.Y., de Leeuw, J.W., 1995. Occurrence and origin of mono-, di-, and trimethylalkanes in modern and holocene cyanobacterial mats from Abu Dhabi, United Arab Emirates. Geochimica et Cosmochimica Acta 59, 2999–3015.

Kenyon, C.N., 1972. Fatty acid composition of unicellular strains of blue-green algae. Journal of Bacteriology 109, 829–834.

Klock, J.-H., Wieland, A., Seifert, R., Michaelis, W., 2007. Extracellular polymeric substances (EPS) from cyanobacterial mats: characterisation and isolation method optimisation. Marine Biology 152, 1077–1085.

Neunlist, S., Bisseret, P., Rohmer, M., 1988. The hopanoids of the purple non-sulfur bacteria *Rhodopseudomonas palustris* and *Rhodopseudomonas acidophila* and the absolute configuration of bacteriohopanetetrol. European Journal of Biochemistry 171, 243–252.

Pearson, A., Leavitt, W.D., Sáenz, J.P., Summons, R., Tam, M.C.-M., Close, H.G., 2009. Diversity of hopanoids and squalene-hopene cyclases across a tropical land-sea gradient. Environmental Microbiology 11, 1208–1223.

Peiseler, B., Rohmer, M., 1992. Prokaryotic triterpenoids. Bacteriohopanetetrols with new side-chain configuration from two *Acetobacter* species: structure determination and biogenetic implications. Journal of Chemical Research (S) 298–299, 2353–2369.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35, 7188–7196.

Rashby, S.E., Sessions, A.L., Summons, R.E., Newman, D.K., 2007. Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. Proceedings of the National Academy of Sciences USA 104, 15099–15104.

Rasmussen, B., Fletcher, I.R., Brocks, J.J., Kilburn, M.R., 2008. Reassessing the first appearance of eukaryotes and cyanobacteria. Nature 455, 1101–1104.

Renoux, J.M., Rohmer, M., 1985. Prokaryotic triterpenoids: new bacteriohopanetetrol cyclitol ethers from the methylotrophic bacterium *Methylobacterium organophilum*. European Journal of Biochemistry 151, 405–410.

Rohmer, M., Bouvier-Navé, P., Ourisson, G., 1984. Distribution of hopanoid triterpenes in prokaryotes. Journal of General Microbiology 130, 1137–1150.

Rontani, J.-F., Volkman, J.K., 2005. Lipid characterization of coastal hypersaline cyanobacterial mats from the Camargue (France). Organic Geochemistry 36, 251–272.

Sachse, D., Sachs, J.P., 2008. Inverse relationship between D/H fractionation in cyanobacterial lipids and salinity in Christmas Island saline ponds. Geochimica et Cosmochimica Acta 72, 793–806.

Sáenz, J.P., Wakeham, S.G., Eglinton, T.I., Summons, R.E., 2011. New constraints on the provenance of hopanoids in the marine geologic record: bacteriohopanepolyols in marine suboxic and anoxic environments. Organic Geochemistry 42, 1351–1362.

Scherf, A.-K., 2008. Geochemical Characterization of Microbial Mats from the Intertidal Area of the United Arab Emirates and Incubation Experiment for Investigating the Microbial Community. PhD Thesis, University of Oldenburg, Germany.

Scherf, A.-K., Rullkötter, J., 2009. Biogeochemistry of high salinity microbial mats – part 1: lipid composition of microbial mats across intertidal flats of Abu Dhabi, United Arab Emirates. Organic Geochemistry 40, 1018–1028.

Shiea, J., Brassell, S.C., Ward, D.M., 1990. Mid-chain branched mono- and dimethyl alkanes in hot spring cyanobacterial mats: a direct biogenic source for branched alkanes in ancient sediments? Organic Geochemistry 15, 223–231.

Sinninghe Damsté, J.S., Rijpstra, I.C., Schouten, S., Fuerst, J.A., Jetten, M.S.M., Strous, M., 2004. The occurrence of hopanoids in planctomycetes: implications for the sedimentary biomarker record. Organic Geochemistry 35, 561–566.

Summons, R.E., Jahnke, L.L., Hope, J.M., Logan, G.A., 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature 400, 554–557.

Talbot, H.M., Coolen, M.J.L., Sinninghe Damsté, J.S., 2008a. An unusual 17α,21β(H)-bacteriohopanetetrol in Holocene sediments from Ace Lake (Antarctica). Organic Geochemistry 39, 1029–1032.

Talbot, H.M., Summons, R.E., Jahnke, L.L., Cockell, C.S., Rohmer, M., Farrimond, P., 2008b. Cyanobacterial bacteriohopanepolyol signatures from cultures and natural environmental settings. Organic Geochemistry 39, 232–263.

Talbot, H.M., Summons, R., Jahnke, L.L., Farrimond, P., 2003. Characteristic fragmentation of bacteriohopanepolyols during atmospheric pressure chemical ionisation liquid chromatography/ion trap mass spectrometry. Rapid Communications in Mass Spectrometry 17, 2788–2796.

Trichet, J., Défarge, C., Tribble, J., Tribble, G., Sansone, F., 2001. Christmas Island lagoonal lakes, models for the deposition of carbonate–evaporite–organic laminated sediments. Sedimentary Geology 140, 177–189.

Vilcheze, C., Llopiz, P., Neunlist, S., Poralla, K., Rohmer, M., 1994. Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria Azotobacter vinelandii, Beijerinckia indica and Beijerinckia mobilis. Microbiology 140, 2749–2753.

Wakeham, S.G., Amann, R., Freeman, K.H., Hopmans, E.C., Joergensen, B.B., Putnam, I.F., Schouten, S., Sinninghe Damsté, J.S., Talbot, H.M., Woebken, D., 2007. Microbial ecology of the stratified water column of the Black Sea as revealed by a comprehensive biomarker study. Organic Geochemistry 38, 2070–2097.

Welander, P.V., Coleman, M.L., Sessions, A.L., Summons, R.E., Newman, D.K., 2010. Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. Proceedings of the National Academy of Sciences USA 107, 8537–8542.

Wieland, A., Pape, T., Möbius, J., Klock, J.H., Michaelis, W., 2008. Carbon pools and isotopic trends in a hypersaline cyanobacterial mat. Geobiology 6, 171–186.

Zundel, M., Rohmer, M., 1985. Prokaryotic triterpenoids 1. 3β-methylhopanoids from *Acetobacter* species and *Methylococcus capsulatus*. European Journal of Biochemistry 150, 23–27.

2 Phylogenetic Analysis of a Microbialite-Forming Microbial Mat from a Hypersaline Lake of the Kiritimati Atoll, Central Pacific

Dominik Schneider[1], Gernot Arp[2], Andreas Reimer[2], Joachim Reitner[2], Rolf Daniel[1]

PLoS ONE (2013), Volume 8, Issue 6, e66662

[1]Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany

[2]Geoscience Centre, Georg-August University Göttingen, Göttingen, Germany

---

Author contributions:

Conceived and designed the experiments: DS GA JR RD.

Performed the experiments: DS GA AR.

Analyzed the data: DS GA AR RD.

Wrote the paper: DS GA RD.

OPEN ACCESS Freely available online

PLOS ONE

Phylogenetic Analysis of a Microbialite-Forming Microbial Mat from a Hypersaline Lake of the Kiritimati Atoll, Central Pacific

Dominik Schneider[1], Gernot Arp[2], Andreas Reimer[2], Joachim Reitner[2], Rolf Daniel[1]*

1 Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany, **2** Geoscience Centre, Georg-August University Göttingen, Göttingen, Germany

Abstract

On the Kiritimati atoll, several lakes exhibit microbial mat-formation under different hydrochemical conditions. Some of these lakes trigger microbialite formation such as Lake 21, which is an evaporitic, hypersaline lake (salinity of approximately 170‰). Lake 21 is completely covered with a thick multilayered microbial mat. This mat is associated with the formation of decimeter-thick highly porous microbialites, which are composed of aragonite and gypsum crystals. We assessed the bacterial and archaeal community composition and its alteration along the vertical stratification by large-scale analysis of 16S rRNA gene sequences of the nine different mat layers. The surface layers are dominated by aerobic, phototrophic, and halotolerant microbes. The bacterial community of these layers harbored *Cyanobacteria* (*Halothece* cluster), which were accompanied with known phototrophic members of the *Bacteroidetes* and *Alphaproteobacteria*. In deeper anaerobic layers more diverse communities than in the upper layers were present. The deeper layers were dominated by *Spirochaetes*, sulfate-reducing bacteria (*Deltaproteobacteria*), Chloroflexi (Anaerolineae and Caldilineae), purple non-sulfur bacteria (*Alphaproteobacteria*), purple sulfur bacteria (*Chromatiales*), anaerobic *Bacteroidetes* (*Marinilabiacae*), *Nitrospirae* (OPB95), *Planctomycetes* and several candidate divisions. The archaeal community, including numerous uncultured taxonomic lineages, generally changed from *Euryarchaeota* (mainly *Halobacteria* and *Thermoplasmata*) to uncultured members of the *Thaumarchaeota* (mainly Marine Benthic Group B) with increasing depth.

Citation: Schneider D, Arp G, Reimer A, Reitner J, Daniel R (2013) Phylogenetic Analysis of a Microbialite-Forming Microbial Mat from a Hypersaline Lake of the Kiritimati Atoll, Central Pacific. PLoS ONE 8(6): e66662. doi:10.1371/journal.pone.0066662

Editor: Jose Luis Balcazar, Catalan Institute for Water Research (ICRA), Spain

Received February 12, 2013; **Accepted** May 8, 2013; **Published** June 10, 2013

Funding: This study has been supported by the Deutsche Forschungsgemeinschaft (DFG-FOR 571 "Geobiology of Organo- and Biofilms"; DA 374/5-2, AR335/5-3, publication 61) and the Courant Center Geobiology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rdaniel@gwdg.de

Introduction

Microbial mats are one of the best existing examples to study early life on Earth, as their fossil counterparts, stromatolites, could be dated back to approximately 3.5 Ga [1–3]. These very first ecosystems might have played a major role in forming our present atmosphere and thereby have paved the way for oxygen-dependent life [4,5]. In general, microbial mats are vertically multilayered carpets of microbial assortments including specialized consortia of *Bacteria*, *Archaea*, and *Eukarya* that are embedded in self-produced extracellular polymeric substances (EPS). The metabolism of these communities results in physicochemical gradients leading to mat layering and changes in mineral saturation (for review see [6,7]). Microbial mats are found in shallow aquatic environments all over the world such as thermal springs, hydrothermal vents, tidal flats, and hypersaline lakes [8]. Depending on the type, major functional groups in the microbial mats are photolithoautotrophs, aerobic/anaerobic heterotrophs, fermenters, sulfide oxidizers, and methanogens [7]. Some of these mats lead to the formation of microbialites, which are microbially induced mineral precipitations such as aragonite and gypsum. However, the processes of microbialite formation are still not fully understood [6].

Microbial mats were studied from Guerrero Negro (Baja California Sur, Mexico) [9–13], Highborne Cay (Bahamas) [14,15], Shark Bay (Australia) [16,17], and Solar Lake (Eilat, Israel) [18]. These studies revealed that microbial mats harbor a by far larger diversity of microbes than previously assumed. Furthermore, it has been shown that microbial mat communities possess a complex vertical distribution profile, which is mainly driven by light and oxygen penetration.

In this study, the microbial mat of the evaporitic, hypersaline Lake 21 located on Kiritimati atoll (Republic of Kiribati, Central Pacific) was investigated. This unique mineralizing mat is associated with the formation of decimeter-thick reticulate microbialites composed of aragonite and layers of gypsum crystals [19]. So far, initial studies regarding the hydrochemistry, microbial mat population and source of microbialite-formation have been conducted [19–22]. Recent studies analyzed the microbial community composition of the Kiritimati mats using microscopic observations, lipid biomarker analyses, bacteriohopanoids, or classical clonal 16S rRNA gene sequencing [19,20,22]. Initial 16S rRNA gene analysis revealed that bacteria affiliated to several phyla including *Cyanobacteria*, *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, *Chloroflexi*, and *Planctomycetes*, and a small number of *Archaea* (*Euryarchaeota* and *Crenarchaeota*, most crenarchaeal members were

reordered to *Thaumarchaeota* [23]) were present in the analyzed mat. Taking the small survey size of 359 16S rRNA gene sequences of previous studies and the high microbial diversity found in other microbial mats into account, the previous insights into the prokaryotic community composition were limited with respect to spatial distribution and coverage. Consequently, larger surveying efforts are required for the comprehensive description of the microbial communities in these ecosystems [9].

In this study, we investigated the bacterial and archaeal communities in nine different layers collected from a microbial mat of the hypersaline Lake 21. We analyzed approximately 197,000 16S rRNA gene sequences from *Bacteria* (122,000) and *Archaea* (74,000). Analysis revealed highly diverse microbial communities along the vertical stratification of the mat, which mirror the micro-environmental properties. The predominant bacterial and archaeal taxa of this ecosystem were identified. A large amount of the detected microbial taxa showed similarities to 16S rRNA gene sequences of so far uncultured organisms, suggesting the presence of novel species and metabolic traits in this habitat.

Materials and Methods

Site Description and Sample Collection

Kiritimati (formerly Christmas Island) is the world's largest atoll and part of the Northern Line Islands of Republic of Kiribati with a land area of approximately 321 km^2 [24]. Field work permits were kindly given by the responsible Ministry of Environment, Lands and Agricultural Development (MELAD) - Environment and Conservation Division, Republic of Kiribati (Bikenibeu, Tarawa, Republic of Kiribati). As microorganisms were studied, no experiments involving humans or vertebrates were conducted. Sampling of the microbial mat was carried out in March 2011 on the atoll of Kiritimati during a two-week expedition (Figure 1). The island harbors approximately 500 lakes with different salinities ranging from nearly freshwater to hypersaline conditions. The lakes are largely occupied by lithifying and non-lithifiying microbial mats [21,25].

The sampling site was located to the southeast shore line (1°96′N, 157°33′W) of the hypersaline Lake 21 which exhibited an average lake water salinity of 170‰. The lake was highly evaporated as samples were taken after a dry period of 11 months (Figure S1). Additionally, chimneys importing water with lesser salinity were observed in the middle of the lake. Lake water chemistry is depicted in Table S1. The structural profile of the microbial mat is well defined by a clear colored vertical zonation into nine layers (Figure 2). The mat surface was covered by 1 to 4 cm of lake water. The thickness of the mat was approximately 10 cm at the sampling site but reached up to 20 cm at the lake center. For each layer, four mat samples were collected at 3 pm in the described location. The mat was separated in the field into nine different layers, merged with RNALater (Qiagen, Hilden, Germany) to stabilize the nucleic acids. Subsequently, the samples were stored and transported frozen to Germany, where the samples were stored at −80°C until further use.

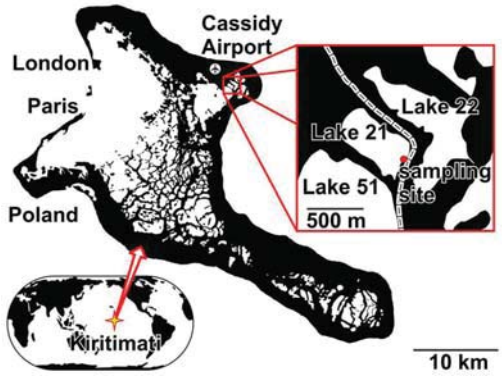


Figure 1. Map of Kiritimati Island and the sampling site at Lake 21.
doi:10.1371/journal.pone.0066662.g001

Extraction of Microbial DNA

Equal amounts (1 g) of the four random samples for each of the nine mat layers were mixed and subsequently homogenized by grinding with liquid nitrogen to minimize potential sample heterogeneity. Environmental DNA was isolated of each pooled sample by employing the MoBio PowerBiofilm DNA isolation kit (MO BIO Laboratories, Carlsbad, USA) following the guidelines of the manufacturer with slight modifications. For each DNA-isolation, 100 mg sample was used. The volume of buffer 3 was increased two-fold as described in the manual for strong contaminated samples. Final purification and concentration of DNA was achieved by employing SureClean (Bioline GmbH, Luckenwalde, Germany) as recommended by the manufacturer. The DNA yield was determined by using a NanoDrop ND-1000 spectrophotometer (Peqlab Biotechnologie GmbH, Erlangen, Germany).

Amplification of 16S rRNA Genes and Pyrosequencing

The hypervariable regions V3 to V5 of the 16S rRNA gene were amplified by PCR. For amplification of 16S rRNA genes the PCR reaction mixture (50 µl) contained 10 µl of 5-fold reaction buffer (Fusion GC buffer, Finnzymes, Vantaa, Finland), 200 µM of each of the four deoxynucleoside triphosphates, 0.2 µM of each primer, 5% DMSO (Finnzymes), 1 U of Phusion hot start high-fidelity DNA Polymerase (Finnzymes) and 50 ng of isolated DNA as template. Primer sequences for amplification of the V3 to V5 region included the Roche 454 pyrosequencing adaptors (underlined), a key (TCAG), and a variable multiplex identifier (MID) consisting of ten bases. For *Bacteria* the forward primer used was V3for_B 5′-CGTATCGCCTCCCTCGCGCCATCAG-MID-TACGGRAGGCAGCAG-3′ [26] and the reverse primer V5rev_B 5′-CTATGCGCCTTGCCAGCCCGCTCAG-MID-CCGTCAATTCMTTTGAGT-3′ [27]. For amplification of archaeal 16S rRNA genes the forward primer V3for_A 5′-CGTATCGCCTCCCTCGCGCCATCAG-MID-CCCTAYGGGGYGCASCAG-3′ [28] and the reverse primer V5rev_A 5′-CTATGCGCCTTGCCAGCCCGCTCAG-MID-GTGCTCCCCCGCCAATTCCT-3′ [29] were used. Taxonomic coverage of selected primer pairs was determined for *Bacteria* and *Archaea* with PrimerProspector [30] (Figure S2). The following thermal cycling scheme was used for amplification of partial bacterial 16S rRNA genes: initial denaturation at 98°C for 5 min, 25 cycles of denaturation at 98°C for 45 s, annealing for 45 s at 60°C, and extension at 72°C for 30 s, followed by a final extension period at 72°C for 5 min. For amplification of the archaeal 16S rRNA genes, the annealing temperature was adjusted to 57°C. Negative controls contained the entire reaction mixture without the template DNA. The resulting PCR products were checked by agarose gel electrophoresis for appropriate size and purified by gel extraction using the peqGold gel extraction kit as recommended by the manufacturer (Peqlab Biotechnologie GmbH). Quantification of the PCR products was performed using the Quant-iT dsDNA BR assay kit and a Qubit fluorometer (Invitrogen GmbH,

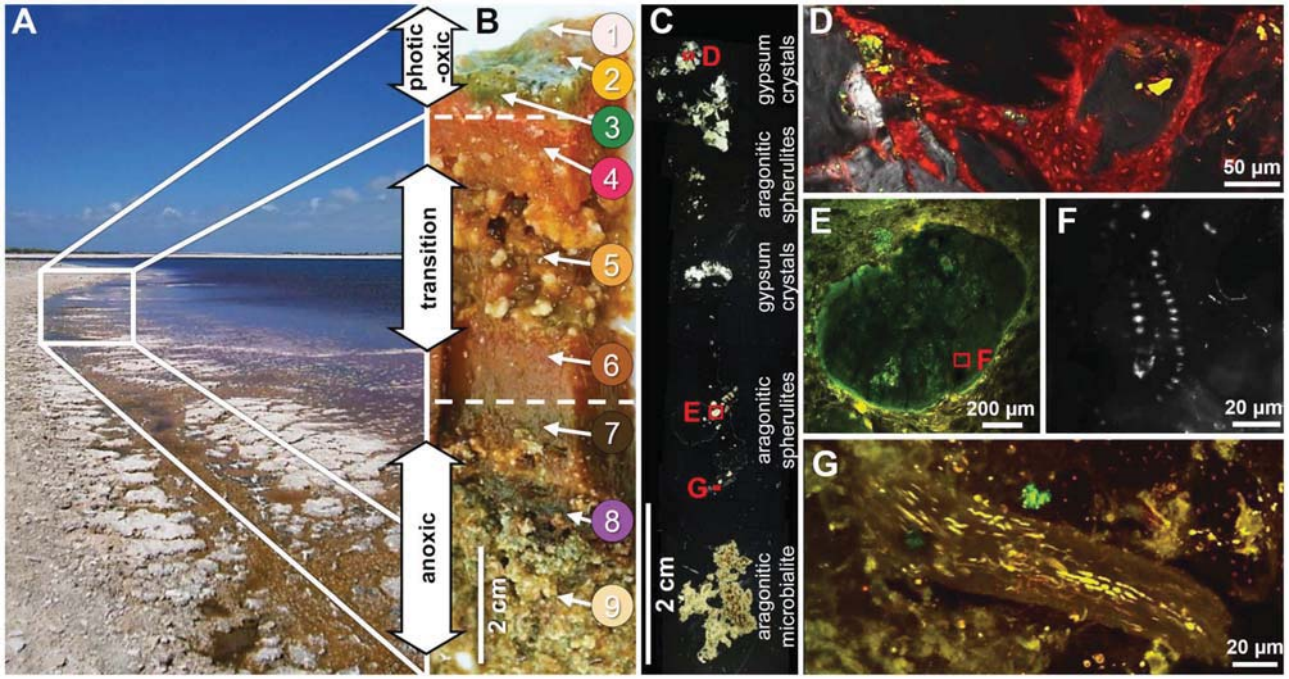


Figure 2. Localization of sampling site, structure, and characteristics of the microbial mat derived from Lake 21. Southwest view of the sampling site (A), cross-section of the microbial mat of Lake 21 (B), and vertical thin-section of the microbial mat and its mineral precipitates viewed under X-nicols (C). Detailed view of (C) showing *Cyanothece* with gypsum crystals from layer 3 (overlay of X-nicols and LSM image, ex 633 nm/em 640–702 nm) (D); spherulite with dead *Johannesbaptistia* from layer 7 (LSM micrograph ex 488, 543 and 633 nm/em 500–530, 565–615, 640–702 nm) (E); magnified subfossil *Johannesbaptistia* (LSM micrograph, ex 488 nm, em 500–550 nm) (F); ceased *Microcoleus* filament from layer 8 (LSM micrograph, ex 488, 543 and 633 nm/em 500–530, 565–615, 640–702 nm) (G). Description of microbial mat layers in (B): Layer 1 was a light pink, loose, and transparent film, scattered with black dots, which covered the mat only at the shoreline (0 to 2 m). Layer 2 was defined by bright orange coloration and could be observed in each sample. Layer 3 was green and layer 4 bright pink to red with small carbonate precipitations on top and was always located below the three surface layers. Layer 5 was a loose, slimy orange to brownish section interstratified with large gypsum crystals. Layer 6 was dark orange-brown with small carbonate precipitates at the bottom side. Layer 7 was dark brown and randomly seeded with small carbonate precipitates. Layer 8 was a dark purple layer, with reticulate microbialites. Layer 9 was creamy layer with sand like texture followed by a reticulate microbialite fabric [19].
doi:10.1371/journal.pone.0066662.g002

Karlsruhe, Germany) as recommended by the manufacturer. All PCR reactions were performed in triplicate and pooled in equal amounts.

The 16S rRNA gene sequences were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under project accession number SRA058120.

Processing and Analysis of 16S rRNA Gene Datasets

The Göttingen Genomics Laboratory determined the sequences of the partial 16S rRNA gene amplicons by using a 454 GS-FLX pyrosequencer (Roche, Mannheim, Germany) and Titanium chemistry following the instructions of the manufacturer for amplicon sequencing. Generated 16S rRNA gene datasets were processed and analyzed employing the QIIME 1.4 software package [31]. Initially, we removed sequences shorter than 250 bp, containing unresolved nucleotides, exhibiting an average quality score lower than 25, harbor mismatches longer than 2 bp in the forward primer, or possessing homopolymers longer than 8 bp with the QIIME script *split_libraries.py*. This also removes forward and reverse primer sequences. Subsequently, pyrosequencing noise was removed by employing the Denoiser 0.91 [32] included in QIIME [31]. We additionally removed unclipped reverse primer sequences (not detected by *split_library.py*) by employing cutadapt [33] with default settings. Chimeric sequences were removed using UCHIME [34] included in USEARCH (6.0.152). As the studied environment might harbor *Bacteria* and *Archaea*, which are not well represented by reference databases, we applied UCHIME in de novo mode. Subsequently, we used UCHIME in reference mode against the Greengenes Gold dataset (gold_strains_gg16S_aligned_19-Mar-2011.fasta) as the reference database [34,35]. Operational taxonomic unit (OTUs) determination was performed with the UCLUST algorithm [36] at a genetic divergence level of 3%, which represents species-level according to Schloss and Handelsman [37]). For this purpose, most abundant sequence within an OTU was picked by employing the QIIME scripts *pick_otus.py* and *pick_rep_set.py* [31]. OTUs clustered at 3% genetic divergence were the basis of all presented results. Taxonomic classification of the picked reference sequences (OTUs) was performed by similarity searches using BLAST (blastall version 2.2.18) [38] against the SILVA SSU database release 111 (SSURef_111_NR_tax_silva_trunc.fasta) [39]. The SILVA taxonomy of the first hit based on E-value was then adopted to infer taxonomy of the representative sequences according to the QIIME script *assign_taxonomy.py*. An OTU table was created using *make_otu_table.py*. Singletons, chloroplast and extrinsic domain OTUs were removed from the table by employing *filter_otu_table.py*. Details on the sequence data processing are given in Table S2. Finally, OTUs that exhibited low coverage (<95%) to their BLAST hit were designated as "Unclassified".

Diversity Indices and Rarefaction Analysis

The alpha-diversity indices such as Chao1 indices, Shannon indices, and Michaelis-Menten fit were calculated employing the QIIME script alpha_diversity.py *(10 replicates per sample) (Table 1). Shannon indices were multiplied by ln 2 to allow comparison with non-QIIME-based studies. Rarefaction curves were calculated from the modified OUT table with alpha_rarefaction.py* (step size 50 for *Bacteria* and 25 for *Archaea*; 10 replicates per sample). *The analyses were performed* at the same level of surveying effort (6,350 bacterial and 2,350 archaeal randomly selected sequences per sample).

Community Comparison of Microbial Mat Layers

To compare the bacterial and archaeal communities across all mat layers a principal component analysis (PCA) was performed using Canoco 4.56 (Microcomputer Power, Ithaca, NY, USA). The relative abundances of bacterial phyla, candidate divisions, proteobacterial subclasses, archaeal families and candidate groups, unclassified groups, and the artificial group "Other" (summary of all phyla, subdivisions and candidate divisions, accounting for less than 0.5% relative abundance in any given sample) were used. The scaling was focused on inter-taxa correlations and taxa scores were divided by standard deviation. The graph was centered by taxa.

For the bacterial community comparison of Guerrero Negro and Kiritimati microbial mat layers, the Guerrero Negro 16S rRNA clone library gene sequences of the ten analyzed layers were obtained from GenBank (Bioproject PRJNA29795). Sequences were trimmed to the corresponding 16S rRNA gene region (V3–V5) of the Kiritimati sequences by employing the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Sequences were further processed as described for the Kiritimati sequences. Weighted principle coordinate analysis (PCoA) was performed at the same level of surveying effort employing the QIIME scripts *beta_diversity_through_plots.py* and *make_3d_plots.py* for plotting OTUs to the weighted PCoA.

Lake and Pore Water Hydrochemistry

Temperature, electrical conductivity, pH, and redox potential of lake and pore waters were recorded *in situ* using portable instruments from WTW GmbH, Germany. The utilized Schott pH-electrode was calibrated against NBS buffers 7.010 and 10.010 (HANNA instruments). Total alkalinity was immediately determined after sampling by acid-base titration using a hand-held titrator and 1.6 N H_2SO_4 cartridges as titrant (Hach Corporation). Dissolved oxygen was analyzed titrimetrically following the Winkler method. Samples for determination of main anions and cations were collected in pre-cleaned PE-bottles. Samples for cation analysis were filtered in the field through 0.8 μm membrane filters (Millipore) and fixed by acidification. Main cations (Ca^{2+}, Mg^{2+}, Na^{+}, and K^{+}) and anions (Cl^{-} and $SO_4{}^{2-}$) were analyzed by ion chromatography with conductivity detection (Dionex Corporation and Metrohm). ICP-OES (Perkin Elmer) was used to determine Sr^{2+} and Fe^{2+}. Dissolved phosphate, silica, and ammonia concentrations were measured by spectrophotometric methods (Unicam). Measured values were processed with the computer program PHREEQC [40] in order to calculate ion activities and PCO_2 of the water samples as well as saturation state with respect to calcite, aragonite and gypsum.

Table 1. Bacterial and archaeal 16S rRNA gene diversity analyses of microbialite-forming Lake 21 mat at 3% genetic divergence level.

Sample	No. high quality reads	No. OTUs	Chao1	Shannon	Michaelis Menten fit	Coverage (%)
Bacteria						
Layer 1	9,445	95	135.8	2.01	111.6	85.4
Layer 2	6,646	67	79.7	2.43	69.4	96.9
Layer 3	6,315	98	130.6	2.77	101.9	96.1
Layer 4	6,279	142	166.2	3.32	154.8	91.7
Layer 5	7,505	163	184.3	3.68	172.6	94.4
Layer 6	12,580	145	174.9	3.36	155.9	93.2
Layer 7	9,741	227	275.1	4.10	248.9	91.2
Layer 8	10,417	273	317.2	4.37	299.6	91.0
Layer 9	11,117	199	242.0	3.34	226.8	87.6
Archaea						
Layer 1	3,107	28	29.1	1.53	30.1	93.1
Layer 2	3,805	22	33.6	1.16	25.0	89.0
Layer 3	3,418	27	32.3	1.33	29.3	91.4
Layer 4	3,524	27	35.6	1.47	31.6	84.9
Layer 5	10,663	26	33.6	1.91	26.1	98.0
Layer 6	7,235	36	52.4	1.65	42.3	86.0
Layer 7	3,080	33	41.9	1.44	37.0	88.6
Layer 8	2,544	41	48.1	2.11	43.0	96.1
Layer 9	2,353	27	33.0	1.37	33.2	81.2

doi:10.1371/journal.pone.0066662.t001

Results and Discussion

Characteristics of the Lake 21 Microbial Mat

The microbialite-forming mat samples were taken from the shoreline of the southeastern part of Lake 21 (Figure 1). The lake water has been analyzed with respect to physical and physico-chemical parameters, cations, anions and nutrients (Table S1). The hydrochemical data of lake and pore waters are depicted in Figure 3. The water chemistry of the lake was characterized by a pH of 7.95, a temperature of 32°C and an average salinity of 170‰. Cation and anion concentrations of the lake water were 2.6 M Na^+, 3.1 M Cl^-, 298.3 mM Mg^{2+}, and 143.1 mM SO_4^{2-}. These data correspond to the data gained by Arp *et al.* in 2002, but ion concentrations were approximately 1.6-fold higher in 2011. This might be a result of strong evaporation of the lake water caused by a dry period of almost one year prior to sampling in 2011 (Figure S1). Lake 21 pore waters generally show high Ca^{2+} and SO_4^{2-} concentrations, which are probably caused by surface water or groundwater [19]. Major ion supply to the lake water is currently due to the influx of less saline groundwater (Table S1) via chimney-like microbial mat structures. With respect to salinity, the hypersaline Lake 21 was more similar to the Solar Lake (salinity 200‰) than to Guerrero Negro (salinity 90‰) [9,11,18].

We observed nine color-separated mat layers with fractional carbonate precipitations (Figure 2B). A reticulate microbialite was observed below the microbial mat, as described earlier in 2002 [19]. In contrast to our study, Arp *et al.* found 2002 a separation in only four layers, which exhibited orange, green, purple, or grey/brown color [19]. The below-described phylogenetic analysis indicated that these four layers were overgrown by a new mat community and represented layers 6 to 9 in the present microbial mat.

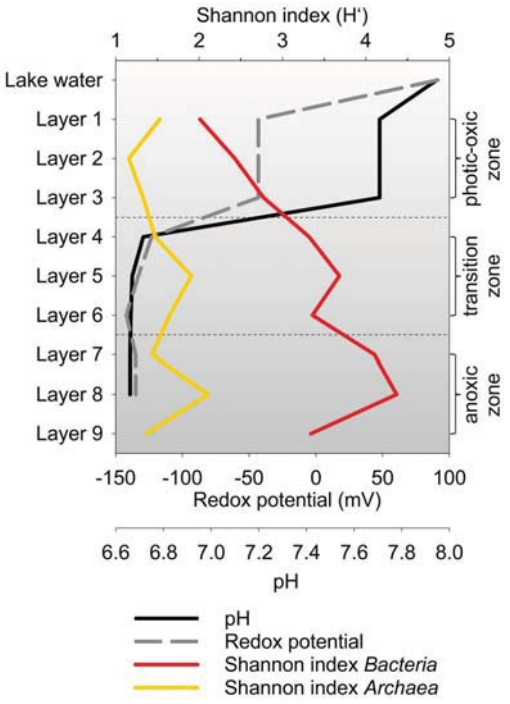


Figure 3. Hydrochemical data of lake and pore water and Shannon indices of sampled microbial mat layers. The redox potential (Eh in mV) and pH (at 32°C) were measured in the field (Table S3). Shannon indices were calculated at 3% genetic divergence and same level of surveying effort (Table 1).
doi:10.1371/journal.pone.0066662.g003

The nine microbial mat layers were separated and analyzed. The chemo-physical properties of the nine mat layers and the arrangement of the detected phylogenetic groups indicated roughly a separation into three major mat zones. These zones were designated photic-oxic zone (layers 1–3), transition zone (dysphotic, low oxygen, layers 4–6), and anoxic zone (layers 7–9) with respect to their main characteristics (Figure 3). The pH measurements during daytime showed slightly alkaline pH value in the lake water (7.95) and the photic-oxic zone (7.7) and slightly acidic pH values in the transition (6.68) and anoxic (6.66) zones. Similar observations were made for the Guerrero Negro mat, but the pH range was broader (pH range of approximately 8.2 to 6.2 from top to bottom) [11]. The redox potential in Lake 21 mat changed from positive in the lake water (+91.4 mV) to negative in the photic-oxic zone (−43 mV). Within the mat layers, a redox potential gradient ranging from −43 mV (photic-oxic zone) to −135 mV (anoxic zone) was recorded. This gradient is typical for microbial mat systems, as the conditions change from oxic to anoxic with increasing depth.

Diversity and General Characteristics of the Bacterial Community

Two studies previously investigated the microbial communities of the Lake 21 mat. One study focused on the analysis of bacteriohopanepolyols (BHP) within the mat, which were classified as cyanobacterial and proteobacterial BHPs [22]. The second study identified microbes by Sanger-based sequencing of clonal 16S rRNA genes [19]. However, due to low phylogenetic resolution of the BHP analysis and small survey size (395 sequences) of the other study, a detailed and comprehensive assessment of the microbial mat community is still lacking. In this study, large-scale next-generation 16S rRNA gene sequencing was used to increase the survey size with respect to the vertical stratification. As indicated by metagenomic and molecular studies of other microbial mats, bacteria are the predominant life forms in these habitats followed by archaea and eukaryotes [10,41–44]. The bacterial 16S rRNA gene analysis resulted in recovery of 80,045 high quality sequences with an average read length of 429 bp. The number of sequences per layer ranged from 6,279 (layer 4) to 12,580 (layer 6). We were able to assign all 16S rRNA gene sequences to the domain *Bacteria* and to classify 78,742 of these sequences below the domain level. The classified sequences were affiliated to 26 bacterial phyla and candidate divisions. Dominant bacterial groups that showed more than 0.5% abundance in any layer belonged to 11 phyla and 4 candidate divisions. In addition, the bacterial community composition showed a well-defined stratification along the depth gradient of the mat.

To determine the bacterial diversity and richness rarefaction analyses were performed. Alpha diversity analysis was performed at the same level of surveying effort. The *Bacteria* encompassed 422 OTUs at 3% genetic divergence level. The OTU number in the layers ranged from 67 (layer 2) to 273 (layer 8). In general, the bacterial diversity increased with depth (Figure 3). This has also been reported from other microbial mat studies such as Guerrero Negro and Solar Lake [9,18]. In the top layers diversity estimates could also be influenced by sedimentation of planktonic bacteria that live in the water column. For example, *Salisaeta longa*, which is the closest related cultured representative of the most abundant OTU in layer 1 (56% of the total bacterial community), was isolated from hypersaline water bodies [45]. Rarefaction curves generally showed saturation for the bacterial communities of all

nine samples at 3% genetic divergence (Figure 4A). The maximal number of expectable OTUs was estimated for each sample by using the Michaelis-Menten fit alpha diversity estimation. The average coverage was 91.9%. Thus, the majority of bacterial phylotypes was covered by the surveying effort. Shannon indices ranged from 2.01 to 4.37 with maxima in layers 5 and 8 (Figure 3). Direct comparison of the bacterial mat layer communities by PCA revealed three clusters (Figure 5). The top cluster the middle cluster, and the bottom cluster consisted of layers 1 to 3, layers 4 to 7, and layers 8 to 9, respectively.

Bacterial Community Composition and Distribution

The bacterial community was dominated by the phyla *Bacteroidetes* (30.5%), *Proteobacteria* (27.3%), *Spirochaetes* (15.3%), *Cyanobacteria* (5.9%), *Chloroflexi* (5.2%), *Nitrospirae* (3.7%), *Planctomycetes* (2.2%), *Firmicutes* (1.9%), CD WS3 (2%), unclassified taxa (1.6%), and *Deferribacteres* (1.4%) (Figure 6). The following phyla and candidate divisions represented each less than 1% of the total bacterial communities *Gemmatimonadetes*, *Fibrobacteries*, *Acidobacteria*, *Actinobacteria*, *Lentisphaerae*, *Deinococcus-Thermus*, *Chlorobi*, *Synergistetes*, and several candidate divisions (OP9, TA06, Hyd24-12, BRC1, NPL-UPA2, BHI80-139, WS6, and JL-ETNP-Z39) (Figure S3).

