The Arctic Mid-Ocean Ridge Vent Fields – A valuable Resource for Marine Bioprospecting?

Juliane Wissuwa

Dissertation for the degree of philosophiae doctor (PhD)
at the University of Bergen

2016

Dissertation date: May 25th 2016

Year: 2016

Title: The Arctic Mid-Ocean Ridge Vent Fields – A valuable Resource for Marine Bioprospecting?

Author: Juliane Wissuwa

Print: AiT Bjerch AS / University of Bergen

Contents

Scientific environment

The work presented in this thesis was carried out at the Centre for Geobiology and Department of Biology at the University of Bergen. Further, the project included a two month research stay at the Norwegian Structural Biology Centre (NorStruct), University of Tromsø and a three month research stay abroad at the Institute of Protein Biochemistry, CNR, Naples, Italy. The project was part of “Mining of a Norwegian biogoldmine through metagenomics” funded by the Norwegian Research Council to IHS (NFR-Biotek 2021-RCN208491).

Acknowledgements

Yes, I did it! But I could not have done this without help:

Most of all, I thank my main supervisor Ida Helene Steen. Thank you that you gave me the opportunity to come to Norway and work on this very interesting and versatile topic. It was great to have all those possibilities, opportunities and freedom while working here on my PhD. I enjoyed all the travel, the work with interesting people in different Labs and go to a variety of Conferences. Your confidence and positivity can be contagious. I also thank my Co-supervisor Runar Stokke for always having an open door. Thank you for all your help and discussions. Without you, the whole bioinformatics analysis would have been very challenging. Further I like to thank my Co-supervisors Hans Torstein Kleivdal and Rolf Birger Pedersen for making this PhD possible.

I like to thank Arne O. Smalås for letting me work in his lab at the Norwegian Structural Biology Centre (NorStruct), University of Tromsø. Big thanks to Bjørn Altermark for the help with the fosmid library construction. And thanks to all the people there who helped me and made my stay pleasant, not only with regards of work. This part of Norway became my "Winter-wonder Land".

My thanks goes to Donatella de Pascale and her research group at the CNR, Naples, Italy, who allowed me to continue my traveling from the north of Europe to the south of Europe. My special thank goes to you Titti, for all your organization, time, help and kindness to make my stay in Naples as pleasant, diversified and enjoyable as possible.

I like to thank Arvind Sundaram from the Norwegian sequencing centre for his help with the fosmid sequencing and assembly. I thank Kjersti Lian, NorStruct, for the protein analysis.

A big thanks goes to Frida Lise. You were a big help in the lab and you had an answer to almost every question. I hope, I will get to that point one day. Thank you, Irene, for all the help and discussions inside and outside the lab. I learned a lot from you. Anita, it was great working with you and thanks for helping me with the protein purification. Gro, even though the periods were short, it was always fun working with you. Thanks for your input and help. Do you know that you are a great company in Conferences? Siv and Bibiana, thank you for giving me a space in you lab.

Thanks to my office mates Jan, Sven, Anders, Rui and Yuangao. Even though I like it quiet, as you may have noticed, it was so much nicer having you as company. To Sven, the one and only! For your appropriate comments, organization talent, supply of good food and crazy Jacuzzi parties. (Happy?!) Together with PartyGuy, Bubble Girl and the other Pebble Boy it was awesome to have a small PhD gang. And our feasting period was just insane. IDIB!
Thank you to everyone at CGB. The parties, the student seminars, the occasional Friday beers and Pizza at Café Special evenings and all the other events made the last 4 years to a very great time.
Thanks to my friends in Bergen, outside the lab. You helped that this period was a great experience. Especially I thank you Thomas, I owe you a lot. Thank you, Frieda and Chrischi for always being there. Stéphane, I want to thank you for your encouragement, your help and all the time you spend in Bergen, btw no more involuntary rainy holidays. ;)

Zum Schluss möchte ich ganz besonders meiner Familie für all die Unterstützung danken. Besonders die mentale, für die Fahrdienste im Heimaturlaub, die vielen kleinen Kehrpakete und Postkarten!

Vielen Dank.

Abstract

The past decade the awareness for sustainability has come more and more into focus of society in terms of a need for eco-friendlier processes and renewable energies, new and better pharmaceuticals, and agricultural- and aquaculture challenges amongst others. Hence, enzyme technology has emerged as one of the most important technologies to meet future demands. A common consensus is that the marine environments represent a largely untapped potential for industrial applicable enzymes. As deep-sea hydrothermal vent fields have revealed a biodiversity of largely uncultivated microorganisms they have been reviewed as hot spots for finding new biocatalysts. In this study, samples from the Arctic Mid-Ocean-Ridge (AMOR) deep-sea hydrothermal vent fields have been investigated as a source for potential carbohydrate degrading biocatalysts, with a focus on starch degrading enzymes, using a multifaceted approach; isolation of a novel starch degrader (**Paper I**), integrated sequenced-based meta-omics study of a microbial biofilm (**Paper II**), genomic analysis of microbial isolates (**Paper I** and **III**), and finally, a functional-based screening of a microbial biofilm was conducted (described in detail in section 3.2).

In general, the sequence-based screening for glycoside hydrolases (GHs) in all sequencing analyses was low, with *Lutibacter profundi* LP1^{T} exhibiting the highest amount with 24 GHs (**Paper III**). However, a purified thermophilic neopullulanase with a melting temperature of 76.4 °C showed a comparable starch degradation capacity as the reference α-amylase from *Bacillus licheniformis* (**Paper I**). In comparison, a higher amount of proteases was identified searching against the MEROPS protein database. For both isolates (**Paper I** and **III**), protease activity was confirmed on agar plates hydrolysing casein and gelatin. Interestingly, the activity-based screening of a microbial biofilm identified 274 clones with starch degrading activity. Selective sequencing of 106 positive clones did not reveal any homologs of known α-amylases or members of the GH13 family. Further investigations of the clones are needed; however, this observation could be the first step in identifying new amylases yet to be characterized.

The biofilm used in this study (**Paper II** and activity screening, section 3.2), attached to a deep-sea black smoker chimney, gave us the opportunity to investigate the trophic relationship between a biopolymeric producing filamentous *Epsilonproteobacteria* with

organotrophic consumption of an epibiotic *Bacteroidetes*. Genomic characterization of an isolated *Bacteroidetes* from the biofilm further strengthened the link between primary production and the role of organotrophic degradation within the biofilm (**Paper III**).

Besides exploiting metagenomics for bioprospecting, the comprehensive data obtained in my thesis enabled an expansive insight to functional and metabolic traits and linking them to taxonomical groups. The individual approaches provided different information, which combined contributed to our understanding of interactions in deep-sea hydrothermal food webs.

Abbreviations

AMOR - Arctic Mid-Ocean Ridge

CBM – Carbohydrate-binding module

COG – Clusters of orthologous groups

GH - glycoside hydrolase

JMVF – Jan Mayen Vent Field

LCVF – Loki's Castle Vent Field

NGS – Next generation sequencing

ORF – open reading frame

R&D – Research & Development

List of Publications

Paper I

Wissuwa J, Stokke R, Fedøy AE, Lian K, Smalås AO, Steen IH. (2016): Isolation and complete genome sequence of the thermophilic *Geobacillus* sp. 12AMOR1 from an Arctic deep-sea hydrothermal vent site, in Standards in Genomic Science 11:16. DOI: 10.1186/s40793-016-0137-y

Paper II

Stokke R, Dahle H, Roalkvam I, Wissuwa J, Daae FL, Tooming-Klunderud A, Thorseth IH, Pedersen RB, Steen IH. (2015): Functional interactions among filamentous *Epsilonproteobacteria* and *Bacteroidetes* in a deep–sea hydrothermal vent biofilm, in Environmental Microbiology 17(10). DOI:10.1111/1462-2920.12970

Paper III

Wissuwa J, Le Moine Bauer S, Steen IH, Stokke R.: Complete genome sequencing of *Lutibacter profundi* LP1^{T} isolated from an Arctic deep-sea hydrothermal vent system, Manuscript in preparation for Standards in Genomic Science

1. Introduction

1.1 Background

In recent years an awareness and need for sustainability, ecological friendly actions and health have grown. With increasing population size and global industrialization there is an expanding demand of food, water, energy, healthcare and new materials produced by environmentally beneficial applications.[1] Earth´s reservoirs of fossil fuels are also decreasing and the use of renewable biomaterials is globally defined as important for the production of energy as well as new biomaterials. We see now a trend where the importance of a geo-based economy, the geoeconomy, is decreasing relative to a continually growing bio-based economy, the bioeconomy.[2] Clearly, the use of biotechnological applications will likely increase more in the future, which means that there will be an increasing need to find new living organisms and identify their potentially useful products.

Biotechnology is defined as the technical application of living organisms, systems and processes as well as their metabolic products.[3] This technology has wide-spread application in various branches of major industrial areas including health care, food, non-food and the environment. For instance, the pharmaceutical industry uses microbial products as antibiotics, antitumor agents and immunosuppressants (Demain and Adrio 2008). The food and agriculture industry uses them for food-processing agents, biopesticides and antiparasite agents. Vitamins, amino acids, organic acids, detergents and bio-catalysts are produced in chemical industry, whereas the environmental industry uses microbial products for bio-remediation and production of bioenergy (Demain and Adrio 2008). In 2009 it was estimated that over 500 commercial products were manufactured using biocatalysts of microbial origin (Singh 2010). The global market for industrial enzymes is growing in various sectors of industry, ranging from industrial application such as for detergent, starch, leather, textile, pulp and paper to food, feed and healthcare. The global market, estimated to $3.3 billion in

[1] OECD (2009). The Bioeconomy to 2030: Designing a Policy Agenda: Main findings and policy conclusions. OECD – Organisation for Economic Co-operation and Development. http://www.oecd.org/futures/bioeconomy/2030

[2] European Commission (2012). Innovating for Sustainable growth: A Bioeconomy for Europe. European Commission, Directorate – General for Research and Innovation B-1049 Brussel, ISBN 978-92-79-25376-8

[3] United Nation (1992). Convention on Biological Diversity. https://www.cbd.int/convention/text/default.shtml

2010, is expected to reach nearly $7.1 billion by 2018.[4,5] The biodegradable nature and cost effectiveness of enzymes let them become financially and ecologically viable alternatives to chemical processes. Thus the need for novel microorganisms and biocatalysts is increasing in line with new demands, technologies and approaches.