The photic-oxic zone was represented by the surface layers 1, 2 and 3. In general, the bacterial communities of these layers were highly similar. The zone harbored *Cyanobacteria* as primary producers of organic matter. One predominant phylotype was detected that showed high 16S rRNA gene similarity (98%) to an unicellular, extreme halotolerant species of the *Halothece* cluster (*Euhalothece* sp. strain MPI 95AH13, AJ000710), which was isolated from a benthic gypsum crust in a solar evaporation pond (Eilat, Israel) [46]. The presence of this phylotype has also been observed in lithifying and non-lithifying hypersaline microbial mats, e.g. in the Guerrero Negro hypersaline mat (JN454085) [9,14,16,47,48]. Other detected cyanobacterial lineages in the photic-oxic zone were *Spirulina* but only in minor abundances (<1%, mainly layer 4). *Cyanobacteria* were detected in significant abundances, but exhibited a lower diversity than in other saline microbial mats such as Guerrero Negro or Highborne Cay [9,14]. Studies of extreme halophile *Cyanobacteria* showed that high salinity precludes growth of certain cyanobacterial species such as *Synechococcus* sp. (AJ000716), as most unicellular *Cyanobacteria* prefer growth at <130‰ salinity [46]. Correspondingly, we observed in a microbial mat from the less saline Kiritimati Lake 2 (salinity approximately 120‰) that the cyanobacterial diversity was higher

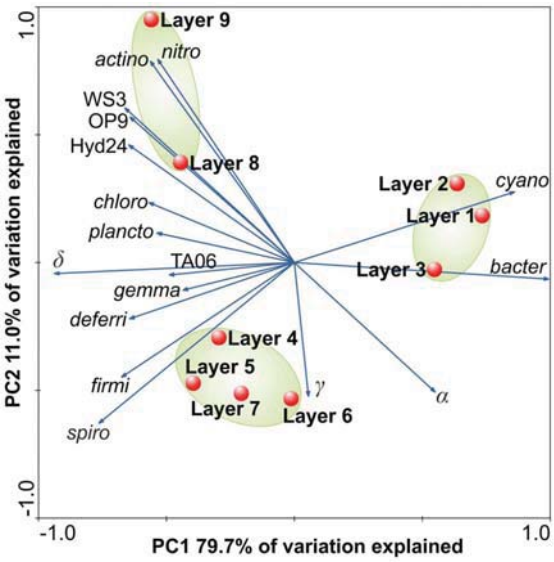


Figure 5. PCA of bacterial communities in the mat layers. PCA was based on relative abundances of bacterial phyla, proteobacterial classes, and candidate divisions. Blue arrows point in direction of increased taxon abundance. Clustering of sampling sites due to similarities in bacterial community composition is highlighted by green ellipses. Abbreviations: *α*, *Alphaproteobacteria*; *β*, *Betaproteobacteria*; *γ*, *Gammaproteobacteria*; *δ*, *Deltaproteobacteria*; *actino*, *Actinobacteria*; *nitro*, *Nitrospirae*; OD1, Candidate division (CD) WS3, CD WS3; Hyd24, CD Hyd24; TA06, CD TA06; *chloro*, *Chloroflexi*; *plancto*, *Planctomycetes*; *gemma*, *Gemmatimonadetes*; *deferri*, *Deferribacteres*; *firmi*, *Firmicutes*; *spiro*, *Spirochaetes*; *bacter*, *Bacteroidetes*; *cyano*, *Cyanobacteria*.
doi:10.1371/journal.pone.0066662.g005

than in the Lake 21 microbial mat (data not shown). The *Euhalothece*-related *Cyanobacteria* were accompanied with known phototrophic or putative phototrophic members of the *Alphaproteobacteria* and *Bacteroidetes*. The *Alphaproteobacteria* of the surface layers were mainly represented by purple non-sulfur bacteria, which were classified as uncultured members of *Rhodobacterales* and *Rhodospirillales*. These were also detected in the surface layers of the Guerrero Negro and Solar Lake mats [9,18]. The 16S rRNA gene sequences of the *Rhodobacterales* showed similarities to uncultured *Rhodovulum*, *Sediminimonas*, and several uncultured *Rhodobacteraceae*. The low similarities (<95%) of these 16S rRNA genes to cultured

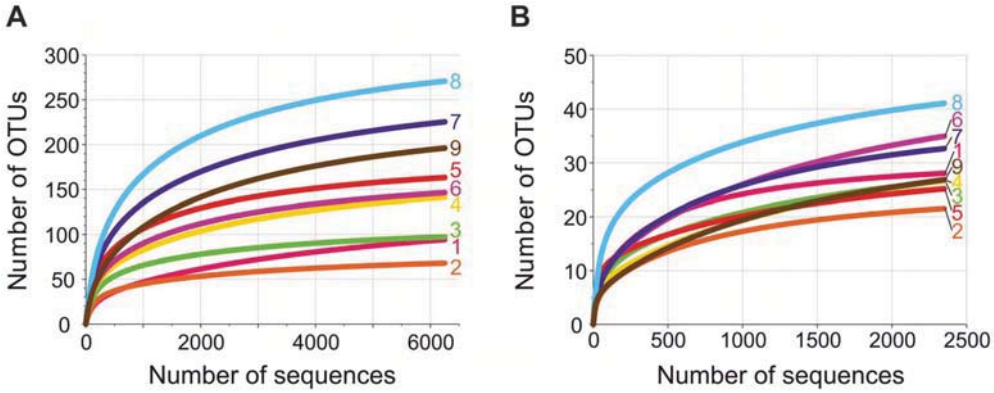


Figure 4. Rarefaction analysis of the bacterial (A) and archaeal (B) communities in the mat layers. Curves were calculated at the same level of surveying effort at 3% genetic divergence employing QIIME [31].
doi:10.1371/journal.pone.0066662.g004

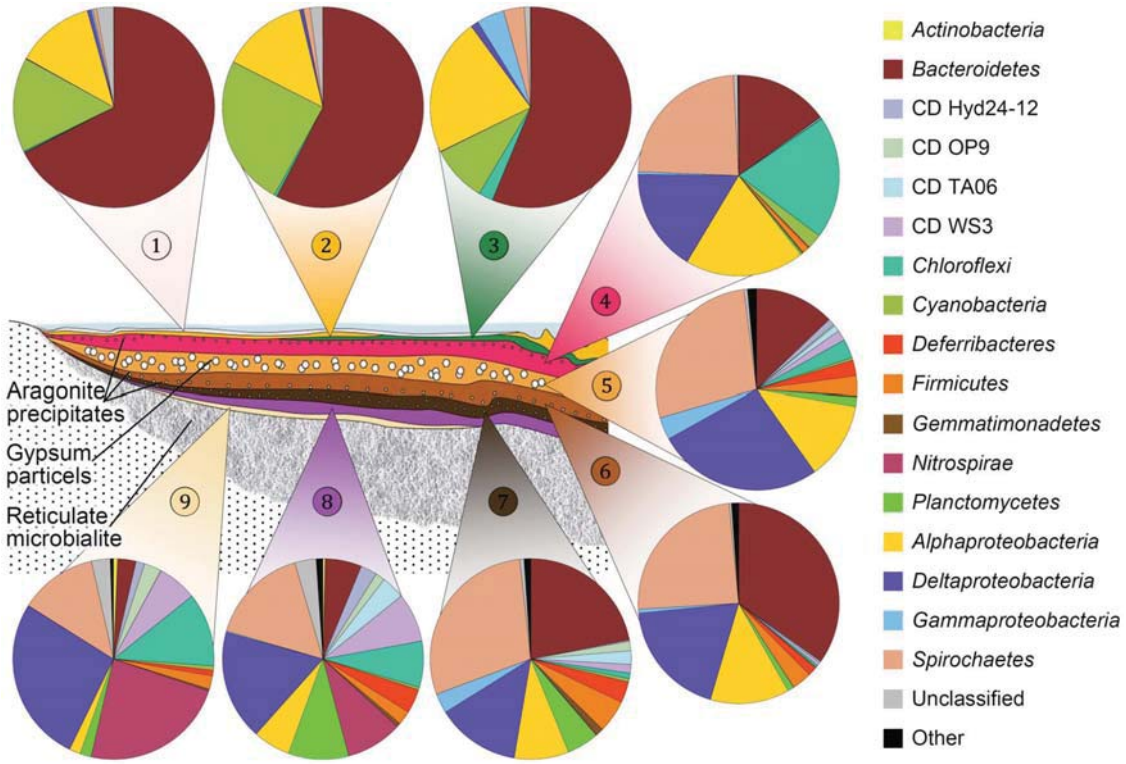


Figure 6. Model of the stratified microbial mat of Lake 21 and relative abundances of bacterial phyla. Illustrated are the bacterial phyla, candidate divisions, and proteobacterial classes found within the different layers. Taxonomic classification was performed according to SILVA SSU rRNA database 111 [39]. Bacterial phyla with a relative abundance lower 0.5% were summarized as the artificial group "Other". The artificial group "Other" includes *Acidobacteria*, CD BHI80-139, CD BRC1, CD JL-ETNP-Z39, CD NPL-UPA2, CD WS6, *Chlorobi*, *Deinococcus-Thermus*, *Fibrobacteres*, *Lentisphaerae*, and *Synergistetes* (for details see Figure S3).
doi:10.1371/journal.pone.0066662.g006

species indicated that several novel *Rhodobacterales* species may thrive in the surface part of the mat. Most *Rhodospirillaceae* prefer anoxic conditions, are photoheterotrophic in light and grow chemotrophically in dark. This suggests that anoxic niches exist in the surface mat matrix (in accordance with the diel cycle) or that the here occurring species differ in their characteristics from known species [49]. *Bacteroidetes*, which predominated the surface layers, consisted of uncultured members of the *Rhodothermaceae* (mainly uncultured *Salisaeta* and *Salinibacter*) and *Saprospiraceae*. *Saprospiraceae* and *Salinibacter* increased in abundance from layer 1 to layer 3, whereas *Salisaeta* decreased. The localization in the upper layers suggests phototrophy for these bacterial groups. *Salinibacter* species have been detected in several hypersaline environments around the world, such as lakes, and coastal or solar salterns [50]. Correspondingly, the genome of the extremely halophilic, photoheterotrophic *Salinibacter ruber* isolated from a microbial mat contains four rhodopsins [51]. Interestingly, three of these were halorhodopsins that typically occur in the domain of *Archaea*. Rhodopsin-containing *Archaea* belonging to the *Halobacteria* have been detected in the surface layers (see below), indicating the possibility of horizontal gene transfer (HGT) events between archaea and bacteria. Recently, considerable numbers of (halo-)viruses have been detected in hypersaline habitats, which could be an additional source for gene transfer events [52].

The transition zone spans approximately layers 4 to 6. In this zone, the most drastic change of the microbial community composition was observed. Simultaneously, the bacterial diversity increased (approximately 1.4-fold) compared to the photic-oxic zone. This alteration was observed on phylum and subgroup level. The community in this zone was dominated by *Proteobacteria*, *Spirochaetes*, *Bacteroidetes*, *Chloroflexi*, *Firmicutes* and *Deferribacteres*. The composition of the *Alphaproteobacteria* and *Bacteroidetes* subgroups changed compared to the photic-oxic zone. The *Alphaproteobacteria* were mainly represented by the *Rhizobiales*. The predominant alphaproteobacterial OTU was an uncultured *Dichotomicrobium*, which was abundant in layers 3 to 8. The 16S rRNA gene sequence of this OTU exhibited high similarity (99%) to *Dichotomicrobium thermohalophilum* (FR733679), which is an aerobic, moderate thermophilic and halophilic heterotroph isolated from the hypersaline Solar Lake (Eilat, Israel). This species builds net-like hyphae and thereby may also be a structural component of the microbial mat fabric [53]. In layer 4, low numbers of 16S rRNA gene sequences affiliated to *Roseospira* (1.5%) and *Rhodovibrio* (1.1%) were observed in addition to uncultered *Rhodospirillales* (2.5%). *Roseospira* and *Rhodovibrio* disappeared in the lower layers. *Bacteroidetes* appeared in layer 4 in low abundance and were represented by *Rhodothermaceae* and *Marinilabiaceae*. The latter group was encountered in layers 4 to 7, with highest abundance in layer 6 (27.6% of the total bacterial community), indicating a preferred anaerobic lifestyle. The dominant OTU of the *Marinilabiaceae* possessed highest sequence similarity (approximately 96%) to a Guerrero Negro mat clone (JN445925). In addition, *Chloroflexi*, *Deltaproteobacteria* and *Spirochaetes* occurred in the transition zone. *Chloroflexi* were a major component of layer 4, and consisted primarily of uncultured members of *Anaerolineaceae*, which exhibited similarities to 16S rRNA gene sequences of *Anaerolineaceae*

derived from the Guerrero Negro mat. Members of this family are strictly anaerobic, filamentous, multicellular, non-pigmented and non-phototrophic [54]. The *Deltaproteobacteria* accumulated in the transition zone and formed a main part of the bacterial community through the deeper mat layers. The majority of the detected representatives of the *Deltaproteobacteria* such as *Desulfovibrionales*, *Desulfobacterales*, and *Desulfarculales* are sulfate-reducing bacteria (SRB). SRB play a major role in carbon cycling and evolved before oxygenic photosynthesis developed [55]. As the reduction of sulfate is an anaerobic process, the distribution of SRB in the transition and anoxic zone corresponds well with the mat properties. However, traces of *Deltaproteobacteria* (0.6 to 1.2%) were observed in the photic-oxic zone. Recently, it has been shown that SRB in microbial mats possess oxygen tolerance [20,55]. Furthermore, it has been demonstrated that some *Desulfovibrio* not only tolerate but also use low oxygen tensions for growth [56,57]. The *Desulfobacterales* were primarily composed of uncultured members of *Desulfobacteraceae* and a small number of *Desulfosalsimonas*, which both have been previously isolated from microbial mats [9,58]. The *Desulfovibrionales* were affiliated to uncultured *Desulfovibrio* and *Desulfovermiculus* (>98% 16S rRNA gene sequence similarity to *Desulfohalobium utahense*). *Desulfohalobium utahense* was isolated from the Great Salt Lake (Utah, USA) and is characterized as an anaerobic chemoorganotrophic SRB [59]. *Spirochaetes* also formed a major part of the center and bottom mat communities and were highly diverse (45 OTUs). All 16S rRNA gene sequences were affiliated to uncultured *Spirochaeta*, which were also detected in the Guerrero Negro mats [9]. Their cork screw-like type of movement might be advantageous in EPS-rich mat habitats, as flagella mediated movement might be less efficient. An interesting result was the unique occurrence of purple sulfur bacteria (PSB) affiliated to the *Chromatiales* in layer 5, which was characterized by abundant spherical gypsum crystals (Figure 2C). Members of the PSB, *Chromatiaceae* (2.3%) and *Ectothiorhodospiraceae* (1%), were detected in low abundances within layer 5. PSB of these families are phototrophic and produce internal (*Chromatiaceae*) and external (*Ectothiorhodospiraceae*) elemental sulfur globules, respectively [60]. The preferential occurrence in one layer in addition with the metabolic features of the PSB might point to an involvement in gypsum precipitation. The relative low abundance of this group could be caused by low recovery of DNA from these organisms due to adherence of cells to gypsum crystal fissures.

Finally, the anoxic light-free zone (layers 7 to 9) harbored mainly *Spirochaetes*, *Deltaproteobacteria*, *Gammaproteobacteria*, *Planctomycetes*, *Nitrospirae*, *Firmicutes*, *Deferribacteres*, and different candidate divisions (WS3, TA06, OP9, Hyd24-12). These phyla showed specific patterns within their composition along the depth gradient. *Deltaproteobacteria* composition changed from *Desulfobacterales* and *Desulfovibrionales* in the transition zone to *Desulfobacterales* (more diverse, including *Nitrospinacaea* and candidate group SEEP-SRB1) and *Desulfarculales* in the anoxic zone. The *Desulfarculales* were primarily represented by an OTU, which was detected in an earlier molecular study of the same lake (HM480235) [19]. The genome of the sulfate-reducing *Desulfarculus barsii* indicates that members of this family are strictly anaerobic, non-fermentative and chemoorganotrophic, and oxidize organic substrates completely to CO_2 [61]. The *Chloroflexi* appeared in low abundances (0.9% to 3.1%) in layers 5 to 7, but increased in abundance in layers 8 and 9 (7.4% and 11.9%, respectively). In detail, low amounts of *Anaerolineales* were present, but mainly *Caldilineales* and several candidate subgroups were detected. These subgroups included MSBL5 (Mediterranean Sea Brine Lake group 5) and GIF9 (Groundwater In Flow clone 9), which were originally found in deep anoxic hypersaline Bannock basin and a contaminated groundwater-treated reactor, respectively [62,63]. In addition, uncultured members of the *Nitrospirae*, in particular the candidate order OPB95 (formerly candidate division OP8), increased in abundance in layers 8 and 9. Members of this group were first encountered in Obsidian Pool (Yellowstone, USA) and could be involved in hydrogen oxidation and sulfate reduction [64]. Interestingly, the most abundant *Nitrospirae* OTU showed high 16S rRNA gene sequence similarity (99%) to a 2002 amplified 16S rRNA gene of the same mat (HM480241). Another group inhabiting the last three layers with highest abundance in layer 8 was *Planctomycetes*, composed of candidate Pla4 lineage and candidate group vadinHA49. The dominant OTU of the Pla4 lineage showed low similarity (<92%) to a Guerrero Negro mat clone (JN510021), which was also found in deeper layers (10–22 mm). The clone PLA4 (AF525959), which is one of the reference sequences for the Pla4 lineage, was found in a biological contactor fed with NH_4^+ wastewater [65], suggesting that the related microbes in Lake 21 could also be involved in (anaerobic) nitrogen cycling. *Firmicutes* were encountered in low abundances in the transition to anoxic zones of the mat and showed highest abundance in layers 7 (5.3%). Nevertheless, with 38 OTUs the *Firmicutes* were a diverse group that was primarily comprised of *Clostridia* belonging to the *Clostridiales* (mainly *Ruminococcaceae*) and *Halanaerobiales* (*Haloanaerobiaceae*). *Halanaerobiales* are anaerobic, halophilic and fermenting microorganisms [66]. Their lifestyle corresponds well with their location in the deeper anoxic parts of the mat. Members of this order have also been detected in deeper zones of hypersaline endoevaporitic microbial mats from the Solar Lake (Eilat, Israel) [18]. *Deferribacteres* showed a similar distribution as the *Firmicutes*. The middle layers were mainly inhabited by members of the candidate group LCP-89, which was also detected in saltmarsh sediment contaminated with mercury and PCB. In the lower layers, also uncultured *Caldithrix* relatives were present. These are known as anaerobic and chemoorganotrophic organisms, which are able to perform nitrate reduction and fermentation of di- and polysaccharides [67,68]. However, the 16S rRNA gene sequence similarity to cultured representatives like *Caldithrix abyssi* was rather low (85%). Several members of the candidate phyla WS3, TA06, OP9, Hyd24-12 were encountered in the deeper layers. Especially WS3 and TA06 were abundant in the last three layers. The increase of candidate phyla in the deeper anoxic layers probably originates from tight metabolic dependencies on spatial surrounding species.

Archaeal Diversity and Stratification in Microbial Mat Communities

Recent studies indicate that *Archaea* account only for a small part of microbial mat communities. Nevertheless, they may have important roles in biogeochemical cycles. The archaeal community analysis was carried out for the nine layers of the Lake 21 mat and resulted in 39,729 high-quality 16S rRNA gene sequences with an average read length of 513 bp. The number of sequences per layer ranged from 2,353 (layer 9) to 10,663 (layer 5). All 16S rRNA gene sequences could be assigned to the domain *Archaea* (Table 1). The trend of increasing diversity downwards the microbial mat persisted, but not as pronounced as for *Bacteria*. Interestingly, Shannon indices for archaeal diversity showed a similar depth profile to that of *Bacteria*, but overall diversity is approximately 2.1-fold lower (Figure 3). In addition, the highest degree of archaeal diversity with respect to Shannon index and OTU count was also detected in layers 5 and 8. A total of 86 different archaeal OTUs at 3% genetic divergence could be observed. The amount of OTUs in the different layers ranged

from 22 (layer 2) to 41 (layer 8). Rarefaction curves generally showed saturation for all samples (Figure 4B). A similar trend was also observed for *Bacteria* (Figure 4A). The coverage of the archaeal communities was 89.8% on average. Thus, the majority of *Archaea* was detected by the surveying effort. Comparison by PCA of the archaeal community composition of all layers showed separation of the surface layers (layers 1 to 3) and the bottom layers (layers 4 to 9), which resembles the change from aerobic to anaerobic archaeal community members (Figure 7B).

The archaeal phyla *Euryarchaeota* and *Thaumarchaeota* were detected and represented 67.6 and 20.8% of all archaeal 16S rRNA gene sequences, respectively (Figure 7A). Additionally, due to low alignment coverage (<95%) several OTUs could not be classified below the domain level. Two of these were high abundant and designated as unclassified archaeal groups (UAG) 1 and 2. These groups occured in layer 5 (UAG1) and layer 9 (UAG2). UAG1 and UAG2 showed highest identities (85 and 82%, respectively) to 16S rRNA gene sequences from uncultured *Euryarchaeota* (DHVEG-6) (AACY020254990) and uncultured *Thaumarchaeota* (MBG-B) (AB644695), respectively. In general, *Euryarchaeota* dominated the layers 1 to 6, whereas *Thaumarchaeota* were generally present from layers 4 to 9 and exhibited highest abundances in layers 7 and 8 (Figure 7B). The archaeal community of the photic-oxic zone almost solely consisted of *Halobacteria*, but the sublevel composition differed between the layers. Layer 1 was dominated by *Halobacteriaceae*, which were composed of OTUs exhibiting high similarities to *Halomarina*, *Halorussus*, and *Halorubrum*. The latter phylogenetic groups represented 69, 15, and 8% of all archaeal sequence in the layer, respectively. These genera comprise aerobic, halophilic and often phototrophic *Archaea*, which is in accordance with their localization [69]. The halobacterial genera detected in layer 1 disappeared in layers 2 and 3, which were dominated by uncultured members of the Deep Sea Hydrothermal Vent Group 6 (DHVEG-6). Sedimentation of these *Archaea* from lake water to the surface layer is possible and might explain the shift of the halobacterial subgroups. In the darker and more oxygen-depleted transition zone, *Halobacteria* strongly decrease in abundance and *Thermoplasmata* proliferate. *Archaea* of this group were affiliated to uncultured members of Marine Benthic Group D (MBG-D), DHVEG-1 and CCA47, which were all frequently found in marine deep anoxic environments such as sediments and hydrothermal vents [70–73]. A recent study showed that members of MBG-D are widely distributed in marine environments, preferentially in deeper anoxic sediments. At least some members of this group are involved in detrital protein degradation [74]. This is in accordance to their localization in the anoxic mat parts (layers 4 to 9) and might explain the in 2002 observed enhanced ammonia concentration of the mat pore water [19]. Therefore, this highly abundant group of the archaeal community might be involved in the recycling of detrital proteins from the EPS matrix. Uncultured members of the *Thaumarchaeota* increased in abundance from layers 6 to 9. In layers 7 and 8 *Thaumarchaeota* were the predominant archaeal group. Most of the corresponding 16S rRNA gene sequences were classified as members of MBG-B, which have been found in anoxic marine habitats such as Black Sea microbial mats and in the Guerrero Negro mat [10,75]. Unfortunately, no cultivated representatives of this group exist to date, but it has been proposed that these organisms are sulfate-reducers [76]. Interestingly, members of *Methanomicrobia* and Group C3 (*Thaumarchaeota*) were only detected in significant abundances in layer 8. The class *Methanomicrobia* is capable of methanogenesis and therefore requires the strictly anoxic milieu of the deepest parts of the mat. This class was composed of two OTUs showing similarities to *Methanohalophilus* and *Methermicoccus*. The absence of methanogens in layer 9 might be due to

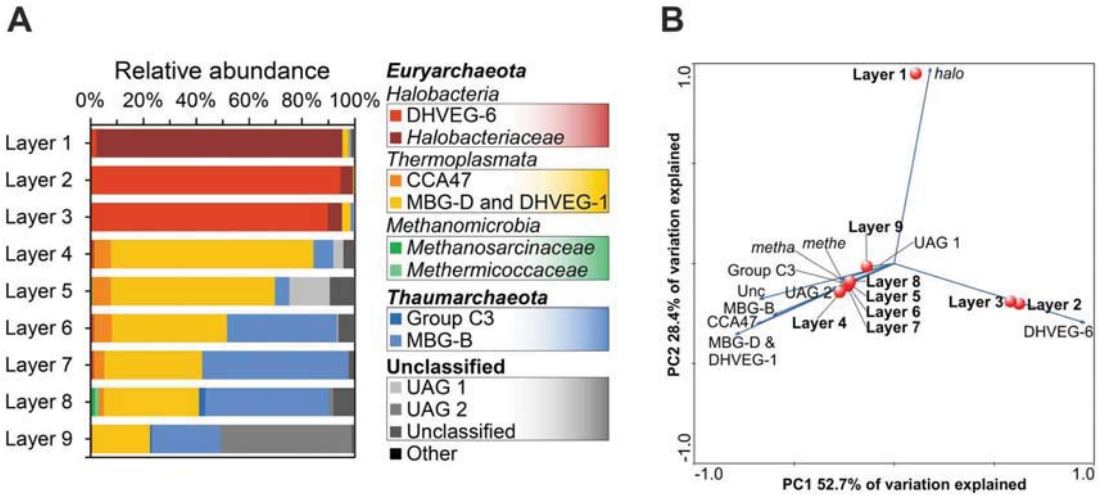


Figure 7. Relative abundances and comparison of the archaeal communities of the Lake 21 microbial mat. Relative abundances of archaeal lineages identified within the different microbial mat layers (A). Taxonomic classification was performed according to SILVA SSU rRNA database 111 [39]. Taxonomic lineages with a relative abundance lower 0.5% were summarized as the artificial group "Other". Abbreviations: DHVEG-6, Deep Sea Hydrothermal Vent Group 6; CCA47, deep anoxic layer of an intertidal pool at Cape Cod; MBG-D and DHVEG-1, Marine Benthic Group D and Deep Sea Hydrothermal Vent Group 1; Group C3, Crenarcheotic Group C3 from marine lakes and sediments; MBG-B, Marine Benthic Group B; Other: Miscellaneous Euryarchaeotic Group; 20c-4, marine sediment (Greece, Aegean Sea); TMEG, Terrestrial Miscellaneous Group; MCG, Miscellaneous Crenarchaeotic Group (for details see Figure S4). Comparison of archaeal communities of microbial mat layers by PCA (B). PCA was based on relative abundances of archaeal lineages. Blue arrows point in direction of increased taxon abundance. Abbreviations: *halo*, *Halobacteriaceae*; DHVEG-6, see (A), MBG-D & DHVEG-1, Marine Benthic Group D and Deep Sea Hydrothermal Vent Group 1; CCA47, see (A); MBG-B, see (A); Group C3, see (A); Unc, unclassified archaeal groups; UAG 1, Unclassified archaeal group 1; UAG 2; unclassified archaeal group 2; *metha*, *Methanosarcinaceae*; *methe*, *Methermicoccaceae*.
doi:10.1371/journal.pone.0066662.g007

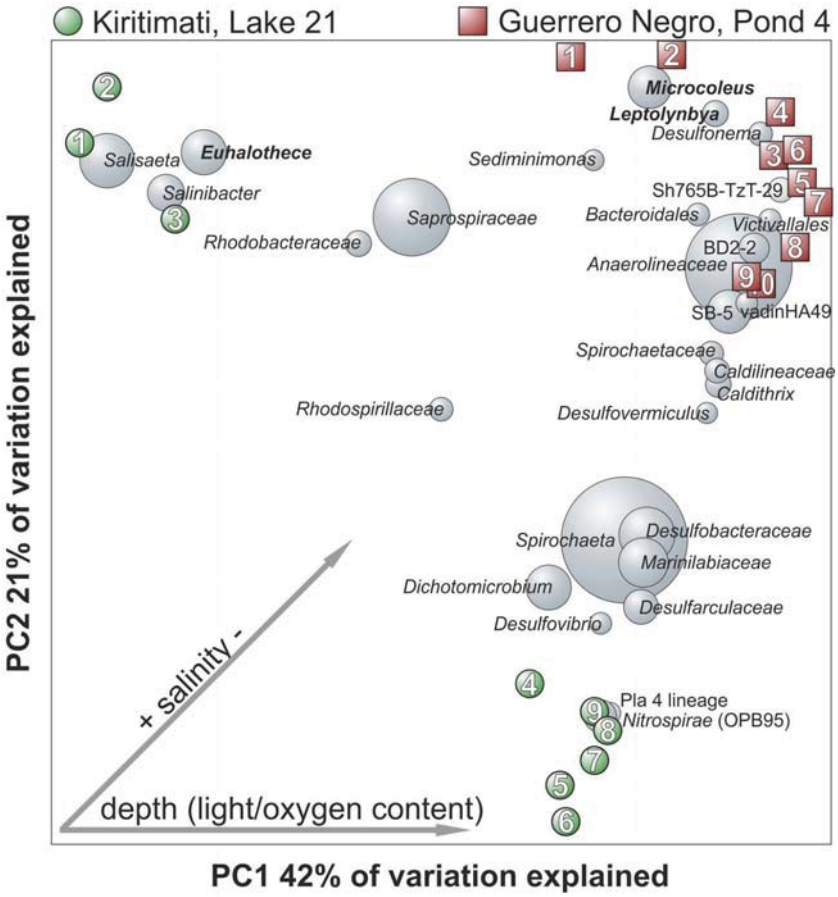


Figure 8. Comparison of Kiritimati and Guerrero Negro microbial mat layers by weighted principal coordinate analysis (PCoA). Numbers represent the layers (samples). The most abundant taxa (weighted average of all samples) are drawn in grey spheres. Sizes of the spheres are proportional to the mean relative abundance of the OTU across all samples. Potential environmental influences on community composition such as salinity, light and oxygen are illustrated by grey arrows.
doi:10.1371/journal.pone.0066662.g008

groundwater-mediated influx of low oxygen levels. The 16S rRNA gene sequences affiliated to Group C3, which is a group of uncultured *Archaea* often found in low-temperature terrestrial and marine habitats [77], exhibited low similarities (<96%) to known 16S rRNA gene sequences. The trend of the archaeal group distribution along the depth gradient of the microbial mat is very similar to the reported distribution of two studies from Guerrero Negro [10,13], pointing to a general occurrence of these archaeal groups in microbial mats. However, our current knowledge of *Archaea* in these habitats is limited, as cultured representatives are often lacking.

Development of the Lake 21 Microbial Mat over the last Decade and Comparison with the Guerrero Negro Mat Community

The analysis of prokaryotic communities in this study revealed that almost all of the bacterial lineages, found in the previous 2002 mat sample of Lake 21 [19], were detected and are in accordance with the BHP analysis of the 2002 sample [22]. The latter study showed that cyanobacterial BHPs were located at the top part and proteobacterial BHPs in the lower part of the 2002 sampled mat [19,22]. In addition, we found in 2011 representatives of different candidate phyla (BRC1, OP9, WS3, WS6, Hyd24-12, TA06), *Deferribacteres*, *Gemmatimonadetes*, and several other phyla and candidate divisions in low abundances (Figure 6 and 7; Figure S3 and S4). The additional bacterial groups detected in the 2011 sample may not have been detected in the prior study due to a smaller survey size. In addition, mat communities might have changed from 2002 to 2011 by overgrowth of the former 2002 mat community. A new mat community establishment is indicated, as the 2002 mat was thinner (1 cm at shoreline [19]) than that of 2011 (approximately 10 cm). In addition, 16S rRNA gene sequences from 2002 matched only sequences derived from deeper layers of the 2011 sampled mat. Correspondingly, the general lamination fabric of the mat changed, as more mat layers were observed in 2011 than in 2002. Microscopic observations showed that in 2002 frequently detected *Cyanobacteria* (*Microcoleus* and *Johannesbaptistia*) were in the present study only found in the lower layers of the mat or even trapped in gypsum crystals (Figure 2D–G). Thus, the bottom layers of the 2011 sampled mat resemble the 2002 mat layers. The observed mat structure in the present study might result from a collapse of ''old'' mat layers by microbial degradation of the EPS fabric and development of a new surface mat community.

Based on similarity searches, the overall prokaryotic community of the 2011 investigated Lake 21 mat showed more similarities to the Guerrero Negro microbial mat [9] than to the 2002 sampled Lake 21 mat [19]. Approximately 61% of the detected bacterial

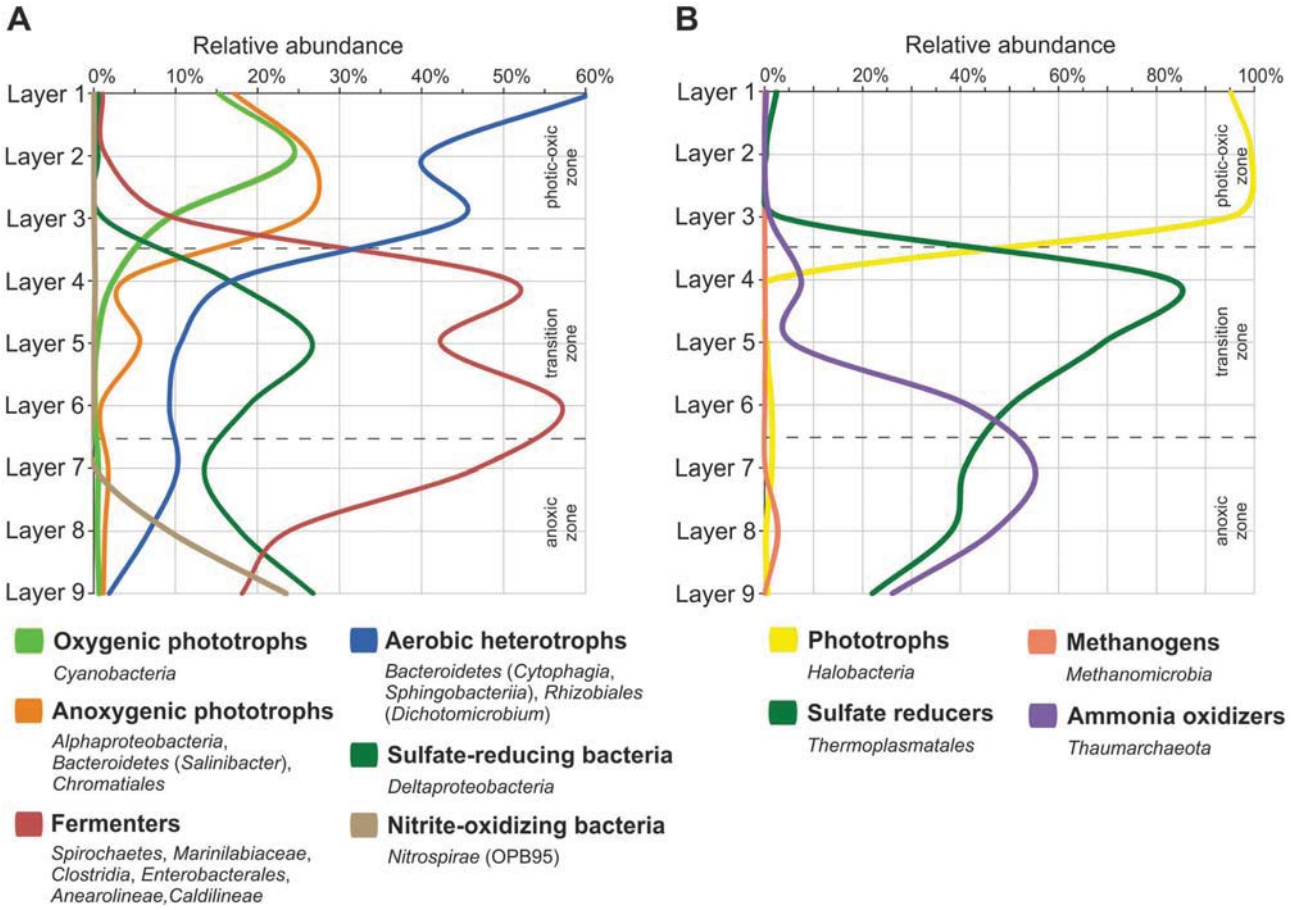


Figure 9. Spatial distribution and relative abundance of bacterial (A) and archaeal (B) functional groups.
doi:10.1371/journal.pone.0066662.g009

OTUs exhibited highest similarities to database entries derived from the Guerrero Negro microbial mat [9]. Generally, several significant similarities of phyla distribution and composition along the depth gradient were encountered, such as the almost ubiquitously distributed SRB and the high diversity of *Spirochaetes*. Comparison of the bacterial mat communities by PCoA revealed also several differences with respect to relative abundances (Figure 8). One example is the different cyanobacterial community in both microbial mat surfaces. *Euhalothece* dominated in Lake 21 mat, whereas *Microcoleus* and *Leptolyngbya* were most abundant in the Guerrero Negro mat. Another example is the abundance of *Nitrospirae* (OPB95), which occurred in high abundance in the anoxic zone (layer 9) of Lake 21, but were almost absent in the Guerrero Negro mat. Reasons for differences in community structures of Lake 21 and Guerrero Negro mat as well as the higher diversity of the Guerrero Negro bacterial community might be related to different salinities (170‰ Lake 21 and 80‰ Guerrero Negro). It is indicated that the filtering effect of higher salinities also affects the remaining bacterial and archaeal community structures towards presence of salt-resistant organisms [46].

Microbialite Formation in Lake 21 Microbial Mat

Microbialite formation is an active process in the investigated Lake 21 mat (Figure 2C). The mat is characterized by different types of mineral precipitates: the surface is covered with gypsum crystals and to a lower amount with aragonite crystals, followed by small aragonite spherulites in layers 4, 6, and 7, large gypsum crystals in layer 5, and aragonitic, reticulate microbialites at the bottom of the mat, which already begin to form in the last two layers.

In general, control of microbialite formation is given at three different levels: (i) the macroenvironmental mineral saturation states, (ii) the metabolic activity of microbes affecting ion activities within the mats, and (iii) extracellular polymeric substances affecting mineral nucleation within the mats (for review see, e.g. [6,19,78–82]).

In Kiritimati Lake 21, $CaCO_3$ mineral supersaturation is high, and the photosynthetic *Cyanobacteria* in the top layers further increase the $Ca^{2+} \times CO_3^{2-}$ ion activity product significantly within the mat. However, microbialite formation largely takes place in deeper mat parts, while the photosynthetic top layers of the mats show only minor precipitates such as scattered aragonite spherulites [19]. This observation has been explained by a strong inhibition of precipitation in the top mat layers, and successive heterotrophic degradation of inhibiting exopolymers with depth [19].

The present data now substantiate the abundance of fermenting bacteria such as *Spirochaetes*, *Marinilabiaceae*, *Clostridia*, *Anaerolineae* and *Caldilineae*, which potentially degrade exopolymers in the transitional zone and anoxic zone. Besides of breakdown of nucleation inhibition, a secondary release of Ca^{2+} from the degraded exopolymers may further contribute to microbialite

formation [79,81]. Sulfate-reducing bacteria were almost absent in oxic top layers (Figure 9), contrary to reports from other microbialite forming settings (e.g., [55]). Instead, the distribution pattern of SRB rather matches the traditional view [83], i.e. high abundances in the transition and anoxic zones. While the metabolic effect of SRB on the $CaCO_3$ mineral saturation is arguable, because pH and $CaCO_3$ saturation states decrease in these zones [19], some of them may degrade exopolymers too [55], thereby contributing to the breakdown of nucleation inhibition. In turn, methanogens (*Methanomicrobia*) were only detected in low abundances and restricted to mat layer 8 - contrary to previous assumptions [19].

A complete decomposition of the mat exopolymer matrix is finally achieved at the mat bottom, where *Nitrospirae* and *Thaumarchaeota* are strikingly abundant (Figure 9). While *Nitrospirae* comprise nitrite-oxidizing as well as sulfate-reducing members, *Thaumarchaeota* are largely considered as ammonia oxidizers [84]. Indeed, high ammonium concentrations in the anoxic pore water of the microbialites [19] apparently foster the growth of these microbial groups at the mat bottom.

Conclusions

Microbial mats act as self-controlling ecosystems by (re-)cycling the essential elements of life. The microbial mats from the Kiritimati lakes are excellent study objects to analyze the high complexity of microbial mat communities due to well-laminated mat layers, which allow high resolution spatial profiling. In addition, mineral precipitation under extreme environmental conditions such as high UV radiation and salinity can be studied. In accordance with the high oxygen level and light intensity, the prokaryotic communities in the surface layers of the Lake 21 mat are mainly composed of halophile oxygenic and anoxygenic phototrophs, and aerobic heterotrophs. In the transition zone, SRB, fermenting bacteria, and potential sulfate-reducing *Archaea* proliferate, while the abundance of phototrophic organisms declines. In the anoxic zone also SRB (with a different sublevel composition) and fermenters build the major part of the bacterial community. An involvement of both in microbialite formation is likely. Ammonia-oxidizing *Archaea* and low numbers of methanogens were also present in the deepest mat layers (Figure 9). In general, heterotrophic prokaryotes dominate the whole mat, as they nourish on organic matter supplied by phototrophic organisms from the photic-oxic zone. In addition, the diversity of bacterial and archaeal communities increased with depth. This result can be generalized for hypersaline microbial mats, as other DNA-based studies of microbial mats encountered a similar trend. The phylogenetic analyses of the mat also illustrates the gap of knowledge with respect to microbes inhabiting these ecosystems, as most of the detected 16S rRNA gene sequences showed low or no significant similarities to known and characterized microbes. Approximately 27% of the bacterial and 61% of the archaeal OTUs showed 16S rRNA gene sequence identities of lower than 95% including OTUs with low coverage values (<95%) to database entries, indicating at least the presence of novel species. In addition, most of these OTUs exhibited highest similarities to 16S rRNA gene sequences from uncultured and uncharacterized organisms. Thus, assessment of ecosystem functions of these microbes is difficult, as cultivated representatives are absent. To shed light on the key metabolic traits of the unknown taxa functional profiling of the microbial mat layers by direct DNA sequencing complemented with isolation and analysis of individual organisms will be performed in further studies. This would also allow linking phylogenetic composition and functional repertoire as well as unraveling lithification processes in these unique and robust ecosystems.

Supporting Information

Figure S1 The daily number of hourly observed precipitation reports during 2010 and 2011.
(PDF)

Figure S2 Primer coverage of bacterial and archaeal phyla.
(PDF)

Figure S3 Relative abundances of rare bacterial phylogenetic groups.
(PDF)

Figure S4 Relative abundances of rare archaeal phylogenetic groups.
(PDF)

Table S1 Hydrochemical data of sea water and Lake 21 waters, Kiritimati atoll.
(PDF)

Table S2 Data processing of 16S rRNA gene sequences.
(PDF)

Table S3 Hydrochemical data of Lake 21 water and pore water of microbial mat layers.
(PDF)

Acknowledgments

We would like to thank Nicole Brinkmann for the organization of the expedition to Kiritimati. We would also like to thank Gerhard Hundertmark for the photograph of the microbial mat cross section and Wolfgang Dröse for the preparation of hardpart microtome sections. We thank Bernd Wemheuer and Sascha Dietrich for maintaining and expanding our in-house scripts for 16S rRNA gene analysis.

Author Contributions

Conceived and designed the experiments: DS GA JR RD. Performed the experiments: DS GA AR. Analyzed the data: DS GA AR RD. Wrote the paper: DS GA RD.