The marine environment is characterized by a rich diversity of organisms, many of which remain undescribed (Kennedy *et al.,* 2010, Dalmaso *et al.,* 2015).[6,7] Because of this high biodiversity, marine ecosystems are particularly well-suited for bioprospecting. Bioprospecting is the systematic search for valuable compounds and commercialization of new products based on biological resources (Kodzius and Gojobori 2015).[6] Marine species that live in extreme environments such as deep-sea hydrothermal vents could be of particular interest, due to their physico-chemical adaptations. Deep-sea hydrothermal vents harbour ecosystems in areas with tectonic activity (Figure 1), and are located along mid-ocean ridges, back-arc basins, volcanic arcs and actives seamounts (Tyler *et al.,* 2002, German *et al.,* 2011). They were discovered in the late 1970s during the exploration of the Galapagos spreading centre (Corliss *et al.,* 1979). Within a deep-sea hydrothermal vent field multiple zones of focused, high-temperature venting and low-temperature diffuse flows may be found (Nakamura and Takai 2014).These represent some of the most physically and chemically diverse habitats on Earth for microbial growth where the geochemical and thermal gradients provide a wide variety of niches for microbial colonization (Takai *et al.,* 2006, Tivey 2007). These ecosystems are fuelled by chemosynthesis where primary production is performed by chemoautotrophic archaea and bacteria that utilize reduced chemical substances such as H_2, H_2S, CH_4, Fe(II) and Mn(IV) (Table 1) (Takai *et al.,* 2006, Takai and Nakamura 2011, Sievert and Vetriani 2012). Different geological settings of the hydrothermal systems influence the chemical composition of the hydrothermal fluids (Kelley *et al.,* 2002, Tivey 2007, Flores and Reysenbach 2011, Orcutt *et al.,* 2011, Nakamura and Takai 2014). Deep-

[4] BBC Research (2011). Report BIO030 F Enzymes in Industrial Applications: Global Markets. BBC Research, Wellesley, MA, USA

[5] BBC Research (2014). Report BIO030 H Enzymes in Industrial Applications: Global Markets. BBC Research. Wellesley, MA, USA

[6] Nasjonal strategi (2009). Marin bioprospecting – en kilde til ny og bærekraftig verdiskaping. Fiskeri- og kystdepartementet, Kunnskapsdepartementet, Nærings- og handelsdepartementet, Utenriksdepartementet, Norway

[7] Global Ocean Commission (2013). Policy Options Paper # 4: Bioprospecting and marine genetic resources in the high seas. Global Ocean Commission. http://www.globaloceancommission.org

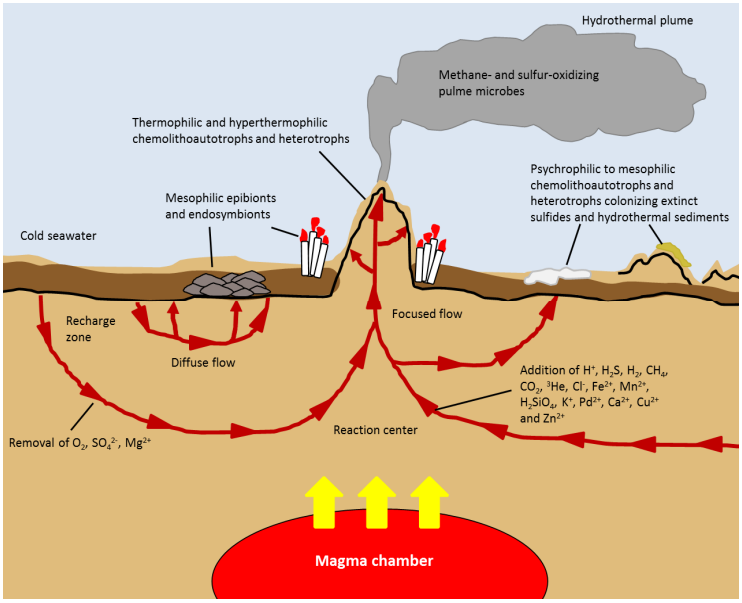

Figure 1: Hydrothermal circulation along Mid-Ocean Ridge illustrating compositional changes of seawater and microbial habitats supported by hydrothermal fluids. The small arrows indicate fluid flow while large arrows indicate heat transfer from magmatic source. Image is not to scale. Adapted from Flores and Reysenbach (2011).

sea hydrothermal microbial communities include chimney structures, plumes, microbial mats, sediments, diffuse flow hydrothermal fluids and epibiontic microflora of deep-sea vent metazoans (Figure 1) and are reviewed elsewhere (Dubilier *et al.,* 2008, Flores and Reysenbach 2011, Orcutt *et al.,* 2011, Sievert and Vetriani 2012). The microorganisms living under such extreme conditions (high pressure, extreme temperatures and pH) are commonly called extremophiles (Podar and Reysenbach 2006, Cavicchioli *et al.,* 2011a). They are well-known to produce enzymes or "biocatalysts" (Antranikian *et al.,* 2005). Industrial processes often include extreme conditions, which are similar to those that exist in extreme habitats. Biocatalysts from extremophiles are particularly interesting as their range of applications is versatile for biotechnology processes (Dalmaso *et al.,* 2015) and fundamental research (Sarmiento *et al.,* 2015). The number of biocatalysts derived from deep-sea hydrothermal vents is however limited.[7] This may be explained by the fact that they remain largely underexplored resources as access to deep-sea hydrothermal systems is difficult (Thornburg *et al.,* 2010). Furthermore, the logistics for studying such environments is extremely challenging, cost-intensive and requires a vast amount of resources and specialized

equipment, including the use of remotely operating vehicles (ROVs) (Thornburg *et al.,* 2010, Cavicchioli *et al.,* 2011a). Thus research explorations to these environments have rather fulfilled the demand of basic research in terms of revealing the phylogenetic, metabolic and physiological microbial diversity of those ecosystems than serve for the search and identification of novelties for biotechnology and industrial applications (Sievert and Vetriani 2012).[7]

Table 1: Examples of thermodynamically favourable redox reactions utilized by microorganisms from marine hydrothermal environments. Modified from Flores and Reysenbach (2011) and complemented (Knittel and Boetius 2009).

Conditions	Electron donor	Electron acceptor	Free energy ΔG (kJ/mol)	Metabolic process
Aerobic	H_2	O_2	-237	Hydrogen oxidation
	HS^-	O_2	-797	Sulphide oxidation
	S^0	O_2	-585	Sulphur oxidation
	$S_2O_3^{2-}$	O_2	-952	Thiosulfate oxidation
	Fe^{2+}	O_2	-44.3	Iron oxidation
	NH^{4+}	O_2	-275	Nitrification
	CH_4 and other C1 compounds	O_2	-810	Methane oxidation
	Organics	O_2	-477	Heterotrophic metabolism
Anaerobic	H_2	NO_3^-	-239	Denitrification
	H_2	S^0	-98.3	Sulphur reduction
	H_2	SO_4^{2-}	-38.1	Sulphate reduction
	H_2	CO_2	-34.7	Methanogenesis
	CH_4	SO_4^{2-}	-16	Methane oxidation
	Organics	SO_4^{2-}	-40.6	Heterotrophic sulphate reduction
	Organics	S^0	-25.1	Heterotrophic sulphur reduction
	Organics	Organics	-38.5	Fermentation

Since 2005 researchers at the Centre for Geobiology, University of Bergen, have discovered deep-sea hydrothermal vent systems at the ultraslow-spreading Arctic Mid-Ocean ridge (AMOR) in Norwegian territorial waters (Figure 2). AMOR extends over 4000 km north of the Arctic Circle (66°N) in the Norwegian-Greenland Sea and consists of several smaller ridge sections (Pedersen *et al.,* 2010b). The Jan Mayen Vent Fields (JMVF) and Loki's Castle Vent Field (LCVF) were the selected study sites for this PhD thesis. They are both basalt hosted hydrothermal vent systems located along AMOR.

Explorations of the microbial biodiversity have identified a large diversity of novel microbial taxa with a multiplicity of uncultivated microbial linages, in addition to microorganisms found widely distributed in deep-sea hydrothermal vent systems (Lanzen *et al.,* 2011, Jaeschke *et al.,* 2012, Jorgensen *et al.,* 2012, Dahle *et al.,* 2013, Urich *et al.,* 2014, Dahle *et al.,* 2015, Stokke *et al.,* 2015, Schouw *et al.,* 2016, Steen *et al.,* 2016). These results point towards an untapped genetic reservoir encoding potentially valuable biocatalysts.

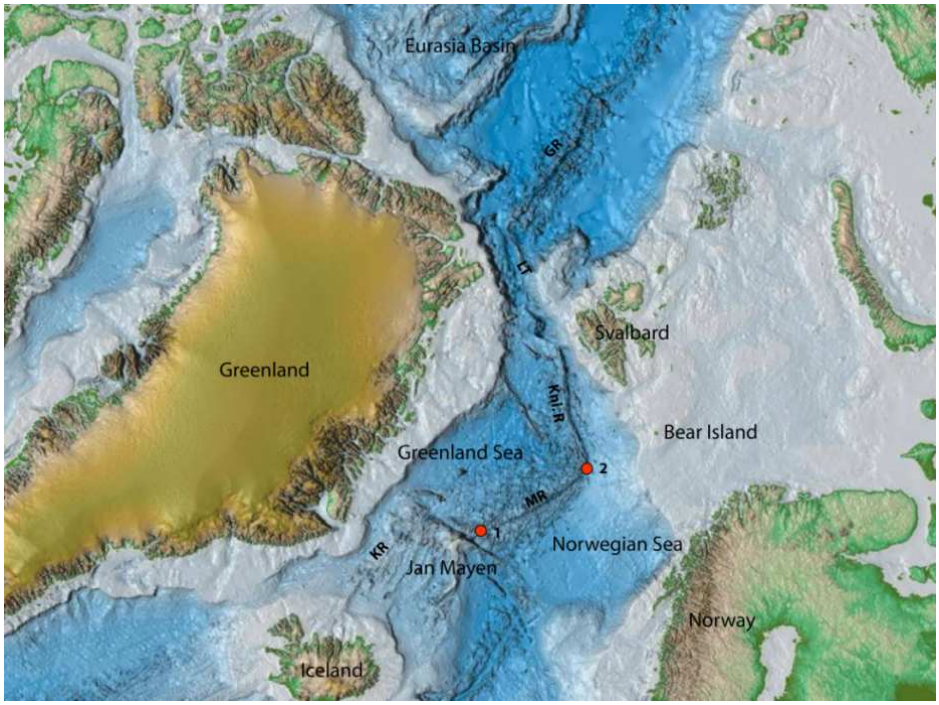

Figure 2: Arctic Mid-Ocean Ridge with the location of the hydrothermal systems 1) Jan Mayen vent field and 2) Loki's Castle vent field. KR, Kolbeinsey Ridge; MR, Mohns Ridge; Kni. R, Knipovich Ridge; LT, Lena Through and GR, Gakkel Ridge. Map after Pedersen *et al.* (2010b).

1.2 Enzymes from extremophiles and their biotechnological application

Microbial enzyme resources are provided by nature in an immense amount. These biocatalysts exhibit high specific activities toward substrates, are biodegradable, have a low demand for energy and lead to limited production of wastewater and by-products (Adrio and Demain 2014).

Comparative analyses of homologous enzymes from psychrophiles, mesophiles and (hyper-) thermophiles have revealed a variety of characteristics promoting thermostability (Russell *et al.,* 1998, Knapp *et al.,* 1999, Mallick *et al.,* 2002, D'Amico *et al.,* 2003). In general, (hyper)-thermophilic enzymes are intrinsically stable and have a rigid conformation (Vieille and Zeikus 2001). To promote this stability "loose structures" like the protein ends (N- and C-terminus) and loops are either shortened or anchored to the protein surface (Vieille and Zeikus 2001, Bell *et al.,* 2002, Leiros *et al.,* 2012). Hydrophobic amino acids are shifted to the interior of the protein, promoting elevated hydrophobic interactions (Knapp *et al.,* 1999). The amino acids arginine and proline are found in higher frequency in thermophilic proteins, where they facilitate stabilization of the protein structure. The number of disulphide bonds formed by cysteines increases with temperature elevation (Mallick *et al.,* 2002). A higher number of charged amino acids are located on the protein surface and large networks of ionic interactions and hydrogen bonds can be found in active site regions and at subunit interfaces (Knapp *et al.,* 1999). In contrast, psychrophilic proteins are intrinsically more disordered, leading to more flexible structures (Feller 2010). The compactness of the protein interior is reduced, due to a smaller hydrophobic core. Fewer and weaker molecular interactions lead to unfolding of proteins at moderate temperatures. A higher proportion of non-polar residues exposed on the surface and the increased state of loose or relaxed of protein extremities are considered as additional destabilizing factors (Feller 2010). In comparison with mesophilic and thermophilic homologs, psychrophilic proteins show a higher occurrence of glycine clusters, especially in functional regions, but have a lower proline content and disulphide bonds (Feller 2010). The increased flexibility of psychrophilic enzymes results in a modified active site with elevated heat-labile properties. Although the catalytic residues are conserved among the extremophilic enzymes, the catalytic cavity seems larger and more accessible to ligands (Russell *et al.,* 1998). The weaker substrate affinity reduces the activation energy and results in higher enzyme activity (D'Amico *et al.,* 2003).

Interest in psychrophilic enzymes with industrial potential has increased over the last years (Trincone 2011, Dalmaso *et al.,* 2015). Various enzymes, such as lipases, proteases, amylases and ureases, were obtained from isolates or metagenomic studies from a diverse range of cold habitats (Prasad *et al.,* 2014, Vester *et al.,* 2014, De Santi *et al.,* 2016a). These are applied in a wide variety of industries, including food and beverages, detergents, textiles, research and cosmetics (Cavicchioli *et al.,* 2011b, Dalmaso *et al.,* 2015). For example, in molecular biology, there is particular interest in the use of cold-active DNA modifying enzymes, like alkaline phosphatases, DNases and nucleases (Sarmiento *et al.,* 2015). The increasing numbers of available cold-active enzymes is not only due to the growing interest and efforts of reducing energy consumption in various applications, but also to due to the fact that diverse psychrophilic environments are easier to exploit than other extreme habitats, such as the deep-sea. The Earth biosphere is considered to be cold, with over 80% being permanently subjected to temperatures below 5 °C (Cavicchioli *et al.,* 2011a). Therefore the quantity of habitats comprising bacteria producing cold-adapted enzymes is versatile.

Enzymes from (hyper)-thermophilic organisms are also of interest in terms of industrial applications. In contrast to their mesophilic counterparts, these enzymes are thermostable and resistant to irreversible heat inactivation. Other advantages are enhanced biomass conversion, minimized contamination and reduction of process costs (Antranikian *et al.,* 2005). (Hyper) -thermophilic enzymes have gained importance in biorefineries for bioethanol production, where the temperatures are elevated to degrade complex compounds. Industrial markets using thermophilic enzymes comprise food and beverages, detergents, pulp and paper and biofuels, among others (Adrio and Demain 2014, Sarmiento *et al.,* 2015).

The broad ranges of physicochemical boundaries found in hydrothermal vents allow the discovery of distinct enzymes covering a wide temperature range for activity. However, as Podar and Reysenbach stated, “the path from extremophile to successful commercial application is not documented in peer reviewed scientific publications in most cases, due to highly competitive nature of industrial R&D. These information’s can be partially followed through patents, biotechnology meetings and company websites” (Podar and Reysenbach 2006). This clearly presents a major challenge to obtain and present a complete overview of the amount and types of enzymes, originating from hydrothermal vents, which have biotechnological applications.

1.3 Glycoside hydrolases

The carbohydrate-active enzyme (CAZy) database (http://www.cazy.org/) describes structurally-related catalytic domains and carbohydrate-binding modules of enzyme families that degrade, modify or create glycosidic bonds (Cantarel *et al.,* 2009). The currently included enzyme classes are: glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs), and non-catalytic carbohydrate-binding modules (CBMs) (Levasseur *et al.,* 2013). The CAZy classification system is based on amino acid sequence comparison, and members of the same glycoside hydrolase (GH) family should exhibit sequence similarities, share catalytic machinery, employ the same reaction mechanism and adopt the same type of catalytic domain fold (Henrissat 1991).

1.3.1 Amylases

Amylases catalyse the hydrolysis of starch and related α-glucan molecules to glucose monomers (Figure 3). Starch and related polysaccharides occur in great abundance in nature and serve as energy storage molecules in plants (van der Maarel *et al.,* 2002, Kelly *et al.,* 2009). Starch is composed of two different glucose polymers: 1) amylose, a linear α-1,4-linked glucose molecule chain and 2) amylopectin with the additional association of α-1,6 branches. Glycogen is the storage form of glucose in mammals and is analogous to starch in plants. It is structurally similar to amylopectin, but with a higher degree of α-1,6 branches (Kelly *et al.,* 2009). Enzymes, that are active towards α-glucosidic bonds, present in starch and related poly- and oligosaccharides, have been designated as “amylolytic enzymes” (Janecek *et al.,* 2014). They catalyse the enzymatic reactions hydrolysis, transglycosylation and isomerization. Amylases are mainly represented in the GH family 13, although members probably exist in the families GH57, GH119 and GH126 (Janecek *et al.,* 2014).

The endo-acting enzyme α-amylase (EC 3.2.1.1) is the most widely studied amylolytic enzyme. It catalyses the hydrolysis of internal α-1,4-glucosidic bonds of amylose and amylopectin chains in a random fashion (MacGregor and Svensson 1989) (Figure 3). Exo-acting amylases release terminal glucose monomers from non-reducing ends of polysaccharide chains. Glucan 1,4-α-glucosidase (EC 3.2.1.3, glucoamylase) and α-

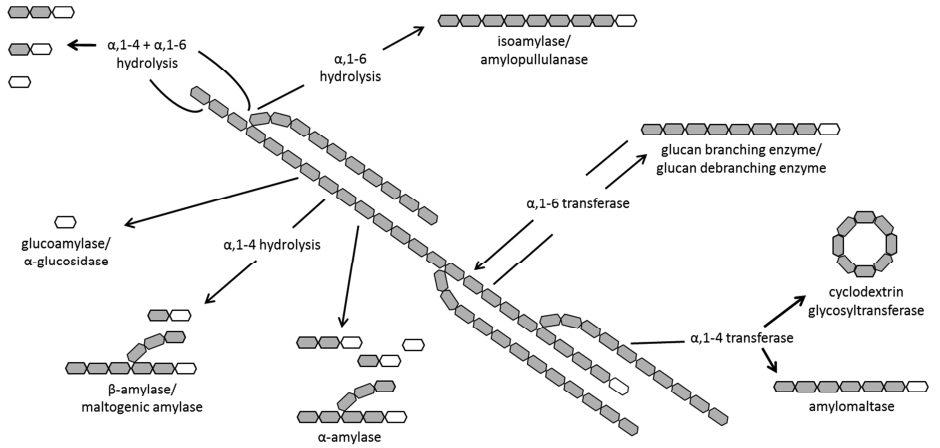

Figure 3: Different enzymes involved in starch degradation. The white ring structure marks the reducing end of a glucan molecule. Adapted from van der Maarel, et al. (2002)

glucosidase (EC 3.2.1.20) are able to cleave both α-1,4 and α-1,6 bonds to release single glucose residues (van der Maarel *et al.,* 2002). The β-amylase (EC 3.2.1.2, maltogenic amylase) hydrolyses maltose units instead of glucose residues from exclusively α-1,4 glucosidic polysaccharides. It is the only known enzyme of the GH14 family. Other exoamylases are glucan α-1,4 maltohydrolase (EC 3.2.1.133) and diverse maltooligosaccharide forming amylases (van der Maarel *et al.,* 2002). The debranching enzymes isoamylase (EC 3.2.1.68) and pullulanase type I (3.2.1.41) mediate the hydrolysis of α-1,6-linked branching points in amylopectin and related polysaccharides, whereas amylopullulanase (EC 3.2.1.1/41) can additionally hydrolyse α-1,4 glycosidic bonds (Hii *et al.,* 2012). Another group of starch converting enzymes are transferases. They cleave an α-1,4 glucosidic bond from a donor molecule and transfer a part to a glucosidic acceptor by forming a new glucosidic bond (Figure 3). Amylomaltase (EC 3.2.1.25) and cyclodextrin glycosyltransferase (EC 2.4.1.19) catalyse a new α-1,4 glucosidic bond with the acceptor, whereas the branching enzyme (EC 2.4.1.18) forms new α-1,6 glucosidic bonds (van der Maarel *et al.,* 2002).

The wide range of industrial applications for amylases includes food and feeding production (e.g. fruit juice clarification, baking and beer industries), detergent, textiles, pulp and paper, bio-remediation and bioethanol production (van der Maarel *et al.,* 2002, Yang *et al.,* 2014).

The liquefaction and saccharification process of starch for sugar syrups (e.g. glucose, fructose or maltose syrups) or ethanol production requires a variety of thermophilic amylases at different production steps (Prakash and Jaiswal 2010, Hii *et al.,* 2012, Homaei *et al.,* 2016). An α-amylase is applied during liquefaction, which has to withstand temperatures of 105 °C, with optimum activity at 95 °C, and pH 6.0 to 6.2 for the hydrolysis of starch into oligosaccharides. In the following saccharification step, glucoamylases and pullulanases are utilized to produce glucose at 60 °C and pH of 4.2 – 4.5. A last step in fructose syrup production is isomerization of glucose into fructose by glucose isomerase at 60 °C and elevated pH of 7.8 (Crabb and Shetty 1999). In the detergent application amylases must withstand elevated temperatures (60 °C), alkaline conditions and be resistant to oxidants (Cherry and Fidantsef 2003). Pullulanases are applied as effective additives in detergents and in the production of high-amylose starches, diverse syrups and cyclodextrins (Hii *et al.,* 2012). A variety of starch-active enzymes has been suggested as anti-staling agents in the baking industry to improve texture, volume and flavour of bakery products, such as α-amylase, branching and debranching enzymes, maltogenic amylases, β-amylases and amyloglucosidases (van der Maarel *et al.,* 2002). Transferases like cyclodextrin glycosyltransferase find applications in the production of novel glycosylated compounds or cyclodextins for food, pharmaceutical and cosmetic industry (Biwer *et al.,* 2002, van der Maarel *et al.,* 2002, Radu *et al.,* 2016).