References

1. Van Kranendonk MJ, Philippot P, Lepot K, Bodorkos S, Pirajno F (2008) Geological setting of Earth's oldest fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. Precambrian Res 167: 93–124.
2. Allwood AC, Walter MR, Kamber BS, Marshall CP, Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. Nature 441: 714–718.
3. Bontognali TR, Sessions AL, Allwood AC, Fischer WW, Grotzinger JP, et al. (2012) Sulfur isotopes of organic matter preserved in 3.45-billion-year-old stromatolites reveal microbial metabolism. Proc Natl Acad Sci U S A 109: 15146–15151.
4. Hoehler TM, Bebout BM, Des Marais DJ (2001) The role of microbial mats in the production of reduced gases on the early Earth. Nature 412: 324–327.
5. Tice MM, Lowe DR (2004) Photosynthetic microbial mats in the 3,416-Myr-old ocean. Nature 431: 549–552.
6. Dupraz C, Reid RP, Braissant O, Decho AW, Norman RS, et al. (2009) Processes of carbonate precipitation in modern microbial mats. Earth-Science Reviews 96: 141–162.
7. Van Gemerden H (1993) Microbial mats: A joint venture. Marine Geology 113: 3–25.
8. Seckbach J, Oren A (2010) Microbial mats: modern and ancient microorganisms in stratified systems; Seckbach J, OrenA, Dordrecht; New York: Springer. 606.
9. Kirk Harris J, Gregory Caporaso J, Walker JJ, Spear JR, Gold NJ, et al. (2013) Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. Isme J 7: 50–60.

10. Robertson CE, Spear JR, Harris JK, Pace NR (2009) Diversity and stratification of archaea in a hypersaline microbial mat. Appl Environ Microbiol 75: 1801–1810.
11. Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, et al. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl Environ Microbiol 72: 3685–3695.
12. Feazel LM, Spear JR, Berger AB, Harris JK, Frank DN, et al. (2008) Eucaryotic diversity in a hypersaline microbial mat. Appl Environ Microbiol 74: 329–332.
13. Jahnke LL, Orphan VJ, Embaye T, Turk KA, Kubo MD, et al. (2008) Lipid biomarker and phylogenetic analyses to reveal archaeal biodiversity and distribution in hypersaline microbial mat and underlying sediment. Geobiology 6: 394–410.
14. Mobberley JM, Ortega MC, Foster JS (2012) Comparative microbial diversity analyses of modern marine thrombolitic mats by barcoded pyrosequencing. Environ Microbiol 14: 82–100.
15. Myshrall KL, Mobberley JM, Green SJ, Visscher PT, Havemann SA, et al. (2010) Biogeochemical cycling and microbial diversity in the thrombolitic microbialites of Highborne Cay, Bahamas. Geobiology 8: 337–354.
16. Allen MA, Goh F, Burns BP, Neilan BA (2009) Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. Geobiology 7: 82–96.
17. Goh F, Allen MA, Leuko S, Kawaguchi T, Decho AW, et al. (2009) Determining the specific microbial populations and their spatial distribution within the stromatolite ecosystem of Shark Bay. Isme J 3: 383–396.
18. Sørensen KB, Canfield DE, Teske AP, Oren A (2005) Community composition of a hypersaline endoevaporitic microbial mat. Appl Environ Microbiol 71: 7352–7365.
19. Arp G, Helms G, Karlinska K, Schumann G, Reimer A, et al. (2012) Photosynthesis versus Exopolymer Degradation in the Formation of Microbialites on the Atoll of Kiritimati, Republic of Kiribati, Central Pacific. Geomicrobiology Journal 29: 29–65.
20. Bühring SI, Smittenberg RH, Sachse D, Lipp JS, Golubic S, et al. (2009) A hypersaline microbial mat from the Pacific Atoll Kiritimati: insights into composition and carbon fixation using biomarker analyses and a 13C-labeling approach. Geobiology 7: 308–323.
21. Trichet J, Défarge C, Tribble J, Tribble GW, Sansone FJ (2001) Christmas Island lagoonal lakes, models for the deposition of carbonate-evaporite-organic laminated sediments. Sedimentary Geology 140: 177–189.
22. Blumenberg M, Arp G, Reitner J, Schneider D, Daniel R, et al. (2013) Bacteriohopanepolyols in a stratified cyanobacterial mat from Kiritimati (Christmas Island, Kiribati). Organic Geochemistry 55: 55–62.
23. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. Nat Rev Microbiol 6: 245–252.
24. Woodroffe CD, McLean RF (1998) Pleistocene morphology and Holocene emergence of Christmas (Kiritimati) Island, Pacific Ocean. Coral Reefs 17: 235–248.
25. Saenger C, Miller M, Smittenberg RH, Sachs JP (2006) A physico-chemical survey of inland lakes and saline ponds: Christmas Island (Kiritimati) and Washington (Teraina) Islands, Republic of Kiribati. Saline Systems 2: 8.
26. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res 35: e120.
27. Wang Y, Qian PY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS One 4: e7401.
28. Gantner S, Andersson AF, Alonso-Saez L, Bertilsson S (2011) Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. J Microbiol Methods 84: 12–18.
29. Teske A, Sorensen KB (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? Isme J 2: 3–18.
30. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, et al. (2011) PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27: 1159–1161.
31. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7: 335–336.
32. Reeder J, Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods 7: 668–669.
33. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17: 10–12.
34. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194–2200.
35. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–5072.
36. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461.
37. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71: 1501–1506.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
39. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35: 7188–7196.
40. Parkhurst DL, Appelo CAJ (1999) User's guide to PHREEQC, Version 2: a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. Denver, Colo.: U.S. Geological Survey: Earth Science Information Center, Open-File Reports Section. 312.
41. Breitbart M, Hoare A, Nitti A, Siefert J, Haynes M, et al. (2009) Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Cienegas, Mexico. Environ Microbiol 11: 16–34.
42. Kunin V, Raes J, Harris JK, Spear JR, Walker JJ, et al. (2008) Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. Mol Syst Biol 4: 198.
43. Papineau D, Walker JJ, Mojzsis SJ, Pace NR (2005) Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. Appl Environ Microbiol 71: 4822–4832.
44. Spear JR, Ley RE, Berger AB, Pace NR (2003) Complexity in natural microbial ecosystems: the Guerrero Negro experience. Biol Bull 204: 168–173.
45. Vaisman N, Oren A (2009) *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the *Bacteroidetes*. Int J Syst Evol Microbiol 59: 2571–2574.
46. Garcia-Pichel F, Nubel U, Muyzer G (1998) The phylogeny of unicellular, extremely halotolerant cyanobacteria. Arch Microbiol 169: 469–482.
47. Bebout BM, Carpenter SP, Des Marais DJ, Discipulo M, Embaye T, et al. (2002) Long-term manipulations of intact microbial mat communities in a greenhouse collaboratory: simulating earth's present and past field environments. Astrobiology 2: 383–402.
48. Sahl JW, Pace NR, Spear JR (2008) Comparative molecular analysis of endoevaporitic microbial communities. Appl Environ Microbiol 74: 6444–6446.
49. Garrity GM, Brenner DJ, Krieg NR, Staley JT (2005) Part C The *Alpha-*, *Beta-*, *Delta-*, and *Epsilonproteobacteria*; GarrityGM, New York: Springer. 1388.
50. Anton J, Pena A, Santos F, Martinez-Garcia M, Schmitt-Kopplin P, et al. (2008) Distribution, abundance and diversity of the extremely halophilic bacterium *Salinibacter ruber*. Saline Systems 4: 15.
51. Mongodin EF, Nelson KE, Daugherty S, Deboy RT, Wister J, et al. (2005) The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc Natl Acad Sci U S A 102: 18147–18152.
52. Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM (2012) Global network of specific virus-host interactions in hypersaline environments. Environ Microbiol 14: 426–440.
53. Hirsch P, Hoffmann B (1989) *Dichotomicrobium thermohalophilum*, gen. nov., spec, nov., Budding Prosthecate Bacteria from the Solar Lake (Sinai) and Some Related Strains. Systematic and Applied Microbiology 11: 291–301.
54. Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, et al. (2006) *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae* classis nov. and *Caldilineae* classis nov. in the bacterial phylum *Chloroflexi*. Int J Syst Evol Microbiol 56: 1331–1340.
55. Baumgartner LK, Reid RP, Dupraz C, Decho AW, Buckley DH, et al. (2006) Sulfate reducing bacteria in microbial mats: Changing paradigms, new discoveries. Sedimentary Geology 185: 131–145.
56. Krekeler D, Cypionka H (1995) The preferred electron acceptor of *Desulfovibrio desulfuricans* CSN. FEMS Microbiol Ecol 17: 271–277.
57. Krekeler D, Sigalevich P, Teske A, Cypionka H, Cohen Y (1997) A sulfate-reducing bacterium from the oxic layer of a microbial mat from Solar Lake (Sinai), *Desulfovibrio oxyclinae* sp. nov. Arch Microbiol 167: 369–375.
58. Isenbarger TA, Finney M, Rios-Velazquez C, Handelsman J, Ruvkun G (2008) Miniprimer PCR, a new lens for viewing the microbial world. Appl Environ Microbiol 74: 840–849.
59. Jakobsen TF, Kjeldsen KU, Ingvorsen K (2006) *Desulfohalobium utahense* sp. nov., a moderately halophilic, sulfate-reducing bacterium isolated from Great Salt Lake. Int J Syst Evol Microbiol 56: 2063–2069.
60. Garrity GM, Brenner DJ, Krieg NR, Staley JT (2005) Part B The *Gammaproteobacteria*; GarrityGM, New York: Springer. 1006.
61. Sun H, Spring S, Lapidus A, Davenport K, Del Rio TG, et al. (2010) Complete genome sequence of *Desulfarculus baarsii* type strain (2st14). Stand Genomic Sci 3: 276–284.
62. Daffonchio D, Borin S, Brusa T, Brusetti L, van der Wielen PW, et al. (2006) Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. Nature 440: 203–207.
63. Alfreider A, Vogt C, Babel W (2002) Microbial diversity in an in situ reactor system treating monochlorobenzene contaminated groundwater as revealed by 16S ribosomal DNA analysis. Systematic and Applied Microbiology 25: 232–240.
64. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 180: 366–376.
65. Pynaert K, Smets BF, Wyffels S, Beheydt D, Siciliano SD, et al. (2003) Characterization of an autotrophic nitrogen-removing biofilm from a highly loaded lab-scale rotating biological contactor. Appl Environ Microbiol 69: 3626–3635.
66. Elevi Bardavid R, Oren A (2012) The amino acid composition of proteins from anaerobic halophilic bacteria of the order *Halanaerobiales*. Extremophiles 16: 567–572.
67. Miroshnichenko ML, Kostrikina NA, Chernyh NA, Pimenov NV, Tourova TP, et al. (2003) *Caldithrix abyssi* gen. nov., sp. nov., a nitrate-reducing, thermophilic,

anaerobic bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent, represents a novel bacterial lineage. Int J Syst Evol Microbiol 53: 323–329.
68. Miroshnichenko ML, Kolganova TV, Spring S, Chernyh N, Bonch-Osmolovskaya EA (2010) *Caldithrix palaeochoryensis* sp. nov., a thermophilic, anaerobic, chemo-organotrophic bacterium from a geothermally heated sediment, and emended description of the genus *Caldithrix*. Int J Syst Evol Microbiol 60: 2120–2123.
69. Garrity GM, Brenner DJ, Krieg NR, Staley JT (2001) The *Archaea* and the Deeply Branching and Phototrophic *Bacteria*; GarrityGM, New York: Springer. 721.
70. Takai K, Horikoshi K (1999) Genetic diversity of archaea in deep-sea hydrothermal vent environments. Genetics 152: 1285–1297.
71. Vetriani C, Jannasch HW, MacGregor BJ, Stahl DA, Reysenbach AL (1999) Population structure and phylogenetic characterization of marine benthic *Archaea* in deep-sea sediments. Appl Environ Microbiol 65: 4375–4384.
72. Stoeck T, Epstein S (2003) Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. Appl Environ Microbiol 69: 2657–2663.
73. Ferrer M, Guazzaroni ME, Richter M, Garcia-Salamanca A, Yarza P, 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: 824–837.
74. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, et al. (2013) Predominant archaea in marine sediments degrade detrital proteins. Nature 496: 215–218.
75. Knittel K, Lösekann T, Boetius A, Kort R, Amann R (2005) Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. Appl Environ Microbiol 71: 467–479.
76. Turova TP, Kolganova TV, Kuznetsov BB, Pimenov NV (2002) Phylogenetic diversity of the archaeal component in microbial mats on coral-like structures associated with methane seeps in the Black Sea. Mikrobiologiia 71: 230–236.
77. DeLong EF, Pace NR (2001) Environmental diversity of bacteria and archaea. Syst Biol 50: 470–478.
78. Kempe S, Kazmierczak J (1990) Calcium carbonate supersaturation and the formation of in situ calcified stromatolites. In: Ittekkot VA, Kempe S, Michaelis W, Spitzy A, editors. Facets of Modern Biogeochemistry. New York: Springer. 255–278.
79. Trichet J, Défarge C (1995) Non-biologically supported organomineralization. Monaco: l'Institut Océanographique de Monaco. 203–236.
80. Arp G, Wedemeyer N, Reitner J (2001) Fluvival tufa formation in a hard-water creek (Deinschwanger Bach, Franconian Alb, Germany). Facies 44: 1–22.
81. Arp G, Reimer A, Reitner J (2003) Microbialite formation in seawater of increased alkalinity, Satonda Crater Lake, Indonesia. Society for Sedimentary Geology 73: 105–127.
82. Dupraz C, Visscher PT (2005) Microbial lithification in marine stromatolites and hypersaline mats. Trends Microbiol 13: 429–438.
83. Cohen Y, Castenholz RW, Halvorson HO (1984) Microbial Mats: Stromatolites. Canada: John Wiley & Sons. 498.
84. Pester M, Schleper C, Wagner M (2011) The *Thaumarchaeota*: an emerging view of their phylogeny and ecophysiology. Curr Opin Microbiol 14: 300–306.

2.1 Supplemental information for chapter B2

Contents

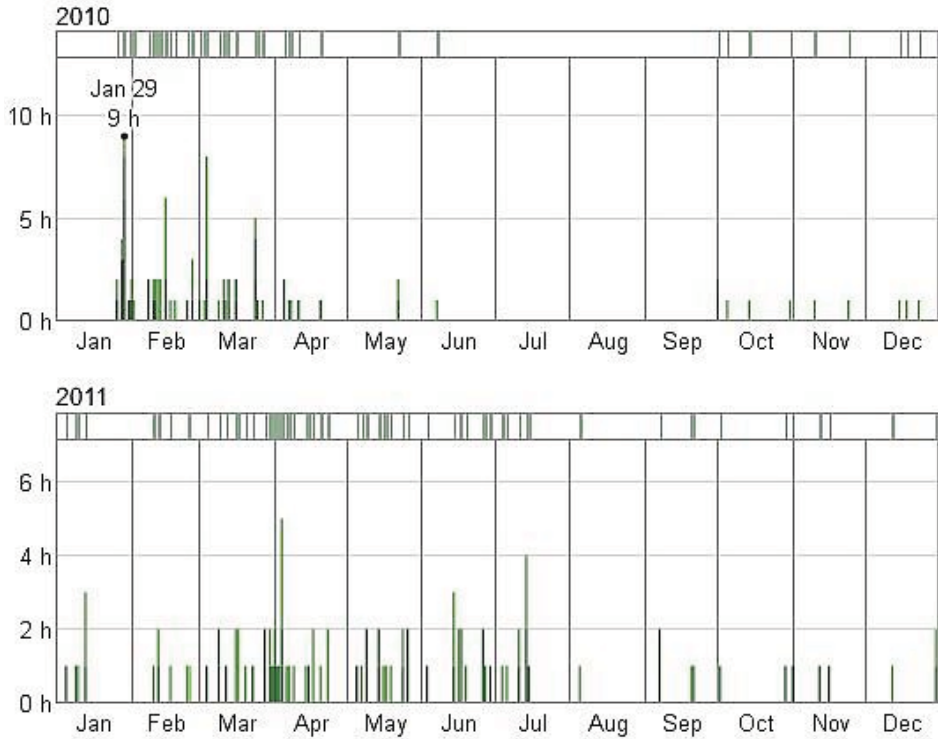
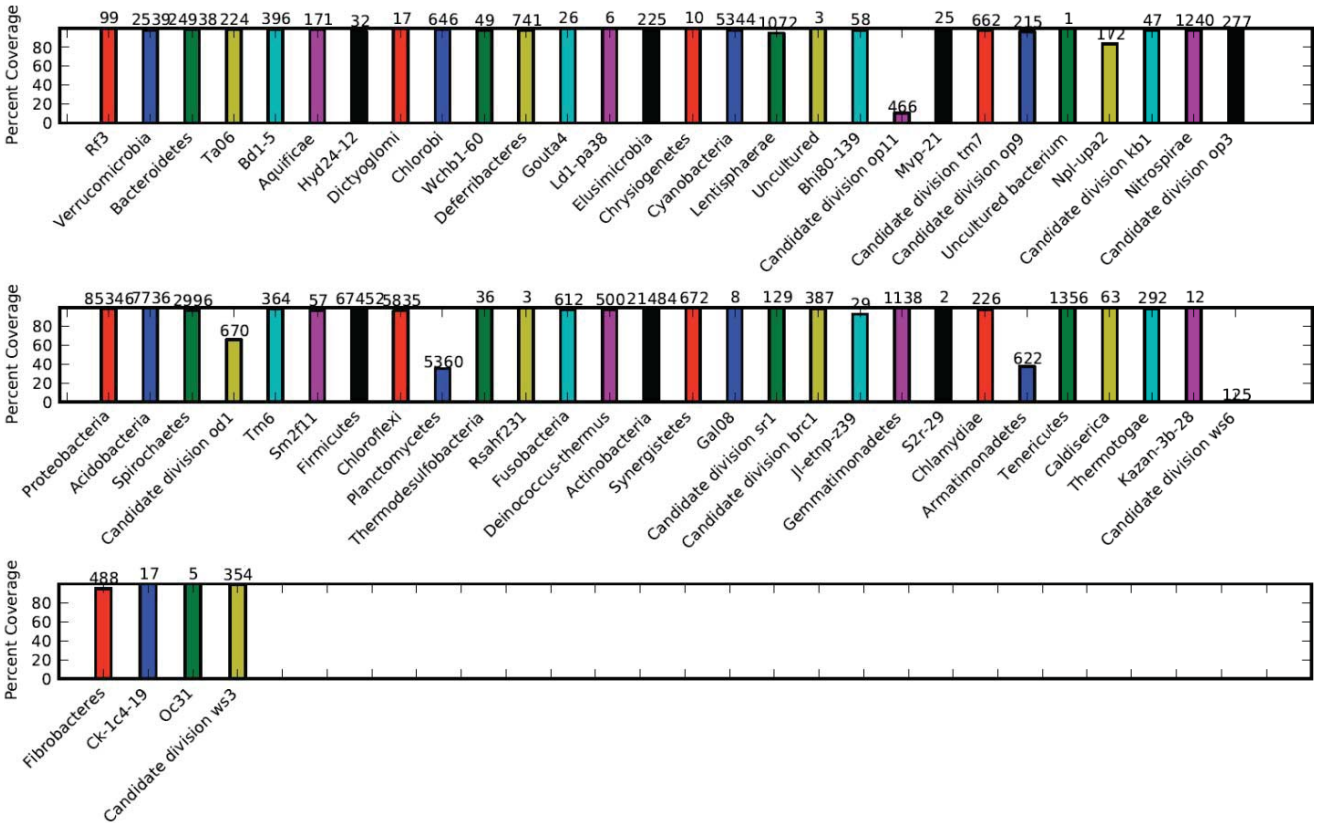
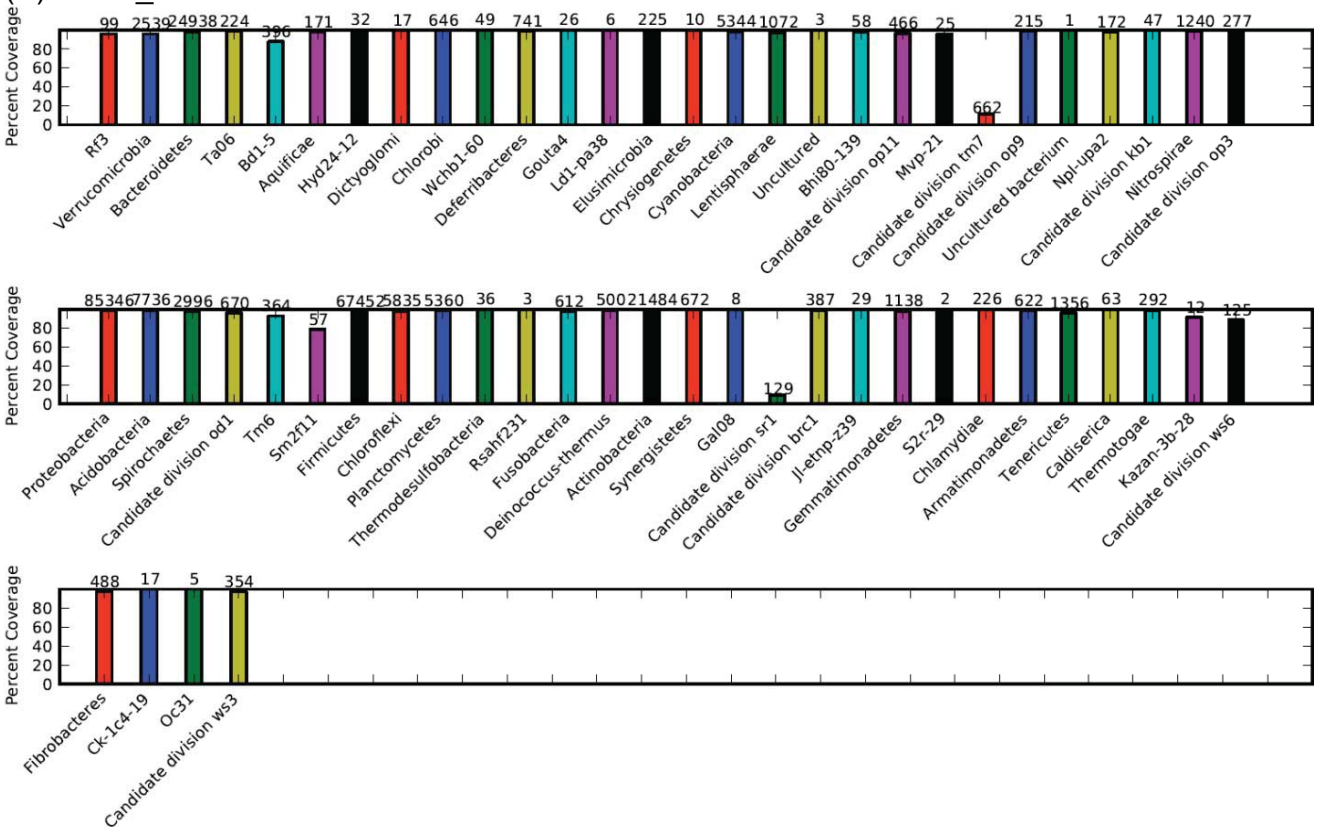


Figure S1. The daily number of hourly observed precipitation reports during 2010 and 2011. Color coded according to precipitation type and stacked in order of severity. From the bottom up, the categories are heavy, moderate, and light rain (dark to light green); and drizzle (lightest green). The faint shaded areas indicate climate normals. The bar at the top of the graph is green if any precipitation was observed that day and white otherwise (http://weatherspark.com).

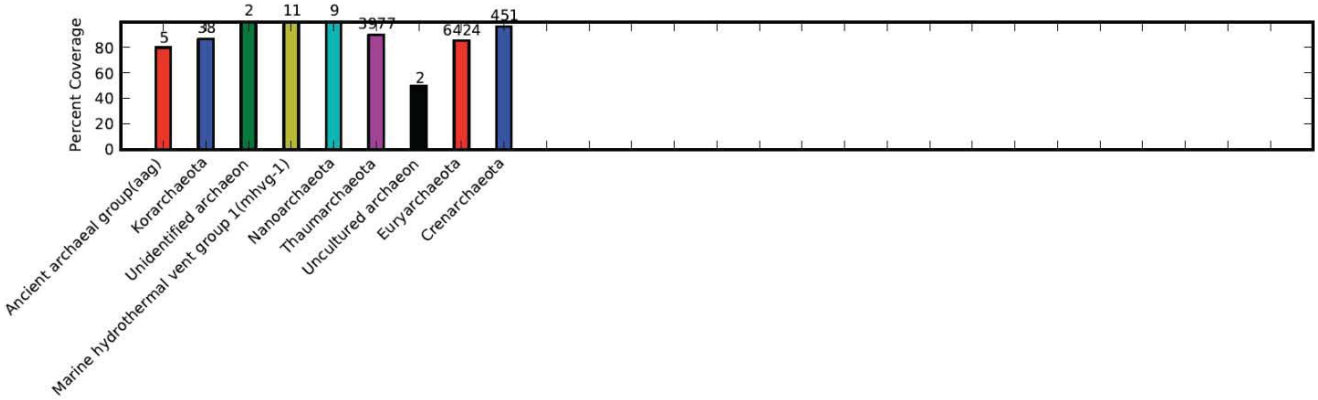
(A) V3for_B


Percent Coverage
0
20
40
60
80
99 25 39 24 9 38 224 396 171 32 17 646 49 741 26 6 225 10 5344 1072 3 58 466 25 662 215 1 172 47 1240 277
Rf3
Verrucomicrobia
Bacteroidetes
Ta06
Bd1-5
Aquificae
Hyd24-12
Dictyoglomi
Chlorobi
Wchb1-60
Deferribacteres
Gouta4
Ld1-pa38
Elusimicrobia
Chrysiogenetes
Cyanobacteria
Lentisphaerae
Uncultured
Bhi80-139
Candidate division op11
Mvp-21
Candidate division tm7
Candidate division op9
Uncultured bacterium
Npl-upa2
Candidate division kb1
Nitrospirae
Candidate division op3
Percent Coverage
0
20
40
60
80
85 346 77 36 29 96 670 364 57 67 45 25 835 5360 36 3 612 500 21 484 672 8 129 387 29 1138 2 226 622 1356 63 292 12 125
Proteobacteria
Acidobacteria
Spirochaetes
Candidate division od1
Tm6
Sm2f11
Firmicutes
Chloroflexi
Planctomycetes
Thermodesulfobacteria
Rsahf231
Fusobacteria
Deinococcus-thermus
Actinobacteria
Synergistetes
Gal08
Candidate division sr1
Candidate division brc1
Jl-etnp-z39
Gemmatimonadetes
S2r-29
Chlamydiae
Armatimonadetes
Tenericutes
Caldiserica
Thermotogae
Kazan-3b-28
Candidate division ws6
Percent Coverage
0
20
40
60
80
488 17 5 354
Fibrobacteres
Ck-1c4-19
Oc31
Candidate division ws3

(B) V5rev_B



(C) V3for_A



(D) V5rev_A

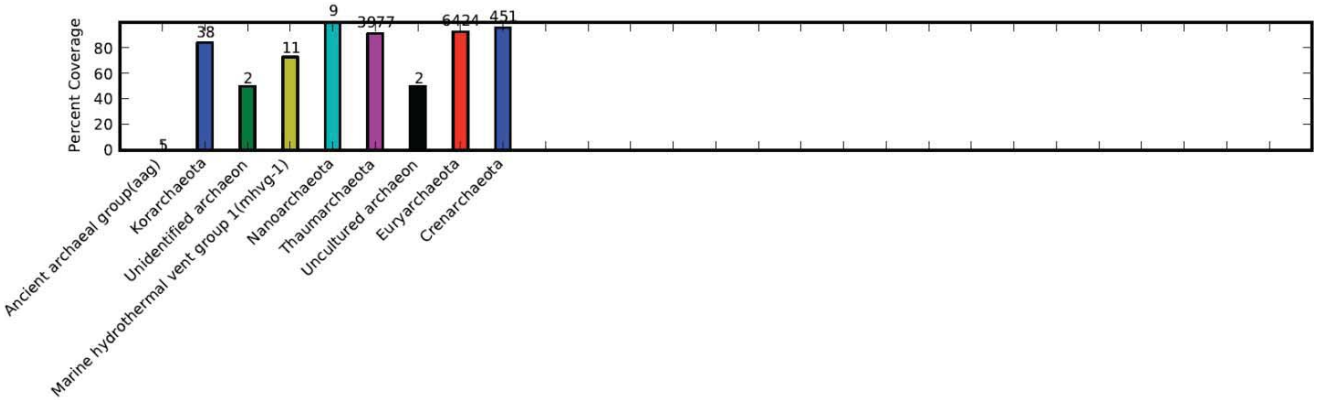


Figure S2. Primer coverage of bacterial and archaeal phyla. Coverage of bacterial phyla of forward (A) and reverse (B) primers and archaeal phyla of forward (C) and reverse (D) primers were determined with PrimerProspector [1]. Colored bars represent percent coverage of the primers and numbers above bars depict total sequences within the SILVA SSU database (release 111 Ref NR) for each phylum [2].

References

1. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, et al. (2011) PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27: 1159-1161.

2. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35: 7188-7196.

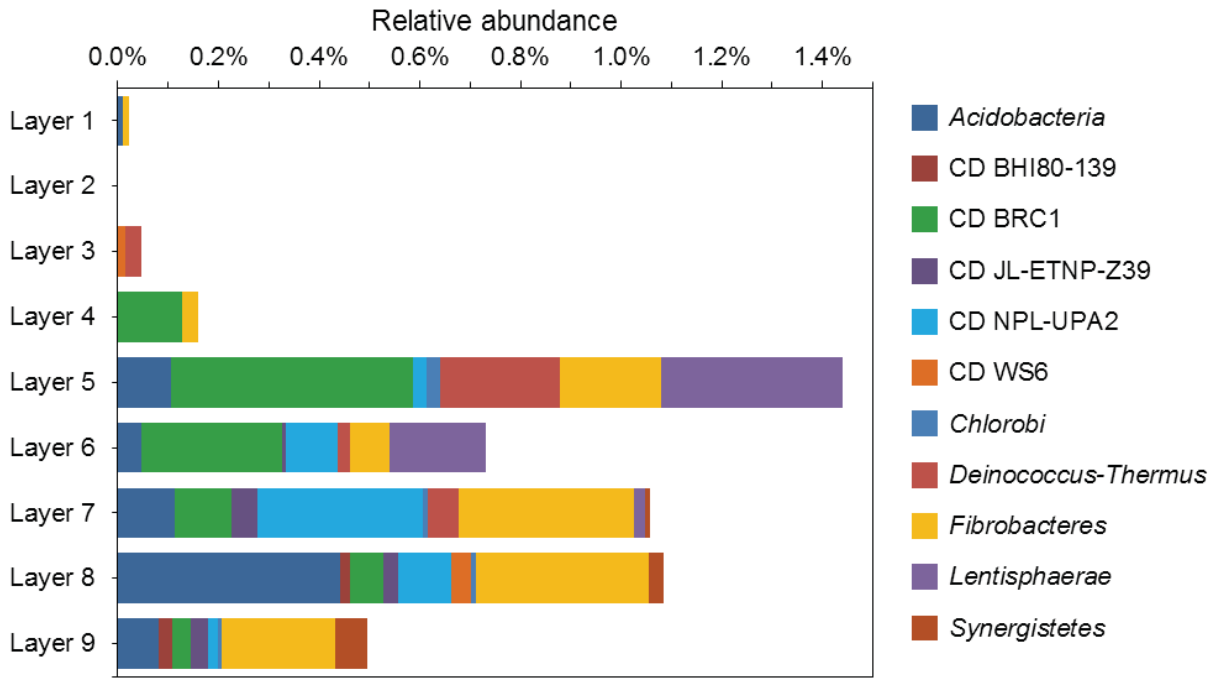


Figure S3. Relative abundances of rare bacterial phylogenetic groups. The artificial group "Other" (Figure 6) includes rare (<0.5%) bacterial phyla and candidate divisions (CD).

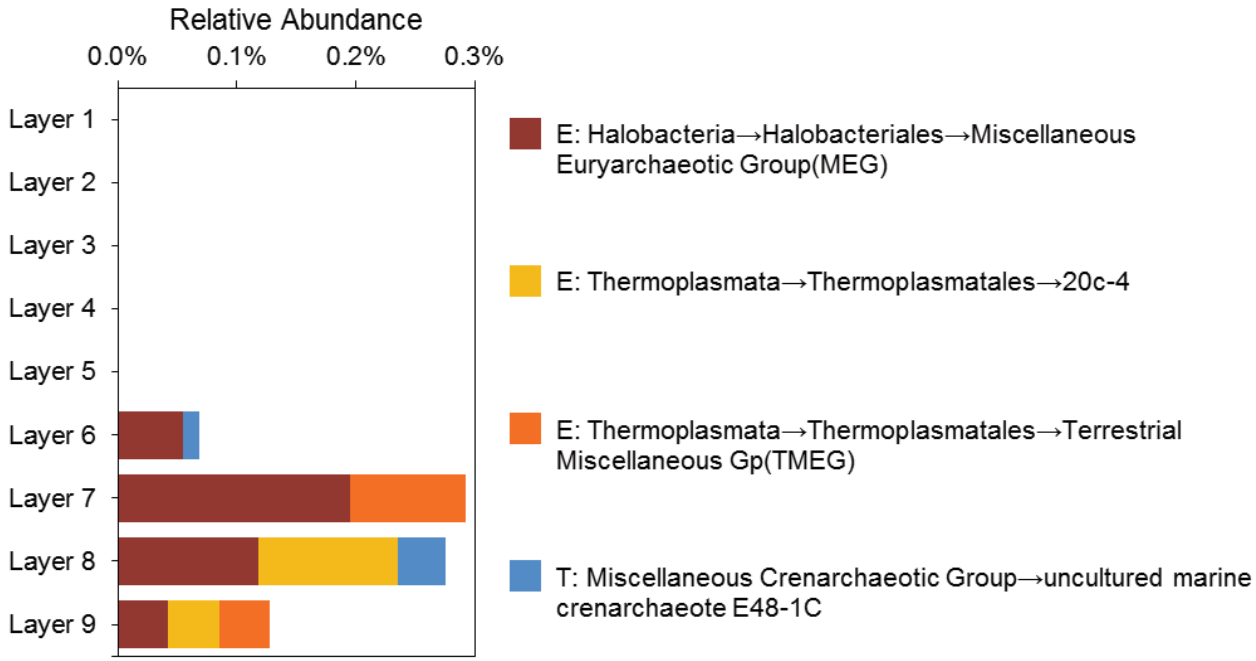


Figure S4. Relative abundances of rare archaeal phylogenetic groups. The artificial group "Other" (Figure 7) includes rare (<0.5%) archaeal lineages and candidate divisions. Abbreviations: E, *Euryarchaeota*; T, *Thaumarchaeota*.

Table S1. Hydrochemical data of sea water and Lake 21 waters, Kiritimati atoll.

Sample	Sampling depth [m]	Date	T [°C]	EC 25°C (mS cm^{-1})	Salinity (‰)	Eh (mV)	pH
Seawater (Captain Cook Hotel)	0.1	19.03.2011	28.4	53.0	34.7	n.d.	8.36
Lake 21 (SE-side 4 m from shore)	0.1	13.03.2011	31.7	189.8	171.8	91	7.95
Lake 21 (seepage-affected lake bottom)	1.0	15.03.2011	28.7	124.8	94.4	n.d.	8.03
Lake 21 (surface, 2002)	0.1	04.09.2002	32.5	140.1	112.9	263	8.37

Sample	O_2	Ca^{2+}	Mg^{2+}	Na^{+}	K^{+} (mmol L^{-1})	Sr^{2+}	Cl^{-}	SO_4^{2-}	TA meq L^{-1}
Seawater (Captain Cook Hotel)	0.198	10.4	54	474	10.3	0.093	552	28.6	2.31
Lake 21 (SE-side 4 m from shore)	0.079	29.3	299	2633	57.1	0.426	3054	143.3	4.84
Lake 21 (seepage-affected lake bottom)	n.d.	19.1	155	1363	29.7	0.232	1588	74.7	3.44
Lake 21 (surface, 2002)	n.d.	38.0	197	1643	34.7	0.312	1841	113.6	4.24

Sample	Si	Fe (μmol L^{-1})	NH^{4+}	PO_4^{3-}	PCO_2 (μatm)	$SI_{Calcite}$[a]	$SI_{Aragonitea}$ (log IAP/KT)	SI_{Gypsum}[a]
Seawater (Captain Cook Hotel)	3.6	1.10	n.d.	0.44	269	0.86	0.72	-0.66
Lake 21 (SE-side 4 m from shore)	54.6	1.35	0.086	0.02	871	1.26	1.12	0.13
Lake 21 (seepage-affected lake bottom)	n.d.	0.23	n.d.	n.d.	676	0.92	0.78	-0.26
Lake 21 (surface, 2002)	31.1	1.97	2.25	0.46	224	1.44	1.30	0.13

[a]Saturation index $SI_{Calcite} = \log \Omega_{Calcite} = \log$ [ion activity product $(Ca^{2+}) \times (CO_3^{2-})$ / solubility product $K_{calcite}$]
n.d., not determined

Table S2. Data processing of 16S rRNA gene sequences.

Sample	No. of sequences												
	Raw	<250 bp	With ambiguous base pairs	Quality score <25	>8 homo-polymers	>2 mismatches in forward primer	After quality filtering	Potential chimeras (UCHIME *de novo*)	Potential chimeras (UCHIME reference mode)	In OTU table	Singletons	Extrinsic domain and chloroplast OTUs	Final OTU table
Bacteria													
Layer 1	13,523	2,997	551	89	0	106	9,780	4	0	9,776	315	16	9,445
Layer 2	9,965	2,809	424	35	0	28	6,669	3	0	6,666	19	1	6,646
Layer 3	10,604	3,297	451	27	0	38	6,791	374	9	6,408	93	0	6,315
Layer 4	10,418	3,080	741	16	0	36	6,545	81	1	6,463	184	0	6,279
Layer 5	11,200	2,705	773	51	0	46	7,625	53	4	7,568	63	0	7,505
Layer 6	17,981	4,283	835	87	2	106	12,668	6	0	12,662	82	0	12,580
Layer 7	15,064	4,086	906	84	3	64	9,921	6	1	9,914	173	0	9,741
Layer 8	16,653	4,268	1,195	106	3	55	11,026	5	2	11,019	602	0	10,417
Layer 9	17,308	4,697	948	133	1	41	11,488	1	0	11,487	370	0	11,117
Archaea													
Layer 1	6,867	998	458	937	9	1,323	3,142	0	0	3,142	35	0	3,107
Layer 2	4,980	710	149	33	0	274	3,814	7	0	3,807	2	0	3,805
Layer 3	4,585	646	223	50	0	219	3,447	12	0	3,435	17	0	3,418
Layer 4	5,101	607	334	43	0	376	3,741	4	0	3,737	213	0	3,524
Layer 5	22,937	5,088	711	3,071	1	2,519	11,547	852	0	10,695	31	1	10,663
Layer 6	13,985	2,345	335	2,060	4	1,956	7,285	28	0	7,257	22	0	7,235
Layer 7	5,933	959	160	819	2	824	3,169	73	0	3,096	16	0	3,080
Layer 8	4,859	745	166	758	0	453	2,737	38	0	2,699	155	0	2,544
Layer 9	5,061	954	137	876	0	617	2,477	54	0	2,423	69	1	2,353

Table S3. Hydrochemical data of Lake 21 water and pore water of microbial mat layers.

Sample	pH	Temperature (°C)	Redox potential (mV)
Lake water	7.948	31.7	91.4
Layer 1[a]	7.708	32	-43
Layer 2[a]	7.708	32	-43
Layer 3[a]	7.708	32	-43
Layer 4	6.716	32	-123
Layer 5	6.670	32	-133
Layer 6	6.664	32	-142
Layer 7[a]	6.661	32	-135
Layer 8[a]	6.661	32	-135
Layer 9	Not determined	Not determined	Not determined

[a]values derive from pooled porewater

3 Metagenomic and metatranscriptomic analyses of bacterial communities derived from a calcifying karst water creek biofilm and tufa

Dominik Schneider[1], Andreas Reimer[2], Angelina Hahlbrock[1,3], Gernot Arp[2], Joachim Reitner[2], Rolf Daniel[1]

Geomicrobiology Journal, submitted (03.09.2013)

[1]Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany,
[2]Geoscience Centre, Georg-August University Göttingen, Göttingen, Germany
[3]Present address: Molecular and Cellular Oncology/Mainzer Screening Center, University Hospital of Mainz, Langenbeckstrasse 1, 55131 Mainz

---

Author contributions:
Conceived and designed the experiments: DS GA JR RD.
Performed the experiments: DS AH AR.
Analyzed the data: DS AH AR RD.
Wrote the paper: DS GA RD.