Various microorganisms (Bacteria, Archaea and Fungi) are sources for amylolytic enzymes implemented in present-day biotechnology (Homaei *et al.,* 2016). The most deployed thermophilic amylolytic enzymes derive from different *Bacillus* species, such as α-amylases (Kindle *et al.,* 1986, Nedwin *et al.,* 2013), pullulanases (Tomimura 1991), β-amylase (Matsunaga *et al.,* 2011) and cyclodextrin glycosyltransferase (Kragh *et al.,* 2010). The leading enzyme producer Novozyme has Toruzyme®, a cyclodextrin glycosyltransferase from *B. licheniformis*, and BAN 480 L, an α-amylase from *B. amyloliquefaciens*, in their thermophilic product range. Other commonly used bacterial sources are the *Firmicutes Thermoanaerobacter* or *Thermoanaerobium* for cyclodextrin glycosyltransferases (Starnes *et al.,* 1996), *Clostridium thermosulfurogenes* for β-amylases (Zeikus and Hyun 1987) and *Geobacillus* for cyclodextrin glycosyltransferases and α-amylases (Shiosaka 1976, Tamuri *et al.,* 1981); the *Bacteroidetes Rhodothermus obamensis* for branching enzyme (Shinohara *et*

al., 2001, Hendriksen and Norman 2002) or *Thermotogales*, such as *Fervidobacterium,* for pullulanases (Antranikian and Rgensen 1996). For processes with higher temperature demands, Archaea represent an adequate resource for hyperthermophile enzymes (Littlechild 2015). Species of *Pyrococcus* and *Thermococcus* among others, are not only used as biocatalyst source for DNA modification (Mathur 1996, Schildkraut and Schildkraut 2008), but also for other hydrolases, such as amylases and pullulanases (Antranikian *et al.,* 1992, Antranikian and Sjoholm 1997).

1.3.2 Carbohydrate binding modules

A variety of carbohydrate-active enzymes have acquired one or more additional non-catalytic domains that interact with poly-, oligo- or monosaccharides. Theses auxiliary domains exhibit an autonomous folding and are known as carbohydrate-binding modules (CBM). Currently, CBMs are grouped into 73 families in the CAZy database (February 2016). Proteins containing CBMs recognize a variety of insoluble polysaccharides (e.g. cellulose, chitin, xylan, β-glucans, starch, glycogen) and other different sugars (e.g. arabinofuranose, mannan, fucose, lactose, galactose, lipopolysaccharides) (Guillen *et al.,* 2010). Thereby CBMs play a key role in substrate hydrolysis, but are also important for metabolisms, energy storage, structural support, attachment, immunological recognition, etc. A fine carbohydrate-binding specificity by CBMs enable the anchoring of the catalytic domain to the polysaccharide surface and enhance enzymatic activity (McLean *et al.,* 2002). Some CBM families are invariant in their substrate specificity, whereas modules of other families bind to a range of different polysaccharides (Michel *et al.,* 2009). In general, the binding specificity of the CBMs are congruent with the catalytic activity of the enzymatic domains (Michel *et al.,* 2009). Additionally, the disruption of diverse polysaccharide structures by CBMs has been described (Giardina *et al.,* 2001, Guillen *et al.,* 2010). The capability of binding starch is found at present in 12 carbohydrate-binding module families (CBM20, CBM21, CBM25, CBM26, CBM34, CBM41, CBM45, CBM53, CBM58, CBM68, CBM69) (Janecek *et al.,* 2011, Peng *et al.,* 2014, Xu *et al.,* 2014).

1.4 Enzyme discoveries by (meta)genomic analysis

The Earths biosphere contains a wide variety of habitats which host vast microbial diversity. Prokaryotes have been determined to comprise 10^6 to 10^8 separate genospecies, representing the largest proportion of distinct organisms (Simon and Daniel 2011). Up to now, over 99% of the total cell counts of a given sample observed under the microscope cannot be isolated by cultivation techniques (reviewed by Torsvik and Ovreas 2002). Over the last few decades, a variety of genomic and metagenomic tools have been developed to study this microbial diversity and ecology, including DNA cloning systems, DNA sequencing, PCR, hybridization techniques and bioinformatics (Rashid and Stingl 2015). Such culture-independent techniques enable the analysis of the phylogenetic, genetic and the functional diversity of a sample, without the need for isolating individual organisms. The terms "metagenome" and "metagenomics" were coined by Handelsman and colleagues and refer to the genomic analysis of uncultured microorganisms (Handelsman *et al.,* 1998). Since then, metagenomic approaches constituted two types of analysis to obtain information from metagenomic libraries. The first approach, called "sequencing-based metagenomics", uses sequencing techniques to screen genetic material for particular DNA sequences. The second approach screens for a heterologous expressed traits and is termed "functional metagenomics" (Schloss and Handelsman 2003). The same techniques are now applied for bioprospecting "microbial dark matter" as a resource for novel biocatalysts and secondary metabolites (Piao *et al.,* 2014, Michalska *et al.,* 2015). As a consequence, new molecular tools and sequencing techniques, genomics and metagenomic-based enzyme discoveries have become increasingly important over the years (Kodzius and Gojobori 2015, Ufarte *et al.,* 2015b, Mirete *et al.,* 2016). In the following, the sequence-based and function-based approaches will be described in detail.

1.4.1 Sequence-based metagenomics

The sequence-driven approach is adequate for ecological studies that address questions such as "who is there" and "what can they do" (Tyson *et al.,* 2004, von Mering *et al.,* 2007, Simon and Daniel 2011). At the beginning of the metagenomic era, the generation of metagenomic libraries from environmental samples was a necessary part of the method. The entire genomic DNA content of a sample was isolated and genomic fragments of a certain size were cloned

into a vector before the transformation into a suitable host. Genome libraries play an important role in e.g. analysing phylogenetic affiliations of organisms by 16S rRNA or other highly conserved genes, characterizing operon structures, to facilitate discovery of genes and whole pathways for potential physiology and ecological function or industrial applications, and even in the assembly of genomes from uncultured organisms (Schloss and Handelsman 2003, Tyson *et al.,* 2004). In the first study using this approach, environmental DNA was cloned into bacteriophage lambda and screened for 16S rRNA gene sequences to analyse the phylogenetic diversity of marine picoplankton, a decade before the term 'metagenome' was coined (Schmidt *et al.,* 1991). De Long *et. al*. demonstrated that sequence-based metagenomics has the potential to link phylogeny with functional aspects of uncultured microorganisms (Beja *et al.,* 2000a). A bacterial rhodopsin gene was identified from a rRNA-containing clone of an uncultivated marine *Gammaproteobacterium* (SAR86 group) from a marine planktonic community metagenomic library (Beja *et al.,* 2000a, Beja *et al.,* 2000b). In the past, bacteriorhodopsin was only attributed to halophilic archaea and the finding expanded the knowledge on how light energy enters marine ecosystems. With the development of next generation sequencing (NGS) technologies, the analyses of samples has been shifted towards a sequencing-only approach, making the generation of clone libraries unnecessary (Margulies *et al.,* 2005, Shendure *et al.,* 2005). Entire genomes or metagenomes of environmental samples can now be sequenced in parallel, leading to an exponential growth of available sequence data and thereby providing valuable insights into diversity, genome size and metabolic functions of microorganisms (Gonzalez *et al.,* 2008, Pope *et al.,* 2012, Albertsen *et al.,* 2013). Recently, the technology has evolved; enabling separation of single cells and mediating sequencing of their amplified genomes (Raghunathan *et al.,* 2005, Marcy *et al.,* 2007). By using single-cell genome sequencing, Rinke and colleagues were able to recover and analyse the genomic data of 201 uncultivated bacterial and archaeal cells of 29 novel phylogenetic branches, revealing unexpected metabolic features (Rinke *et al.,* 2013).

Working with sequence-driven metagenomics requires prior knowledge about the gene sequences of interest. Analyses are referred to as direct studies when making use of conserved DNA sequences to design PCR primers for target genes or hybridization probes to screen for clones in metagenomic libraries that contain the sequences of interest (Schloss and Handelsman 2003). The undirected approach uses random sequencing of metagenomic DNA.

All analyses are based on sequence homology towards known sequences and are limited in terms of finding novel and unique ones. The applications of adequate bioinformatic tools are essential for the analysis of sequencing data, including functional annotation of genes, and assessment of the performance. Specialized databases of specific protein groups (e.g., enzyme classes) can be helpful for gene annotation and are of great relevance for functional prediction to biology in general and to biotechnology in particular. Integrated annotation tools perform sequence comparison against sequence collections and enable the identification of homologues. The batch BLAST tool of the MEROPS peptidase database detects peptidases in newly determined sequences (Rawlings and Morton 2008). Information about enzymes involved in the assembly, modification and breakdown of oligo- and polysaccharides is provided by CAZy, a database for carbohydrate-active enzymes (Cantarel *et al.,* 2009). The database for automated carbohydrate-active enzyme annotation, (dbCAN), and the CAZYmes analysis toolkit, (CAT), incorporate the knowledge of CAZy and provide automated annotation for CAZymes (Park *et al.,* 2010, Yin *et al.,* 2012). Thereby, sequence-based metagenomics enables the identification of protein encoding genes with lower sequence homology towards known genes and also the discovery of modified or new activities, which is an interesting aspect for bioprospecting.

1.4.1 Function-based screening methods

Functional screening methods are based on heterologous expression of target genes present in metagenomic libraries and selection for certain phenotypes, such as colour, inhibition zones or fluorescence (Rashid and Stingl 2015). Clones with the desired functional properties are subsequently sequenced to identify the genes of interest. An advantage of this approach is the identification of novel activities, uncoupled from sequence homology, which enables the discovery of new genes with low or missing sequence homologies to known genes (Simon *et al.,* 2009). The experimental setup is important for functional-based metagenomics regarding the choice of DNA fragment size, vector and expression host (Kakirde *et al.,* 2010). The approach can be coupled with high-throughput technologies, thereby shortening the processing time and allowing for larger clone libraries to improve the screening outcome for an otherwise labour-intensive protocol (Ekkers *et al.,* 2012). Furthermore, a suitable production organism has already been established for the detected gene activities.

Nevertheless, robust and sensitive screening methods need to be applied to overcome low hit rates due to low gene expression levels or incompatible host expression machinery towards target genes.

The basic phenotypical functional screening is based on visual detection of traits, such as colony pigmentation, irregular colony morphology or halo formation on plate overlays (Ekkers *et al.,* 2012). Craig and colleagues used the above mentioned traits to screen for associated small-molecule production using high-throughput screening in multiple hosts (Craig *et al.,* 2010). The application of indicator medium is more common in functional-based screening (Simon and Daniel 2011). Supplements in media will lead to the detection of catabolic, metabolic or antibiotic capabilities as well as chemical reactions or particular small molecules. The use of solid media enables elevated analytical screening throughput of clones, although the detection sensitivity might be reduced due to diffusion of reaction products, which leads to low signal-to-noise ratios (Rashid and Stingl 2015). An extensive study of a human metagenome by solid media-based multi-step functional screening for carbohydrate-active enzymes facilitated the identification of target genes (Tasse *et al.,* 2010). The application of liquid-based high-throughput screening is another alternative. It can be a sensitive and reproducible method in combination with a small reaction volume, in microliter scale, and automated detection systems of chromogenic or fluorescent signals (Nyyssonen *et al.,* 2013). In addition to those aforementioned 'basic' approaches, more sophisticated high-throughput screening technologies have been developed. Functional screening can be performed using heterologous complementation of host strains or mutants for growth under selective conditions (Schipper *et al.,* 2009, Simon *et al.,* 2009) or induced gene expression (Uchiyama *et al.,* 2005, Uchiyama and Miyazaki 2010, Simon and Daniel 2011). Growth of clones can only be observed if foreign DNA inserts promote expression of desired gene products that facilitate viability. Cold-active DNA polymerase activities have been detected using *Escherichia coli* harbouring a cold-sensitive lethal mutation in DNA polymerase I as a host, for screening metagenomic libraries obtained from glacial ice (Simon *et al.,* 2009).

The use of functional-driven metagenomics to search for new biocatalysts was applied for the first time by Healy and colleagues (1995). They constructed a gene library from thermophilic, anaerobic lignocellulose digesters and screened successfully for thermophilic cellulases. Since then, a vast number of studies have been performed screening for a variety

of biocatalysts (Yan *et al.,* 2013, Biver *et al.,* 2014, Lee *et al.,* 2015, De Santi *et al.,* 2016b) and reviewed (Ferrer *et al.,* 2009, Uchiyama and Miyazaki 2009, Ufarte *et al.,* 2015a).

2. Aims

Given the promising biotechnological potential of extremophilic microorganisms from the AMOR, the overall aim of this thesis was to reveal this potential by cultivation, genome analyses and by culture-independent metagenomic analyses. Since amylases and proteinases/peptidase are known to have broad application areas, the focus was put on these enzyme classes.

Specific objectives were to:

- Isolation and characterization of starch-degrading thermophilic microorganisms from the AMOR.
- Identification and characterization of genes encoding carbohydrate-degrading enzymes in genomes and shotgun metagenomic data from AMOR.
- Construction of a metagenomic library of environmental DNA from a microbial mat situated on a hydrothermal chimney and identification of α-amylases by functional screening using plate-based activity assays.

3. Results and discussion

In this thesis, the work included isolation, characterization and genome analysis of a new *Geobacillus* strain from hydrothermal sediments collected from the JMVF (**Paper I**). Moreover, a biofilm (09ROV3-BS) sampled from a black smoker chimney wall at LCVF was studied using meta-omics technologies for the analysis of the community structure and functions (**Paper II**). The genome of *Lutibacter profundi* LP1^T (Le Moine Bauer *et al.,* in review), isolated from the same biofilm, was also sequenced and analysed to assess its ecological role in the biofilm and to identify genes encoding enzymes of industrial relevance (**Paper III**). Finally, a metagenomic library for screening of potential novel amylases was made from the same biofilm. The main results are presented and discussed in the **papers I**, **II** and **III**, respectively. The procedures and results obtained from the work with the metagenomic library are only presented and discussed in this section (see 3.2).

3.1 DNA-sequencing technologies

Three different high-throughput DNA-sequencing technologies were used in this PhD-project. In **Paper II**, the metagenome and transcriptome of the biofilm sampled from the black smoker wall was sequenced using 454-shotgun sequencing in combination with 8 kb paired-end sequencing. The 454 system, later known as Roche 454, was launched in 2005 and was the first commercially used next generation sequencing (NGS) system. This system uses pyrosequencing technology, which is based on the detection of pyrophosphate released during nucleotide incorporation. The reactions are carried out in a picotiter plate, where single DNA strands are captured via specific adaptors by amplification beads (Margulies *et al.,* 2005). This technique can generate a long read length of 700 bp and 0.7 G data per run with 99.9% accuracy within 24 hours (Liu *et al.,* 2012a). Paired-end libraries from genomic DNA consist of short paired tags from the two ends of DNA fragments (3 kb, 8 kb or 20 kb), which are extracted and covalently linked as paired-end constructs for high-throughput sequencing (Shendure *et al.,* 2005, Fullwood *et al.,* 2009). The mapping of the tags to reference genomes improves the DNA sequence efficiency, and relates discrete contigs in genome assemblies due to distance relationships of DNA fragment ends. The combination of shotgun and paired-end sequencing of the biofilm resulted in a total number of ~500 Mbp

with 1´840´364 total reads (943´600 shotgun reads, 896´764 paired-end reads) with an average read length of 344 bp. The direct sequencing of environmental DNA will be referred to as "metagenomic shotgun" data.

In **Paper I** and **III**, the genomes from *Geobacillus* sp. 12AMOR1 and *L. profundi* LP1^T were sequenced using a technology developed by Pacific Bioscience (PacBio). PacBio developed the single-molecule real-time (SMRT) method for long-read DNA sequencing (Eid *et al.,* 2009). The SMRT cell consists of millions of zero-mode waveguides (ZMW), each containing one set of enzymes, where the complementary strand of a long single stranded DNA molecule is synthesized by sequencing. The nucleotides are phospholinked with individual fluorophores that are released during DNA synthesis. By using the zero-mode waveguide, a single-fluorophore detection of incorporated nucleotides can be realized (Eid *et al.,* 2009). Even though the throughput is lower than in second generation sequencing techniques, the turnover rate of SMRT is quite fast with 3 hours per run and read length is by far longer, with 5500 bp in average. The sequencing reactions of the two bacterial genomes were archived with two SMRT cells for each genome. A total number of 595 Mbp was generated for *Geobacillus* sp. 12AMOR1 with 80´512 reads having an average length of 7393 bp. In comparison, the genome sequencing of *L. profundi* LP1^T generated 362.9 Mbp with 63´994 reads and an average length of 5671 bp.

In the functional genomics work (described below), Illumina sequencing technology was used. Illumina has developed a sequencing technique based on sequencing by synthesis. Single stranded DNA is attached to the flowcell by adapters and forms clusters by bridge amplification. After additional DNA separation into single strands, DNA sequencing is detected by synthesis with fluorescent reversible terminator deoxyribonucleotides, one base at a time (Bentley *et al.,* 2008). Different platforms have been launched, such as HiSeq, NextSeq and MiSeq. They vary in read length from (50 to 300 bp), sequencing direction of DNA strands (single read or paired-end), size of data (output up to ~600 Gb) and time (between 1 to 6 days). As part of this work, I sequenced 106 fosmids using 300bp paired-end MiSeq platform and obtained 19´878´063 raw sequence counts comprising 11.9 Gb in total. The metagenomic data obtained in this approach will be referred to as "metagenomic library" data.

3.2 Metagenomic library of the black smoker biofilm (09ROV03-BS)

3.2.1 Methods

3.2.1.1 DNA-extraction

For the metagenomic library, high molecular weight DNA was extracted from 1.2 g of a biofilm, sampled from a black smoker chimney wall at LCVF, using a modified protocol of Tiedje and colleagues (Zhou *et al.,* 1996). Modifications included the following: 15 ml DNA extraction buffer was used for the sample (1.2 g biofilm and 6 ml sea water). The initial incubation time for cell lysis was extended to 2 h at 56 °C by limiting the movement to occasional gentle shaking. The second incubation period was carried out after adding 2 ml of 20% SDS (approx. 1% end concentration). After centrifugation at 4800 x g for 20 minutes at 4 °C the DNA was extracted from the supernatant was extracted in three cycles with equal volume of chloroform (1st cycle 50:50 chloroform/phenol, 2nd and 3rd cycle only chloroform). The DNA was precipitated with 0.6 volume isopropanol at -20 °C overnight and resuspended in 10 mM Tris pH 7.6. As pulse field gel electrophoresis was not available for the verification of purified high molecular weight DNA quality, 20 cm long 0.8% agarose gels were run for 20 hours at 35 V and at 20 °C (data not shown). The high molecular weight DNA was concentrated with Amicon Ultra 100K (Millipore, Germany) and the concentration measured with NanoDrop 2000 (Thermo Scientific).

3.2.1.2 Metagenomic library production

The metagenomic library was constructed using the following fosmid cloning system – the CopyControl™ Fosmid Library Production Kit (Epicentre, Madison, WI, USA). Instructions described by the manufacturer were followed with one exception. In short, shearing of the high molecular weight DNA by pipetting and following size selection was carried out before the DNA end-repair step to increase the yield of modified DNA fragments with appropriate size. The size-selected, end-repaired DNA fragments were purified again using 3 M sodium acetate, as described in the protocol. The size-selected, modified metagenomic DNA fragments were ligated to the pCC2FOS fosmid vector, and packed into bacteriophages. The host strain *E. coli* Epi300-T1^{R} was infected and clones were spread on selective agar plates.

The estimated metagenome size of the biofilm sample 09ROV3-BS by shotgun sequencing was 165 Mbp. The approximate number of clones needed for a complete metagenomic fosmid library was calculated based on the formula provided by the manufacturer's instructions. To cover 99% of the metagenomic DNA in the library, assuming an insert size of 40 kb, a minimum of 18´994 clones were required. The generated fosmid clones were picked using the Qpix2 XT (Molecular Devices) to construct a library of 22´800 clones. DNA insertions of selected fosmid clones were verified by Sanger sequencing using pCC2FOS forward and reverse sequencing primers (Epicentre, Madison, WI, USA) flanking the cloning site of the fosmid vector.

3.2.1.3 Functional screening of the metagenomic library

The whole metagenomic library was subjected to functional screening and analysed for the presence of amylase and xylanase activity. Clones were cultivated for 2 days at 37 °C on LB agar plates supplemented with 1% starch or 0.25% xylan. Zones of substrate degradation were visualized using a modified Gram's iodine reagent (1.0 g KI, 0.5 g iodine in 300 ml distilled Water) or 0.5% Congo red, respectively (Teather and Wood 1982, Kasana *et al.,* 2008). In total, 274 clones exhibited activity for starch degradation were identified by halo formation surrounding the clones. Positive clones for xylan degradation were not observed.

3.2.1.4 Sequencing of amylase-encoding fosmid clones

Hundred and six fosmid clones with positive amylase activity were selected for sequencing to identify the amylase-encoding genes. Purified fosmids, using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, Germany), were pooled into 10 batches of 10 or 11 fosmids, containing equal quantities of DNA for each fosmid.

The size for each fosmid was estimated by restriction analysis using *Bam*HI, which flanks both sides of the cloning site in the fosmid vector. The average insert size was 35 Kbp, yielding approximately 796 Mbp of total cloned genomic DNA. To identify the individual fosmids after sequencing, the ends of each fosmid insert were sequenced in a Sanger reaction using the two flanking primers, pCC2FOS forward and pCC2FOS reverse. Sequencing of the fosmid batches were performed at the Norwegian Sequencing Centre in Oslo, Norway

(www.sequencing.no). There, the batches were tagged individually before combining them into one pool for sequencing; using paired-end (300 bp) MiSeq, Illumina.

The total number of sequence output varied between the batches, ranging from 1.58 to 2.6 million. After splitting the reads, based on their tags, into individual batches, reads belonging to the sequence tags, fosmid vector or to the host *E. coli* were removed. The remaining reads were assembled into contigs using MIRA (Chevreux *et al.,* 2004).

The assembly resulted in 402 large contigs (≥5000 bp) of the total 485 contigs. Contigs larger than 1100 bp were analysed in terms of fosmid insert recovery. Multiple sequence alignments of individual batches were therefore performed in combination with the corresponding Sanger "end-sequences" of the fosmids. In total, complete inserts of 44 fosmids were recovered, of which 36 inserts were fully assembled into single contigs. Open reading frames (ORFs) of assembled contigs were predicted using Prodigal v2.6.2 (Hyatt *et al.,* 2010) with the –meta option, for metagenomes, enabled. In order to identify carbohydrate-degrading enzymes, a HMM search (Finn *et al.,* 2011) of the ORFs was performed using the dbCAN database (Yin *et al.,* 2012). Additionally, a search for proteases was performed using the MEROPS peptide database (Rawlings and Morton 2008).