METAGENOMIC AND METATRANSCRIPTOMIC ANALYSES OF BACTERIAL COMMUNITIES DERIVED FROM A CALCIFYING KARST WATER CREEK BIOFILM AND TUFA

Dominik Schneider[a], Andreas Reimer[c], Angelina Hahlbrock[a,b], Gernot Arp[c], Joachim Reitner[c], and Rolf Daniel[a]

[a]Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany

[b]Present address: Molecular and Cellular Oncology, Mainzer Screening Center, University Hospital of Mainz, Langenbeckstrasse 1, 55131 Mainz

[c]Geoscience Centre, Georg-August University Göttingen, Goldschmidtstraße 3, D-37077 Göttingen, Germany

To whom correspondence should be addressed:

Rolf Daniel, Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany. Phone: 0049-551-3933827, Fax: 0049-551-3912181, rdaniel@gwdg.de

Abstract

Calcification of freshwater streams and the involvement of microorganisms in this process are still not fully understood. Here, we report on the metagenomic and metatranscriptomic analyses of bacterial community structures derived from an actively calcifying karst water creek biofilm and underlying tufa by employing next-generation sequencing technologies.

The persistent and metabolically active bacterial communities were assessed by DNA-based and RNA-based 16S rRNA gene sequence analysis, respectively. We identified filamentous *Cyanobacteria* belonging to the *Oscillatoriales* as the predominant bacterial microorganisms in the biofilm. Cyanobacteria were accompanied by a high diversity of mainly aerobic members affiliated to different bacterial phyla. The second most abundant phylum was *Proteobacteria*, represented by the classes *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, and *Deltaproteobacteria*. In addition, *Chloroflexi*, *Planctomycetes*, *Actinobacteria*, *Acidobacteria*, *Bacteroidetes*, *Spirochaetes*, and *Verrucomicrobia* were present in higher abundances (>0.5%). Several of these phyla included potentially novel subgroups. The tufa stro-

matolite carbonate one centimeter below the surface exhibited a similar prokaryotic community as the superficial biofilm but showed a lower abundance of *Cyanobacteria* and a more diverse microbial community.

We obtained insights into the *in situ* microbial metabolism by employing directed sequencing of enriched mRNA (cDNA) and subsequent taxonomic and functional analysis of biofilm-derived reads. High activities in photosynthesis, carbon and protein metabolism were indicated. Furthermore, we constructed large-insert metagenomic libraries of all sampling sites and identified five novel, mainly cyanobacteria-derived, proteolytic enzymes of the serine protease families S1 and S8.

Introduction

The Westerhöfer Bach is a karst water stream in Germany located in the west of the Harz Mountains. It receives water from a spring area discharging from the Middle Triassic Muschelkalk Group aquifer (Arp *et al.* 2010). Thus, with respect to the water origin it has a relatively high content of bicarbonate, calcium, sulfate, and magnesium (Arp *et al.* 2010). The riverbed is defined by a calcareous laminated tufa (stromatolite) covered by a salient calcareous biofilm, which visibly contains filamentous *Cyanobacteria* (Arp *et al.* 2010). In general, freshwater biofilms harbor diatoms, algae, protozoa, fungi and bacteria, which are adapted to low salinity and nutrient supply and embedded in a matrix consisting of extracellular polymeric substances (EPS) (Dürr and Thomason 2010). Tufa can be described as laminated lithic sediments of calcium carbonate containing organic and inorganic compounds. Calcification, e.g., tufa formation, has been intensively studied recently (Arp *et al.* 2010; Beraldi-Campesi *et al.* 2012; Gonzalez-Munoz *et al.* 2010; Pedley 2013; Perri *et al.* 2012). Generally, four mechanisms were proposed for calcification within these systems (1) trapping of calcite particles to the tufa/biofilm matrix (particles derived from older limestone or by precipitation within the water column) (La Touche 1913; von Buch 1809), (2) physicochemical CO_2-degassing (e.g., Burger 1911; Schürmann 1918; Unger 1861), (3) adsorption to plant surfaces (Klähn 1923), and (4) CO_2-assimilation by photosynthetic organisms (e.g., Shiraishi *et al.* 2008b; Shiraishi *et al.* 2010). Tufa formation of Westerhöfer Bach starts in a distance of approximately 130 to 150 m from the spring site. Microsensor experiments revealed that the activity of photo-

trophic microorganisms (mainly *Cyanobacteria*) significantly influences the hydrochemistry at the biofilm surface on the microscale (Shiraishi *et al.* 2008b). At daytime, photosynthesis-driven removal of CO_2 leads to calcite precipitation around the cell surfaces. This results in formation of cyanobacterial calcite sheaths (Zippel and Neu 2011). However, mass balance calculations suggested that photosynthesis related calcification only accounts for up to 20% of Ca^{2+} loss. The remaining depletion is probably caused by physicochemical precipitation (Arp *et al.* 2010). An additional influence on carbonate precipitation might be the degradation of EPS components by microbes leading to release of Ca^{2+} (Pentecost and Bauld 1988; Shiraishi *et al.* 2008a). Studies, although not from freshwater systems, have demonstrated that several bacterial taxa possess the direct or indirect potential to precipitate a wide range of minerals under certain environmental conditions (Gonzalez-Munoz *et al.* 2010; Jroundi *et al.* 2010; Rusznyak *et al.* 2012; Schneider *et al.* 2013).

Studies of tufa-forming freshwater systems targeting the characterization of microbes by molecular approaches are rare. Studies with respect to bacterial distribution and diversity of these systems were carried out at Iberian Range (Spain) (Beraldi-Campesi *et al.* 2012) and Taroko National Park (Taiwan) (Ng *et al.* 2006). Bacterial diversity was assessed by traditional Sanger-based sequencing of cloned 16S rRNA genes. In these studies, the occurrence and predominance of filamentous *Cyanobacteria* followed by high abundances of *Proteobacteria* was detected. Taking into account the high diversity of freshwater biofilms (Beraldi-Campesi *et al.* 2012; Ng *et al.* 2006) the surveying effort of both studies, encompassing 94 and 381 16S rRNA gene sequences, was too low to cover the bacterial community and diversity.

Previous culture-based studies and small-scale analysis of cloned 16S rRNA genes focused on microorganisms derived from stream water, biofilm, and tufa samples, i.e., *Cyanobacteria* (Arp *et al.* 2010), other *Bacteria* (Cousin and Stackebrandt 2010), and diatoms (Arp *et al.* 2010). Thereby, several novel taxa were detected and described. From the Westerhöfer Bach stream water several novel flavobacterial and pedobacterial species were isolated and described, e.g., *Flavobacterium aquidurense* (Cousin *et al.* 2007), *Flavobacterium rivuli*, *subsaxonicum*, *swingsii*, and *reichenbachii* (Ali *et al.* 2009), *Pedobacter westerhofensis*, *metabolipauper*, *hartonius*, and *steynii* (Muurholm *et al.* 2007). A study by Cousin *et al.* (2010) focused on the spatial distribution of bacteria along a depth gradient of the Westerhöfer Bach

tufa and a biofilm sample (Cousin and Stackebrandt 2010). Their results indicated that members of the phylum *Proteobacteria* dominated in the lower tufa layers, whereas the uppermost tufa layer and biofilm were dominated by *Cyanobacteria*. However, by limitations based on the employed methods (low surveying effort) only a small fraction of the biofilm community has been assessed (Cousin and Stackebrandt 2010). It was shown that a novel *Tychonema* species and several unidentified *Cyanobacteria* lineages constituted a major part of the cyanobacterial community in the biofilm (Arp *et al.* 2010; Brinkmann *et al.* 2013, this issue).

In this study, we report on analysis of the bacterial community composition of the Westerhöfer Bach biofilm along a downstream gradient of three sampling sites and its underlying tufa. For this purpose, we employed large-scale amplicon-based pyrosequencing of 16S rRNA genes. To our knowledge this is the first study of cyanobacterial tufa-systems based on next-generation sequencing techniques. The parallel investigation of 16S rRNA gene sequences derived from DNA and RNA allowed us to observe differences between the entire and active bacterial communities inhabiting the biofilm. We used a metatranscriptomic approach to gain first insights into the active metabolic processes of a biofilm sample. In addition, construction and screening of large-insert metagenomic libraries resulted in identification of five novel genes encoding proteolytic enzymes, which originated from filamentous *Cyanobacteria*. In this way, this study contributed to unravel composition and functions of bacterial biofilm communities in karstic environments and their possible influences on calcification.

Materials and Methods

Site description and sample collection

Samples were collected at three different sites of the Westerhöfer Bach (WB) in July 2009 at approximately 1:00 pm (Figure 1A and 1B). The Westerhöfer Bach is an up to 2 m wide karst water creek, which is located in the Western Harz Mountains, which receives its water from a spring area discharging from the Middle Triassic Muschelkalk Group aquifer (Arp *et al.* 2010). With only small annual variations, water velocity is approximately 2.0 L s^{-1} and the water temperature varies from 9 to 10°C dur-

ing summer (Arp *et al.* 2010). The three sampling sites were located 262 (WB3), 287 (WB4) and 313 m (WB5) downstream of the creek spring area. It should be noted that sampling site WB3 was located behind a small cascade with inactive tufa.

Approximately 6 cm^2 of the up to 320 µm thick biofilm were collected at each sampling site. One tufa sample (area of 5 cm^2 and depth of approximately 1 cm) was acquired at sampling site WB5 after the complete removal of the biofilm. To remove potential contaminations of the tufa sample by rivulet microorganisms, rinsing with TE buffer (10 mM Tris, 1mM EDTA, pH 8) was performed. The samples were immediately frozen in liquid nitrogen and transported on dry ice to the laboratory. Subsequently, samples were stored at -80°C until further use.

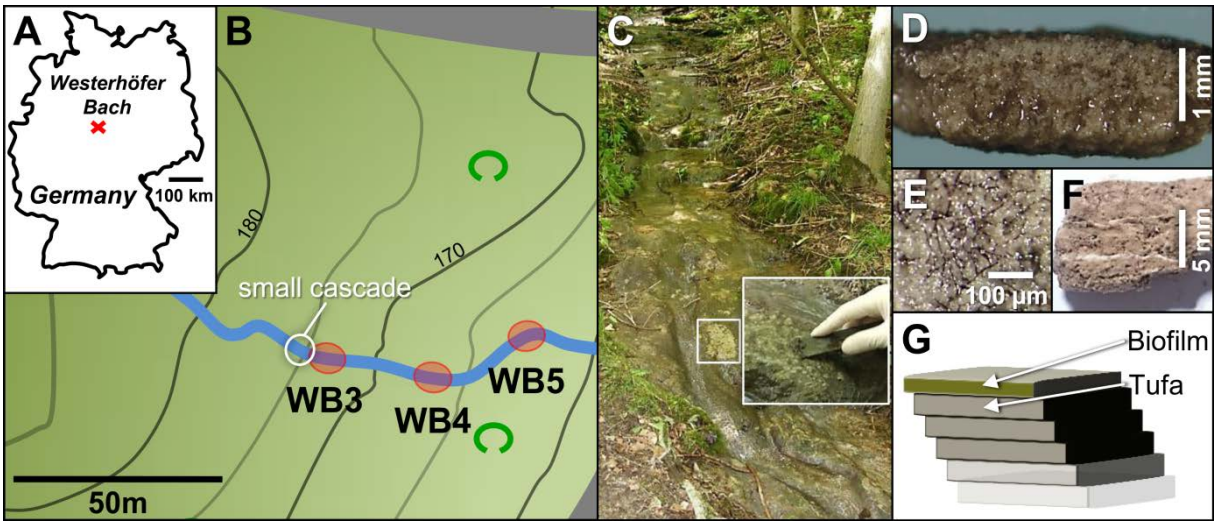


Fig. 1. Location, sampling, and structure of the studied tufa-forming karstwater stream Westerhöfer Bach, Germany. Geographic location of the investigated karstwater creek at Westerhof (A). Map of Westerhöfer Bach lower section including sampling sites (B). The distance between the three sampling sites WB3, WB4 and WB5 is approximately 25 m (see Shiraishi *et al.* (2008b) for whole creek course). The tufa sample was acquired at WB5 after the complete removal of the biofilm. View on WB4 upstream and detail (white square) showing sampling (C). Cross-section of highly calcified biofilm, brown-green coloration originates from cyanobacteria (D). Top-view on biofilm surface with filamentous cyanobacteria (Oscillatoriales) and surrounding calcareous microbialite (E). Cross-section of tufa sample depicting lamination, each lamina encompasses 1 year (white deposits contain high amounts of diatoms) (F). Draft of the Westerhöfer Bach tufa lamination fabric (G).

Isolation of nucleic acids

In order to isolate environmental DNA and RNA, the samples were first grinded with liquid nitrogen to reduce potential sample heterogeneity. Environmental DNA and RNA were isolated from each sample by employing the MoBio PowerBiofilm DNA

isolation kit (MO BIO Laboratories, Carlsbad, USA) and the MoBio PowerBiofilm RNA isolation kit (MO BIO Laboratories), respectively, as recommended by the manufacturer. Isolations were initiated by using 250 mg sample material. In the case of RNA isolation, contaminating DNA was removed by treatment with TurboDNase (Applied Biosystems, Darmstadt, Germany) as suggested by the manufacturer. The reaction mixture was subsequently purified and concentrated by using the RNeasy MinElute Cleanup kit (Qiagen, Hilden, Germany) as recommended by the manufacturer. To verify the complete removal of DNA, a PCR reaction targeting the 16S rRNA gene was performed as described below. The DNA and RNA yields were estimated by employing NanoDrop ND-1000 spectrophotometer (Peqlab Biotechnologie GmbH, Erlangen, Germany).

Amplification of 16S rRNA genes and transcripts and pyrosequencing

To assess the bacterial community composition the V2 to V3 region of the 16S rRNA gene was amplified by PCR. For amplification of 16S rRNA genes from DNA template the PCR reaction mixture (50 µl) contained 10 µl of 5-fold reaction buffer (Fusion GC buffer, Finnzymes, Vantaa, Finland), 200 µM of each of the four deoxynucleoside triphosphates, 0.2 µM of each primer, 5% DMSO (Finnzymes), 1 U of Phusion hot start high-fidelity DNA Polymerase (Finnzymes) and 50 ng of isolated DNA as template. The targeted V2 to V3 region was amplified with the following set of barcoded primers containing the Roche 454 pyrosequencing adaptors (underlined), a key (TCAG), a variable multiplex identifier (MID) consisting of ten bases, which was followed by the primer: V2for 5'-GCCTCCCTCGCGCCATCAG-MID-AGTGGCGGACGGGTGAGTAA-3' and V3rev 5'-GCCTTGCCAGCCCGCTCAG-MID-CGTATTACCGCGGCTGCTG-3'. Primers were modified from Schmalenberger *et al.* (2001) and have been successfully applied to characterize bacterial communities in different types of soil (Nacke *et al.* 2011a). The following thermal cycling scheme was used: initial denaturation at 98°C for 5 min, 25 cycles of denaturation at 98°C for 45 s, annealing for 45 s at 68°C, and extension at 72°C for 30 s, followed by a final extension period at 72°C for 5 min. Negative controls contained the entire reaction mixture without the template DNA. The resulting PCR products were purified by using the peqGold gel extraction kit (Peqlab Biotechnologie GmbH) as recommended by the manufacturer.

RNA was converted to cDNA followed by amplification of the 16S rRNA genes with the Qiagen OneStep RT-PCR kit (Qiagen) according to the manufacturer's instructions. To amplify the V2 to V3 region of the 16S rRNA genes the above-described primers were used. The reaction mixture (25 µl) contained 5 µl of 5x Qiagen RT-PCR Buffer (Qiagen), 200 µM of each of the four deoxynucleoside triphosphates, 0.1 µM of each primer, 1 µl of Qiagen OneStep RT-PCR enzyme mix (Qiagen) and 100 ng of isolated total RNA. The following thermal cycling scheme was used: reverse transcription was performed at 55°C for 30 min, initial PCR activation step at 95°C for 15 min, 25 cycles of denaturation at 94°C for 1 min, annealing for 30 s at 58°C and extension at 72°C for 30 s, followed by a final extension period at 72°C for 10 min. Negative controls, size control, and purification of the RNA-derived amplicons were performed as described for the DNA-derived 16S rRNA gene amplicons.

All PCR reactions (including RT-PCR) were performed in triplicate and pooled in equal amounts. Quantification of the PCR products was performed using the Quant-iT dsDNA BR assay kit and a Qubit fluorometer (Invitrogen, Karlsruhe, Germany) as recommended by the manufacturer.

Processing and analysis of 16S rRNA gene datasets

The Göttingen Genomics Laboratory determined the sequences of the partial 16S rRNA genes by using a Roche GS-FLX 454 pyrosequencer (Roche, Mannheim, Germany) and Titanium chemistry following the instructions of the manufacturer for amplicon sequencing. The 16S rRNA genes sequence datasets were processed and analyzed employing the QIIME 1.4 software package (Caporaso *et al.* 2010). Initially, sequences shorter than 300 bp, containing unresolved nucleotides, exhibiting an average quality score lower than 25, possessing homopolymers longer than 8 bp, or harbor mismatches longer than 2 bp in the forward primer were removed from the datasets employing the QIIME script *split_libraries.py*. Afterwards, pyrosequencing noise was removed by employing the Denoiser 0.91 (Reeder and Knight 2010) included in QIIME (Caporaso *et al.* 2010). We additionally removed unclipped reverse primer sequences by employing cutadapt (Martin 2011) with default settings. Chimeric sequences were removed using UCHIME (Edgar *et al.* 2011) in reference mode with the Greengenes gold dataset (gold_strains_gg16S_aligned.fasta) (DeSantis *et*

al. 2006) as reference database. Operational taxonomic unit (OTU) determination was performed using UCLUST (Edgar 2010) at genetic divergence level of 1% employing pick_otus.py. This high phylogenetic resolution was of crucial importance, especially for differentiation of cyanobacterial 16S rRNA genes. For this purpose, the most abundant sequence within an OTU was picked by employing the QIIME script *pick_rep_set.py* (Caporaso *et al.* 2010). OTUs clustered at 1% genetic divergence were the basis of all presented results. Taxonomic classification of the picked reference sequences (OTUs) was performed by similarity searches using BLAST (blastall version 2.2.18) (Altschul *et al.* 1990) against the SILVA database version 111 (SSURef_111_NR_tax_silva_trunc.fasta) (Pruesse *et al.* 2007). The SILVA taxonomy of the first hit based on E-value was then adopted to infer taxonomy of the representative sequences according to the QIIME script *assign_taxonomy.py*. An OTU table was created using *make_otu_table.py*. Sequences of chloroplasts and OTUs with low alignment coverage (<95%) with respect to the corresponding BLAST hit were removed for further analysis. Rarefaction curves, Shannon indices and Michaelis-Menten fit were calculated from the OTU table by employing the QIIME script *alpha_diversity.py* at same level of survey effort (10,400 16S rRNA gene sequences; step size 100; 10 replicates per sample) (Caporaso *et al.* 2010).

Principal component analysis (PCA) of the biofilm communities was performed with CANOCO 4.56 for Windows (Microcomputer Power, Ithaca, NY, USA). Relative abundances of bacterial phyla, candidate divisions, proteobacterial subclasses, and the artificial group “Other” (summary of all phyla and candidate divisions, which account for less than 0.5% in any given sample) were used. The scaling was focused on inter-taxa correlations and taxa scores were divided by standard deviation. The graph was centered by taxa.

Analysis of the WB3 metatranscriptome

Microbial community RNA from sampling site WB3 was isolated and DNA removal was performed as described above. In addition, mRNA was enriched by using the Ribominus transcriptome isolation kit for bacteria (Invitrogen) following the manufacturer’s protocol with one modification: the denaturation of the RNA was performed at 70°C for 10 min. The enriched mRNA was purified and concentrated by using the RNeasy MinElute Cleanup kit (Qiagen). The success of mRNA enrichment was moni-

tored by using an Agilent 2100 Bioanalyzer (Agilent Technologies, Böblingen, Germany) according to the protocol of the manufacturer.

The purified mRNA was reverse-transcribed to cDNA by using random hexamer primers pd(N)6 (Roche) and the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen) as recommended by the manufacturer. The preparation of the cDNA library for pyrosequencing was conducted as described by the manufacturer (Roche). Sequencing was performed by the Göttingen Genomics Laboratory employing a 454 GS-FLX pyrosequencer (Roche) and Titanium chemistry. This resulted in a total of 175,391 sequences with an average read length of 313 bp.

Analysis of WB3 metatranscriptome dataset

To remove remaining rRNA sequences of the pyrosequencing-derived dataset, we performed BLAST searches against the most recent SILVA SSU and LSU database (Pruesse *et al.* 2007) as described by Stewart *et al.* (Stewart *et al.* 2010). This step is crucial for the following analysis as public databases are known to contain misannotations (Tripp *et al.* 2011). The rRNA-free dataset was submitted to the MG-RAST server (Meyer *et al.* 2008) for a detailed analysis of the expressed genes and their taxonomic affiliation using the default settings of the pipeline.

Bacterial strains and construction of fosmid libraries

Bacterial strains and vectors used in the present study are shown in Supplementary Table S1. *Escherichia coli* strain EPI300-T1R (Epicentre Biotechnologies, Madison, WI, USA) was used as a host for the cloning of metagenomic DNA. In addition, *E. coli* strain TOP10 (Invitrogen) was employed for subcloning of the target genes. The large-insert metagenomic fosmid libraries WB3, WB4, WB5 and WB5tufa (see Supplementary Table S2) were constructed by using the CopyControl Fosmid Library Production kit (Epicentre Biotechnologies) as described by Nacke *et al.* (2011b) resulting in approximately 36,800 library-containing clones (approximately 9,200 per sample), which were arrayed and stored in 96-well microtiter plates.

Activity-based screening for proteolytic enzymes

Escherichia coli strains were routinely grown in Luria-Bertani (LB) medium at 37°C. For activity-based screening, the arrayed library-containing *E. coli* clones were streaked on LB agar plates containing skim milk (2% [wt/vol]) as indicator substrate. To maintain the presence of recombinant fosmids and increase the copy number of the fosmids the indicator agar contained 12.5 mg chloramphenicol L^{-1} and 0.001% arabinose, respectively. Proteolytic activity exhibiting clones were identified by the formation of halos on indicator agar after incubation for 1 to 14 days. Initially, 37 proteolytic clones were detected and controlled for unique fosmid inserts by restriction analysis with BamHI (Fermentas, St. Leon-Rot, Germany) as recommended by the manufacturer.

Sequence analysis of proteolytic activity conferring genes

To avoid time-consuming sequencing of complete fosmid inserts and to enable rapid detection of proteolytic activity conferring genes the target genes were subcloned. The recombinant fosmids from positive clones were isolated, sheared by sonication for 3 s at 30% amplitude, cycle 0.5 (UP200S Sonicator, Dr. Hielscher GmbH). Subsequently, the resulting DNA fragments were separated by agarose gel electrophoresis. Appropriate fragments (1.5 to 3.5 kbp) were excised and purified from gels by using the peqGold gel extraction kit (Peqlab Biotechnologie GmbH). The resulting DNA fragments were ligated into pCR-2.1-TOPO (Invitrogen), and used to transform *E. coli* TOP10 (Invitrogen) as recommended by the manufacturer. The resulting recombinant *E. coli* strains were re-screened on the indicator agar. The recombinant plasmids derived from positive clones were sequenced. The initial prediction of ORFs located on the plasmids-inserts was performed by using the ORF-finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) provided by the National Center for Biotechnology Information (NCBI) and finally annotated with Artemis (Rutherford *et al.* 2000). The results were verified and improved manually by using criteria such as the presence of a ribosome-binding site, GC frameplot analysis, and similarity to known genes. Classification of the identified proteases was performed by BLAST searches implemented in the MEROPS peptidase database (http://merops.sanger.ac.uk) (Rawlings *et al.* 2012). Similarity searches of related protein sequences were performed by employing BLAST against the GenBank database. Domain structures

were analyzed by submitting protein sequences to InterProScan (Quevillon *et al.* 2005). Multiple alignments of deduced protein sequences and construction of a phylogenetic tree were performed by employing MEGA5 (Tamura *et al.* 2011).

Hydrochemistry analyses

Measurements of water chemistry data (e.g., temperature, pH, anion and cation concentrations) at the three sampling sites were performed as described by Shiraishi *et al.* (2008b). A detailed list of all measured values is shown in Table 1.

Nucleotide sequence accession numbers

The sequence reads from the 454 GS-FLX pyrosequencing of 16S rRNA gene sequences and metatranscriptomic cDNA were deposited in the NCBI Sequence Read Archive (SRA) under project accession number SRA061945. The nucleotide sequences of proteolytic activity conferring genes (*pwb1* to *pwb5*) were deposited in GenBank under accession numbers KC589113 to KC589117.

Results and Discussion

General characteristics of samples

The Westerhöfer Bach (Figure 1A and B) originates from the Muschelkalk Group aquifer (Arp *et al.* 2010) showing karst water's typical high partial pressure of carbon dioxide (pCO_2) (Table 1). The pCO_2 rapidly decreased from the spring site (15,136 µatm) by 11-fold to the sampling sites (average of 1,339 µatm). Correspondingly, CO_2 degassing increased the pH from spring site (7.22) to the sampling sites (average 8.21). Saturation index of calcite (SIC) was almost 1 at all three sampling sites, which is approximately 10-fold higher than SIC (0.08) of the spring site. In addition, it is above the known requirement (SIC >0.8) for $CaCO_3$ precipitation (Kempe and Kazmierczak 1994). The creek water is enriched in Ca^{2+} (average 2.94 mmol L^{-1}), Mg^{2+} (average 1.55 mmol L^{-1}), and SO_4^{2-} (average 2.23 mmol L^{-1}). Specifically, the high sulfate content originates from the dissolution of evaporites from the Muschelkalk Group within the catchment area. The conditions with respect to the chemi-

cal properties of the surrounding water were highly similar at all sampling sites. In addition, the calcifying biofilm is an uninterrupted framework along the creek and environmental properties are highly similar between the biofilm sampling sites (Table 1 and Figure 2A).

Sampling and nucleic acid isolation

Biofilm samples at three different sampling sites (WB3, WB4, and WB5) and one tufa sample (acquired below the biofilm at WB5) were recovered in summer (June) 2009 at noon (Figure 1B). The DNA and RNA isolations from the three biofilm samples (Figure 1D and E) yielded approximately 2.5 µg DNA and 2.3 µg RNA per 250 mg sample material. The DNA isolation from the tufa (Figure 1F) resulted in approximately 1.2 µg per 250 mg sample material. RNA isolation from tufa yielded 0.5 µg per 250 mg, which is approximately 4.6-fold lower than the RNA yield of the biofilm. Variations of isolation procedure and amount of starting material did not significantly increase the RNA yield from tufa (data not shown). The low recovery of RNA from tufa might be due to low metabolic activity of embedded microbes persisting in dormant states (Lennon and Jones 2011). Amplification of cDNA from the tufa RNA isolations yielded low cDNA concentrations, which were beyond the detection limit of Qubit fluorometer. All subsequent pyrosequencing efforts from the tufa cDNA failed. This might be a result of co-extracted inhibiting substances of the tufa fabric such as EPS components or different microbialite constituents hindering cDNA generation.

Overview of the bacterial diversity Westerhöfer Bach biofilm and tufa

A total of 250,833 bacterial 16S rRNA gene sequences (Table 2) were acquired by the amplification of the V2 to V3 16S rRNA gene region from isolated DNA and RNA. After quality filtering, denoising, removal of potential chimeras, sequencing artifacts (low alignment coverage OTUs), and chloroplasts 139,062 high quality bacterial 16S rRNA gene sequences were recovered (Table 1). These comprised a total of 4,802 OTUs at a genetic divergence level of 1%. For diversity analysis the sequence amount was trimmed to same surveying effort per sample (10,400 sequences per sample). As depicted by the rarefaction curves and OTU numbers (Figure 3 and Table 2), the observed and the estimated richness (see Table 1) of all three sampling sites increased downstream by approximately 1.3-fold. DNA-based and RNA-based

Table 1. Hydrochemical data of stream waters from Westerhöfer Bach.

Sampling site (Coordinates)	Distance spring area (m)	EC 25°C	pH	T_{pH}	(mmol l^{-1})												Alkalinity (meq l^{-1})	pCO_2 (µatm)	$SI_{Calcite}$[a] (log IAP/KT)
					Ca^{2+}	Mg^{2+}	Na^{2+} *10^{-1}	K^{+} *10^{-2}	Sr^{2+} *10^{-2}	Ba^{2+} *10^{-4}	Si *10^{-1}	B *10^{-3}	Cl^{-} *10^{-1}	SO_4^{2-}	NO_3^{-} *10^{-2}	PO_4^{-} *10^{-4}			
Spring site	3	892	7.22	9.0	3.44	1.54	3.15	4.98	1.55	2.3	1.64	7.1	2.90	2.23	69.8	0.89	5.51	15,136	0.08
WB3 (51°45'44", 10°5'35")	262	809	8.21	12.7	2.98	1.55	3.16	5.02	1.53	2.2	1.67	5.7	2.92	2.23	6.93	1.76	4.69	1,349	0.98
WB4 (51°45'43", 10°5'37")	287	813	8.19	13.0	2.94	1.55	3.15	4.96	1.56	2.2	1.69	5.7	2.92	2.23	6.90	1.67	4.59	1,380	0.95
WB5 (51°45'44", 10°5'38")	313	804	8.22	13.6	2.90	1.55	3.14	4.93	1.57	2.3	1.72	5.3	2.92	2.22	6.92	1.71	4.50	1,288	0.97

[a]Saturation index $SI_{Calcite} = \log \Omega_{Calcite} = \log$ (*ion activity product* $[Ca^{2+}] \times [CO_3^{2-}]$/*solubility product* $K_{Calcite}$)

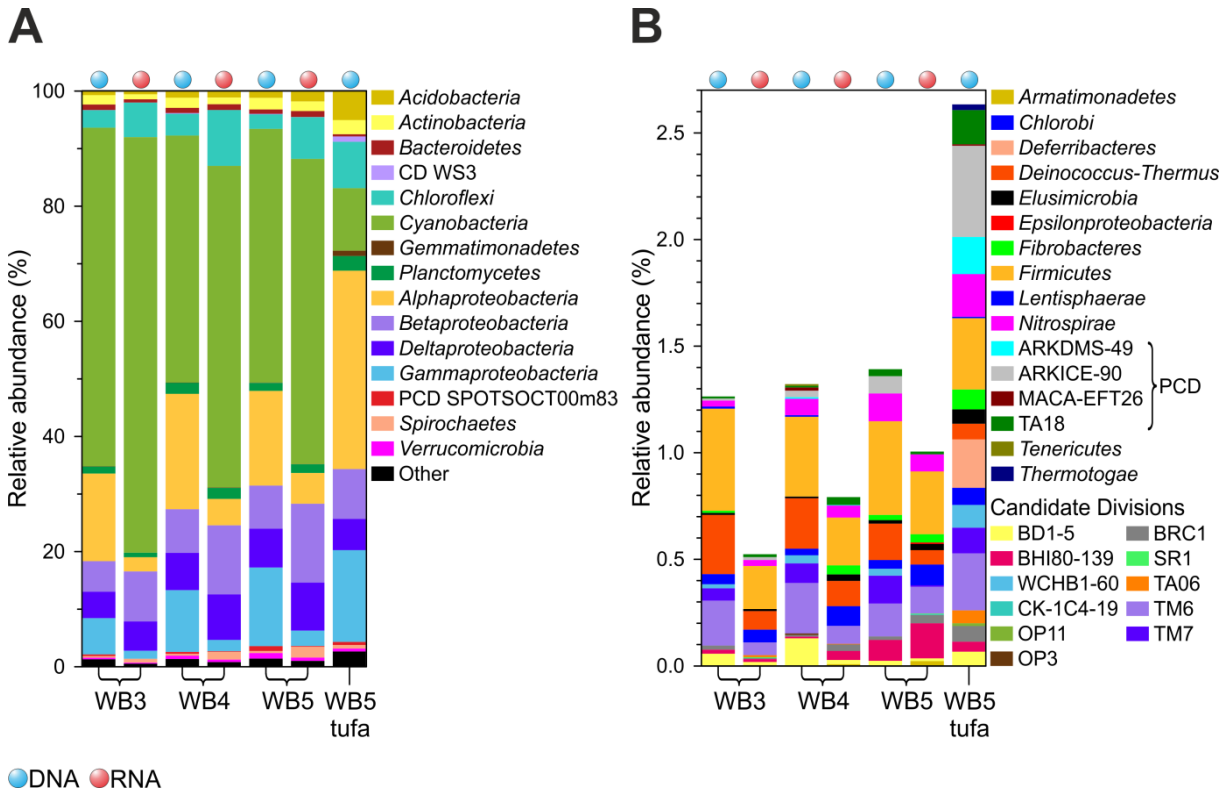


Fig. 2. Bacterial community composition of the Westerhöfer Bach biofilm and tufa samples. Relative abundances of phylogenetic groups detected by 16S rRNA analysis of biofilm samples derived from Westerhöfer Bach (A). The samples are separated by template (DNA, blue circles; RNA, red circles above bar charts). Phylogenetic groups accounting for <0.5% of all classified sequences were summarized in the artificial group "Other". Relative abundances of rare phyla and candidate divisions, which were grouped to the artificial group "Other" (B). Abbreviations: CD, Candidate division; PCD, proteobacterial candidate division.

Table 2. Bacterial 16S rRNA gene diversity analyses of Westerhöfer Bach biofilm and tufa samples at 1% genetic divergence level.

Sample	No. raw reads	No. high quality reads	Average read length	No. OTUs	Chao1	Shannon	Michaelis Menten fit	Coverage (%)
WB3 DNA	19,983	10,443	357	649	1,829	3.57	875	74.1
WB3 RNA	40,404	21,744	362	522	1,137	2.88	718	72.7
WB4 DNA	23,738	13,091	359	774	1,797	4.14	1,009	76.7
WB4 RNA	61,268	35,619	367	685	1,054	3.65	889	77.0
WB5 DNA	23,372	12,285	359	818	2,267	4.31	1,071	76.3
WB5 RNA	51,295	30,920	361	796	1,298	3.86	1,046	76.1
WB5tufa DNA	30,773	14,960	367	1,297	3,016	5.73	1,702	76.2

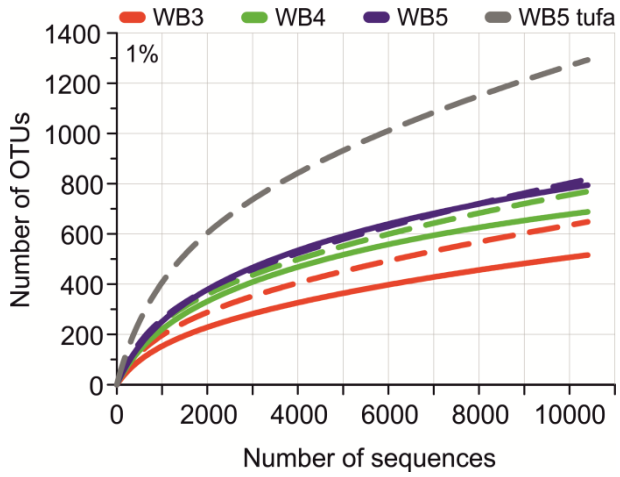


Fig. 3. Rarefaction analysis of entire and active bacterial communities in the biofilm and tufa samples at genetic divergence level of 1%. Curves were generated from 10,400 randomly picked sequences of each sample. DNA-born sequences are indicated by dashed lines, and RNA-derived sequences by solid lines.

estimations of OTUs at 1% genetic diversity resulted in 649 to 818 (DNA) and 522 to 796 (RNA) OTUs for the biofilm (Table 2). Shannon indices for biofilms based on DNA-derived 16S rRNA sequences ranged from 3.57 to 4.31. On DNA level, tufa showed highest bacterial diversity of all samples (Shannon index 5.73). The RNA-derived Shannon indices of the biofilm samples were lower and ranged from 2.88 to 3.86. Comparable biofilms as those from Iberian Range rivers (Spain) displayed similar levels of diversity as measured by Shannon index (ranging from 1.34 to 3.83) (Beraldi-Campesi *et al.* 2012). However, the Shannon indices from the Iberian Range samples were calculated from DGGE band patterns and therefore might miss considerable diversity. The higher bacterial diversity on DNA level compared to RNA level was expected, as not all bacteria of the entire community are active. A higher richness of the entire compared to the active bacterial community was also observed in other studies comparing DNA-derived and RNA-derived bacterial community structures (Baldrian *et al.* 2012; Yoshida *et al.* 2012). At higher taxonomic resolution several differences were apparent, i.e., *Rhodobacterales* (uncultured *Rhodobacteraceae*) dominated DNA-derived datasets whereas *Burkholderiales* (mostly *Comamonadaceae*) dominated RNA-based datasets.

The highest bacterial diversity was observed in the tufa sample (1,297 OTUs). This can be attributed to temporal-conditioned community variations as the tufa sample encompassed three tufa layers (see Figure 1F), which roughly represent a time span of three years. Thus, the analyzed bacterial community reflects time-averaging including seasonal variations of three years. The saturation extent of the rarefaction curves indicated that major parts of the bacterial communities in biofilm samples were covered by the surveying effort (Figure 3). The coverage per sample was predicted by Michaelis Menten fit and ranged from 73-77%.

Comparison of entire and active bacterial biofilm communities

The abundance of the phylogenetic groups detected by 16S rRNA gene analysis derived from DNA and RNA of the three biofilm samples were compared by PCA (Figure 4). We observed differences in the bacterial community composition with respect to template type and location of the sampling sites. Axis PC1 explained the major part of the observed variance (80.9%). The samples were mainly separated by template type and, thus, by entire and active bacterial community. PC2 accounted for nearly one fifth of the explained variance (17%). This axis apparently describes influences of the physical properties of the sampling sites. The biofilm at WB3 had a fluffy to sand-like texture, whereas biofilms at WB4 and WB5 were more solid. Upstream of WB3 a small cascade is located, which may have influence on the biofilm community. The tufa of the cascade was inactive as a biofilm was absent and no active tufa formation was observed (see Shiraishi *et al.* 2008b). Alterations in water velocity and O_2 concentrations could impact the downstream microbial community. Additionally, sluicing of soil-born bacteria from the river shore, followed by sedimentation and adhesion to the biofilm matrix is a possible explanation for the increase of bacterial diversity at the downstream sampling sites (WB4 and WB5).

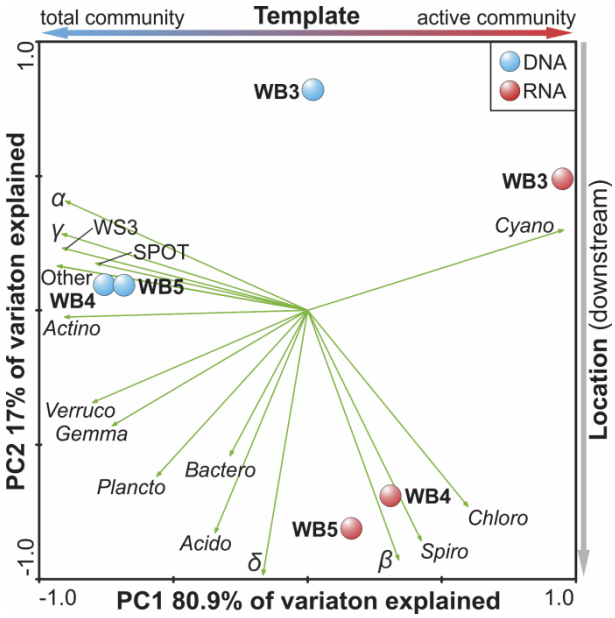


Fig. 4. Comparison of bacterial community composition of biofilm samples using principal component analysis (PCA). PCA is based on relative abundances of bacterial phyla and proteobacterial classes. The biofilm samples are separated by template (DNA, blue circles; RNA, red circles above bar charts). Abbreviations: *Acido*, *Acidobacteria*; *Actino*, *Actinobacteria*; *Bactero*, *Bacteroidetes*; WS3, Candidate division WS3; *Chloro*, *Chloroflexi*; *Cyano*, *Cyanobacteria*; *Gemma*, *Gemmatimonadetes*; *Plancto*, *Planctomycetes*; *Spiro*, *Spirochaetes*; *Verruco*, *Verrucomicrobia*; α, *Alphaproteobacteria*; β, *Betaproteobacteria*; γ, *Gammaproteobacteria*; δ, *Deltaproteobacteria*; SPOT, proteobacterial candidate group SPOT-SOCT00m83 (San Pedro Ocean Time Series October, referring to clone SPOTSOCT00_5m83, DQ009140).