3.2.1.5 Fosmid supernatant activity screening

To confirm the positive amylase activity identified in the functional screening of the library, an additional activity screen of supernatants from selected clones was performed. Therefore, 5 fosmid clones, which have shown high activity in the library screen, and a negative control, containing the vector with the fosmid control insert (Epicentre, Madison, WI, USA), were cultured in 200 ml LB at 37 °C shaking at 170 rpm overnight. The supernatants were collected by centrifugation and filtration (0.2 µm) to remove cells, and concentrated approximately 200 times in 3K Centricon Plus-70 devices (Merck-Millipore, Ireland) at 4 °C. The concentrated supernatants were incubated on a 1% starch agar plate together with 5 µg of α-amylase from *B. licheniformis* (Sigma-Aldrich, Germany) as positive control for 4.5 h at 37 °C.

3.2.2 Results and discussion of the metagenomic library screening

The initial manual analysis of the 09ROV3-BS metagenomic shotgun sequencing data was based on ORFs associated with carbohydrate metabolism of COG functional category. Two ORFs were identified encoding a fragmented α-amylase belonging to *Bacteroidetes*. This was one of the rationales for the metagenomic library generation.

3.2.2.1 Fosmid screening

Screening of the metagenomic library for amylase activity resulted in a relatively high number i.e. 274 of positive clones, comprising 1.2% of the library (Figure 4). The average hit rate for activity screenings are generally less than one clone with activity per 1'000 clones screened (Lammle *et al.,* 2007, Kakirde *et al.,* 2010), due to a low specialized enzyme frequency in natural habitats, low gene expression levels or incompatible host expression machinery (Ekkers *et al.,* 2012). The obtained high fraction of putative positive clones for amylase activity could therefore indicate the presence of false positives.

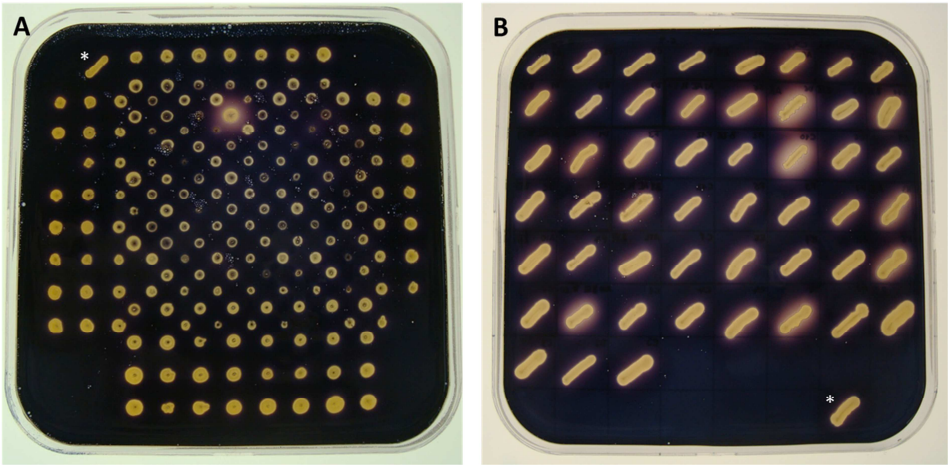

Figure 4: Primary (A) and secondary (B) screening of the biofilm metagenomic library for amylase activity on 1% starch agar colored with iodine. Starch hydrolysis is visible as clear halo around clones. Negative control *E. coli* Epi300-T1^{R} containing the pCC1FOSvector with the fosmid control insert is marked with asterisk.

The use of the CopyControl™ Fosmid Library Production Kit (Epicentre, Madison, WI, USA) is an established procedure and has been applied for the generation of many fosmid libraries (Neveu *et al.,* 2011, Liu *et al.,* 2012b, Nyyssonen *et al.,* 2013, Lee *et al.,* 2015, De

Santi *et al.,* 2016b, Maruthamuthu *et al.,* 2016). Although *E. coli* Epi300-T1^{R} carries genes with amylase activity in its genome (Nyyssonen *et al.,* 2013), new α-amylases have been identified using *E. coli* Epi300-T1^{R} as host strain (Tasse *et al.,* 2010, Liu *et al.,* 2012b, Maruthamuthu *et al.,* 2016). Furthermore, the screening procedure used for the metagenomic library in this project did not differ significantly from other studies with positive screening outcome, except for different incubation times and temperatures prior to screening, as well as the modified Gram's iodine reagent (Lammle *et al.,* 2007, Liu *et al.,* 2012b). Hence, the observed activity on the plates does not necessarily indicate that all of the clones were false positives.

Indeed, the change of substrate to AZCL-linked or other chromogenic substrates (Kracun *et al.,* 2015) could be used to verify the activities of the identified clones. The use of such substrates visualizes enzymatic degradation directly and can also simplify the screening procedure by allowing multiple screens simultaneously and eliminates additional potential error-prone colour steps. The activity screening using the supernatant of selected fosmid clones revealed amylase activity of the negative control on starch agar plates (Figure 5). This result was not observed in the functional metagenomic library screening on agar plates and could be due to cell lyses during liquid cultivation that resulted in the release of intracellular amylases into the supernatant. Expression of a gene with amylase activity located on the

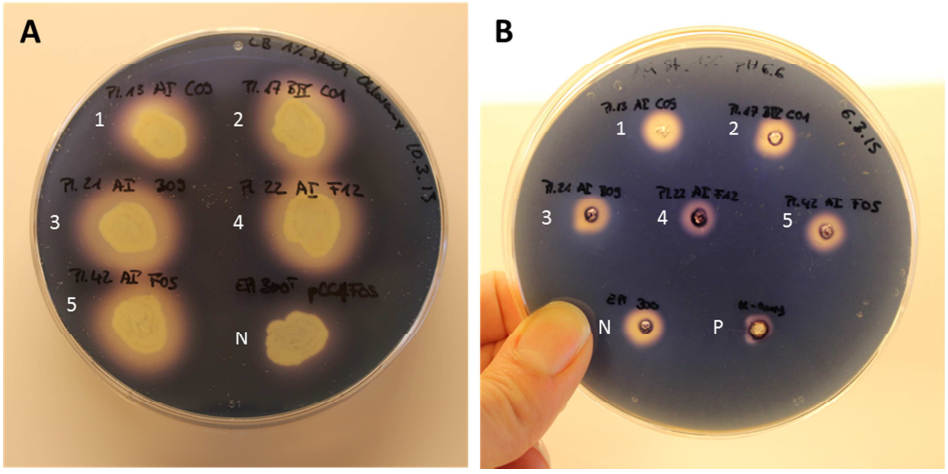

Figure 5: Amylase activity screening and supernatant activity assay of selected fosmid clones on starch agar plates colored with Gram's solution. *E. coli* Epi300-T1^{R} containing the vector with the fosmid control insert was used for negative control (N). Positive control (P) was α-amylase from *B. licheniformis*.

fosmid control insert of the negative control would have led to a constant positive amylase activity in the functional screening, which was not the case (Figure 4). Future verification experiments could include a retransformation of the plasmid to *E. coli* Epi300-T1^{R} or alternative host strains (Liu *et al.,* 2012b). Alternatively, analysis of α-glucosidase activity using the crude protein extract as described by Maruthamuthu (2016) could be applied.

It should be noted that the performed screening assays targeted secreted enzymes. To include the screening for intracellular enzymes, a lysis step has to be added or an alternative method should be applied. This could increase the number of positive hits. It is further recommended to increase the fosmid copy number, which is inducible by the addition of L-arabinose in the CopyControl system, thus enhancing the protein expression in the cells, which could be useful to distinguish activity from native host activity.

3.2.2.2 Sequencing and analysis of fosmids

The 106 sequenced metagenomic DNA inserts from fosmids contained in total 3017 annotated ORFs, whereof 51.6% could be assigned to *Sulfurovum*, the dominating genus of *Epsilonproteobacteria* within the biofilm (**Paper II**). A sequence comparison between metagenomic shotgun data and metagenomic library data was performed using blastn (Altschul *et al.,* 1990) and resulted in 633 scaffold matches. For the metagenomic library, 235 out of the 485 unique scaffolds matched scaffolds from the metagenomic shotgun data (**Paper II**) ranging from 68 to 23´986 bp with sequence identities between 73 and 100%. The majority of the metagenomic library scaffolds matched against scaffolds to the dominating population of *Sulfurovum* in the biofilm (**Paper II**). Interestingly, a higher number of scaffolds from the metagenomic library showed highest sequence identity with scaffolds of the *Sulfurovum* Bin2 (94), rather than the more complete and dominating *Sulfurovum* Bin1 (45). Congruent with the lower abundance of *Bacteroidetes* in the biofilm community (Dahle *et al.,* 2013), only 11 metagenomic library scaffolds had highest similarity to the *Bacteroidetes* Bin3. The remaining 85 scaffolds hit none of the main three Bins and showed sequence identity with sequences in the Bin_REST fraction. Alignments of selected longer scaffold from the metagenomic library with *Sulfurovum* scaffolds from the shotgun sequenced metagenome identified preponderantly the same ORF's and annotations, as the example Batch5_c7 vs. Bin_Sc1 shown in Table 2.

Approximately one third of the ORFs from the sequenced metagenomic DNA library were annotated as hypothetical proteins. The CAZyme database (Cantarel *et al.,* 2009) was used to identify amylase encoding genes in the metagenomic library data. The search for CAZymes identified 61 hits containing 7 unique ORFs each, for glycoside hydrolases (GH) and carbohydrate binding modules (CBM), respectively. Annotation and dbCAN search (Yin *et al.,* 2012) of the sequencing data from the metagenomic library did not reveal any homologs of known α-amylase or members of the GH13 family. The identified GH families were mainly associated with peptidoglycan modification (GH23, GH103) or hydrolysis of different β-glycosidic bonds (GH3, GH5) (Figure 6). CBMs might indicate potential

Table 2: Concordant sequence match of scaffolds from metagenomic library sequencing and metagenomic shotgun sequencing

Fosmid Batch10_bbm_c5		**Metagenome Bin1_Sc3**	
ORF	**Annotation**	**ORF**	**Annotation**
batch10_bbm_c5_1	hypothetical protein	Bin1_Sc3_00242	hypothetical protein
batch10_bbm_c5_2	MULTISPECIES: hypothetical protein	Bin1_Sc3_00243	hypothetical protein DEFDS_0394
batch10_bbm_c5_3	peptide ABC transporter substrate-binding protein	Bin1_Sc3_00244	oligopeptide ABC transporter substrate-binding protein
batch10_bbm_c5_4	sulfide-quinone reductase	Bin1_Sc3_00245	sulfide-quinone reductase
batch10_bbm_c5_5	hypothetical protein	Bin1_Sc3_00246	hypothetical protein
batch10_bbm_c5_6	formylmethionine deformylase	Bin1_Sc3_00247	formylmethionine deformylase
batch10_bbm_c5_8	hypothetical protein	Bin1_Sc3_00249	hypothetical protein
batch10_bbm_c5_9	hypothetical protein	Bin1_Sc3_00250	hypothetical protein
batch10_bbm_c5_10	30S ribosomal protein S9	Bin1_Sc3_00251	30S ribosomal protein S9
batch10_bbm_c5_11	50S ribosomal protein L13	Bin1_Sc3_00252	50S ribosomal protein L13
batch10_bbm_c5_12	recombinase RecB	Bin1_Sc3_00253	recombinase RecB
batch10_bbm_c5_13	beta-lactamase	Bin1_Sc3_00254	metallo-beta-lactamase family protein
batch10_bbm_c5_14	hypothetical protein	Bin1_Sc3_00256	hypothetical protein
batch10_bbm_c5_15	2-hydroxyacid dehydrogenase	Bin1_Sc3_00257	2-hydroxyacid dehydrogenase
batch10_bbm_c5_16	succinate dehydrogenase	Bin1_Sc3_00258	succinate dehydrogenase
batch10_bbm_c5_17	hypothetical protein	-	
batch10_bbm_c5_18	putative protein-disulfide isomerase	Bin1_Sc3_00259	putative protein-disulfide isomerase
-		Bin1_Sc3_00260	NO HIT
-		Bin1_Sc3_00261	NO HIT
batch10_bbm_c5_21	hypothetical protein	Bin1_Sc3_00262	hypothetical protein, partial
batch10_bbm_c5_22	hypothetical protein	-	

glycoside hydrolases, as they can be found as additional domain mediating substrate binding or in the vicinity as part of a gene cluster mediating certain hydrolytic activity (Guillen *et al.,* 2010). The ORFs containing members of four different CBM families (CBM11, CBM44, CBM50, CBM66) were annotated as hypothetical proteins or based on containing domains, mostly polycystic kidney disease I (PKD) domain (Figure 7). Many of those CBMs are associated to binding of different β-1-4-linked saccharide polymers according to the CAZy database. Hence, the presence of CBMs there might be an indication for unknown GHs with β-1-4 saccharide polymers degrading activity. In the sequenced metagenomic DNA from fosmid inserts a 1784 amino acid long ORF was identified, containing a dockerin domain, a cohesin domain and a PorSec sorting domain located C-terminal. Since cohesin and dockerin are part of the cellulosome, a multi-enzyme complex produced by anaerobic bacteria for degradation of cellulosic substrates (Bayer *et al.,* 1994, Smith and Bayer 2013), may be associated with this function. Missing hits towards known amylases (GH13) or associated CBMs, such as CBM48, CBM20 and CBM34, in annotation and CAZy analyses of

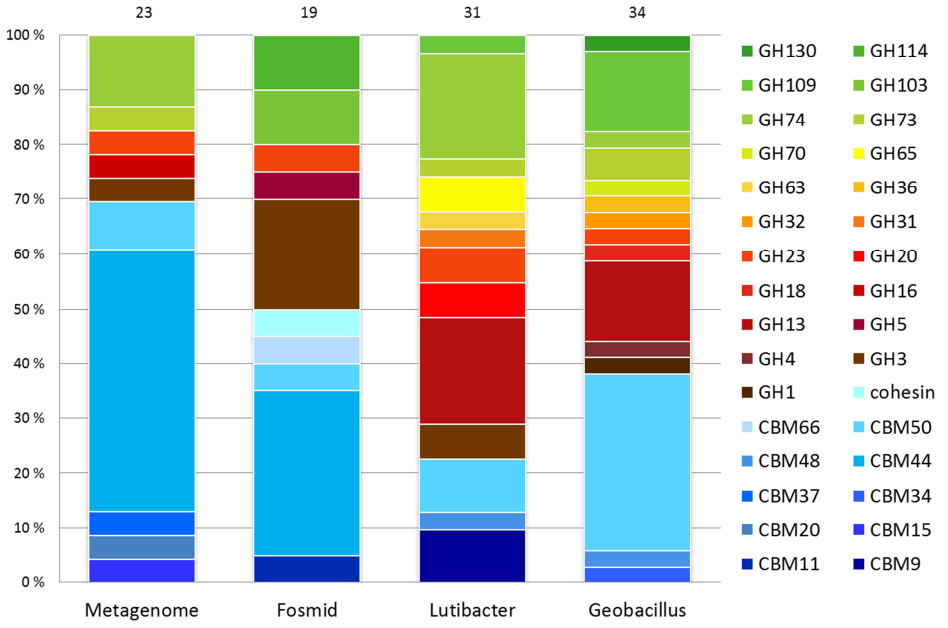

Figure 6: Comparison of GH and CBM family profiles of AMOR (meta-) genomes. The total number is placed above the individual columns.

sequenced fosmids could be an indication for novel amylase sequences. To verify these assumptions, further investigation is needed.

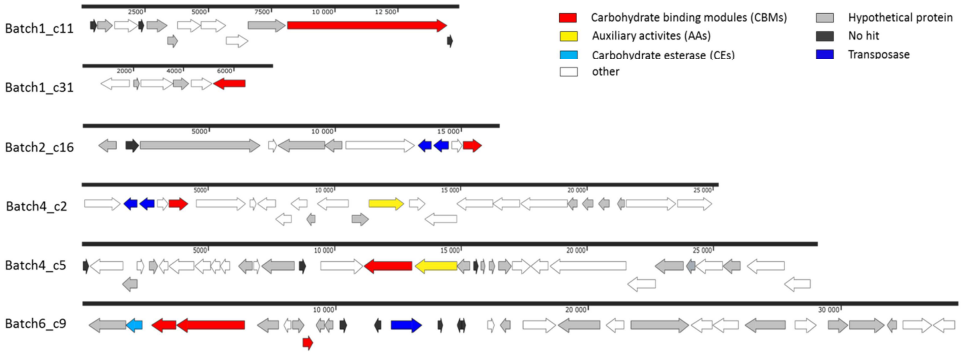

Figure 7: Graphical representation of metagenomic DNA fosmid inserts with detected CBMs. Detected and annotated genes are represented as arrows.

3.2.3 Conclusion – metagenomic library

Apart from the amylase screening, the establishment of a metagenomic fosmid library from a biofilm, originated from a deep-sea hydrothermal vent system, can now serve as a resource for future functional screenings. As approximately 36% of the identified ORFs are of unknown function, this library represents a great source for finding new biocatalysts accompanied by novel sequences and functions.

3.3 New insight into the biotechnological potential of the AMOR vent fields

The results obtained in this project have revealed that the hydrothermal systems located at AMOR represent a valuable source of new microbial isolates and biocatalysts.

An advantage of new microbial isolates is that the cultivation on a diversity of substrates enables the direct screening for hydrolytic activity. The selective cultivation of specialized degrading microorganisms was one essential part of this thesis, resulting in the identification of *Geobacillus* sp. 12AMOR1 (**Paper I**). Research on *Geobacillus* species has been carried out since the 1920s (Donk 1920). The genus is known for its versatile hydrolytic capacity at elevated temperatures and diverse enzymes have been applied for decades in biotechnology

(Zeigler 2014). Genome sequencing of *Geobacillus* sp. 12AMOR1 enabled the identification of five GH13 genes, encoding a trehalase, a neopullulanase, a glycogen-branching enzyme and two α-amylases (**Paper I**). Sequence similarities with GH13 from other *Geobacillus* strains were high (99%), and all five GH13 were expressed and purified in *E. coli*. The starch degradation capacity of the neopullulanase was comparable with a commercially available α-amylase from *B. licheniformis* (Sigma-Aldrich) at 60 °C (**Paper I**). With a melting temperature of 76.4 °C, this thermophilic enzyme could be suitable for processes at elevated temperatures, such as the saccharification step in the starch industry (Prakash and Jaiswal 2010). One of the *Geobacillus* sp. 12AMOR1 α-amylases (GARCT_00683) showed a strong hydrolytic potential, however the enzyme could not be purified in sufficient amount to conduct further experiments. Unfortunately, a high conformity of genes between *Geobacillus* strains deteriorates the potential of the identified biocatalysts from *Geobacillus* sp. 12AMOR1. The ~26% of hypothetical genes in the genome gives scope for the discovery of new enzymes.

The genus *Lutibacter* was first described 10 years ago and research has so far focused on isolation and characterization of novel strains (Choi and Cho 2006, Nedashkovskaya *et al.,* 2015), not on the biotechnological potential. *L. profundi* LP1^T represents to our knowledge the first sequenced and characterized genome of this genus (**Paper III**). Annotation analysis using Prokka (Seemann 2014) identified six GH13 as sucrose phosphorylase, malto-oligosyltrehalose hydrolase, two cyclomaltodextrinases and two α-amylases. It should be noted that cyclomaltodextrinases, neopullulanases and maltogenic amylases are not easily distinguishable from each other and need detailed characterization (Lee *et al.,* 2002). Interestingly, the automated NCBI Prokaryotic Genome Annotation Pipeline (Angiuoli *et al.,* 2008) identified only three of the six GH13 as α-amylases. The remaining genes were annotated as hypothetical proteins (**Paper III**). Blastp analyses of the six GH13 revealed amino acid sequence similarities between 60 to 80% with homologous glycosylases, which have been mostly identified by sequencing without biochemical characterization. Altogether, this strengthens the potential biotechnological value of enzymes from *L. profundi* LP1^T.

The frequency of identified CAZymes varied from 1.99% to 5.73% in the different (meta-) genome samples (Table 3). Both isolates, *Geobacillus* sp. 12AMOR1 and *L. profundi* LP1^T revealed a small selected arsenal of glycoside hydrolases in comparison to known complex

polysaccharide degraders (Xu *et al.,* 2003, Izquierdo *et al.,* 2012, Mann *et al.,* 2013). They comprised glycoside hydrolase families able to hydrolyse α- and β-glucosidic poly- and oligosaccharides (Figure 6). Similar results have been obtained from both 09ROV3-BS metagenomic studies, with even lower numbers of identified glycoside hydrolases, still covering different, mostly β-linked hydrolytic activities. A CBM20-type ORF, known for starch binding, with an E-value close to cut-off threshold e-05 has been found in the 09ROV3-BS metagenomic shotgun sequencing data (Figure 6). The corresponding ORF comprised a hypothetical protein of 379 amino acids with only 39% identity to a hypothetical protein from a *Bacteroidales* bacterium TBC1. This gene represents an interesting candidate for future amylase activity screenings.

Table 3: Number of genes associated with certain enzyme classes found in the AMOR ecosystem applying diverse –omic approaches.

		Metagenome 09ROV03 Bacteroidetes	**Fosmid library 09ROV03**	**Lutibacter genome**	**Geobacillus Genome**
	Scaffolds	60	486	1 (Isolate)	1 (Isolate)
	Assembly (Mbp)	~1.77	3.65	~2.967	~3.41
	Total ORF #	2617	3408		
	Annotated ORF	2373	3017	2601	3323
dbCAN	AA	9	5	3	3
	CBM	16	9	7	13
	GH	7	10	24	21
	GT	80	21	45	28
	CE	20	15	24	23
	PL	4	0	1	0
	Total	**136**	**60**	**104**	**88**
MEROPS	**Total**	**88**	**70**	**130**	**127**
Frequency (%)	CAZy	5.73	1.99	4.00	2.65
	GH	0.29	0.33	0.92	0.63
	GH13	0.00	0.00	0.23	0.15
	MEROPS	3.71	2.32	5.00	3.82

AA – auxiliary activities; CBM – carbohydrate binding module; GH – glycoside hydrolases; GT – glycosyl transferases; CE – carbohydrate esterases; PL – polysaccharide lyases; CAZy – includes all carbohydrate-active enzymes.