Bacterial community structure of the Westerhöfer Bach biofilm

Taxonomic classification of 16S rRNA gene sequences was based on SILVA taxonomy (Pruesse *et al.* 2007). Plastid originated 16S rRNA gene sequences from the DNA and RNA derived datasets indicated that diatoms and red algae thrive in the biofilm and tufa (Supplementary Figure S1). This is partly in contrast to previous reports that showed that besides diatoms mainly green algae are part of the microbiota of the Westerhöfer Bach (Arp *et al.* 2010). However, as the used primers have not been optimized for detection of chloroplasts, the chloroplast related data might be incomplete. This is further emphasized by the current studies of Hodač *et al.* (2013, this issue) and Brinkmann *et al.* (2013, this issue), who detected a high diversity of green algae and diatoms by 18S rRNA gene analysis, respectively. The tufa chloroplast-affiliated OTUs showed similarities (94-96%) to members of *Rhodophyta* (*Palmaria palmate*), which possess an endolithic lifestyle (Campbell 1980). Chloroplasts exhibited only low metabolic activity as indicated by the low amounts of RNA-derived 16S rRNA gene sequences (Supplementary Figure S1). In accordance with Arp *et al.* (2010), chloroplast 16S rRNA gene sequences of the biofilm showed mainly similarities to chloroplasts from *Nitzschia* and *Porphyridium*.

Most bacterial 16S rRNA gene sequences could be classified as members of the phyla *Cyanobacteria* (42-72%) and *Proteobacteria* (18-45%) from DNA and RNA-derived datasets of all biofilm samples (Figure 2A), whereas tufa showed the opposite relation (11% *Cyanobacteria*, 66% *Proteobacteria*). The dominance of *Cyanobacteria* and *Proteobacteria* was observed previously (Arp *et al.* 2010; Cousin and Stackebrandt 2010) and seems to be a general setup in biofilms of calcified freshwater rivers, as similar community structures were also observed in corresponding habitats (Beraldi-Campesi *et al.* 2012; Heath *et al.* 2010; Ng *et al.* 2006). The main trend of the bacterial biofilm community composition downstream was the increase of members of the *Proteobacteria* accompanied with the decrease of *Cyanobacteria*. In general the DNA-based community assessment revealed a higher abundance of *Proteobacteria* than the RNA-based analysis, indicating that only certain proteobacterial groups encountered in the biofilm were active during the day. Additionally, the biofilm community comprised *Chloroflexi*, *Acidobacteria*, *Actinobacteria*, *Planctomycetes*, *Bacteroidetes*, *Spirochaetes*, and *Verrucomicrobia* in lower abundances. The comparison of active and entire biofilm community revealed that *Cyanobacteria*, certain

proteobacterial subgroups, *Chloroflexi*, and *Spirochaetes* were more abundant in the RNA-derived datasets indicating them as part of the metabolic active bacterial community (Figure 2A).

Filamentous Cyanobacteria are the main inhabitants of Westerhöfer Bach biofilm

The vast majority of DNA-derived and RNA-derived 16S rRNA gene sequences were classified as *Cyanobacteria*, indicating that these organisms are not only major but also highly active constituents of the biofilm communities (Figure 5A). Members of the *Cyanobacteria* were mainly affiliated to *Oscillatoriales*, which are filamentous cyanobacteria. They perform cell division in one plane and are able to glide by rotation of their trichome. Heterocysts and akinetes known from other cyanobacterial groups are absent (Anagnostidis and Komárek 1988). At higher taxonomic resolution (genus level), the *Cyanobacteria* divided into three dominant lineages. The most prominent lineages on DNA and RNA level were affiliated to the *Tychonema*/*Phormidium*/*Microcoleus* clade (TPM), which exhibited an average abundance of approximately 50% of the overall biofilm community. Unfortunately, differentiation of *Cyanobacteria* based on 16S rRNA gene sequences or by morphology is difficult and sometimes erroneous due to high similarities of 16S rRNA genes and variable morphotypes of different species. This is the case for the *Oscillatoriales* (Arp *et al.* 2010; Casamatta *et al.* 2005). However, the TPM clade further divided in two major groups: One (TPM 1) showed highest similarity (99%) to the cultured members *Tychonema* sp. K27 (GQ324965), *Phormidium corium* PMC299.07 (GQ859649), *Phormidium autumnale* SAG 78.79 (EF654084), and *Microcoleus antarcticus* UTCC 474 (98%, AF218373) and the other (TPM 2) to *Oscillatoriales* cyanobacterium WBK15 (99%, GQ324967). Additionally, uncultured members with high similarity (99%) to TPM 1 were detected in Antarctic microbial mats (JQ310416, HQ827417), river periphyton (KC682959), and alpine soil (Himalaya) (HQ189056). Members of TPM 2 were derived from submerged sinkholes (e.g., FJ866619), river sediment (EF667676), Antarctic microbial mats (HQ827250), and New Zealand streams (FJ203933). These data suggest that members of these cyanobacterial lineages are widely distributed, preferentially in colder environments. It is most likely that TPM 1 is closely related to *Tychonema* sp. K27, whereas TPM 2 might represent *Oscillatoriales* cyanobacterium

WBK15, as these strains were originally isolated from Westerhöfer Bach. The predominance of *Cyanobacteria* in biofilms of the Westerhöfer Bach was previously analyzed by classical sequencing of 16S rRNA gene clone libraries (Arp *et al.* 2010; Cousin and Stackebrandt 2010) and microscopic analysis (Cousin and Stackebrandt, 2010). In contrast to our study, Cousin *et al.* (2010) identified a distant relative of *Gloeothece* sp. KO68DGA (93% identity, AB067580), which is an unicellular cyanobacterium, as predominating cyanobacterial group. This might be an indication of seasonal and temporal (sampling of Cousin and Stackebrandt (2010) in May 2006) bacterial community changes of the bacterial biofilm community, which were recently observed for Antarctic cyanobacterial mats (Kleinteich *et al.* 2012). However, Brinkmann *et al.* (2013, this issue) also observed mainly members of the *Oscillatoriales* in the Westerhöfer Bach biofilm and could link them to occurrence in the highly calcified creek sites. *Cyanobacteria* are known to be the primary producers of organic matter by fixing CO_2 via photosynthesis and thereby generating biomass, which serves as nutrition source for other microorganisms, especially in the low-nutrient environment of Westerhöfer Bach. In particular, members of the *Oscillatoriales* are the most abundant inhabitants of calcifying freshwater biofilms (Beraldi-Campesi *et al.* 2012; Bissett *et al.* 2008; Cadel-Six *et al.* 2007; Perri *et al.* 2012; Wood *et al.* 2012). The induction of calcification in Westerhöfer Bach by members of the *Oscillatoriales* was recently described (Shiraishi *et al.* 2008a; Zippel and Neu 2011). Members of *Limnothrix* (99% similarity to *Limnothrix redekei* NIVA-CYA 227/1, AB045929) represented the third most dominant cyanobacterial lineage. This lineage was also found as part of Antarctic cyanobacterial mats. Additionally, similar cyanobacteria were found in freshwater from Lake Ekoln (e.g., HQ386319), and a freshwater biofilm grown under artificial conditions (FJ662718). However, the specific role of this group remains unclear but it may be involved in the calcification process, as it secretes complex xanthan-like EPS (Khattar *et al.* 2010).

Aerobic Proteobacteria are abundant commensals in Westerhöfer Bach biofilm

The second most abundant and also highly diverse phylum of the biofilm community (2,420 OTUs) was represented by the *Proteobacteria*. All detected members of the *Proteobacteria* with respect to known characteristics of related cultured members were aerobic. This is not surprising, as the photosynthetic cyanobacteria produce an

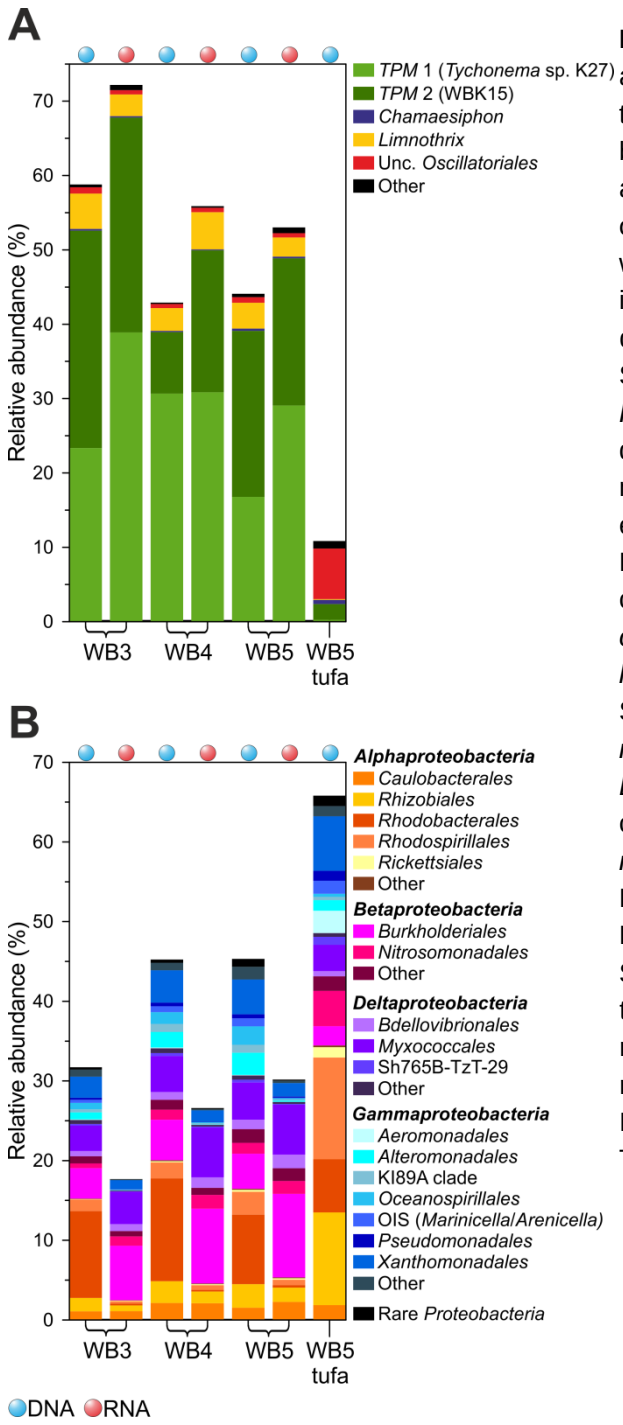


Fig. 5. Relative abundances of *Cyanobacteria* (A) and *Proteobacteria* (B) in biofilm and tufa at high taxonomic resolution. The samples are separated by template (DNA, blue circles; RNA, red circles above bar charts). Relative abundances of the cyanobacterial lineages accounting for ≤0.5% were summarized in the artificial group "Other", including *Microcystis*, *Calothrix*, and several candidate clades (ML635J-21, MLE1-12, SHA-109, SM1D11, and SM2F09). Relative abundances of *Proteobacteria* at order level (B). Orders and candidate groups that accounted for <1% were summarized for each class in the artificial group "Other". These included for the *Alphaproteobacteria* DB1-14, MNG3, OCS116, *Parvulculales*, SAR11 clade, and *Sphingomonadales*; for the *Betaproteobacteria* B1-7BS, Hot Creek 32, *Hydrogenophilales*, *Methylophilales*, *Neisseriales*, *Rhodocyclales*, SC-I-84, TRA3-20, and UCT N117; for the *Deltaproteobacteria* *Desulfarculales*, *Desulfobacterales*, *Desulfuromonadales*, GR-WP33-30, SAR324 clade, and *Syntrophobacterales*; for the *Gammaproteobacteria* *Acidithiobacillales*, *Chromatiales*, EC3, *Enterobacteriales*, *Legionellales*, MD2904-B13, *Methylococcales*, NKB5, *Pasteurellales*, SS1-B-09-64, and *Thiotrichales*. Rare proteobacterial class and candidate divisions were summarized under the artificial group "Rare Proteobacteria" including *Epsilonbacteria*, ARKDMS-49, ARKICE-90, MACA-EFT26, SPOTSOCT00m83, and TA18. Abbreviations Order Incertae Sedis (OIS).

aerobic milieu by constant O_2 production during daytime. We observed several differences of the relative abundances of the proteobacterial orders between DNA-derived and RNA-derived 16S rRNA gene sequences (Figure 5B). *Alphaproteobacteria* and *Gammaproteobacteria* were more dominant in the DNA-derived dataset, whereas *Betaproteobacteria* and *Deltaproteobacteria* showed an opposite trend (see also Figure 4). This indicated that the latter are more active members of the biofilm communi-

ty. Beraldi-Campesi *et al.* (2012) analyzed similar tufa-forming communities of three rivers in Spain. They also observed a high abundance of *Alphaproteobacteria* and *Gammaproteobacteria* within their DNA-derived 16S rRNA gene datasets (Beraldi-Campesi *et al.* 2012).

The *Alphaproteobacteria* within the biofilm encompassed in descending order *Rhodobacterales*, *Rhizobiales*, *Caulobacterales*, *Rhodospirillales*, and *Rickettsiales*. DNA-born datasets were dominated by sequences derived from purple non-sulfur bacteria (PNSB) of the *Rhodobacterales* (mainly uncultured *Rhodobacter*), which showed only relatively low abundances in the RNA-derived datasets (Figure 3B). Members of the genus *Rhodobacter* are known to inhabit freshwater environments and are diverse regarding their metabolism and respiration. They can grow photoheterotrophic in the light, and in the dark chemotrophic by aerobic respiration, denitrification, fermentation, or oxidant-dependent metabolism (Garrity *et al.* 2005a). The *Rhizobiales* were mainly represented by candidate division A0839 of which information on metabolic traits is lacking. In addition, members of *Caulobacterales* such as *Caulobacteraceae* (e.g., *Brevundimonas* and *Phenylobacterium*) on DNA level, and *Hyphomonadaceae* (e.g., *Woodsholea*, *Hirschia*, and *Hyphomonas*) on RNA level were present. *Brevundimonas* and *Phenylobacterium* are chemoorganotrophic and possess a strictly respiratory metabolism (Garrity *et al.* 2005a). Members of the *Caulobacterales* (especially *Brevundimonas*) were recently detected in calcium carbonate precipitating communities in porous calcareous building stones (Jroundi *et al.* 2010). Furthermore, members of the *Rhodospirillales* (*Rhodospirillaceae* and *Acetobacteraceae*) and parasitic/mutualistic *Rickettsiales* (several candidate divisions) were detected in lower abundances.

Members of the *Betaproteobacteria* were in general more abundant in RNA-derived datasets, indicating an active part in the biofilm community. Within the *Betaproteobacteria*, members of the *Burkholderiales* that mainly comprised *Comamonadaceae* dominated the betaproteobacterial 16S rRNA gene sequences. These organisms exhibited a 2-fold higher relative abundance on RNA level than on DNA level. In addition, this diverse class showed a slight increase in abundance downstream the creek. Most members of the *Comamonadaceae* are motile and chemoorganotrophic or facultative chemolithotrophic. These organisms use O_2 as terminal electron acceptor, and are able to degrade complex organic compounds (Garrity *et al.* 2005a).

These properties make them to ideal commensals for cyanobacteria by using their produced O_2 and EPS compounds. Furthermore, *Nitrosomonadales* (mostly uncultured *Nitrosomonadaceae*) that are lithoautotrophic and able to oxidize ammonia (Garrity *et al.* 2005a) were observed in low abundances. Traces of partly known freshwater inhabitants, e.g., *Methylophilales*, *Rhodocyclales*, and several candidate divisions were also detected.

The *Gammaproteobacteria* generally showed higher abundances in DNA-derived 16S rRNA. Especially, the orders *Alteromonadales* (mainly *Haliea* and OM60/NOR5 clade) and *Oceanspirillales* (mainly *Pseudospirillum*) that were often observed in marine environments were detected in the DNA-derived datasets, but were almost absent in RNA-derived datasets. This suggests that these members of the *Gammaproteobacteria* might be less active at day or generally in freshwater settings. However, the *Xanthomonadales* formed the largest phylogenetic group of the *Gammaprobteobacteria* on DNA and RNA level and almost all gammaproteobacterial 16S rRNA sequences belonged to this group. In both cases, the corresponding 16S rRNA gene sequences were mainly affiliated to uncultured members of *Sinobacteraceae* (mainly *Steroidobacter* and JTB255 marine benthic group) and *Xanthomonadaceae* (mainly uncultured *Xanthomonadaceae*, *Arenimonas*, and *Aquimonas*). Members of these families are obligate aerobes with strictly respiratory metabolism employing O_2 as terminal electron acceptor (Garrity *et al.* 2005b). Additionally, members of the *Xanthomonadaceae* are known to produce xanthan, which could be a structural component of the biofilm/tufa fabric (Garrity *et al.* 2005b). Uncultured members of the *Xanthomonadales* were also detected in the river tufa of the Iberian Range (Beraldi-Campesi *et al.* 2012).

Deltaproteobacteria were slightly more abundant (average of approximately 1.4-fold) in RNA-derived datasets. The order *Myxococcales* comprising uncultured *Nannocystineae* and *Sorangiineae*, as well as candidate division (CD) 0319-6G20 and *Cystobacterineae*, were the predominating deltaproteobacterial group and exhibited average relative abundance of 4.4% (DNA-based) and 9% (RNA-based). A slight increased relative abundance downstream the creek was recorded. The *Myxococcales* were highly diverse with 434 OTUs and could also be detected in low abundances in a former study (Cousin and Stackebrandt 2010) and in the biofilms of the Iberian Range rivers (Beraldi-Campesi *et al.* 2012). The OTUs detected within this

order refer to uncultured bacteria of the *Nannocystineae* (*Haliangiaceae*), except one OTU that exhibited high similarity (99%) to *Kofleria* sp. DSM 53797 (HF586693). In general, members of this group use proteolytic and bacteriolytic strategies for growth (Garrity *et al.* 2005a). Due to the low sequence similarities to cultured species (<95%), the here detected members of the *Nannocystineae* might represent novel genera. Similar 16S rRNA gene sequences (99%) have been detected in leaf litter incubated in stream water (HM534303) and in a biofilm formed during continuous supply of water with added carbohydrates (JQ791750). The *Soriangiineae* consisted of uncultured *Polyangiaceae*, *Phaselicystidaceae*, and *Sandaracinaceae*. Additionally, members of *Bdellovibrionales* were detected, which were mainly composed of members of the OM27 clade. Overall, the *Deltaproteobacteria* were mainly represented by aerobic predator bacteria with nutrition depending on degradation of biofilm microorganisms and possibly EPS.

The here encountered richness of *Proteobacteria* demonstrates that benthic tufa ecosystems harbor a high amount of different phylotypes, which due to low nutrient availability all are potentially dependent on primary production by *Cyanobacteria*.

Low abundant bacterial phyla within the biofilm

The third most abundant phylum was *Chloroflexi*. The RNA-derived datasets encompassed more 16S rRNA gene sequences (average of 7.7%) than the DNA-derived sequences (average of 3.1%) (Figure 2A). The predominant OTU (5.2% on RNA level and 0.4% on DNA level) within the *Chloroflexi* belonged to candidate group vadinBA26. Surprisingly, the most similar 16S rRNA gene sequence corresponded to an uncultured bacterium found in a hot spring (EF470476, 97% identity), suggesting a wider temperature tolerance of members from this taxa. The low similarity of this OTU to cultured bacteria (86% to Bacterium Kaz4, AB491168) indicates a previously unrecognized subgroup within the *Chloroflexi*, which is active within the Westerhöfer Bach biofilm community. In addition, members of the *Anaerolineales*, which showed highest similarities to 16S rRNA gene sequences derived from aquatic moss pillars of a freshwater lake (99%, AB630549) and the former survey of Westerhöfer Bach (99%, FM175331), were detected. However, members of the *Anaerolineales* are known to be anaerobic, non-phototrophic, and thermophilic (Yamada *et al.* 2006). As

the here detected 16S rRNA gene sequences showed low similarities to cultured representatives, the detected taxa might represent novel subgroups. In summary, the *Chloroflexi* related 16S rRNA gene sequences showed generally low similarities to those of cultured representatives. Thus, solid predictions on their functions in this ecosystem cannot be drawn. *Actinobacteria*, known from soil and freshwater habitats, were present in low abundance (average 1.5%) but exhibited a high diversity and comprised mainly *Micrococcales*, *Frankiales*, *Rubrobacterales*, *Propionibacteriales*, and *Acidimicrobiales*. Also the *Planctomycetes*, which were mainly composed of uncultured members of candidate group OM190 (OM, Ocean Margins) and *Phycisphaerae* showed low abundance on DNA and RNA level (average 1.5%). Members of OM190 were first detected in marine coastal picoplankton at 10 m depth (Rappé *et al.* 1997) and later in marine environments and soil, but due to lack of cultured representatives their metabolism remains unclear (Bengtsson and Ovreas 2010). Additional phyla identified were in descending abundance across all biofilm samples *Acidobacteria* (*Acidobacteria* and *Holophagae*), *Bacteroidetes* (*Sphingobacteria* and *Cytophagia*) with only negligible amounts of *Flavobacteria*, which thrive in the creek waters (see Cousin 2009), *Spirochaetes* (*Leptospiraceae*), and *Verrucomicrobia* (*Verrucomicrobiaceae*) (see Figure 2A for phyla accounting for <0.5%).

DNA-based assessment of the bacterial community in tufa

The bacterial community in the tufa sample of WB5 showed several differences compared to the DNA-derived biofilm community of WB5. Major differences were that only 11% of the sequences accounted to *Cyanobacteria* and 66% belonged to *Proteobacteria*. A previous study by Cousin *et al.* (2010) showed that the relative abundance of *Proteobacteria* increased in the deeper tufa layers and only the topmost tufa layer contained significant amounts of *Cyanobacteria*. Thus, we assume that the *Cyanobacteria* detected here also derive from the topmost of the three analyzed tufa layers. The predominance of proteobacterial groups might be accounted to their widely heterotrophic lifestyles enabling chasmolithic life within the tufa. Further, the absence of light within the deeper tufa might additionally push the community into direction of chemolithotrophic bacteria after degradation of EPS remains. In general, the distribution of the proteobacterial community was similar to that recorded during DNA-based assessment of the biofilm community composition (Figure 3B). Interest-

ingly, members of the *Rhodospirillales*, showed a more than 6-fold higher relative abundance in the tufa. Additionally, the tufa harbored high abundances of *Rhizobiales* (mainly CD A0839, CD MNG7, and *Hyphomicrobiaceae*). This might indicate an advantage of a chemotrophic growth mode in micro-oxic dark tufa environments. The biofilm predominating cyanobacterial groups were poorly represented by TPM 2 (Figure 5A). Additionally, unclassified *Oscillatoriales* were found in higher abundance in the tufa layer than in the biofilm. The genus *Pleurocapsa* (1.1%) was detected in the tufa but was absent in the biofilm samples, suggesting that optimal growth conditions were not given at time of sampling. Other phyla detected in tufa comprised *Chloroflexi* (8%), *Acidobacteria* (5.1%), *Planctomycetes* (2.5%), *Actinobacteria* (2.4%), CD WS3 (0.9%), *Gemmatimonadetes* (0.9%), and *Spirochaetes* (0.6%). The *Chloroflexi* showed a similar composition as in biofilm of WB5, most abundant classes were *Anaerolineae*, CD KD4-96, *Caldilineae*, CD TK10, and CD S085. Interestingly, the distantly related members of the vadinBA26 clade detected in the biofilm samples were almost absent in the light-depleted tufa sample. This might indicate a phototrophic lifestyle of the detected group members. However, the activity of all via DNA-based 16S rRNA gene analysis detected lineages in tufa remains unclear, as RNA isolation indicated general low activity of the tufa inhabiting bacteria (see above).

Metatranscriptomic analysis of WB3 biofilm

To gain insights into the in situ metabolic functions of the biofilm community during daytime, a small-scale metatranscriptomic analysis of biofilm sample WB3 was performed. Filtering of the 175,391 reads resulted in 8,978 non-rRNA sequences with an average read length of 235 bp which have been subjected for further analyses.

The mRNA sequences have been taxonomically classified by employing a taxonomic binning approach provided by MG-RAST (Meyer *et al.* 2008). The majority of the sequences belonged to *Bacteria* followed by *Eukaryota*. In addition, low amounts of archaeal and virus-related sequences were detected (Figure 6A). In direct comparison, the bacterial community composition showed several similarities on phylum level to the RNA-based 16S rRNA gene dataset of the same sample (Figure 6B). *Cyanobacteria* and *Proteobacteria* were the most abundant phylogenetic groups. Thus, metatranscriptome analysis illustrated on both investigated levels (metatranscriptomic RNA and 16S rRNA) the high activity of *Cyanobacteria* and *Pro-*

teobacteria in the biofilm. The taxonomic assignment of eukaryotic reads suggested that several different eukaryotic taxa populate the biofilm (data not shown). Classification of eukaryotic reads as *Arthropoda*, *Ascomycota*, *Chlorophyta*, and *Rotifera* seemed reasonable for the biofilm but other classifications were highly doubtful. For example, classification indicated the presence of *Chordata* affiliated to *Actinopterygii* and *Mammalia*, which hardly thrive in the biofilms of Westerhöfer Bach. Due to missing genomic reference data for a wide range of eukaryotic groups and potential high similarities of eukaryotic genes these assignments have to be interpreted with caution.

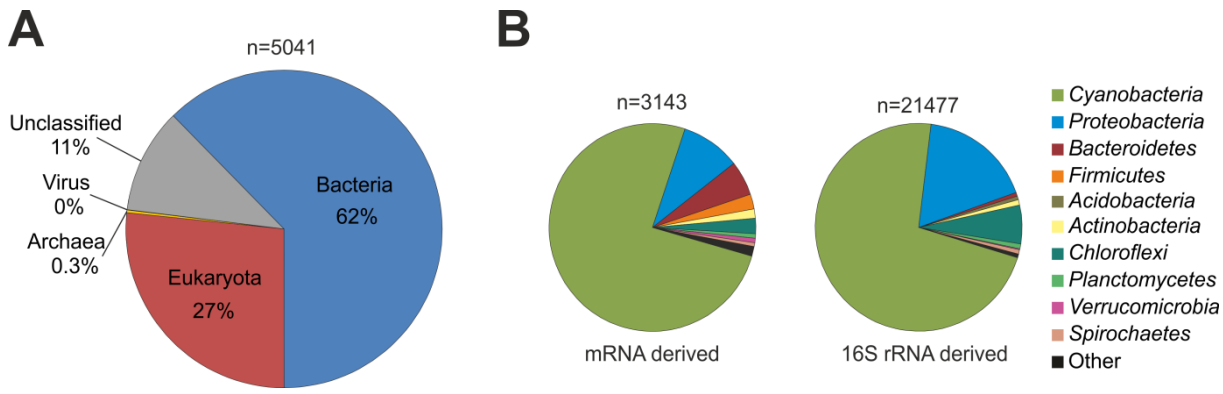


Fig. 6. Taxonomic classification of metatranscriptomic mRNA sequences of sampling site WB3. Annotation of mRNA sequences was performed using MG-RAST (annotation source M5NR) (Meyer *et al.*, 2008). Taxonomy is shown on domain level (A) and on phylum level for bacteria in comparison to 16S rRNA gene sequences derived from WB3 (B). Bacterial phyla with a relative abundance <0.5% were summarized under the artificial group "Other" and included *Deinococcus-Thermus*, *Acidobacteria*, *Chlamydiae*, *Thermotogae*, *Deferribacteres*, *Nitrospirae*, *Aquificae*, *Lenthisphaerae*, and unclassified phyla (B).

Functional analysis was based on approximately non-rRNA sequences, which could be functionally classified by MG-RAST's subsystem technology. Most of the expressed genes were assigned to photosynthesis, protein metabolism, clustering-based subsystems (genes of which function is yet not clearly known), carbohydrate metabolism, miscellaneous (mainly plant-prokaryote DOE project), RNA metabolism, and respiration. Photosynthetic genes comprised different genes assigned to light-harvesting complexes including components of the phycobilisome and electron transport such as components of photosystems I and II. Also small amounts of rhodopsin-related genes were detected. This also reflects the results of the 16S rRNA gene analysis, as only *Cyanobacteria*, plants and algae contain photosystem II. Ex-

pressed genes affiliated to protein and carbohydrate metabolism were mainly genes encoding proteins responsible for basic metabolic cell machinery such as genes for biosynthesis of proteins, protein degradation and folding, and central carbohydrate metabolism. In accordance with the presence of genes encoding components of photosystems and photosynthetic bacteria, expressed genes involved in CO_2 fixation via the Calvin Benson cycle were detected. However, the mRNA data of this small-scale study represents only a glimpse into the expressed genes of the highly diverse biofilm community.

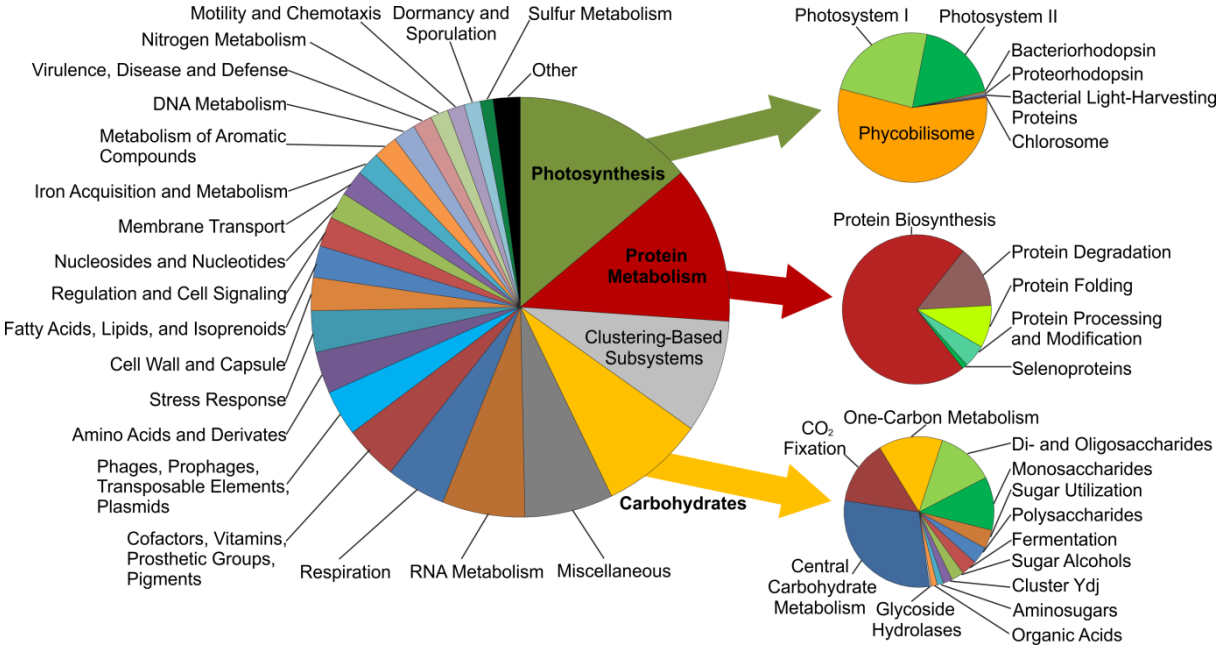


Fig. 7. Metabolic functions detected in the biofilm community at sampling site WB3. Annotation of classifiable protein sequences was performed by MG-RAST's subsystem technology with default settings (Meyer *et al.*, 2008). Low abundant groups (<1%) were summarized under the artificial group "Other" including cell division and cell cycle, phosphorus metabolism, secondary metabolism, and potassium metabolism.

Screening for novel proteases

The functional analysis revealed that protein metabolism including protein-degradation is of importance in the biofilm community. To gain insights into the latter group of enzymes we constructed four large-insert fosmid libraries from biofilm and tufa samples (Supplementary Table S2) and screened the generated libraries for genes conferring proteolytic activity. The average insert sizes of the fosmid libraries ranged from 25.7 to 30.1 kbp. Thus, approximately 1 Gbp of biofilm and tufa DNA

has been cloned. The function-driven screen for genes conferring proteolytic activity was based on the ability of the library-containing *E. coli* clones to form halos when grown on indicator agar medium containing skim milk. Halo formation is caused by hydrolysis of the milk proteins. The screening of approximately 36,800 clones yielded 37 positive clones conferring a stable proteolytic phenotype. To identify the responsible genes on the large-fosmid insert, fosmids were digested. Subsequently, the resulting fragments were subcloned into a plasmid vector and screened for proteolytic activity. In this way, five unique proteolytic activity-conferring genes (*pwb1* to *pwb5*) were identified. The deduced protein sequences revealed 50 to 87% amino acid identity to amino acid sequences from known proteases (Table 3). Classification of

Table 3. Characteristics of protease-encoding genes *pwb1* to *pwb5* and the corresponding gene products.

Gene name	Size (aa)	Family (MEROPS)	Closest similar protein, accession no. of similar protein (no. of encoded amino acids)	Organism	Alignment length/total no. aa query (% identity)
pwb1	321	S1B	V8-like Glu-specific endoprotease, ZP_18909438 (319)	*Leptolyngbya* sp. PCC 7375	201/319 (63)
pwb2	381	S1C	Hypothetical protein, XP_002992168 (413)	*Selaginella moellendorffii*	187/372 (50)
pwb3	425	S8A	Subtilisin, YP_007113547 (425)	*Oscillatoria nigro-viridis* PCC 7112	367/426 (86)
pwb4	424	S8A	Subtilisin, YP_007113547 (425)	*Oscillatoria nigro-viridis* PCC 7112	369/425 (87)
pwb5	440	S8A	Subtilisin-like serine protease, EKV00144 (426)	*Leptolyngbya* sp. PCC 7375	294/436 (67)

the deduced amino acid sequences according to the MEROPS protease database (Rawlings *et al.* 2012) revealed that the enzymes belong to S1 (chymotrypsin family) and S8 (subtilisin family) of the serine proteases. Proteases of family S1 are endoproteases (Rawlings *et al.* 2012), which are often characterized by the presence of an N-terminal signal peptide and a propeptide. Interestingly, Pwb1 belongs to subfamily S1B and the corresponding gene encodes no signal peptide, which has been also observed for a protease (CBM43238) detected in a study focused on screening of proteases from different environmental samples of Germany (Niehaus *et al.* 2011). Pwb2 belongs to subfamily S1C and is a putative periplasmic enzyme with a signal

peptide showing highest identity to a hypothetical protein of eukaryotic origin. Pwb3 and Pwb5 grouped into family subfamily S8A, which is represented by the endoprotease subtilisin Carlsberg.

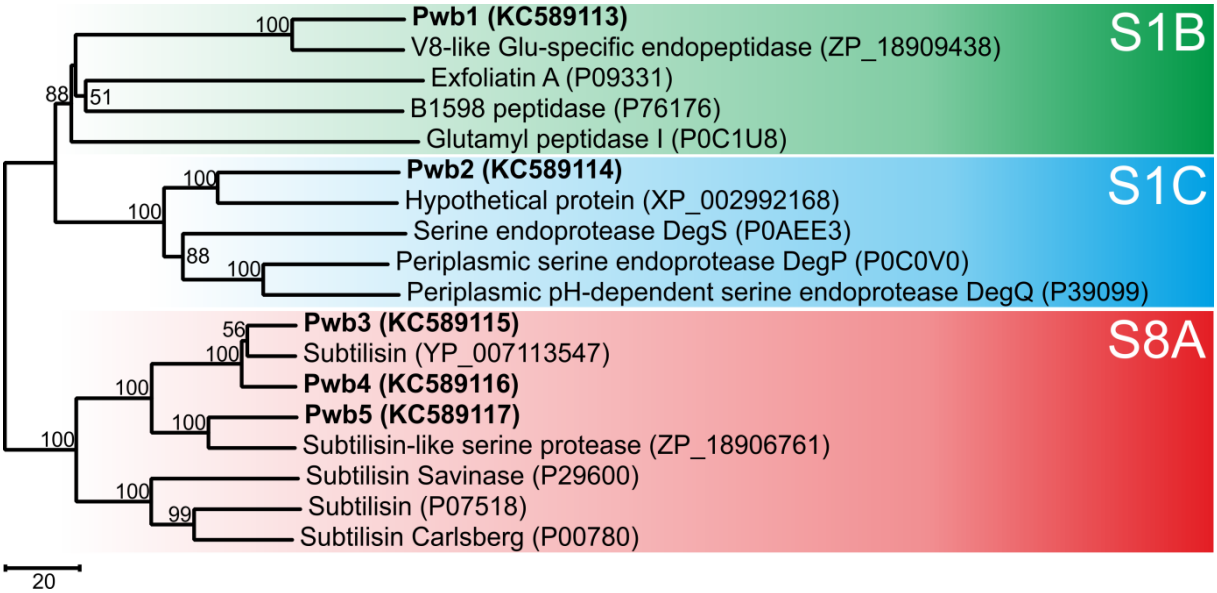


Fig. 8. Unrooted neighbor-joining tree of proteolytic enzymes obtained from karst water biofilm and tufa. The tree includes representative members of serine protease families S1B, S1C, and S8A retrieved from MEROPS database (Rawlings *et al.*, 2012). Amino acid sequences of published proteases were retrieved from GenBank and accession numbers are given in brackets. The length of branches indicates the number of amino acid substitutions per site. Evolutionary analyses were conducted using MEGA5 (Tamura *et al.*, 2011). Bootstrap values were calculated from 1,000 resamplings.

All deduced amino acid sequences, except Pwb2 exhibited highest similarities to cyanobacterial proteases (Figure 8). Pwb2 showed highest identity (50%) to a hypothetical protein of the fern *Selaginella moellendorffii*. Due to the low similarity of Pwb2 to the related database entry it is likely that this gene originated from cyanobacteria or chloroplasts. The cyanobacterial-related proteases (Pwb1, Pwb3, Pwb4 and Pwb5) were affiliated to filamentous *Oscillatoriales* (*Leptolynbya* sp. PCC 7375 and *Oscillatoria nigro-viridis*) and are probably derived from members of the high abundant TPM clade detected by 16S rRNA gene analysis of the biofilm (Figure 5A).

Conclusions

This study is the first report on richness and diversity of the entire and active bacterial communities of an actively growing tufa stromatolite assessed by employing deep sequencing-based metagenomic and metatranscriptomic approaches. The calcifying biofilm samples and the tufa stromatolite sample of the Westerhöfer are inhabited by a highly diverse bacterial community. The up to 320 µm thick tufa-forming bacterial biofilm comprises phototrophs (i.e., *Cyanobacteria*), chemotrophs (i.e., *Comamonadaceae*), bacterial predators (i.e., *Myxococcales*, *Bdellovibrionales*), sulfide-oxidizers (i.e., *Rhodobacteraceae*), sulfate-reducers (i.e., *Xanthomonadales*), and nitrate-reducers (i.e., *Oceanospiraceae*), suggesting the existence of anaerobic micro-niches. The bacterial diversity increased downstream the creek. This is likely due to a higher maturation status of the biofilm matrix in the lower creek parts and continuous sluicing of soil-born bacteria. Several differences in relative abundances of the proteobacterial orders between entire and active biofilm community were observed, indicating temporal shifts in the community. The bacterial 16S rRNA gene analysis showed that the Westerhöfer Bach biofilm is dominated by oxygenic phototrophs (*Oscillatoriales*) and a wide range of diverse (partial facultative) aerobic heterotrophs (*Proteobacteria*, *Chloroflexi*, *Actinobacteria*, *Planctomycetes*, *Acidobacteria*, and *Spirochaetes*), which are known from freshwater tufa environments. The bacterial community of the tufa sample roughly resembles the biofilm community with a shift to the heterotrophic organisms, probably due to the absence of upward moving *Cyanobacteria*. This is caused by poorer light availability in the tufa. In addition, the differences are due to a large time span covered by the tufa (3 years), which includes seasonal changes and changing environmental conditions. However, as we were only able to analyze the bacterial tufa community on DNA level, conclusions on active bacteria within the tufa cannot be drawn.

In spite of the overall highly similar composition of the microbial biofilm communities on phylum level, the fact that several active bacterial groups exhibited low abundance in the DNA-derived 16S rRNA gene analysis demonstrates the limits of DNA-based 16S rRNA gene surveys. DNA-based 16S rRNA gene studies might miss active microbial members derived from low-abundant phylogenetic groups or might lead to false conclusions concerning the activity of inherent microbial community

members. The observed variances between both approaches could derive from differences in metabolic states, dormancy states, genome size, 16S rRNA operon number (Lee *et al.* 2009; Rastogi *et al.* 2009), or presence of DNA from dead cells, which adhere at structural components of biofilms (Bayles 2007).

The small-scale metatranscriptomic analysis of expressed genes and screening for novel proteolytic enzymes supports the results of the 16S rRNA gene analysis and led to further insights into the functions of biofilm-inhabiting microbes. In accordance with the high abundance of active cyanobacteria and other phototrophic organisms, genes involved in photosynthesis and carbohydrate metabolism (e.g., CO_2 fixation) were expressed in the biofilm. The taxonomic affiliation of the bacterial mRNA-derived sequences was roughly in agreement with the 16S rRNA gene-based assessment of community composition. Finally, we detected five novel proteolytic enzymes belonging to three serine protease families; four of these were related to genes originated from cyanobacteria of the *Oscillatoriales*, the dominating cyanobacterial order in the biofilm community. Thus, function-driven analyses of the biofilm community complemented the results of 16S rRNA gene-based assessment of the bacterial biofilm community.

Acknowledgments

We would like to thank Bernd Wemheuer and Sascha Dietrich for maintaining and expanding our 16S rRNA gene analysis scripts. This study has been supported by a grant from the Deutsche Forschungsgemeinschaft (DA 374/5-2, DFG-FOR 571).

References

Ali Z, Cousin S, Fruhling A, Brambilla E, Schumann P, Yang Y et al. 2009. Flavobacterium rivuli sp. nov., Flavobacterium subsaxonicum sp. nov., Flavobacterium swingsii sp. nov. and Flavobacterium reichenbachii sp. nov., isolated from a hard water rivulet. International journal of systematic and evolutionary microbiology 59: 2610-2617.

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

Anagnostidis K, Komárek J. 1988. Modern approach to the classification system of cyanophytes. 3 - Oscillatoriales. Algological Studies/Archiv für Hydrobiologie 50-53: 32-472.

Arp G, Bissett A, Brinkmann N, Cousin S, de Beer D, Friedl T et al. 2010. Tufa-forming biofilms of German karstwater streams: microorganisms, exopolymers, hydrochemistry and calcification. The Geological Society of London 336: 83-118.

Baldrian P, Kolarik M, Stursova M, Kopecky J, Valaskova V, Vetrovsky T et al. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. Isme J 6: 248-258.

Bayles KW. 2007. The biological role of death and lysis in biofilm development. Nat Rev Microbiol 5: 721-726.

Bengtsson MM, Ovreas L. 2010. Planctomycetes dominate biofilms on surfaces of the kelp Laminaria hyperborea. BMC Microbiol 10: 261.

Beraldi-Campesi H, Arenas-Abad C, Garcia-Pichel F, Arellano-Aguilar O, Auque L, Vazquez-Urbez M et al. 2012. Benthic bacterial diversity from freshwater tufas of the Iberian Range (Spain). FEMS microbiology ecology 80: 363-379.

Bissett A, Reimer A, de Beer D, Shiraishi F, Arp G. 2008. Metabolic microenvironmental control by photosynthetic biofilms under changing macroenvironmental temperature and pH conditions. Appl Environ Microbiol 74: 6306-6312.

Brinkmann N, Hodač L, Mohr KI, Hodačová A, Jahn R, Ramm J et al. 2013, this issue. Cyanobacteria and diatoms in biofilms of two karstic streams in Germany and changes of their communities along calcite saturation gradients. Geomicrobiology Journal.

Burger O. 1911. Über schwäbische Kalktuffe insbesondere des Echaztales, Universität Tübingen, Tübingen.

Cadel-Six S, Peyraud-Thomas C, Brient L, de Marsac NT, Rippka R, Mejean A. 2007. Different genotypes of anatoxin-producing cyanobacteria coexist in the Tarn River, France. Applied and environmental microbiology 73: 7605-7614.

Campbell SE. 1980. Palaeoconchocelis starmachii, a carbonate boring microfossil from the Upper Silurian of Poland (425 million years old): implications for the evolution of the Bangiaceae (Rhodophyta). Phycologia 19: 25-36.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7: 335-336.

Casamatta DA, Johansen JR, Vis ML, Broadwater ST. 2005. Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cyanobacteria). Journal of Phycology 41: 421-438.

Cousin S, Pauker O, Stackebrandt E. 2007. Flavobacterium aquidurense sp. nov. and Flavobacterium hercynium sp. nov., from a hard-water creek. International journal of systematic and evolutionary microbiology 57: 243-249.

Cousin S. 2009. Flavobacterial community structure in a hardwater rivulet and adjacent forest soil, Harz Mountain, Germany. Curr Microbiol 58: 409-415.

Cousin S, Stackebrandt E. 2010. Spatial Bacterial Diversity in a Recent Freshwater Tufa Deposit. Geomicrobiology Journal 27: 275-291.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069-5072.

Dürr S, Thomason J. 2010. Biofouling, 1st edn. Wiley-Blackwell: Chichester, West Sussex England ; Ames, Iowa.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194-2200.

Garrity GM, Brenner DJ, Krieg NR, Staley JT. 2005a. Part C The Alpha-, Beta-, Delta-, and Epsilonproteobacteria, vol. 2. Springer: New York.

Garrity GM, Brenner DJ, Krieg NR, Staley JT. 2005b. Part B The Gammaproteobacteria, vol. 2. Springer: New York.

Gonzalez-Munoz MT, Rodriguez-Navarro C, Martinez-Ruiz F, Arias JM, Merroun ML, Rodriguez-Gallego M. 2010. Bacterial biomineralization: new insights from Myxococcus-induced mineral precipitation. Geological Society, London 336: 31-50.

Heath MW, Wood SA, Ryan KG. 2010. Polyphasic assessment of fresh-water benthic mat-forming cyanobacteria isolated from New Zealand. FEMS microbiology ecology 73: 95-109.

Hodač L, Brinkmann N, Mohr KI, Arp G, Hallmann C, Ramm J et al. 2013, this issue. Diversity of microscopic green algae (Chlorophyta) in calcifying biofilms of two karstic streams in Germany. Geomicrobiology Journal.

Jroundi F, Fernandez-Vivas A, Rodriguez-Navarro C, Bedmar EJ, Gonzalez-Munoz MT. 2010. Bioconservation of deteriorated monumental calcarenite stone and identification of bacteria with carbonatogenic activity. Microb Ecol 60: 39-54.

Kempe S, Kazmierczak J. 1994. The role of alkalinity in the evolution of ocean chemistry, organization of living systems, and biocalcification processes. Bull. Inst. Oceanogr.: Monaco. pp 61-117.

Khattar JI, Singh DP, Jindal N, Kaur N, Singh Y, Rahi P et al. 2010. Isolation and characterization of exopolysaccharides produced by the cyanobacterium Limnothrix redekei PUP-CCC 116. Appl Biochem Biotechnol 162: 1327-1338.

Klähn H. 1923. Die Petrogenese der Kalktuffe nebst einigen sich daraus ergebenden geologischen Problemen. Geologisches Archiv. pp 298–316.

Kleinteich J, Wood SA, Küpper FC, Camacho A, Quesada A, Frickey T et al. 2012. Temperature related changes in polar cyanobacterial mat diversity and toxin production. Nature Climate Change 2: 356-360.

La Touche THD. 1913. Geology of the Northern Shan States, vol. 39. Memoirs of the Geological Survey of India.

Lee ZM, Bussema C, 3rd, Schmidt TM. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic acids research 37: D489-493.

Lennon JT, Jones SE. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9: 119-130.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17: 10-12.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M et al. 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9: 386.

Muurholm S, Cousin S, Pauker O, Brambilla E, Stackebrandt E. 2007. Pedobacter duraquae sp. nov., Pedobacter westerhofensis sp. nov., Pedobacter metabolipauper sp. nov., Pedobacter hartonius sp. nov. and Pedobacter steynii sp. nov., isolated from a hard-water rivulet. International journal of systematic and evolutionary microbiology 57: 2221-2227.

Nacke H, Thurmer A, Wollherr A, Will C, Hodač L, Herold N et al. 2011a. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. PLoS One 6: e17000.

Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. 2011b. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. FEMS Microbiol Ecol 78: 188-201.

Ng CC, Huang WC, Chang CC, Tzeng WS, Chen TW, Liu YS et al. 2006. Tufa microbial diversity revealed by 16S rRNA cloning in Taroko National Park, Taiwan. Soil Biology & Biochemistry 38: 342-348.

Niehaus F, Gabor E, Wieland S, Siegert P, Maurer KH, Eck J. 2011. Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. Microb Biotechnol 4: 767-776.

Pedley M. 2013. The morphology and function of thrombolitic calcite precipitating biofilms: A universal model derived from freshwater mesocosm experiments. Sedimentology accepted article.

Pentecost A, Bauld J. 1988. Nucleation of calcite on the sheaths of cyanobacteria using a simple diffusion cell. Geomicrobiology Journal 6: 129-135.

Perri E, Manzo E, Tucker ME. 2012. Multi-scale study of the role of the biofilm in the formation of minerals and fabrics in calcareous tufa. Sedimentary Geology 263-264: 16-29.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J et al. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35: 7188-7196.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R et al. 2005. InterProScan: protein domains identifier. Nucleic Acids Research 33: W116-120.

Rappé MS, Kemp PF, Giovannoni SJ. 1997. Phylogenetic Diversity of Marine Coastal Picoplankton 16S rRNA Genes Cloned from the Continental Shelf Off Cape Hatteras, North Carolina. Limnology and Oceanography 42: 811-826.

Rastogi R, Wu M, Dasgupta I, Fox GE. 2009. Visualization of ribosomal RNA operon copy number distribution. BMC Microbiol 9: 208.

Rawlings ND, Barrett AJ, Bateman A. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 40: D343-350.

Reeder J, Knight R. 2010. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods 7: 668-669.

Rusznyak A, Akob DM, Nietzsche S, Eusterhues K, Totsche KU, Neu TR et al. 2012. Calcite biomineralization by bacterial isolates from the recently discovered pristine karstic herrenberg cave. Applied and environmental microbiology 78: 1157-1167.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16: 944-945.

Schmalenberger A, Schwieger F, Tebbe CC. 2001. Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. Appl Environ Microbiol 67: 3557-3563.

Schneider D, Arp G, Reimer A, Reitner J, Daniel R. 2013. Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the Kiritimati atoll, central pacific. PLoS One 8: e66662.

Schürmann E. 1918. Die chemisch-geologischen Vorgänge bei der Bildung des Uracher Wasserfalls. In: Württemberg JdVfvNi (ed): Württemberg. pp 58–68.

Shiraishi F, Bissett A, de Beer D, Reimer A, Arp G. 2008a. Photosynthesis, Respiration and Exopolymer Calcium-Binding in Biofilm Calcification (Westerhöfer and Deinschwanger Creek, Germany). Geomicrobiology Journal 25: 83-94.

Shiraishi F, Reimer A, Bissett A, de Beer D, Arp G. 2008b. Microbial effects on biofilm calcification, ambient water chemistry and stable isotope records in a highly supersaturated

setting (Westerhöfer Bach, Germany). Palaeogeography, Palaeoclimatology, Palaeoecology 262: 91-106.

Shiraishi F, Okumura T, Takahashi Y, Kano A. 2010. Influence of microbial photosynthesis on tufa stromatolite formation and ambient water chemistry, SW Japan. Geochimica et Cosmochimica Acta 74: 5289–5304.

Stewart FJ, Ottesen EA, DeLong EF. 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. Isme J 4: 896-907.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution 28: 2731-2739.

Tripp HJ, Hewson I, Boyarsky S, Stuart JM, Zehr JP. 2011. Misannotations of rRNA can now generate 90% false positive protein matches in metatranscriptomic studies. Nucleic Acids Research 39: 8792-8802.

Unger F: Beiträge zur (Anatomie und) Physiologie der Pflanzen. VII. Über den anatomischen Bau des Moosstammes. VIII. Über die kalkausscheidenden Organe der Saxifraga enistala. Sitzungsberichte der Akademie der Wissenschaften Wien; Akademie der Wissenschaften Wien. 1861.

von Buch L. 1809. Geognostische Beobachtungen auf Reisen durch Deutschland und Italien, vol. Band II. Haude und Spener: Berlin.

Wood SA, Smith FM, Heath MW, Palfroy T, Gaw S, Young RG et al. 2012. Within-mat variability in anatoxin-a and homoanatoxin-a production among benthic Phormidium (cyanobacteria) strains. Toxins (Basel) 4: 900-912.

Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H et al. 2006. Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. nov., sp. nov. and Leptolinea tardivitalis gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. International journal of systematic and evolutionary microbiology 56: 1331-1340.

Yoshida M, Ishii S, Fujii D, Otsuka S, Senoo K. 2012. Identification of active denitrifiers in rice paddy soil by DNA- and RNA-based analyses. Microbes Environ 27: 456-461.

Zippel B, Neu TR. 2011. Characterization of glycoconjugates of extracellular polymeric substances in tufa-associated biofilms by using fluorescence lectin-binding analysis. Appl Environ Microbiol 77: 505-516.

3.1 Supplemental information for chapter B3

Contents

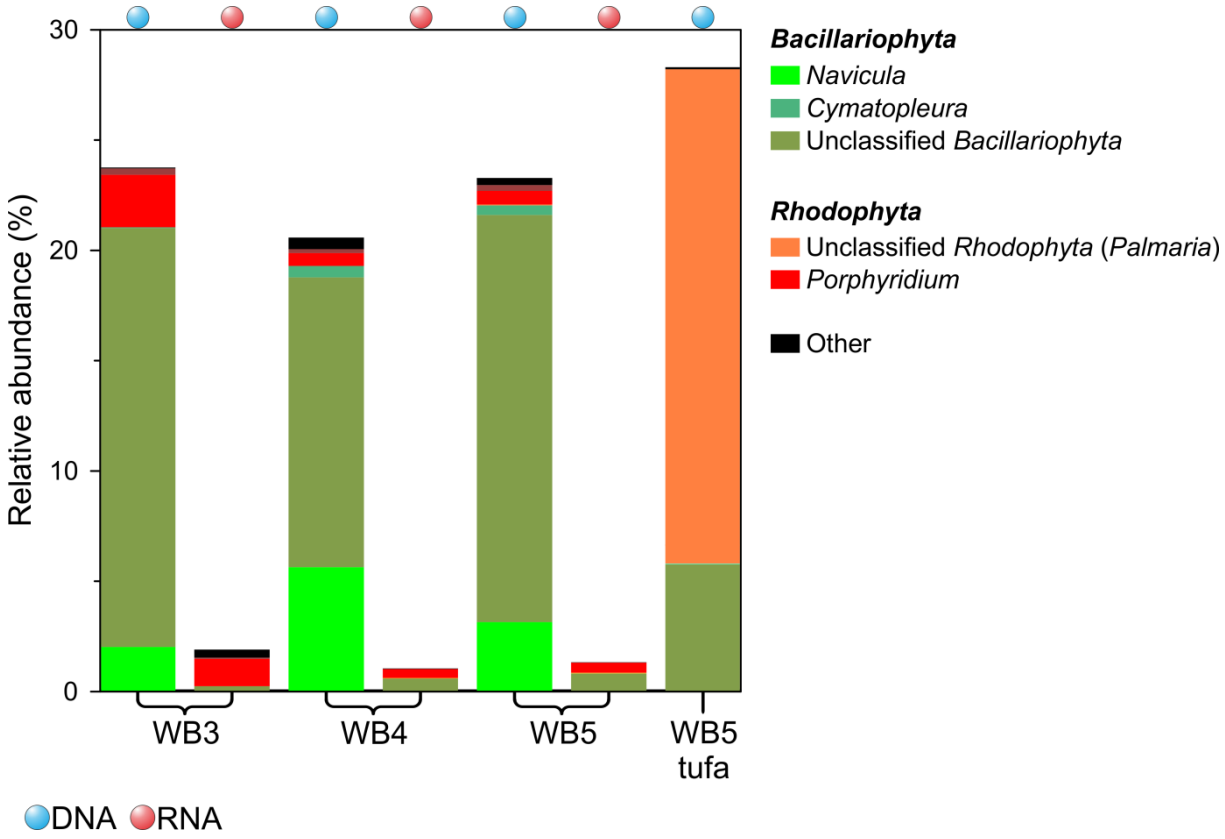


Figure S1. Relative abundance of chloroplast sequences and closest related chloroplast 16S rRNA gene sequences of cultured organisms clustered at a genetic divergence level of 1%. Green colors were used to highlight diatoms (*Bacillariophyta*) and red colors were used for red algae (*Rhodophyta*). All OTUs with abundance <0.5% were summarized in the artificial group “Other”.

Table S1. Vectors used in this study.

Type of vector	Designation	Relevant characteristics	Source
Fosmid	pCC1FOS	Chlr, T7 promoter, FI and pMB1 replicon	Epicentre
Fosmid	fPEP1	pCC1FOS: 41,000 bp fragment of cloned meta-genomic DNA	This study
Fosmid	fPEP2	pCC1FOS: 21,600 bp fragment of cloned meta-genomic DNA	This study
Fosmid	fPEP3	pCC1FOS: 28,000 bp fragment of cloned meta-genomic DNA	This study
Fosmid	fPEP4	pCC1FOS: 29,750 bp fragment of cloned meta-genomic DNA	This study
Fosmid	fPEP5	pCC1FOS: 26,500 bp fragment of cloned meta-genomic DNA	This study
Plasmid	pCR-2.1-TOPO	Kanr, Zeor, lac promoter, pMB1 replicon	Invitrogen

Table S2. Characteristics of fosmid libraries of Westerhöfer Bach biofilm and tufa

Library	Number of clones	Average insert size (kb)	Total DNA (Gbp)
FOS WB3	9,216	27.3	0.26
FOS WB4	9,216	30.1	0.28
FOS WB5	9,216	25.7	0.24
FOS WB5tufa	9,216	29.6	0.27
Total	36,864	28.2	1.05

3.2 Identification and analysis of cellulolytic genes from Westerhöfer Bach biofilm and tufa

One recombinant *Escherichia coli* strain has been detected out of 38,600 screened fosmid-library bearing clones, which conferred cellulolytic activity (Figure 1). The preparation of the fosmid libraries from DNA isolated from the Westerhöfer Bach samples is described in Schneider *et al*. (2013, submitted). The screening was performed on agar plates containing hydroxy-ethylcellulose labeled with Cibacron Brilliant Red 3B-A (HEC_{red}) (see (Nacke *et al.*, 2012)). The fosmid insert was sheared, subcloned and plated on screening media HEC_{red}. Nine positive subclones and 87 randomly picked clones were arrayed on a microtiter plate and subjected to Sanger sequencing. After assembly of all reads and analysis of the approximately 28 kb fosmid insert the genetic organization depicted in Figure 2B was determined. By sequencing of nine positive subclones and sequence analysis, the ORFs responsible for cellulolytic activity were identified (ORFs 9, 10, and 11, Table 1). Phylogenetic affiliation designated by the best BLAST (Altschul *et al.*, 1990) hit suggests that the cloned DNA fragment derives from a yet unknown proteobacterium. Some ORFs had similarities to *Cyanobacteria* of the *Oscillatoriales*, the main group encountered within the Westerhöfer Bach biofilm. Additionally, the genomic content of the whole fragment was determined with Taxy (Meinicke *et al.*, 2011), which gave further evidence for a proteobacterial affiliation, because 44% (17% *Cyanobacteria*) of the DNA content were predicted to be of gammaproteobacterial origin, which is in accordance with the suggested ancestry of the BLAST (Altschul *et al.*, 1990) results. The protein domain structure of the potential cellulolytic gene products of ORFs 9, 10, and 11 is shown in Figure 2A.

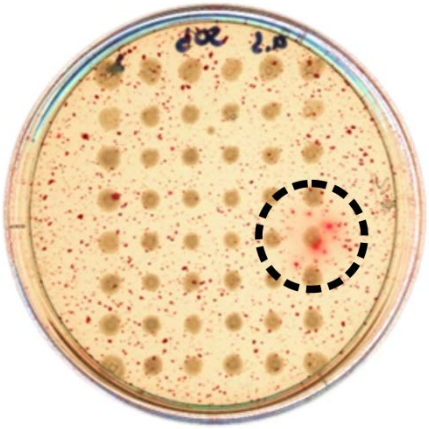


Figure 1. Fosmid clone f50E12 (fosmid library WB 4), which shows activity by the diffusion of the indicator substrate HEC_{red} into the surrounding agar.

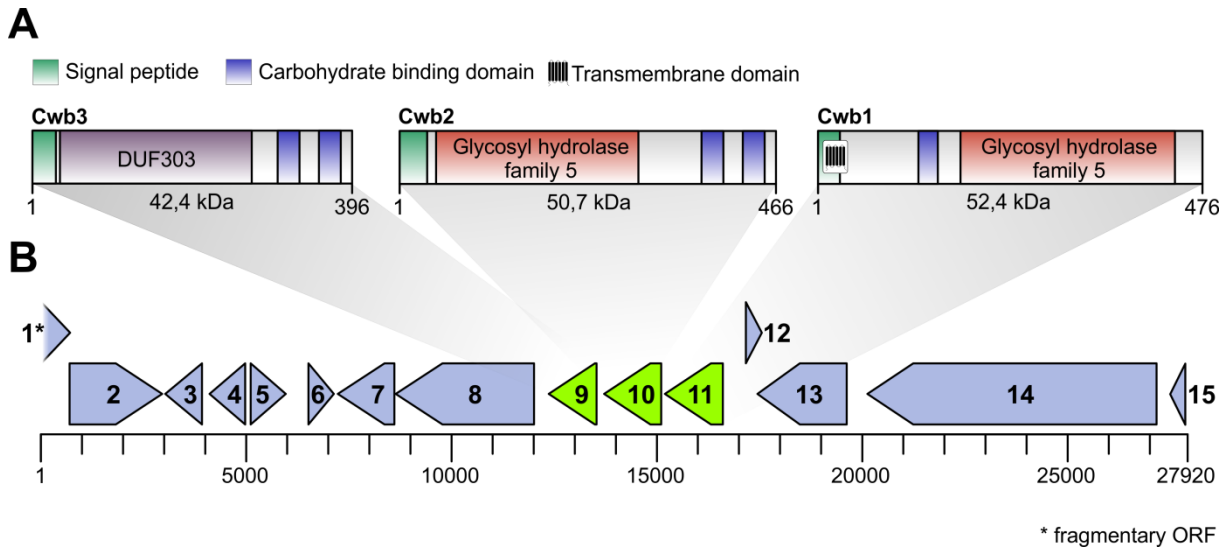


Figure 2. Protein domain structure of the three potentially activity conferring cellulolytic ORFs (A) as predicted by CDD (Marchler-Bauer *et al.*, 2013), length is given in amino acids. Genetic organization of the sequenced fosmid insert (B), length is given in base pairs, characteristics of ORFs are given in Table 1.

Table 1. Characteristics of genes depicted in Figure 2B and their corresponding gene products.

ORF name	Size (bp/aa)	Closest similar protein, accession no. of similar protein (no. of encoded amino acids)	Organism	Alignment length/total no. aa query (% identity)
1	684/227	Nuclease (partial), WP_019099102 (318)	*Pseudomonas putida*	96/229 (42)
2	2286/761	ATP-dependent DNA helicase, WP_008156323 (758)	*Pseudomonas sp.* GM41(2012)	364/755 (48)
3	906/301	Methyltransferase (cyclopropane fatty acid synthase), WP_007865786 (319)	*Polaromonas sp.* CF318	112/279 (40%)
4	891/296	LysR family transcriptional regulator, ACY24824 (290)	Uncultured organism	242/290 (83%)
5	858/285	Zn-dependent hydrolase, ACY24825 (293)	Uncultured organism	228/264 (86%)
6	630/209	Hypothetical protein (pot. IclR transc. regulator), WP_017447001 (257)	*Gayadomonas joobiniege*	94/209 (45%)
7	1371/456	Hypothetical protein, YP_003266061 (443)	*Haliangium ochraceum* DSM 14365	175/442 (40%)
8	3402/1133	Hypothetical protein (pot. cytochrome), YP_003266060 (579)	*Haliangium ochraceum* DSM 14365	132/368 (36%)
9	1191/396	Hypothetical protein (Carbohydrate-binding protein), WP_019600826 (1085)	*Teredinibacter turnerae*	134/253 (53%), 17/34 (50%), 18/32 (56%)
10	1401/466	Endo-1,4-beta-mannanase, YP_001982936 (585)	*Cellvibrio japonicus* Ueda107	167/413 (40%)
11	1431/476	Cellulase, YP_001983414 (556)	*Cellvibrio japonicus* Ueda107	176/464 (38%)
12	390/129	TonB-dependent receptor, YP_001983526 (1155)	*Cellvibrio japonicus* Ueda107	54/124 (44%)
13	2295/764	Conserved repeat domain protein (partial), WP_006636028 (1294)	*Microcoleus vaginatus*	492/750 (66%)
14	7065/2354	Conserved repeat domain protein (pot. involed in pilus biogenesis), WP_006635102 (2339)	*Microcoleus vaginatus*	706/1378 (51%), 416/593 (70%)
15	378/125	DNA-binding protein (pot transc. regulator HTH), WP_009207780 (125)	*Sulfuricella denitrificans*	87/125 (70%)

Additionally, a first biochemical characterization showed that the gene product of ORF 11 (Cwb1) was able to hydrolyze β-D-glucan and carboxymethyl-cellulose (CMC), whereas those of ORF 9 and 10 had no activity on any of the tested substrates (β-D-glucan, CMC, hydroxyethyl-cellulose, xylan from oak and beech, and cellulose). With CMC as substrate, temperature optimum of the reaction was at approximately 40°C and pH-optimum at approximately 7.

References

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

Marchler-Bauer, A., Zheng C., Chitsaz F., Derbyshire M.K., Geer L.Y.*, et al.* (2013). CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res **41**: D348-352.

Meinicke, P., Asshauer K.P., Lingner T. (2011). Mixture models for analysis of the taxonomic composition of metagenomes. Bioinformatics **27**: 1618-1624.

Nacke, H., Engelhaupt M., Brady S., Fischer C., Tautzt J.*, et al.* (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett **34**: 663-675.

Schneider, D., Reimer A., Hahlbrock A., Arp G., Reitner J.*, et al.* (2013, submitted). Metagenomic and metatranscriptomic analyses of bacterial communities derived from a calcifying karst water creek biofilm and tufa. Geomicrobiology Journal.

4 Prokaryotic community composition of four Kenyan soda lakes as revealed by amplicon sequencing

Different sample types (soil, dry and wet sediment, water, and microbial mat) from the Kenyan soda lakes Bogori, Sonachi, Elmenteita, and Magadi have been analyzed with respect to their bacterial and archaeal community composition and diversity. Samplings, DNA isolations, and amplicon PCRs were carried out by Dr. Romano Mwirichia Kachiuru.

Table 1: Diversity of prokaryotic 16S rRNA genes at 3% genetic divergence level.

Sample ID	Number of sequences	OTUs	Chao1	Michaelis Menten fit	Shannon
Bacteria					
B-S	7,732	652	1488.64	914.28	4.56
B-SD	5,122	438	1113.28	558.49	4.64
B-W	2,700	88	198.5	135.92	1.19
S-S	802	264	746.76	556.45	6.82
S-SW	12,382	782	1020.84	866.68	8.01
S-W	14,807	198	423.16	236.07	3.26
E-S	10,099	601	1271.41	692.12	7.19
E-SD	15,238	717	1249.17	875.08	6.76
E-SW	9,370	644	1174.2	728.5	7.67
E-W	7,377	139	232.88	165.46	1.76
M-SW-1	8,750	221	494.19	246.75	4.59
M-SW-2	3,459	141	256.56	177.58	3.5
M-SW-3	3,390	166	302.5	216.16	4.44
M-MM	9,404	164	282.59	179.51	3.99
Archaea					
B-S	5,983	63	68.14	64.71	3.06
B-SD	1,837	133	223.46	159.51	4.84
B-W	2,860	185	286.67	220.61	5.26
S-S	6,881	116	266.27	149.84	1.76
S-SW	28,298	408	485	432.89	5.16
S-W	5,567	222	248.64	238.78	6.16
E-S	977	85	115	96.44	5.23
E-SD	1,456	118	321.33	150.64	4.8
E-SW	2,495	243	412.58	327.46	5.75
E-W	1,290	119	350.25	193.89	3.9
M-SW-1	6,434	129	133	134.54	5.1
M-SW-2	2,344	108	125.14	112.16	5.78
M-SW3	3,345	142	163.38	150.48	5.93
M-MM	1,549	81	156	89.01	4.56

Abbreviations: B, Lake Bogoria; S, Lake Sonachi, E, Lake Elmenteita; M, Lake Magadi; S, soil sample; SD, dry sediment sample; SW, wet sediment sample; W, water sample; MM, microbial mat

Diversity and general characteristics of the bacterial community

The prokaryotic community compositions of 14 samples (Table 1) from four different soda lakes in Kenya were assessed. From all samples, 110,632 high quality bacterial sequences with an average read length of 450 bp were assigned to the domain *Bacteria*, 204 sequences had no BLAST match. The number of bacterial sequences per sample ranged from 802 (sample S-S) to 15,238 (sample E-SD) as shown in Table 1. The classified sequences were affiliated to 43 bacterial phyla and 8 candidate divisions (BRC1, OD1, KB1, OP3, OP9, TM7, WS3 and WS6). The bacterial sequences were clustered into 4,709 OTUS at 3% genetic distance. The OTU number in the samples ranged from 88 (B-W) to 782 (S-SW).

Bacterial community composition

Overall, members of the phyla *Cyanobacteria* (23.1%), *Gammaproteobacteria* (14.6%), *Alphaproteobacteria* (12.2%), *Bacteroidetes* (11.5%), *Firmicutes* (11.3%), *Actinobacteria* (8.7%), *Gemmatimonadetes* (4.6%), *Deltaproteobacteria* (2.8%), *Chlorobi* (2.3%), *Acidobacteria* (2.2%), *Chloroflexi* (2%), *Deinococcus-Thermus* (1.3%), and *Spirochaetes* (0.8%) were present in all samples (Figure 1). These phyla constituted 97.3% of all sequences with the remaining 2.7% distributed among the less abundant phyla and several candidate divisions (see Figure 1 description). Some of the groups were detected in one or two samples only. For example, the candidate divisions KB1, WS6 and OP3 were only detected in the samples E-SW and S-S, respectively. Some lineages, e.g., candidate division Hyd24-12 and *Aquificae* were only detected in one or two samples (E-SW, B-S, M-SW-3).

Twenty phyla accounted for over 98% of all sequences in all samples, whereas only a small percentage indicated as the artificial group Others belonged to the remaining phyla. In the wet sediment sample M-SW-2 from lake Magadi and the three different water samples (B-W, S-W, E-W) *Cyanobacteria* constituted over 50% of the bacterial community. In the other samples, *Proteobacteria* were the predominant phylum with a maximum at 73% of all sequences in the microbial mat (M-MM) from Magadi. A slight difference is noted in the wet sediment sample M-SW-1 in which the phylotypes affiliated with the *Firmicutes* accounted for 50% of the sequences. The sequences were affiliated to 100 classes of which 50% had a representation above 0.1%. The relative abundance of the different phylotypes varied in

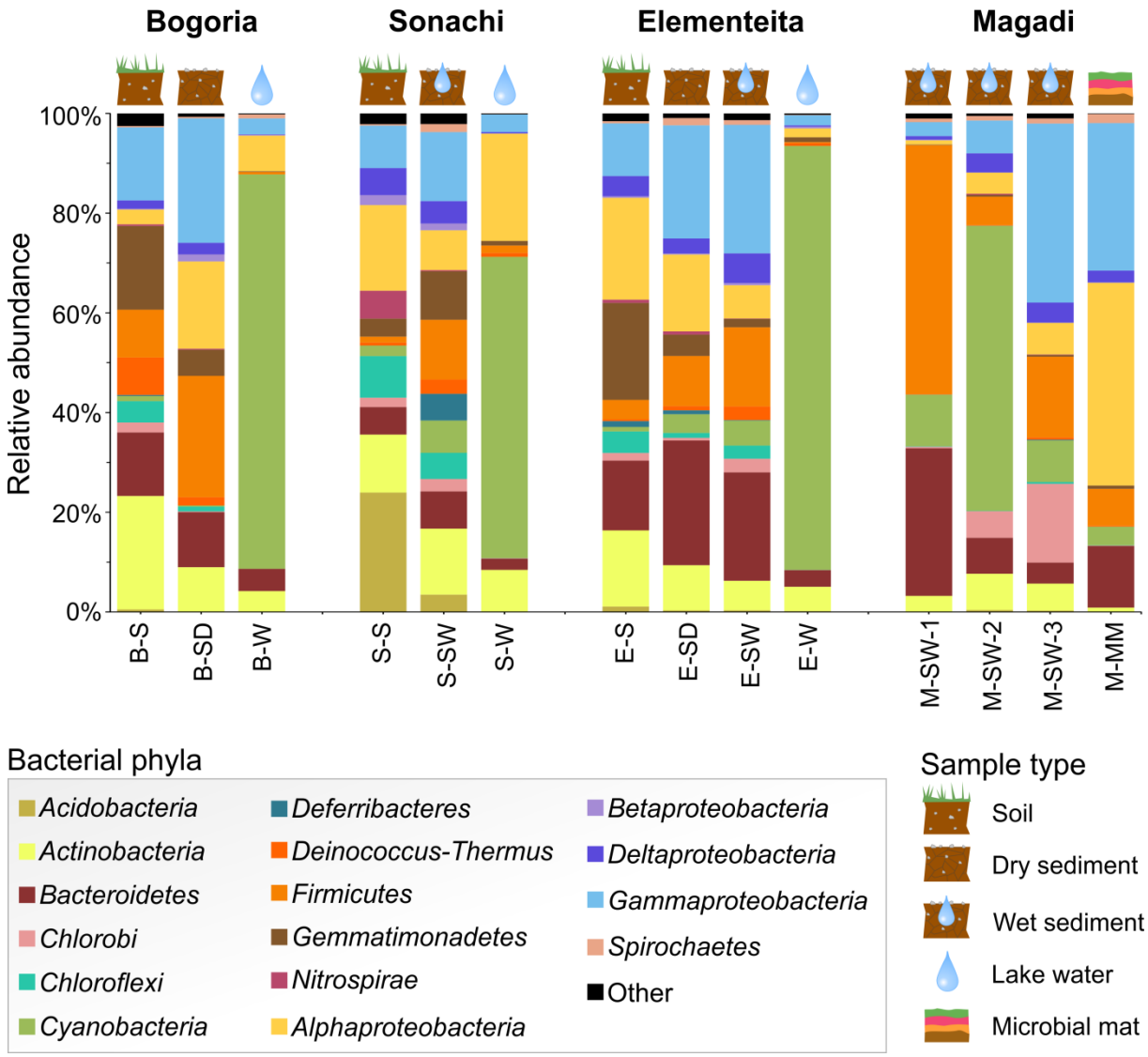


Figure 1. Relative abundance of the bacterial phyla and proteobacterial classes detected in the different soda lake samples. Values are as a percentage of the total number of sequences for each individual sample. Less abundant phylogenetic groups accounting for <1% of all classified sequences were summarized in the artificial group "Other". Other comprised the phyla *Aquificae*, *Armatimonadetes*, *Caldiserica*, *Chrysiogenetes*, *Dictyoglomi*, *Elusimicrobia*, *Fibrobacteres*, *Fusobacteria*, *Lentisphaerae*, *Planctomycetes*, *Synergistetes*, *Tenericutes*, *Thermodesulfobacteria*, *Thermotogae*, *Verrucomicrobia*, and candidate divisions BD1-5, BHI80-139, BRC1, Hyd24-12, Kazan-3B-28, KB1, MVP-21, NPL-UPA2, OD1, OP3, OP9, RF3, SM2F11, TA06, TM6, TM7, WCHB1-60, WS3, and WS6. Abbreviations: B, Lake Bogoria; S, Lake Sonachi, E, Lake Elmenteita; M, Lake Magadi; S, soil sample; SD, dry sediment sample; SW, wet sediment sample; W, water sample; MM, microbial mat

the different samples. For example, whereas the *Cyanobacteria* were more abundant in the water samples, their numbers decreased tremendously in the sediment samples. Notable is the absence of *Chloroflexi* in the water samples. In the Magadi mat sample M-MM, the *Gammaproteobacteria* and *Alphaproteobacteria* represented 40 and 29%, respectively, whereas in the mud sample (M-SW-1) *Clostridia* that belong

to the *Firmicutes* accounted for 50% of bacterial community. The sequences were clustered into 1,016 genera, more than half were without cultured representative. BLAST analysis against the SILVA database resulted in 198 genera with existing cultivated members.

Archaeal diversity and community composition

A total of 71,316 high quality 16S rRNA gene sequences with an average read length of 450 bp were assigned to the domain *Archaea*. In addition, approximately 5,000 16S rRNA gene sequences were obtained that could not be classified, indicating the presence of entirely novel microorganisms. The number of classified archaeal sequences per sample ranged from 977 (E-S) to 28,298 (S-SW) as summarized in Table 1 above. The sequences clustered into 1,397 OTUs at 3% genetic distance with the lowest and highest number in samples B-S and S-SW, respectively.

The phyla *Euryarchaeota* and *Thaumarchaeota* were predominant (89.6 and 7.8%, respectively) as shown in Figure 2. The phylum *Crenarchaeota* constituted 2.1% of the archaeal sequences, whereas four sequences from the phylum *Korarchaeota* were detected, but solely in the dry sediments from Lake Bogoria (0.2%, sample B-SD). In four samples (B-S, S-S, E-SD and E-W) the *Euryarchaeota* constituted over 99% of the total sequences. At the class level, the *Halobacteria* constituted 63.7% of all sequences followed by *Thermoplasmata* (11.7%). In 13 samples, *Halobacteria* were the predominant class with maximal abundances in samples S-S and E-W (>96%). *Thermoplasmata* and *Methanomicrobia* predominated in sample B-S. Sequences affiliated to the classes *Halobacteria*, *Thermoplasmata* and *Thermoprotei* were detected in all samples, whereas the '*Candidatus* Korachaeum' was detected only in sample B-SW. At order level the sequences were clustered into 35 orders of which 24 were represented by uncultured groups. Similarly, of the 147 genera represented in the dataset, only 52 have known species.

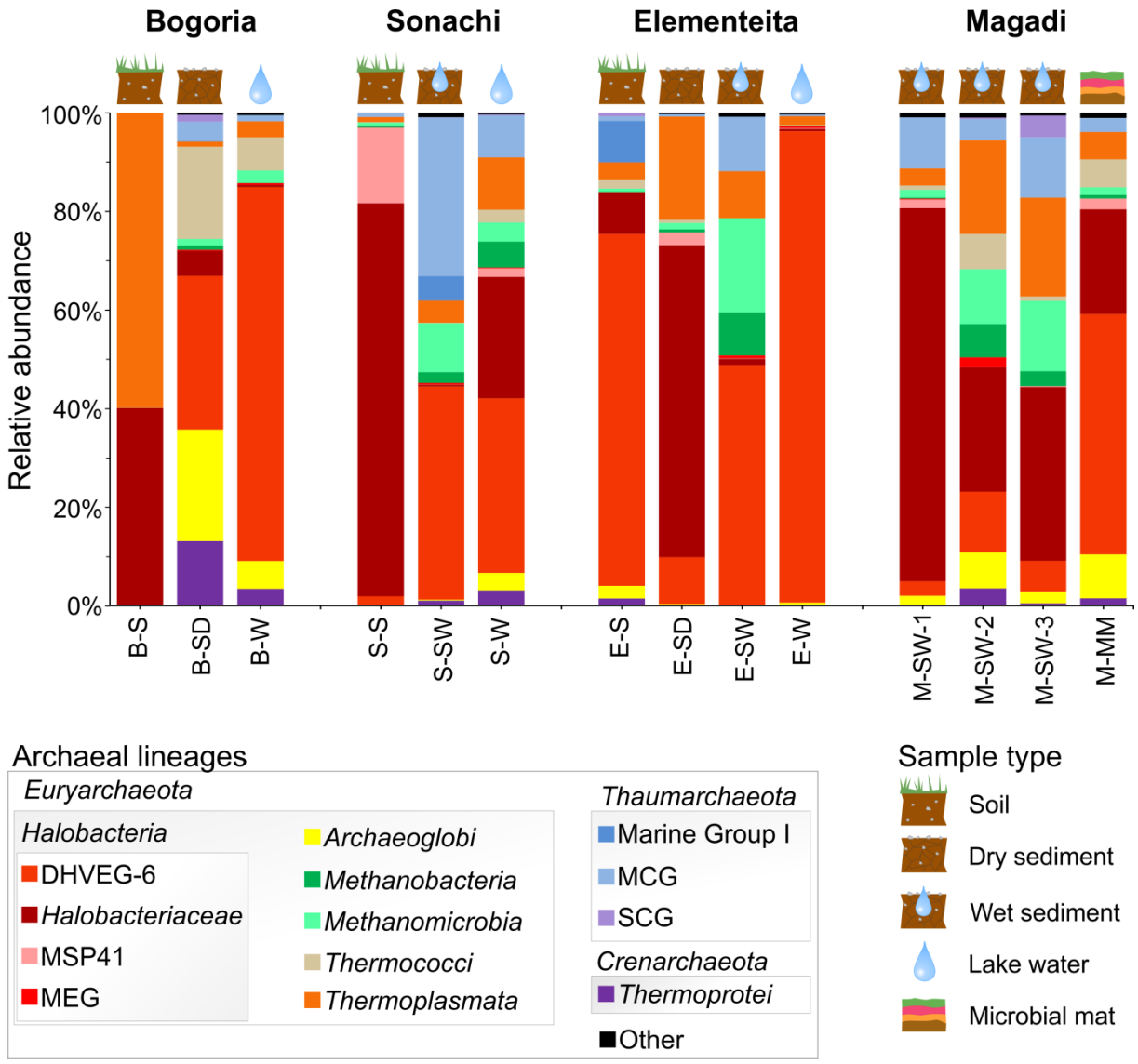


Figure 2. Relative abundances of archaeal classes and halobacterial orders of the different soda lake samples. Values are as a percentage of the total number of sequences for each individual sample. Less abundant phylogenetic groups accounting for <1% of all classified sequences were summarized in the artificial group "Other". Other comprised the phyla *Korarchaeota* (Candidatus *Korarchaeum*), *Thaumarchaeota* (Group C3, Marine Benthic Group B, candidate group MC2A209), and *Euryarchaeota* (Deep Sea Euryarchaeotic Group, Marine Hydrothermal Vent Group, and candidate group SM1K20). Abbreviations: DHVEG-6, Deep Sea Hydrothermal Vent Group 6; MSP, Magadi salt pond; MEG, Miscellaneous Euryarchaeotic Group; MCG, Miscellaneous Crenarchaeotic Group; SCG, Soil Crenarchaeotic Group. Abbreviations: B, Lake Bogoria; S, Lake Sonachi, E, Lake Elmenteita; M, Lake Magadi; S, soil sample; SD, dry sediment sample; SW, wet sediment sample; W, water sample; MM, microbial mat

5 Electroactive biofilm and plankton communities

The bacterial community of plankton and biofilm of ten differentially incubated microbial fuel cells (MFC) were analyzed via pyrosequencing of 16S rRNA gene amplicons as described in Schneider *et al.* (2013). The study encompassed 178,500 high quality DNA-based and RNA-based 16S rRNA gene sequences. The MFCs (two biological replicates) were inoculated with wastewater and incubated at 30°C with different substrates to test for potential applications (Figure 1). After 45 days, planktonic and biofilm samples were collected and immediately frozen at -80°C (for details see (Baudler, 2012)).

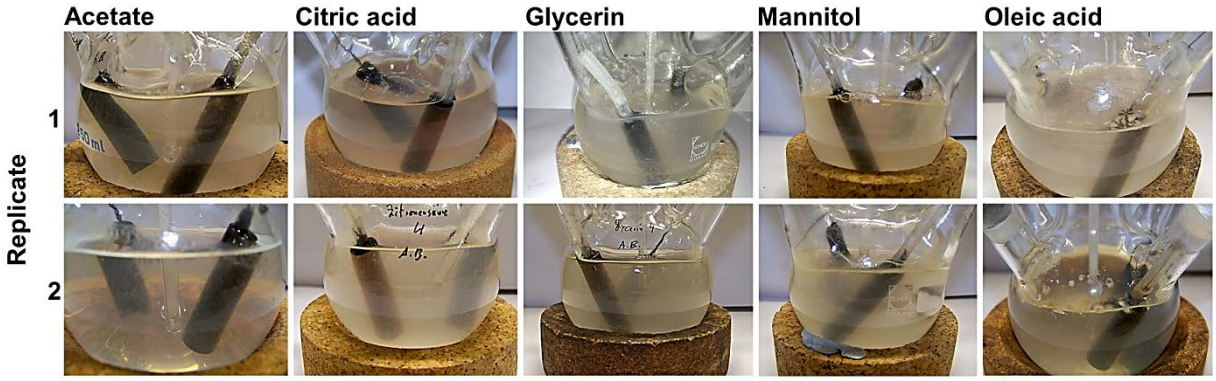


Figure 1. Microbial fuel cells fed with different substrates before collection. Photographs courtesy of André Baudler (Baudler, 2012).

General characteristics and bacterial diversity of the microbial fuel cells

The optical density of the media within each MFC was measured and showed large differences in cell content of the established planktonic communities (see Figure 4.11 (Baudler, 2012)). Highest optical densities of the media were measured for MFCs grown with citric acid, mannitol, and oleic acid, whereas acetate had the lowest. Additionally, the current efficiency (*j*) and Columbic-efficiency (CE) were measured at time point of sampling by André Baudler (2012) (see Figure 2). Furthermore, the diversity of each sample was analyzed at 3% genetic divergence and is represented by the Shannon index of diversity (Shannon, 1948). The Shannon index of all samples ranged between 0.44 (acetate, active biofilm community) to 4.29 (inoculum, entire sewage water community). Generally, the entire bacterial community showed a higher diversity when compared to the corresponding active bacterial community. In most

cases, the planktonic community exhibited a higher diversity as the biofilm community.

Bacterial community composition

The composition of the entire (DNA-based) and active (RNA-based) bacterial communities of plankton and biofilm of each MFC are summarized in Figure 2.

The MFCs grown with acetate showed two different bacterial community profiles. Biofilm of replicate 1 was dominated by *Deltaproteobacteria* (*Geobacter* sp.) and produced electric current; whereas replicate 2 had a more diverse community and a low electric current production.

Citric acid fed MFCs also led to two different communities. The first replicate's plankton was dominated by *Gammaproteobacteria* (*Pseudomonas* sp.) and the biofilm by *Bacteroidetes* (*Sphingobacteriales*, WCHB1-69) and *Deltaproteobacteria* (*Geobacter* sp.). Contrary, plankton of replicate 2 was dominated by *Betaproteobacteria* (*Petrobacter* sp.) and *Gammaproteobacteria*, and the biofilm showed similarities with the acetate grown biofilm, but harbored *Geobacter* sp. of the candidate group AKYG597.

Glycerol as substrate led to a more diverse bacterial community. However, electric current production was only mentionable for replicate 2. Here, *Deltaproteobacteria* were the predominant group of the active biofilm community. Interestingly, replicate 1 harbored the *Geobacter* OTU (AKYG597), which occurred in citric acid fed MFCs and replicate 2 the "productive" *Geobacter* sp. from acetate fed MFC.

MFCs, which were incubated with mannitol showed distinct community compositions when compared to acetate, citric acid, and glycerol fed MFCs. The productive MFC (replicate 1) was dominated by *Bacteroidetes* (WCHB1-69), a mixture of all proteobacterial classes, and *Firmicutes* (*Peptostreptococcaceae*). Within the active community, *Firmicutes* (*Clostridium* sp.) predominated in plankton and biofilm, which were less abundant in DNA-based analysis. Replicate 2 was less productive and showed a lower abundance of *Firmicutes* (*Clostridium* sp.), but a higher abundance of *Bacteroidetes* (WCHB1-69).

The most complex substrate oleic acid resulted in the highest bacterial diversity, but also in the lowest electric current production. Dominant groups in both replicates were in descending order *Betaproteobacteria* (*Comamonadaceae*, *Rhodo-*

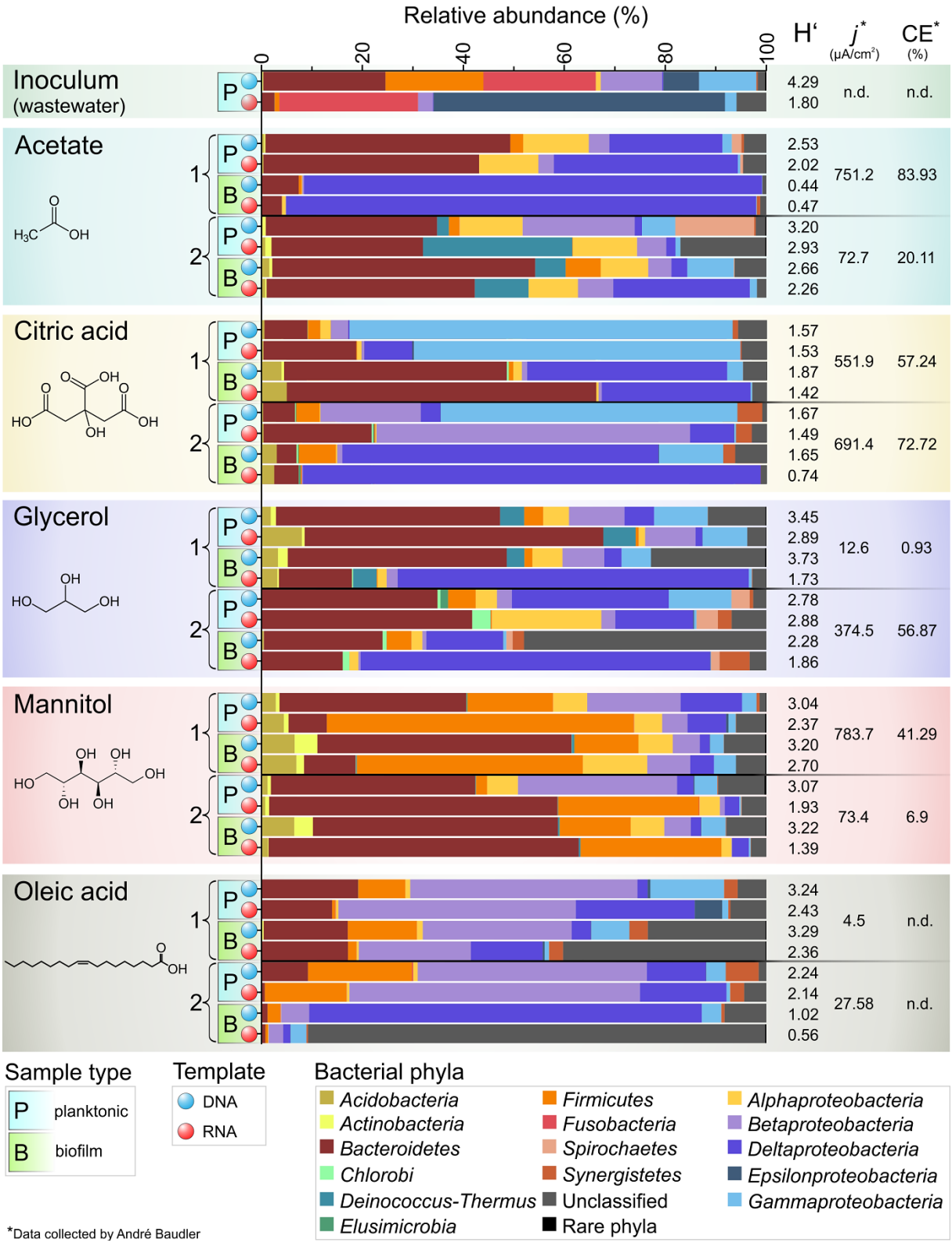


Figure 2. Entire and active bacterial community composition of inoculum (sewage plant Steinhof, Braunschweig), planktonic and biofilm of the with different substrates incubated MFCs (two biological replicates). Analysis was carried out at a genetic divergence level of 3%. Abbreviations: H', Shannon index (calculated at same survey effort of 1,195 sequences); *j*, Current efficiency; CE, Coulumbic-efficiency.

cyclaceae, *Neisseriaceae*), *Deltaproteobacteria* (*Desulfarculaceae*, *Geobacteraceae*), *Firmicutes* (different *Clostridiales*), *Bacteroidetes* (*Rikenellaceae*, *Porphyromonadaceae*), and *Gammaproteobacteria* (*Pseudomonadales*).

References

Baudler, A. (2012). Elektrochemisch aktive Biofilme auf Basis mikrobieller Konsortien: Grundlagenuntersuchung zur Verwendung von Mono - und Mischsubstraten (Master thesis). Braunschweig: Technische Universität Braunschweig. 106 p.

Schneider, D., Arp G., Reimer A., Reitner J., Daniel R. (2013). Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the kiritimati atoll, central pacific. PLoS One **8**: e66662.

Shannon, C.E. (1948). A mathematical theory of communication. The Bell System Technical Journal **27**: 379-423 and 623-656.

C DISCUSSION

Metagenomics and metatranscriptomics have proven to be powerful tools that led to the discovery of novel bioactive molecules and biocatalysts, provided insights into entire (DNA-based) and active (RNA-based) prokaryotic communities, and revealed metabolic key functions of many habitats (Shestakov, 2012; Simon and Daniel, 2011). The advent of cost-effective next-generation sequencing contributed significantly to the success of metagenomic and metatranscriptomic research. Nowadays, the generation of large and representative datasets is feasible, which results in an unprecedented coverage of the genetic extent of microbial communities. Future tasks within this field will be the development of ingenious bioinformatics tools for the analysis of the overwhelming information produced by the current (454, Illumina, PacBio) and even more by the oncoming sequencing techniques (e.g., Illumina's Moleculo Long Read technology).

In this study, four different prokaryotic consortia have been analyzed by metagenomic and partially by metatranscriptomic approaches. The employed techniques allowed the assessment of entire and active prokaryotic communities. The analyzed samples comprised a calcifying freshwater creek biofilm, a calcifying microbial mat from a hypersaline lake, soil, sediment and water samples from four soda lakes, and microbial fuel cells (MFC). For all these samples, DNA and partly RNA were isolated and prokaryotic composition and diversity were investigated by 454 pyrosequencing of hypervariable regions of amplified 16S rRNA genes. The obtained sequencing data was analyzed with a modified QIIME (Caporaso *et al.*, 2010) pipeline workflow, which is exemplary shown in Figure 6. This workflow results in a reliable, comparable, and detailed analysis of next-generation generated 16S rRNA gene sequence data derived from entire and active prokaryotic communities of the investigated microbial assemblages (Sahm *et al.*, 2013; Wemheuer *et al.*, 2013; Wemheuer *et al.*, 2012). However, it has to be considered that phylogenetic analyses based on 16S rRNA genes can be biased, e.g., by the primer pairs employed or the varying number of 16S rRNA operons (and sequence divergence therein) within different phylogenetic groups (Lee *et al.*, 2009; Sun *et al.*, 2013).

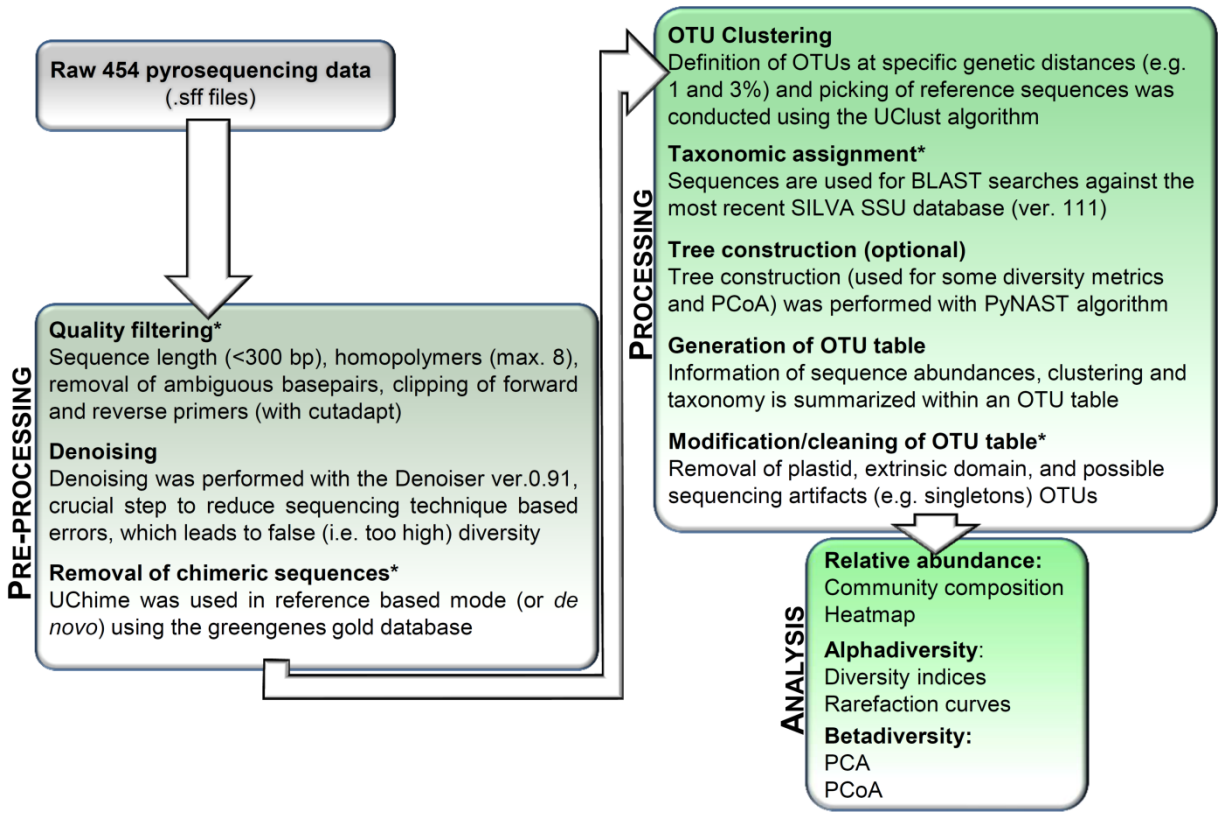


Figure 6. Draft of the 16S rRNA gene data processing pipeline employed to assess the phylogenetic composition of different habitats. Variations to the QIIME (Caporaso *et al.*, 2010) pipeline are marked with asterisks.

Next to the 16S rRNA gene sequence analyses, one of the biofilm samples of the Westerhöfer Bach was subjected to a metatranscriptomic analysis (Schneider *et al.*, 2013, submitted). For this purpose, total biofilm RNA was isolated and reverse transcribed to cDNA using random hexamer primers. Subsequently, cDNA was sequenced by 454 pyrosequencing. This revealed the most abundant gene transcripts of the microbial biofilm community, and thereby the biological functions, which were active at sampling time.

Furthermore, metagenomic large-insert fosmid libraries were constructed with metagenomic DNA from Westerhöfer Bach biofilm and tufa. These libraries were mined for genes conferring hydrolytic activity on peptide and (hemi)cellulose substrates by a function-driven plate assay. The screening led to the discovery of five novel proteolytic and one novel (hemi)cellulolytic enzymes.

1 Metagenomic and metatranscriptomic analyses of different samples

1.1 Diversity within phototrophic biofilm communities

Recently, phototrophic microbial communities in microbial mats of pond 4 from Guerrero Negro (Mexico) (Harris *et al.*, 2013), Highborne Cay (Bahamas) (Mobberley *et al.*, 2012), Shark Bay (Australia) (Allen *et al.*, 2009), Solar Lake (Eilat, Israel) (Sorensen *et al.*, 2005), or biofilms from freshwater settings like Iberian Range (Spain) (Beraldi-Campesi *et al.*, 2012), and Taroko National Park (Taiwan) (Ng *et al.*, 2006) were investigated. Biofilm communities can vary in their microbial community composition and diversity. This is mainly attributed to the surrounding environmental factors (e.g., salinity, pH, temperature, UV-radiation) and the available nutrition and energy sources.

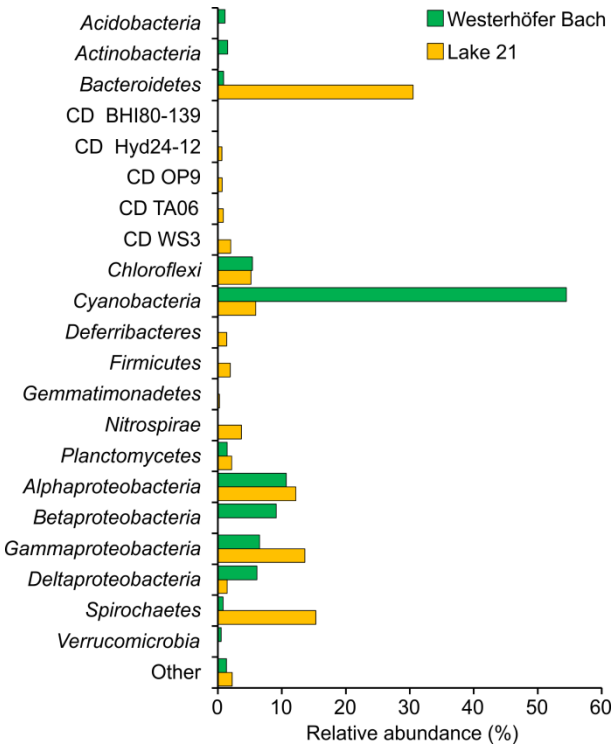


Figure 7. Comparison of Westerhöfer Bach and Lake 21 bacterial communities. For Westerhöfer Bach all sampling spots have been combined and averaged, for Lake 21 microbial mat layers were combined. The artificial group Other includes low abundant phyla and unclassified bacterial 16S rRNA gene sequences.

The phototrophic microbial biofilms (or microbial mat) of Westerhöfer Bach and Kiritimati Lake 21 showed the before mentioned differences in their prokaryotic community compositions. Main differences between these two sampling sites are nutrient availability, salinity, strength of UV-radiation, temperature, and water movement. Although different 16S rRNA gene primer sets were used to assess the bacterial communities (Westerhöfer Bach V2-V3, Lake 21 V3-V5), potentially leading to shifts in relative abundances and inconsistencies in acquisition of certain bacterial groups (Liu *et al.*, 2007), a rough comparison of these two biofilms is possible. The highly oxygenated freshwater stream of Westerhöfer Bach is dominated by filamentous cyanobacteria and aerobic heterotrophs (Schneider *et al.*, 2013, submitted). The laminated microbial mat from Lake 21 consists of a highly organized microbial community (Schneider *et al.*, 2013) that is structured along a light and oxygen gradient, including

all trophic levels (see Figure 7 for an overall comparison of the two biofilms). Despite of differences in community structure and EPS content and composition, the bacterial communities showed similarities in their general prokaryotic community setup. For example, the main production of organic matter is fulfilled by *Cyanobacteria*, which convert light energy and CO_2 to organic matter and oxygen. These are accompanied by anoxygenic phototrophs, which are more pronounced in the microbial mat of Lake 21. This organic matter can be used by heterotrophic (micro)organisms and enables their growth, while their metabolic by-products can be further used by other microbes. Additionally, archaea might perform important metabolic functions such as methanogenesis and ammonia oxidation, although they seem to be less abundant in the analyzed biofilms as indicated by other studies (Breitbart *et al.*, 2009; Kunin *et al.*, 2008; Papineau *et al.*, 2005; Robertson *et al.*, 2009; Spear *et al.*, 2003). Ultimately, this leads to very complex metabolic interactions, exemplified by the distinct archaeal and bacterial communities, which lead to layers of different color and structure in the microbial mat from Lake 21, Kiritimati (Schneider *et al.*, 2013). While the Westerhöfer Bach biofilm was dominated by filamentous cyanobacteria, the microbial mat of Lake 21 harbored coccoid and halotolerant cyanobacteria. However, in Lake 21 cyanobacteria mainly occurred in the upper three layers of the mat, suggesting light penetration is reduced in the lower layers. The phototrophic layer is followed by a mixture of anaerobes, mainly composed of fermenters and sulfur bacteria. Nevertheless, defined microbial architecture might also be present within the Westerhöfer Bach biofilm, which is extremely challenging to sample appropriately because of the low biofilm's thickness (average of 3 mm). In this case, application of other techniques (e.g., FISH, transmission electron microscopy) might be more promising. The comparison highlights the advantages when separation of a solid environmental sample is possible, in this case due to size and clear coloration of different microbial mat layers. Homogenization of sample material, e.g., in soil, is crucial to circumvent potential heterogeneities on the microscale. However, if separation of a sample (e.g., into a gradient) is possible, this can lead to a remarkable gain in information of the microbial community structure in a specific environment.

Prior studies of the Westerhöfer Bach biofilm indicated *Cyanobacteria* and *Proteobacteria* as the predominant parts of the bacterial community (Arp *et al.*, 2010; Cousin and Stackebrandt, 2010). Additionally, a current study of the cyanobacterial

community composition based on traditional 16S rRNA gene sequencing using the Sanger technique yielded similar results as both studies showed that filamentous cyanobacteria of the *Oscillatoriales* (*Tychonema*/*Phormidium*/*Microcoleus* clade) constitute the main fraction of the bacterial communities in the creek biofilm (Brinkmann *et al.*, 2013, submitted). In comparison to other studies of similar habitats, the Westerhöfer Bach biofilm community was assessed at an unprecedented sampling depth. For example, Beraldi-Campesi *et al*. (2012) detected mainly *Cyanobacteria* (43%, also mainly *Oscillatoriales*) and *Proteobacteria* (21%) within the tufa systems of the Iberian Range (Spain). These were accompanied with *Firmicutes*, *Acidobacteria*, *Planctomycetes* and *Actinobacteria* at lower abundances, similar to the tufa building biofilm of Westerhöfer Bach. Ng *et al*. (2006) analyzed tufa at three different sites and found *Proteobacteria* (26-30%), *Cyanobacteria* (17-29%), and *Acidobacteria* (7-13%) to dominate in these environments, which is in accordance with the Westerhöfer Bach tufa sample. However, these studies showed that a comparatively low sequencing effort (94 and 381 analyzed 16S rRNA gene sequences, respectively) might miss less-abundant bacterial phyla. Generally, organisms detected in this environmental setting seem to be adapted to colder environments, as, for example, similar lineages within the *Betaproteobacteria* (*Comamonadaceae*) were also found in glacier ice (Simon *et al.*, 2009), and filamentous cyanobacteria were detected in cold aquatic habitats (Strunecký *et al.*, 2012). Furthermore, taxonomic binning of mRNA from Westerhöfer Bach biofilm showed high similarities to the phylogenetic composition of the 16S rRNA gene analysis (see Figure 6 in Chapter B3) and, therefore, emphasized the reliability of 16S rRNA gene studies.

In 2002, the microbial mat of Lake 21 was sampled by Arp *et al*. (2012) and revealed first insights into its prokaryotic community composition. Additionally, an analysis of the bacteriohopanepolyols was carried out (Blumenberg *et al.*, 2013), which further emphasized the results by Arp *et al*. (2012). Besides the obvious maturation of the microbial mat over one decade, their findings were in accordance to the results obtained in this study. The maturation was reflected by the structural setup of the current microbial mat, which showed more layers (9 compared to 4) than the microbial mat of 2002 investigated by Arp *et al*. (2012). These nine layers were separated into three major mat zones with respect to their main characteristics. The prokaryotic community composition was in accordance with the pH and redox potential

measurements. In the photic-oxic zone mainly aerobic and phototrophic prokaryotes were detected, whereas in the transition and anoxic mat zones a lower pH and redox potential was determined, which was accompanied by the presence of anaerobic prokaryotes. It was hypothesized that the deepest layers resemble the old and overgrown mat from 2002 (Schneider *et al.*, 2013). In comparison to the microbial mat of salt pond 4 from Guerrero Negro (Harris *et al.*, 2013) the prokaryotic community structures showed many similarities. However, the microbial mat of Lake 21 showed an overall lower diversity when compared to pond 4, which might be mainly due to the difference in salinity (80‰ versus 170‰) of the surrounding environment. This is exemplified by the cyanobacterial populations, which are in direct contact to the saline water. Only highly halotolerant *Cyanobacteria* (*Halothece*) were able to grow under the high salinities observed (Garcia-Pichel *et al.*, 1998), whereas a more diverse cyanobacterial community existed in Guerrero Negro (*Microcoleus*, *Leptolyngbya*, *Halothece*). This was also true for the other phyla occurring in the microbial mats. Generally, lower salinities seem to support growth of a more diverse microbial community spectrum. Taking the current studies on microbial mat communities into account (see (Schneider *et al.*, 2013) and references therein), the following phyla seem to be part of the main setup of the analyzed biofilm ecosystems: *Cyanobacteria*, *Proteobacteria*, *Bacteroidetes*, *Spirochaetes*, *Chloroflexi*, *Firmicutes*, *Halobacteria*, *Thermoplasmata*, and *Thaumarchaeota* (Marine benthic group B).

1.2 Metatranscriptomic analysis of a karst water biofilm sample

Metatranscriptomics opens a new view on microbial communities. While metagenomics tells us ‘who is there?’ and enables suggestions on ‘what might they do?’, metatranscritpomics enables us to see ‘who is active?’ and also ‘what are they doing?’. However, metatranscriptomics is a challenging task when it comes to the enrichment of mRNA, as the removal of rRNA, making up to 95% of the total RNA, causes several problems (Jansson *et al.*, 2012). Additionally, it has to be considered that the abundance of gene transcripts from a metatranscriptome at a certain time point and the actual protein abundance within a microbial community (and therefore the biological activity), might not correlate directly due to drastically different half-life periods between RNA and proteins (Moran *et al.*, 2013). Nevertheless, the abun-

dance of rRNA and mRNA provides important information on the genetic activity of whole communities at certain time points and environmental conditions, and, thus, lead to conclusions on the active metabolic features of a community.

In this study, 175,391 RNA sequences from the Westerhöfer Bach (sampling site 3) were analyzed after reverse transcription into cDNA and subsequent sequencing (Schneider *et al.*, 2013, submitted). After RNA isolation, a subtractive hybridization RNA removal method (Ribominus) was used. Besides the removal of several rRNA fragments (verified with an Agilent Bioanalyzer), only 5.1% of the metatranscriptomic dataset were in fact real non-rRNA sequences. This is in the range of other metatranscriptomic studies. For example, a metatranscriptome from the human gut exhibited a non-rRNA content of 6.8%, but no removal of rRNA was performed (Gosalbes *et al.*, 2011). In another study of artificial biofilms the use of MICROBExpress (Ambion) led to an average mRNA enrichment of 25.5% (Yergeau *et al.*, 2012). The different rRNA removal efficiencies might be mainly attributed to fragmentation of rRNA molecules during downstream processes, i.e., by a too harsh RNA isolation step (e.g., bead beating, heating), and subsequent purification and concentration steps (shearing by filter membrane). Subtractive rRNA removal techniques use small (approximately 20 bp) oligonucleotide probes targeting conserved regions of the ribosomal RNA genes. Due to too harsh isolation procedures RNA gets fragmented and, consequently, only rRNA fragments corresponding to the probe can be removed, whereas other fragments remain in the sample. When the RNA is fragmented even further, the described subtractive technique gets continuously less efficient. Additionally, He *et al*. (2010) showed for a simple artificial community of five microorganisms that by the subtractive method employed rRNA-probes might be too specific for metatranscriptomic analyses.

However, when trying to assess a metatranscriptome, the goal is to obtain RNA from all cells within a given environmental sample, which requires harsh extraction methods. Possibilities to overcome this bottleneck could be higher sequencing efforts, i.e., using Illumina sequencing instead of 454 pyrosequencing, to get more transcriptional information at same costs (while sacrificing read length). Another promising approach is the application of self-made, sample specific rRNA probes for the subtractive removal of unwanted rRNA genes. This has been successfully em-

ployed for a planktonic microbial community; nearly 55% rRNA reads could be removed (Stewart *et al.*, 2010).

The analysis of mRNA showed that, besides the problematic removal of rRNA, information on specific active biological functions within a microbial community can be attained. This encompassed gene transcripts involved in photosynthesis, protein metabolism, carbohydrate metabolism, RNA metabolism, and respiration. This was in accordance with the bacterial community analysis, as the biofilm was mainly populated by *Oscillatoriales*. Additionally, the taxonomic affiliation of the mRNA sequences confirmed the RNA-based 16S rRNA gene analysis and *vice versa*.

1.3 Calcification in biofilm systems – linked to metabolic activity?

Microbialites encompass all mineral deposits such as calcium carbonate, gypsum, or iron oxides that are formed by microbial activity. Special layered forms of microbialites are stromatolites. They provide currently the best evidence of early (microbial) life and arose sometime before 3.5 Ga (Allwood *et al.*, 2006; Bontognali *et al.*, 2012; Van Kranendonk *et al.*, 2008). Since the first stromatolites emerged, evolution has led to highly diverse microbial consortia. Today's biofilm systems bear the potential to gain insights into ancient microbial life, as different kinds of biomineralization processes are catalyzed within biofilms, which are good analogs of fossil findings. Many microbialite-producing systems are studied (see (Arp *et al.*, 2010; Arp *et al.*, 2012; Schneider *et al.*, 2013; Schneider *et al.*, 2013, submitted) and references therein). These comprise prominent natural ecosystems such as the stromatolites of Shark Bay (Allen *et al.*, 2009), tufa systems and microbial mat systems. Correspondingly, directed biomineralization by eukaryotes (e.g., corals, mussels, diatoms) is known, whereas the bacterial mechanisms are less studied. Although, magnetotactic bacteria can control biomineralization, it is assumed that bacterial-driven mineralization might be an indirect process, and a side reaction of metabolism under specific conditions (Gonzalez-Munoz *et al.*, 2010).

There are several studies, which show that many species are able to form microbialites (Gonzalez-Munoz *et al.*, 2010; Jroundi *et al.*, 2010; Pedley, 2013; Rusznyak *et al.*, 2012). For example, approximately 20% of the calcium carbonate precipitation in Westerhöfer Bach could be linked to photosynthesis of cyanobacteria,

which is induced by removal of CO_2 (Arp *et al.*, 2010) (Figure 8). This increases pH values in the direct proximity of photosynthetic microorganisms and promotes carbonate precipitation. Also other microorganisms such as green and red algae or diatoms may contribute to tufa formation, but also non-phototrophic bacteria seem to be

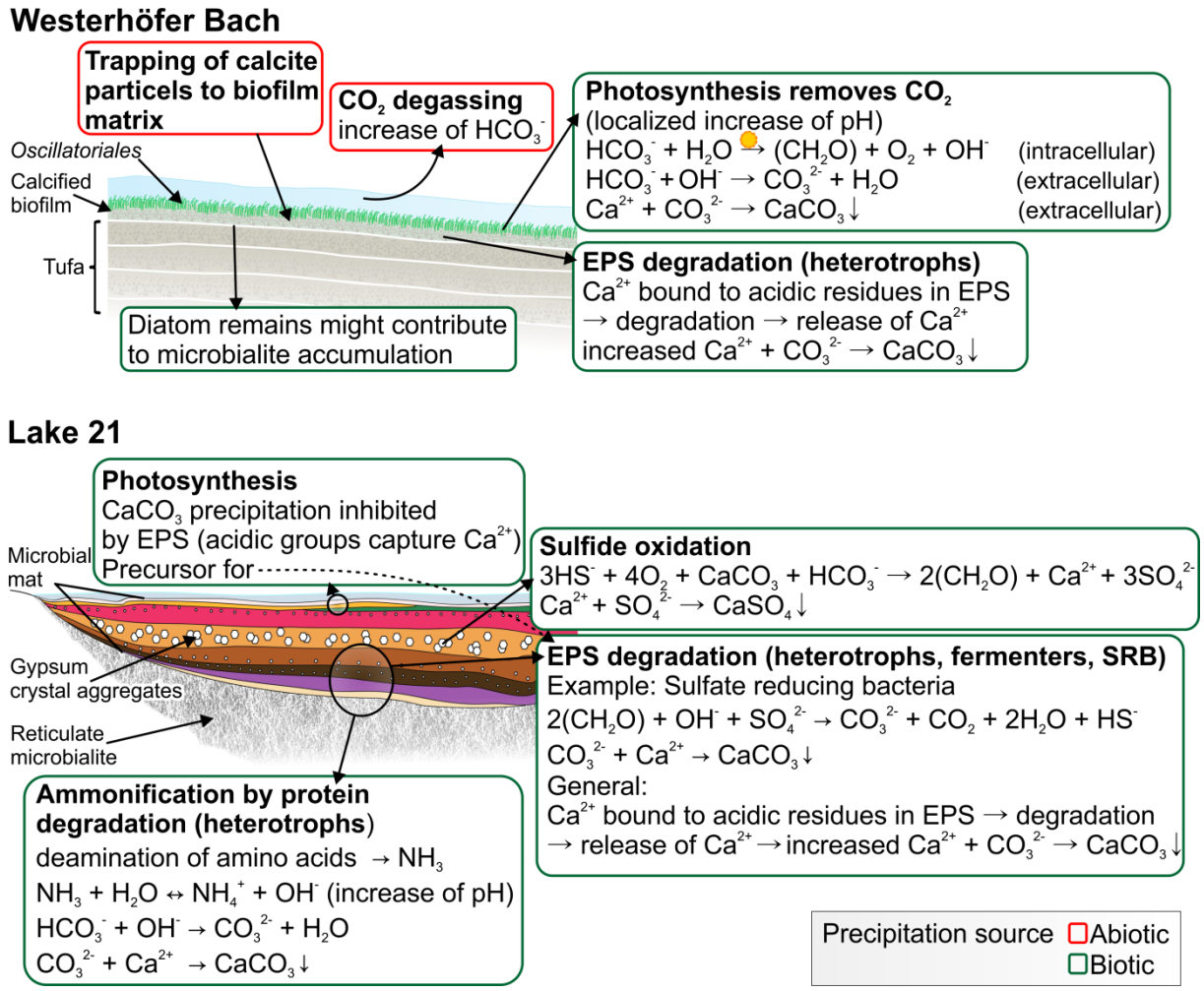


Figure 8. Model of microbialite formation in Westerhöfer Bach biofilm and Lake 21. In Westerhöfer Bach carbonate is precipitated by photosynthesis and abiotic processes (for details see (Arp *et al.*, 2010; Gonzalez-Munoz *et al.*, 2010)). In the microbial mat of Lake 21, photosynthesis serves as a precursor for microbialite formation. Ca^{2+} is captured by EPS. In the lower mat sections EPS is degraded by microorganisms, leading to the release of Ca^{2+} and microbialite precipitation (for details see (Arp *et al.*, 2012; Dupraz *et al.*, 2009; Schneider *et al.*, 2013)).

involved as revealed by a recent study of artificial freshwater tufa systems (Pedley, 2013). In these artificial systems microbialite formation was observed in direct proximity of bacteria in cloud-like EPS, which have also been observed in Westerhöfer Bach (Zippel and Neu, 2011). It is likely that components of the EPS promote or serve as nucleation sites for biomineralization (Dupraz *et al.*, 2009). The identification

of microorganisms incorporated in these cloud-like EPS fabrics would be interesting, but small scale *in situ* sampling and subsequent 16S rRNA gene analysis currently seems impossible. Additionally, for members of the *Myxococcales* it was shown that they possess the ability to precipitate a wide range of minerals under specific environmental situations (Gonzalez-Munoz *et al.*, 2010). Members of this order were also detected in Westerhöfer Bach biofilm. Generally, the construction of a physical robust extracellular biofilm skeleton as an extension of the EPS seems advantageous to protect the prokaryotic community from harms such as shearing forces or eukaryotic grazers.

In the microbial mat of Lake 21, immediate photosynthesis driven carbonate precipitation is inhibited by amorphous EPS where acidic residues trap Ca^{2+}. In the lower mat sections, degradation of Ca^{2+} chelating EPS components by heterotrophic, fermenting bacteria, and sulfate-reducing bacteria might trigger microbialite formation (Arp *et al.*, 2012). Another example for microbialite formation are the large gypsum crystal aggregates in layer 5 of the microbial mat of Lake 21 (see Schneider *et al*. (2013) Figure 2B). These calcium sulfate precipitates could be initiated by hydrogen sulfide oxidation of *Chromatiaceae*.

Apparently, more research is still needed on this topic owing to the plethora of different types of minerals and microorganisms. Recently, potential applications of microbialite formation for remediation and restoration of buildings have been considered (Dhami *et al.*, 2013; Raut *et al.*, 2013). Finally, it seems that under specific environmental conditions a broad range of microbes comprising different metabolic life styles can directly or indirectly trigger biomineralization (Gonzalez-Munoz *et al.*, 2010).

1.4 Bacterial and archaeal communities in Kenyan Soda Lakes

The results of this large scale investigation of prokaryotic community composition of soda lakes corresponds to previous studies, which revealed the existence of all major trophic groups (Duckworth *et al.*, 1996; Grant *et al.*, 1990; Jones *et al.*, 1998; Mesbah *et al.*, 2007; Zavarzin, 2007; Zavarzin *et al.*, 1999). Most culture-independent approaches were based on small survey sizes due to traditional 16S rRNA gene clone libraries (Mwirichia *et al.*, 2010, 2011; Rees *et al.*, 2004). In this study, the extent of

prokaryotic diversity in and around the soda lakes Bogori, Sonachi, Elmenteita, and Magadi has been shown. These were mostly represented by yet uncultured prokaryotic microorganisms.

Overall, the prokaryotic community composition of the different lakes and the different sampling types showed several similarities as depicted by Figure 9. The sample type obviously had the most impact on the bacterial communities. The PCA shows that soil and sediment samples possessed a more related bacterial community composition than the water samples. Wet sediment samples tend in direction of the water samples, indicating an influence of humidity on the bacterial community composition. All water samples were dominated by *Cyanobacteria*. This was expected since it is known that the cyanobacteria constitute the major fraction of the microbial diversity within the East African soda lakes and, for example, serve as food to the Flamingos. Interestingly, different groups of *Cyanobacteria* (Bogoria: *Arthrospira platensis*; Sonachi and Elmenteita: *Synechococcus* sp.) predominate the water of the different lakes, possibly due to different physicochemical characteristics (TDS of 1 mg/l in Bogoria compared to 16 and 19 mg/l in Sonachi and Elmenteita, respectively). Finally, the 16S rRNA gene analysis shows similarities to previous reports arguing that certain groups may be predominant in similar but geographically

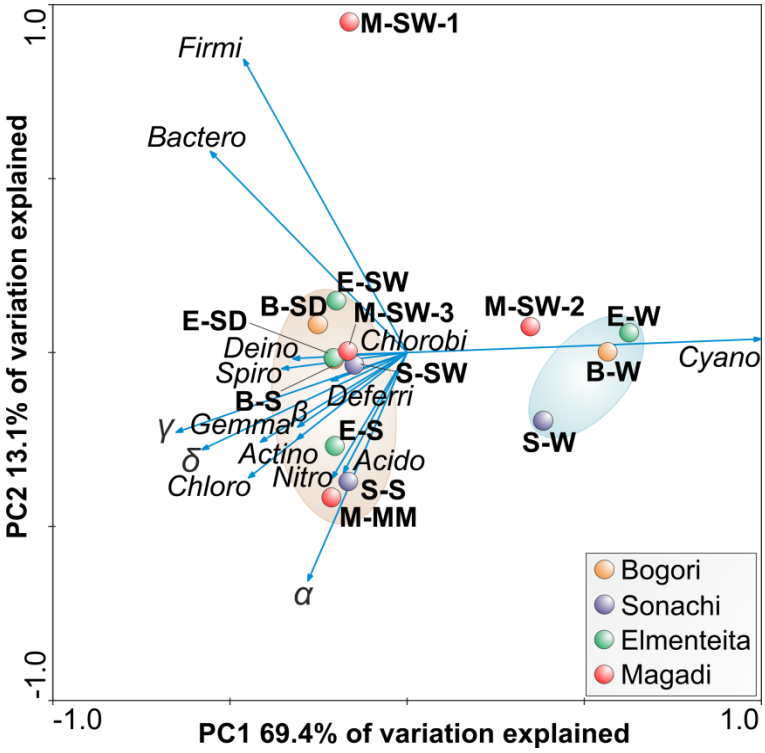


Figure 9. PCA of the bacterial community composition at phylum and proteobacterial class level of the investigated Kenyan soda lakes. Abbreviations: *α*, *Alphaproteobacteria*; *β*, *Betaproteobacteria*; *γ*, *Gammaproteobacteria*; *δ*, *Deltaproteobacteria*; *Acido*, *Acidobacteria*; *Actino*, *Actinobacteria*; *Deino*, *Deinococcus-Thermus*; *Nitro*, *Nitrospirae*; *Chloro*, *Chloroflexi*; *Gemma*, *Gemmatimonadetes*; *Deferri*, *Deferribacteres*; *Firmi*, *Firmicutes*; *Spiro*, *Spirochaetes*; *Bactero*, *Bacteroidetes*; *Cyano*, *Cyanobacteria*. B, Lake Bogoria; S, Lake Sonachi, E, Lake Elmenteita; M, Lake Magadi; S, soil sample; SD, dry sediment sample; SW, wet sediment sample; W, water sample; MM, microbial mat.

distant environments, although several differences on order level and lower were observed. A 16S rRNA gene analysis on RNA level would be interesting to reveal which of the observed prokaryotes are potentially active or persist in dormant states.

1.5 Electroactive biofilm and plankton communities

Bioelectrical systems (BES) are microbial consortia with the potential application as renewable energy source (Borole *et al.*, 2011) due to their ability to produce electricity by metabolizing specific substrates. Biofilm-forming electrochemical active taxa are able to utilize the biomass in wastewater to produce an electric current, while cleaning the water simultaneously. Two main types of bioelectrical systems are known: the microbial fuel cell (MFC) and the microbial electrolysis cell (MEC), which is additionally able to produce fuels (e.g., hydrogen, ethanol, etc.) (Borole *et al.*, 2011). A BES consists of a biological anode and/or a biological cathode in which microorganisms function as catalysts by reducing substrates. In these systems the microorganisms are present as planktonic cells and biofilms. The microorganisms themselves can be electrochemically active or inactive. Even electrochemically inactive microorganisms can be indirectly involved in electricity production by metabolically supporting electrochemically active bacteria (Franks and Kelly, 2010). However, current research on these systems has clearly shown that major efforts for improvement of the techniques are required before BES will be practically relevant (Borole *et al.*, 2011; Harnisch and Rabaey, 2012).

Within the context of this thesis, the bacterial community of plankton and biofilm of ten differentially incubated MFCs were analyzed via pyrosequencing of the 16S rRNA gene as described in Schneider *et al*. (2013). The active fraction within wastewater was mainly constricted to *Epsilonproteobacteria* (*Arcobacter*) and *Fusobacteria* (*Leptotrichiaceae*), whereas the entire bacterial community showed a broader spectrum and additionally included *Firmicutes* (*Bacilli* and *Clostridia*), *Betaproteobacteria* (*Burkholderiales*), *Gammaproteobacteria* (*Enterobacteriales*, *Aeromonadales*, *Pseudomonadales*), and *Bacteroidetes* (*Prevotella*). This example shows clearly the differences between active (RNA-based) and entire (DNA-based) bacterial communities, demonstrating the advantages of comparing both approaches. This was further emphasized by the level of diversity, as indicated by the Shannon indices

of 1.8 (RNA) and 4.29 (DNA) for wastewater. This suggests that under “natural” conditions only a few bacteria are able to perform an active metabolism in wastewater, whereas the majority of the highly diverse bacterial community persists in dormant states (Lennon and Jones, 2011). However, Blazewicz *et al*. (2013) argued the careful interpretation of RNA-derived 16S rRNA gene data, due to different correlations between rRNA ratios of active, growing, dormant, and deceased prokaryotic cells.

As expected the bacterial community compositions of MFCs grown with different substrates differed widely. For example, the planktonic communities of acetate-, citric acid-, glycerol-, and mannitol-grown MFCs were dominated by members of the *Bacteroidetes* and *Proteobacteria*, whereas the plankton of citric acid grown MFC was mainly dominated by *Gammaproteobacteria*, and the oleic acid grown MFC by *Betaproteobacteria*. Also the corresponding biofilm communities differed in several cases, e.g., the acetate biofilm (replicate 1) was dominated by *Deltaproteobacteria* (unclassified *Geobacter* sp.). Another observation was that the biological replicates of many samples led to different bacterial communities and diversities. The acetate grown MFCs differed in their community composition and also in the measured diversity and electric activity. Whereas experiment 1 supplied high electric current at high efficiency, the second experiment was not successful. Obviously, despite similar conditions different bacterial biofilm communities established.

In this study, high electric current was linked to high abundance of members of *Geobacteriaceae*, which are known for high electroactive potential (Kimura and Okabe, 2013). However, not always an efficient electroactive system has established in this survey, maybe due to the low amount of *Geobacter* species within the inoculum. Furthermore, in the glycerol-grown biofilm of MFC (replicate 1) high abundance of *Deltaproteobacteria* was detected, but only low electric current was produced. This might be due to an electroinactive *Geobacter* species (from candidate group AKYG597), which was generally associated with low electric currents. Additionally, uncultivated *Clostridium* sp. were observed in the active bacterial communities of mannitol-grown MFCs, indicating a connection between these bacteria and electric current production. Indeed several studies show the potential application of members of the *Clostridiales* for BES (Finch *et al.*, 2011; Mathuriya and Sharma, 2009; Niessen *et al.*, 2005).

Overall the analysis of MFC communities showed that the substrates acetate, citric acid, and mannitol can be converted most efficiently to electric current by the established microbial communities. Furthermore, the results of the phylogenetic analysis illustrate a nearly positive linear connection between the abundance of *Geobacteraceae* and electric current production on DNA and RNA level (see Figure 4.14 in (Baudler, 2012)). Interestingly, this particular family was nearly absent in the inoculum (neither DNA nor RNA, abundance <0.1%), indicating a preferential growth of this taxon in these systems. Similar observations were made in another study of a BES (Dennis *et al.*, 2013). However, as the biological replicates show high dissimilarities in their prokaryotic community composition and (in presumably linked) electric current production, additional tests should be performed to get more meaningful and representative results. Moreover, an optimization of the inoculum should be taken into account.

2 Discovery of novel biocatalysts

Several industrial relevant enzymes have been identified by metagenomic techniques. These include lipases, esterases, amylases, glucanases, cellulases, xylanases, chitinases, proteases and many more (Shestakov, 2012). The success of these approaches shows the inherent potential of metagenomics as a resource for novel biocatalysts.

In this study, biofilm and tufa of the Westerhöfer Bach have been used for the construction of metagenomic large-insert fosmid libraries containing directly cloned environmental DNA that can be screened to identify genes encoding novel biocatalysts. Fosmids can be used to clone large (approximately between 20 to 40 kb) environmental DNA fragments (Simon and Daniel, 2011) with the advantage (when compared to plasmids) that they can cover whole gene clusters and operons (Daniel, 2005).

2.1 Identification and analysis of proteolytic genes

In order to isolate genes encoding proteolytic activity from the metagenomic libraries of Westerhöfer Bach, a simple activity-based agar plate assay using skim milk as a

substrate was chosen. From the initially detected 37 positive fosmid clones, five unique clones that harbor partially novel genes have been identified from biofilm and tufa fosmid libraries, indicating a sufficient coverage of the genetic content of the Westerhöfer Bach biofilm and tufa. Bacterial proteolytic enzymes are hierarchically subdivided into 225 different families (including subfamilies) and 44 clans based on amino acid homology and biochemical properties (Rawlings *et al.*, 2012). Other studies with the aim to discover proteases of metagenomes from diverse environments (e.g., soil, compost, goat skin, mud, arctic and hot spring sediment), encompassed metalloproteases (Lee *et al.*, 2007; Waschkowitz *et al.*, 2009), but mainly serine proteases (Neveu *et al.*, 2011; Niehaus *et al.*, 2011; Purohit and Singh, 2013; Pushpam *et al.*, 2011; Wemheuer *et al.*, 2013; Zhang *et al.*, 2011) have been detected. In this study, all of the deduced protein sequences of the proteolytic genes were classified as members of the serine families S1B, S1C, and S8A (Schneider *et al.*, 2013, submitted) (Figure 10). Additionally, within the course of a master thesis (Engelhaupt, 2011) a serine protease (family S8A) from a microbial mat (Lake 2, Kiritimati) has been detected, further emphasizing the high abundance of serine proteases in diverse environments. However,

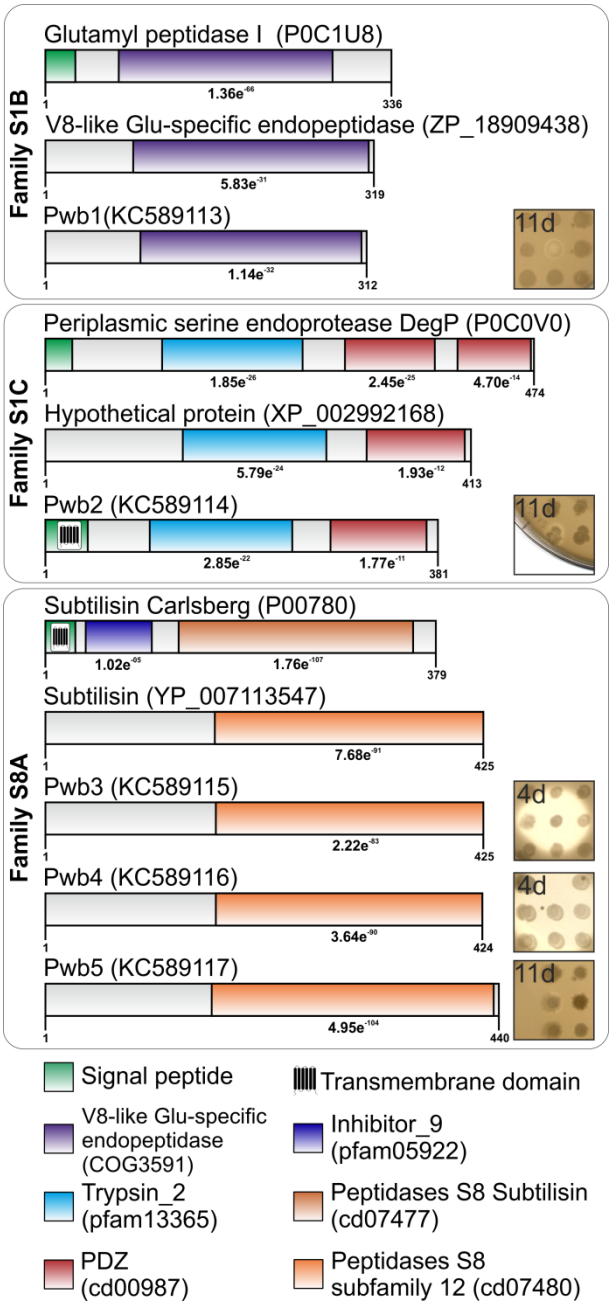


Figure 10. Domain structure of proteolytic gene products Pwb1 – Pwb5 detected within this thesis. Additionally, the domain structures of the representative protein for each family and the next related proteins are shown. Initial screening pictures with halo formation are shown on the right site, including days of incubation until activity was noticed. Further details can be found in Schneider *et al.* [2013, submitted].

the high occurrence of serine proteases in metagenomic screenings remains speculative. It might be linked to the high possibility of heterologous expression due to this family's primordial occurrence along the domain *Bacteria*. Interestingly, Gabor *et al*. (2012) detected microdiversity, i.e., small differences (up to 6 amino acids) in the amino acid sequence of serine proteases of the Subtilisin Carlsberg type within a species. They hypothesized a general genetic dynamic of genes in microbial consortia, which then would increase the chance of detecting different versions of this kind of protease. This would explain the high similarity on amino acid level (85%) observed for the proteases Pwb3 and Pwb4. Generally, bacterial subtilases are non-specific serine endopeptidases involved in nutrient supply (Rawlings *et al.*, 2012), which might be another reason of the ubiquity of this protease type.

2.2 Identification and analysis of cellulolytic genes

Currently, metagenome-derived cellulases are more or less rare. This might be attributed to the low hit rate when identifying cellulolytic enzymes by function-based screenings, possibly linked to the host's inability of secretion or assembly of the foreign genes. For example, in this study screening of approximately 1.05 Gbp of DNA led to the identification of one cellulolytic gene(cluster) from Westerhöfer Bach biofilm. However, screening of soil metagenomic libraries recently led to the detection of one cellulolytic enzyme out of 0.36 Gbp (Nacke *et al.*, 2012), indicating that the high bacterial diversity of soil might be a more promising habitat to screen for cellulolytic enzymes. Even more suitable, compared to soil, are bacterial communities, which utilize cellulosic substrate (e.g., wood) as main nutrient source, for example, the microflora of the beaver gut. Thereby, six unique (hemi)cellulolytic fosmid clones were detected from only 5,800 clones (0.18 Gbp) screened (Wittenberg, 2009).

The initial biochemical characterization of the gene product of ORF 11 (Cwb1) showed its ability to hydrolyze β-D-glucan and carboxymethyl-cellulose (CMC). ORF 9 and 10 showed no activity on the tested substrates, probably due to substrate specificity. Interestingly, the temperature optimum of Cwb1 was at 40°C, which is 30°C above the environmental temperature. This deviance in temperature was also observed for other metagenome-derived enzymes. Several enzymes with unexpected characteristics have been detected by metagenomic studies (Nacke *et al.*,

2012; Waschkowitz *et al.*, 2009). The genetic organization suggests a synergistically mode of substrate degradation, which is known for many cellulolytic systems that are composed of several proteins performing different tasks (Lynd *et al.*, 2002). However, only a small substrate spectrum was tested and, possibly, products of the genes detected use a certain substrate occurring in the biofilm. Therefore, further experiments should be carried out to determine the specific function of this novel cellulolytic apparatus and its potential use for industrial applications.

3 Conclusions

In conclusion, this study revealed the prokaryotic community compositions of different, partially extreme, habitats. A broad spectrum of bacterial and archaeal phyla and candidate divisions was discovered, though many of them are currently uncharacterized. Thereby, this study provided a comparatively detailed idea of the complex prokaryotic communities thriving in the investigated environments. Nevertheless, further research should be conducted to obtain additional insights into the functional repertoire within these microbial assemblages. This could be accomplished by random shotgun sequencing of metagenomic DNA, for example, by employing the new promising Moleculo Long Read technique (Illumina). Additionally, it has been shown that the enormous prokaryotic diversity of these partially extreme habitats can be successfully exploited for the discovery of novel biocatalysts.

4 References

Allen, M.A., Goh F., Burns B.P., Neilan B.A. (2009). Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. Geobiology **7**: 82-96.

Allwood, A.C., Walter M.R., Kamber B.S., Marshall C.P., Burch I.W. (2006). Stromatolite reef from the Early Archaean era of Australia. Nature **441**: 714-718.

Arp, G., Bissett A., Brinkmann N., Cousin S., de Beer D.*, et al.* (2010). Tufa-forming biofilms of German karstwater streams: microorganisms, exopolymers, hydrochemistry and calcification. The Geological Society of London **336**: 83-118.

Arp, G., Helms G., Karlinska K., Schumann G., Reimer A.*, et al.* (2012). Photosynthesis versus Exopolymer Degradation in the Formation of Microbialites on the Atoll of Kiritimati, Republic of Kiribati, Central Pacific. Geomicrobiology Journal **29**: 29–65.

Baudler, A. (2012). Elektrochemisch aktive Biofilme auf Basis mikrobieller Konsortien: Grundlagenuntersuchung zur Verwendung von Mono - und Mischsubstraten (Master thesis). Braunschweig: Technische Universität Braunschweig. 106 p.

Beraldi-Campesi, H., Arenas-Abad C., Garcia-Pichel F., Arellano-Aguilar O., Auque L.*, et al.* (2012). Benthic bacterial diversity from freshwater tufas of the Iberian Range (Spain). FEMS Microbiol Ecol **80**: 363-379.

Blazewicz, S.J., Barnard R.L., Daly R.A., Firestone M.K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. Isme J.

Blumenberg, M., Arp G., Reitner J., Schneider D., Daniel R.*, et al.* (2013). Bacteriohopanepolyols in a stratified cyanobacterial mat from Kiritimati (Christmas Island, Kiribati). Organic Geochemistry **55**: 55–62.

Bontognali, T.R., Sessions A.L., Allwood A.C., Fischer W.W., Grotzinger J.P.*, et al.* (2012). Sulfur isotopes of organic matter preserved in 3.45-billion-year-old stromatolites reveal microbial metabolism. Proc Natl Acad Sci U S A **109**: 15146-15151.

Borole, A.P., Reguera G., Ringeisen B., Wang Z., Feng Y.*, et al.* (2011). Electroactive biofilms: Current status and future research needs. Energy & Environmental Science 4813-4834.

Breitbart, M., Hoare A., Nitti A., Siefert J., Haynes M.*, et al.* (2009). Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Cienegas, Mexico. Environ Microbiol **11**: 16-34.

Brinkmann, N., Hodač L., Mohr K.I., Hodačová A., Jahn R.*, et al.* (2013, submitted). Cyanobacteria and diatoms in biofilms of two karstic streams in Germany and changes of their communities along calcite saturation gradients. Geomicrobiology Journal.

Caporaso, J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D.*, et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7**: 335-336.

Cousin, S., Stackebrandt E. (2010). Spatial Bacterial Diversity in a Recent Freshwater Tufa Deposit. Geomicrobiology Journal **27**: 275-291.

Daniel, R. (2005). The metagenomics of soil. Nat Rev Microbiol **3**: 470-478.

Dennis, P.G., Harnisch F., Yeoh Y.K., Tyson G.W., Rabaey K. (2013). Dynamics of cathode-associated microbial communities and metabolite profiles in a glycerol-fed bioelectrochemical system. Appl Environ Microbiol **79**: 4008-4014.

Dhami, N.K., Reddy M.S., Mukherjee A. (2013). *Bacillus megaterium* mediated mineralization of calcium carbonate as biogenic surface treatment of green building materials. World J Microbiol Biotechnol.

Duckworth, A.W., Grant W.D., Jones B.E., van Steenbergen R. (1996). Phylogenetic diversity of soda lake alkaliphiles. FEMS Microbiol Ecol **19**: 181-191.

Dupraz, C., Reid R.P., Braissant O., Decho A.W., Norman R.S.*, et al.* (2009). Processes of carbonate precipitation in modern microbial mats. Earth-Science Reviews **96**: 141-162.

Engelhaupt, M. (2011). Microbial diversity in a hypersaline microbial mat from Kiritimati (Christmas Island) based on 16S ribosomal RNA gene analysis (Master thesis). Göttingen: Georg-August Universität. 112 p.

Finch, A.S., Mackie T.D., Sund C.J., Sumner J.J. (2011). Metabolite analysis of *Clostridium acetobutylicum*: fermentation in a microbial fuel cell. Bioresour Technol **102**: 312-315.

Franks, A.E., Kelly P.N. (2010). Microbial Fuel Cells, A Current Review. Energies **3**: 899-919.

Gabor, E., Niehaus F., Aehle W., Eck J. (2012). Zooming in on metagenomics: molecular microdiversity of Subtilisin Carlsberg in soil. J Mol Biol **418**: 16-20.

Garcia-Pichel, F., Nubel U., Muyzer G. (1998). The phylogeny of unicellular, extremely halotolerant cyanobacteria. Arch Microbiol **169**: 469-482.

Gonzalez-Munoz, M.T., Rodriguez-Navarro C., Martinez-Ruiz F., Arias J.M., Merroun M.L.*, et al.* (2010). Bacterial biomineralization: new insights from *Myxococcus*-induced mineral precipitation. Geological Society, London **336**: 31-50.

Gosalbes, M.J., Durban A., Pignatelli M., Abellan J.J., Jimenez-Hernandez N.*, et al.* (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS One **6**: e17447.

Grant, W.D., Mwatha W.E., Jones B.E. (1990). Alkaliphiles: Ecology, diversity and applications. FEMS Microbiol Lett **75**: 255–269.

Harnisch, F., Rabaey K. (2012). The diversity of techniques to study electrochemically active biofilms highlights the need for standardization. ChemSusChem **5**: 1027-1038.

Harris, J.K., Caporaso J.G., Walker J.J., Spear J.R., Gold N.J.*, et al.* (2013). Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. Isme J **7**: 50-60.

He, S., Wurtzel O., Singh K., Froula J.L., Yilmaz S.*, et al.* (2010). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Methods **7**: 807-812.

Jansson, J.K., Neufeld J.D., Moran M.A., Gilbert J.A. (2012). Omics for understanding microbial functional dynamics. Environ Microbiol **14**: 1-3.

Jones, B.E., Grant W.D., Duckworth A.W., Owenson G.G. (1998). Microbial diversity of soda lakes. Extremophiles **2**: 191-200.

Jroundi, F., Fernandez-Vivas A., Rodriguez-Navarro C., Bedmar E.J., Gonzalez-Munoz M.T. (2010). Bioconservation of deteriorated monumental calcarenite stone and identification of bacteria with carbonatogenic activity. Microb Ecol **60**: 39-54.

Kimura, Z., Okabe S. (2013). Acetate oxidation by syntrophic association between *Geobacter sulfurreducens* and a hydrogen-utilizing exoelectrogen. Isme J **7**: 1472-1482.

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

Lee, D.G., Jeon J.H., Jang M.K., Kim N.Y., Lee J.H.*, et al.* (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. Biotechnol Lett **29**: 465-472.

Lee, Z.M., Bussema C., 3rd, Schmidt T.M. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res **37**: D489-493.

Lennon, J.T., Jones S.E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol **9**: 119-130.

Liu, Z., Lozupone C., Hamady M., Bushman F.D., Knight R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res **35**: e120.

Lynd, L.R., Weimer P.J., van Zyl W.H., Pretorius I.S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev **66**: 506-577, table of contents.

Mathuriya, A.S., Sharma V.N. (2009). Bioelectricity production from paper industry waste using a microbial fuel cell by *Clostridium* species. Journal of Biochemical Technology **1**: 49-52.

Mesbah, N.M., Abou-El-Ela S.H., Wiegel J. (2007). Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. Microb Ecol **54**: 598-617.

Mobberley, J.M., Ortega M.C., Foster J.S. (2012). Comparative microbial diversity analyses of modern marine thrombolitic mats by barcoded pyrosequencing. Environ Microbiol **14**: 82-100.

Moran, M.A., Satinsky B., Gifford S.M., Luo H., Rivers A.*, et al.* (2013). Sizing up metatranscriptomics. Isme J **7**: 237-243.

Mwirichia, R., Cousin S., Muigai A.W., Boga H.I., Stackebrandt E. (2010). Archaeal diversity in the haloalkaline Lake Elmenteita in Kenya. Curr Microbiol **60**: 47-52.

Mwirichia, R., Cousin S., Muigai A.W., Boga H.I., Stackebrandt E. (2011). Bacterial diversity in the haloalkaline Lake Elmenteita, Kenya. Curr Microbiol **62**: 209-221.

Nacke, H., Engelhaupt M., Brady S., Fischer C., Tautzt J.*, et al.* (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett **34**: 663-675.

Neveu, J., Regeard C., DuBow M.S. (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. Appl Microbiol Biotechnol **91**: 635-644.

Ng, C.C., Huang W.C., Chang C.C., Tzeng W.S., Chen T.W.*, et al.* (2006). Tufa microbial diversity revealed by 16S rRNA cloning in Taroko National Park, Taiwan. Soil Biol Biochem **38**: 342-348.

Niehaus, F., Gabor E., Wieland S., Siegert P., Maurer K.H.*, et al.* (2011). Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. Microb Biotechnol **4**: 767-776.

Niessen, J., Schroder U., Harnisch F., Scholz F. (2005). Gaining electricity from in situ oxidation of hydrogen produced by fermentative cellulose degradation. Lett Appl Microbiol **41**: 286-290.

Papineau, D., Walker J.J., Mojzsis S.J., Pace N.R. (2005). Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. Appl Environ Microbiol **71**: 4822-4832.

Pedley, M. (2013). The morphology and function of thrombolitic calcite precipitating biofilms: A universal model derived from freshwater mesocosm experiments. Sedimentology.

Purohit, M.K., Singh S.P. (2013). A metagenomic alkaline protease from saline habitat: cloning, over-expression and functional attributes. Int J Biol Macromol **53**: 138-143.

Pushpam, P.L., Rajesh T., Gunasekaran P. (2011). Identification and characterization of alkaline serine protease from goat skin surface metagenome. AMB Express **1**: 3.

Raut, S.H., Sarode D.D., Lele S.S. (2013). Biocalcification using *B. pasteurii* for strengthening brick masonry civil engineering structures. World J Microbiol Biotechnol.

Rawlings, N.D., Barrett A.J., Bateman A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res **40**: D343-350.

Rees, H.C., Grant W.D., Jones B.E., Heaphy S. (2004). Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. Extremophiles **8**: 63-71.

Robertson, C.E., Spear J.R., Harris J.K., Pace N.R. (2009). Diversity and stratification of archaea in a hypersaline microbial mat. Appl Environ Microbiol **75**: 1801-1810.

Rusznyak, A., Akob D.M., Nietzsche S., Eusterhues K., Totsche K.U.*, et al.* (2012). Calcite biomineralization by bacterial isolates from the recently discovered pristine karstic herrenberg cave. Appl Environ Microbiol **78**: 1157-1167.

Sahm, K., John P., Nacke H., Wemheuer B., Grote R.*, et al.* (2013). High abundance of heterotrophic prokaryotes in hydrothermal springs of the Azores as revealed by a network of 16S rRNA gene-based methods. Extremophiles **17**: 649-662.

Schneider, D., Arp G., Reimer A., Reitner J., Daniel R. (2013). Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the kiritimati atoll, central pacific. PLoS One **8**: e66662.

Schneider, D., Reimer A., Hahlbrock A., Arp G., Reitner J.*, et al.* (2013, submitted). Metagenomic and metatranscriptomic analyses of bacterial communities derived from a calcifying karst water creek biofilm and tufa. Geomicrobiology Journal.

Shestakov, S.V. (2012). Impact of Metagenomics on Biotechnological Development. Applied Biochemistry and Microbiology **48**: 705-715.

Simon, C., Daniel R. (2011). Metagenomic analyses: past and future trends. Appl Environ Microbiol **77**: 1153-1161.

Simon, C., Wiezer A., Strittmatter A.W., Daniel R. (2009). Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. Appl Environ Microbiol **75**: 7519-7526.

Sorensen, K.B., Canfield D.E., Teske A.P., Oren A. (2005). Community composition of a hypersaline endoevaporitic microbial mat. Appl Environ Microbiol **71**: 7352-7365.

Spear, J.R., Ley R.E., Berger A.B., Pace N.R. (2003). Complexity in natural microbial ecosystems: the Guerrero Negro experience. Biol Bull **204**: 168-173.

Stewart, F.J., Ottesen E.A., DeLong E.F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. Isme J **4**: 896-907.

Strunecký, O., Komárek J., Elster J. (2012). Biogeography of *Phormidium autumnale* (*Oscillatoriales*, *Cyanobacteria*) in western and central Spitsbergen. Polish Polar Research **33**: 369-382.

Sun, D.L., Jiang X., Wu Q.L., Zhou N.Y. (2013). Intragenomic Heterogeneity of 16S rRNA Genes Causes Overestimation of Prokaryotic Diversity. Appl Environ Microbiol **79**: 5962-5969.

Van Kranendonk, M.J., Philippot P., Lepot K., Bodorkos S., Pirajno F. (2008). Geological setting of Earth's oldest fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. Precambrian Research **167**: 93-124.

Waschkowitz, T., Rockstroh S., Daniel R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. Appl Environ Microbiol **75**: 2506-2516.

Wemheuer, B., Taube R., Akyol P., Wemheuer F., Daniel R. (2013). Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. Archaea **2013**: 136714.

Wemheuer, B., Wemheuer F., Daniel R. (2012). RNA-based assessment of diversity and composition of active archaeal communities in the German Bight. Archaea **2012**: 695826.

Wittenberg, S. (2009). Metagenomanalysen von zwei Habitaten mit (hemi-)cellulolytischen mikrobiellen Gemeinschaften (Dissertation). Göttingen: Georg-August Universität Göttingen. 206 p.

Yergeau, E., Sanschagrin S., Waiser M.J., Lawrence J.R., Greer C.W. (2012). Sub-inhibitory concentrations of different pharmaceutical products affect the meta-transcriptome of river biofilm communities cultivated in rotating annular reactors. Environ Microbiol Rep **4**: 350-359.

Zavarzin, G.A. (2007). Alkaliphilic microbial communities. Trans Winogradsky Inst Microbiol V, XIV, Nauka, Moscow (in Russian).

Zavarzin, G.A., Zhilina T.N., Kevbrin V.V. (1999). The alkaliphilic microbial community and its functional diversity. Microbiology (Moscow, English Translation) **68**: 503–521.

Zhang, Y., Zhao J., Zeng R. (2011). Expression and characterization of a novel mesophilic protease from metagenomic library derived from Antarctic coastal sediment. Extremophiles **15**: 23-29.

Zippel, B., Neu T.R. (2011). Characterization of glycoconjugates of extracellular polymeric substances in tufa-associated biofilms by using fluorescence lectin-binding analysis. Appl Environ Microbiol **77**: 505-516.

D SUMMARY

In this thesis, the prokaryotic communities in different ecosystems were analyzed by metagenomic and metatranscriptomic approaches. This was accomplished by amplicon-based sequencing and analyses of different hypervariable regions of the prokaryotic 16S rRNA gene, metatranscriptomics, and function-based screening of metagenomic large-insert libraries. To analyze the next-generation sequencing derived 16S rRNA gene sequences, a pipeline based on the QIIME software package, has been established. The pipeline was used to perform quality-filtering, denoising, primer cutting, clustering, taxonomic classification, calculation of diversity indices, and multivariate analysis.

The entire (DNA-based) and active (RNA-based) bacterial community composition and diversity of the Westerhöfer Bach biofilm and tufa were examined using 139,062 quality-filtered 16S rRNA gene sequences. The bacterial diversity in the biofilm increased downstream the creek. Changing physicochemical conditions might explain the observed differences. The Westerhöfer Bach biofilm was dominated by filamentous *Cyanobacteria*, aerobic members of all classes within the *Proteobacteria*, and *Chloroflexi*. Due to low UV-radiation below the biofilm, the tufa harbored more *Proteobacteria* than *Cyanobacteria*. Moreover, the relative abundances of different proteobacterial classes showed obvious alterations between the entire and active bacterial biofilm communities.

The active community functions of one Westerhöfer Bach biofilm sample was studied in more detail. For this purpose a metatranscriptomic survey was carried out. This resulted in the discovery of the most abundant transcripts within the microbial biofilm community and comprised genes involved in photosynthesis, protein metabolism, carbohydrate metabolism, RNA metabolism, and respiration. The results obtained by taxonomic classification of the metatranscriptomic data corresponded to results of the 16S rRNA gene analysis.

To exploit the metagenomic potential of the Westerhöfer Bach biofilm, four metagenomic large-insert DNA libraries were constructed from environmental DNA extracted from the biofilm and tufa samples. The fosmid libraries contained approximately 36,800 clones with average insert sizes ranging from 25.7 to 30.1 kbp. The libraries harbored approximately 1.05 Gbp of environmental DNA. Function-based

screening of the metagenomic libraries for proteolytic and cellulolytic genes resulted in the identification of five unique proteolytic genes and one (hemi)cellulolytic gene cluster.

For the second habitat, Lake 21 on Kiritimati, the prokaryotic community composition of a microbial mat was studied. The mat consisted of nine well-laminated, differently colored mat layers, which were analyzed separately. The bacterial and archaeal composition was analyzed by DNA-based sequencing of the 16S rRNA genes. The dataset comprised 80,045 bacterial and 39,729 archaeal sequences. The prokaryotic community composition was separated into three major mat zones, which were mainly influenced by oxygen concentration and UV-radiation. Overall, the prokaryotic diversity increased, whereas redox potential and pH value decreased with depth. In accordance with the hydrochemical data, the prokaryotic community structure changed from halophile oxygenic and anoxygenic phototrophs and aerobic heterotrophs in the photic-oxic zone to sulfate-reducing bacteria (SRB), fermenting bacteria, and potential sulfate-reducing *Archaea* in the transition zone. In the anoxic zone SRB, fermenters, ammonia-oxidizing *Archaea*, and low numbers of methanogens were detected. The low similarity of the 16S rRNA gene sequences to known and characterized microbes illustrates the presence of novel species in this habitat.

For the Kenyan soda lakes Bogoria, Sonachi, Elmenteita, and Magadi, the prokaryotic diversity of soil-, sediment-, water-, and microbial mat samples was assessed by DNA-based sequencing of 16S rRNA genes. The dataset comprised 110,632 bacterial and 71,316 archaeal high quality sequences. The low similarities of 16S rRNA gene sequences to known species highlighted the existence of novel bacterial and archaeal taxa within these extreme environments.

Finally, the bacterial biofilms and corresponding planktonic communities of ten microbial fuel cells were analyzed by DNA and RNA-based sequencing of 16S rRNA genes. The datasets comprised 178,500 quality-filtered 16S rRNA gene sequences. The obtained data demonstrated that active and entire bacterial communities mostly showed small variations. In general, diversity showed that current-generating cells exhibited lower bacterial diversity than non-producing cells. The recorded data indicated that mainly uncultured members of *Geobacter* and *Clostridium* were linked to the production of electric currents.

E APPENDIX

Westerhöfer Bach, Germany:

Data is deposited on the enclosed CD under Drive:\Westerhöfer_Bach\ and includes a summary of all detected bacterial lineages (Drive:\Westerhöfer Bach/taxa_summary/bar_charts.html), OTU sequences clustered at 1% genetic divergence and corresponding OTU tables (Drive:\Westerhöfer_Bach\OTUs\), the insert of the cellulolytic activity conferring fosmid (Drive:\Westerhöfer_Bach\f50E12.fasta), the metatranscriptom sequences (Drive:\Westerhöfer Bach/metatranscriptom\), and the submitted manuscript (Drive:\Westerhöfer_Bach\manuscripts\).

Lake 21, Kiritimati:

Data is deposited on the enclosed CD under Drive:\Kiritimati\ and includes a summary of all detected bacterial lineages (Drive:\Kiritimati\taxa_summary_bacteria\bar_charts.html), archaeal lineages (Drive:\Kiritimati\taxa_summary_archaea\bar_charts.html), OTU sequences clustered at 3% genetic divergence and corresponding OTU tables for bacteria and archaea (Drive:\Kiritimati\OTUs\), and the corresponding manuscripts (Drive:\Kiritimati\manuscripts\).

Soda lakes Bogoria, Sonachi, Elmenteita, and Magadi, Kenya:

Data is deposited on the enclosed CD under Drive:\Soda_lakes\ and includes a summary of all detected bacterial lineages (Drive:\Soda_lakes\taxa_summary_bacteria\bar_charts.html) and archaeal lineages (Drive:\Soda_lakes\taxa_summary_archaea\bar_charts.html), OTU sequences clustered at 3% genetic divergence and corresponding OTU tables for bacteria and archaea (Drive:\Soda_lakes\OTUs\).

Microbial fuel cells:

Data is deposited on the enclosed CD under Drive:\MFC\ and includes a summary of all detected bacterial lineages (Drive:\MFC\taxa_summary_bacteria\bar_charts.html), OTU sequences clustered at 3% genetic divergence and corresponding OTU table for bacteria (Drive:\MFC\OTUs\), and the corresponding master thesis of André Baudler (Drive:\MFC\Master_thesis_Andre_Baudler.pdf).

ACKNOWLEDGMENTS

First, I would like to express my gratitude to my doctor father Prof. Dr. Rolf Daniel for the opportunity to work on this fascinating and comprehensive topic (including the thrilling trip to Kiritimati). Also many thanks for the constant support during the thesis, the careful reading of my manuscripts and thesis, and the chance to attend several conferences.

I would also like to thank PD Dr. Michael Hoppert for the straightforward acceptance to be my co-examiner and his interest in the topics covered within this thesis.

Special thanks to Silja, Heiko, and Bernd for proofreading parts of this thesis. I would like to thank Steffi and Mechthild for constant support with sampling and whenever I had issues in the lab. I want to thank Bernd for his adoption to informatics and the maintenance of our 16S rRNA gene in-house pipeline. I would also like to thank Martin, Angelina, Ann-Kristin and Carl for their performance and contributions to this thesis. My gratitude goes to all my current and former lab mates Silja, Heiko, Birgit, Bernd, Martin, Sarah, Kristin, Amélie, Ming-Ji, Peggi, Leo, Genis, Vera, Tanja, and Biswas for the pleasant working atmosphere and help during my work. I would also like to thank Daniela and Petra for their always friendly support whenever I had questions.

Big thanks to all members of the G_2L, especially Andrea (our sequencing-machine), Sonja, Sascha, Heiko, Anja, Andreas (thanks for keeping me up-to-date!), Jörg, and John for the helpful discussions during my work. Additionally, I would like to thank Denis for always giving quick support und strong nerves with us Windows-proven biologists. Thanks to Frauke and Kathleen for keeping the sequencers running.

I would also like to thank Nicole, Andreas, Barbara, Stefan, Ladislav, Danny, Svenja, Hubertus, and Joachim of the DFG Research Unit FOR 571 for the interesting and enlightening discussions and the nice atmosphere during our awesome trip to Kiritimati. Special thanks to Gernot for bringing me closer to calcification processes, the open ears whenever I had questions, and the proofreading of the manuscripts.

Furthermore, I would like to thank the “Werkstatt Team” Olaf, Patrick, Jarek, and Gerd for professional and fast support. Also thanks to the workshop of the Faculty of Chemistry for the construction of sampling split-tubes.

For the nice cooperation during my work on bacteriohopanepolyols, microbial fuel cells, and soda lakes I would like to thank Martin, Volker, Romano, Falk, and André.

For the always good atmosphere when it is time for “eat and pray” in the Nordmensa, I would like to thank Basti, Josh, Sasse, Clara, Karl, Olli, and all random guests.

Finally, I would like to thank my family Christa, Jessica, Franziska, and Kristin for their support, no matter what happened. My biggest thanks go to my girlfriend Jenny, thank you so much for your endless support, the always encouraging words, understanding, and love during the last years <3.