Proteases have a wide application range in industry such as detergents, feed, textile and the leather industry. The ability to hydrolyse recalcitrant proteinous material, such as keratin, has recently become gained increased interest (Brandelli 2008, Gupta *et al.,* 2013). In the early

process of characterizing *Geobacillus* sp. 12AMOR1, the bacterium was also tested for proteinase activity on plates using selected substrates, among them were substrates of industrial relevance and strong hydrolytic activity could be observed. The identification of the corresponding proteases is now part of another PhD project. Blast analysis against the MEROPS protein database identified a high number of putative proteases, with a frequency of 3.82%, compared to CAZymes with a frequency of 2.65% (Table 3). This also provided further motive to analyse additional samples for putative proteases. The results revealed a generally higher frequency of proteases compared to carbohydrate-active enzymes in the AMOR hydrothermal vent environments (Table 3). An exception is the shotgun sequenced metagenome, where a higher number of carbohydrate-active enzymes (136) were identified in comparison to proteases (88). Both isolates contain a similar number of predicted proteases with 127 and 130 for *Geobacillus* sp. 12AMOR1 and *L. profundi* LP1^T, respectively. Proteolytic activity has been verified for *L. profundi* LP1^T on agar plates (Le Moine Bauer *et al.,* in review). The largest number of putative proteases was found in *L. profundi* LP1^T with a frequency of 5%. The isolate encoded several industrial applied endopeptidase families, such as chymotrypsin, subtilisin or thermolysin. A comparative genome study of different *Bacteroidetes* revealed a higher number of proteases in isolates from deep-sea marine sediments than from surface waters, suggesting a role in organic carbon and nitrogen cycling by degradation of proteinaceous organic nitrogen (Qin *et al.,* 2010). Since the 09ROV3-BS library originated from the same biofilm as *L. profundi* LP1^T, this makes it an ideal source to screen for protease activity in the future. Preliminary analysis of the sequenced metagenomic DNA library against the MEROPS peptide database identified 70 unique ORF's as proteases (Table 3). Again, one third of those proteases are annotated as hypothetical proteins or proteins with a general function, such as α/β hydrolase or protease. In addition, potential proteases from all sequencing data contained members of both bacterial keratinase MEROPS families S8 and M14 (Lange *et al.,* 2016).

Typically, enzymes withstand conditions similar to the growth conditions of the microorganism they originate from. In the case of the mesophilic *L. profundi* LP1^T, it corresponds to temperatures between 13 and 34 °C, pH between 5.2 and 7.5 and NaCl concentration of up to 3% (Le Moine Bauer *et al.,* in review). Applications using low to moderate process temperatures could therefore be suitable for *L. profundi* LP1^T derived

enzymes, such as in textile or detergent industry. Biocatalysts from the thermophilic *Geobacillus* sp. 12AMOR1 is expected to be active under a range of temperatures from 40 to 70 °C, pH range of 5.5 to 9.0 and NaCl concentration of up to 5%. This makes the GH13 enzymes feasible for applications involving elevated temperatures, such as starch conversion, paper or detergent industry.

AMOR deep-sea hydrothermal vents represent a source for microorganisms and biocatalysts covering a wide range of physiochemical parameters, which are interesting for various applications. Hence, a combination of functional screening and sequencing-based approaches have implemented the identification of amylases and putative proteases in new isolates, which could encourage the search for interesting biocatalysts from deep-sea hydrothermal vents even further.

3.4 New insight into the ecology of the AMOR vent fields

The results presented in this project have increased the understanding of the function and interaction within microbial communities in deep-sea hydrothermal systems.

Epsilonproteobacteria comprise a significant proportion of the microbial population in deep-sea hydrothermal vents and are found to dominate in communities associated with chimney structures, microbial mats, diffuse flow hydrothermal fluids and epibiotic microflora of deep-sea vent metazoans (Moyer *et al.,* 1995, Campbell *et al.,* 2001, Huber *et al.,* 2003, Flores *et al.,* 2011). As they are mesophilic to moderately thermophilic, they are important primary producers, capable of oxidizing sulphur compounds and hydrogen with oxygen, nitrate or sulphur compounds as terminal electron acceptors (Campbell *et al.,* 2006, Nakagawa *et al.,* 2007, Sievert and Vetriani 2012). Vent fluids of the LCVF are enriched in high concentrations of H_2S, H_2 and CO_2 (Pedersen *et al.,* 2010a). Furthermore, high concentrations of CH_4 and NH_4^+ in the fluids together with a pH of 5.5 suggest an additional sedimentary influence of the system. Thermodynamic models of the LCVF suggest that this vent field represents an extremity in terms of its energetic potential for hosting anaerobic and aerobic methane oxidizers, aerobic ammonium oxidizers as well as sulphide oxidizers (Dahle et al., 2015).

In the biofilm studied in this project, primary production by *Epsilonproteobacteria* appears to involve simultaneous utilisation of sulphur species, hydrogen, formate with nitrate and oxygen coupled to CO_2 fixation using the reverse TCA cycle (Dahle *et al.,* 2013). The dominating genus *Sulfurovum* is affiliated with the biofilm formation by producing unique extracellular polymeric substances (EPS) in form of long filamentous sheaths surrounding the cells (**Paper II**). The EPS represents thermotolerant, recalcitrant polysaccharides resembling chitin or cellulose (**Paper II**). Furthermore, *Bacteroidetes* were identified to colonize the microbial filaments as epibionts. The comprehensive -omics data obtained for the 09ROV3-BS biofilm, together with microscopy, enabled the evaluation of metabolic capability of the two individual taxonomic groups in the community (**Paper II**). The link between the primary production of lithotrophic *Sulfurovum* and organotrophic consumption by *Bacteroidetes* within the food web has been addressed. This is not only beneficial for understanding the hydrolytic potential of the microbial community, but also crucial for the energy flow within the biofilm food web. A corresponding culture-dependent approach contributed to gain insight to the metabolic potential of a *Bacteroidetes* biofilm member and its potential function within the community (**Paper III**). Characterization of the mesophilic *L. profundi* LP1^T (Le Moine Bauer *et al.,* in review) and genome analysis (**Paper III**) revealed interesting metabolic features, albeit the isolate may not be the predominant *Bacteroidetes* member within the biofilm. Some features found in -omics studies could be confirmed though genome analysis, such as the presence of a gliding apparatus. *L. profundi* LP1^T does not represent all features found for *Bacteroidetes* in the –omics approach, e.g. Type IV pilus or other different potential adhesion genes. However, the following metabolic traits have been revealed through the genome analysis, adding new knowledge that was not identified in the previous –omics study. *L. profundi* LP1^T harbours a complete denitrification pathway, whereas in the –omics study only a partial denitrification pathway for reduction of nitrite to nitrous oxide was identified. Growth experiments with the isolate confirmed denitrification of nitrate to nitrite (Le Moine Bauer *et al.,* in review), strengthening the hypothesis of heterotrophic denitrification by *Bacteroidetes* under limited oxygen levels rather than under aerobic conditions. Furthermore, genes involved in sulphide oxidation were identified, such as sulphide:quinone reductase (SQR), which may contribute to sulphide detoxification (**Paper III**). The information obtained from the genome confirms the

organotrophic lifestyle of degrading sugar polymers and proteins, and further provides an overview of the adaption strategies towards its habitat (**Paper III**).

The genus *Geobacillus* comprises a group of Gram-positive, obligate thermophilic, chemo-organotrophic bacteria, including obligate aerobes, denitrifiers, and facultative anaerobes that can grow over a range of 45-75 °C (Coorevits *et al.,* 2012, Hussein *et al.,* 2015). Members of this genus have shown versatile hydrolytic traits at elevated temperatures, and the organisms themselves, or their enzymes, are applied in various biotechnological processes (Zeigler 2001, Hussein *et al.,* 2015). A thermophilic, aerobic, organotrophic *Geobacillus*, belonging to the phylum *Firmicutes*, was isolated from hydrothermal sediments (~90 °C) at JMVF (**Paper I**). The genome of *Geobacillus* sp. 12AMOR1 and its phenotypical properties have been characterized with respect to its biotechnological potential to hydrolyse carbohydrates (**Paper I**). As environmental knowledge about the JMVF hot sediments is limited, the genome information has not been used to assess the functional role of the isolate in the habitat. However, the information provided by genome sequencing can be used in future ecological studies of this environment. In this way, the obtained –omics data, including genome sequencing, contributed additionally to the knowledge of the AMOR hydrothermal vent microbial ecology.

3.5 Future research - Mining potential for other enzymatic groups originating from AMOR

The combinatory approach in this project, involved cultivation and genome analysis of a new *Geobacillus* strain from the JMVF as well as sequence-based and functional metagenomics of a biofilm community attached to a black smoker in the LCVF. The approaches have created a huge data set that is of significance for future mining of new biocatalysts from the AMOR vent fields.

In terms of revealing GH13 amylases, selection by functional screening using cultivation techniques of isolates has been a better choice for identifying amylases in comparison to metagenomic sequencing. Furthermore, amylase activity has been observed in functional screening of a metagenomic library against starch without revealing genes with sequence similarities to known amylases or their auxiliary CBMs in sequence analyses. In my opinion,

future analyses to verify those identified activities in library clones would be an exciting task. Suitable strategies could include retransformation of the plasmids to *E. coli* host strains, the use of different chromogenic substrates or better functional assays. The usage of environmental DNA libraries still enables the identification of novel amylases, despite the intensive research for many decades (Delavat *et al.,* 2012, Liu *et al.,* 2012b, Maruthamuthu *et al.,* 2016). Research continually improves functional assays for carbohydrate-active enzymes in terms of high throughput analyses of clones (Nyyssonen *et al.,* 2013), development of new detection methods (Vidal-Melgosa *et al.,* 2015), application of multiple, simultaneous screenings (Maruthamuthu *et al.,* 2016) and development of new chromogenic substrates (Kracun *et al.,* 2015). These improvements should be applied in further studies in order to achieve the best possible results.

Besides the amylase screening, the established metagenomic library is and remains a resource for other functional screenings in the near future. By using the sequenced metagenomic DNA library as reference, it can be estimated that approximately 36% of the gene sequences may have a hypothetical function. This “dark matter” has a great potential for the discovery of new and potentially novel biocatalysts.

Sequencing of genomes and metagenomes present an overview of the genomic content of an organism or a microbial community. Bioinformatic resources enable the mining for biocatalysts of genomes and metagenomes, based on sequence similarity. Specialized databases, such as the CAZy and MEROPS database, are especially helpful in revealing gene functions. In this study, a variety of carbohydrate-active enzymes, including enzymes belonging to GH13, and proteases have been identified using bioinformatical tools. In future projects, cloning, expression and characterization of those enzymes would be interesting in order to confirm their enzymatic activity and biochemical properties in an industrial context.

Annotation of genes in genomes and metagenomes relies on homology to known and, in the best case, characterized protein sequences. In the current sequencing era the amount of sequenced and annotated data increases rapidly in available databases. The majority of these annotated genes have not been confirmed in wet lab studies, which may cause incorrect annotations. Furthermore, considering the vast amount of genes with unknown function, there is a huge potential for discoveries of novel functions, also in terms of biocatalysts. Piao

et al. successfully mined “genomic dark matter” for novel cellulases by combining sequence similarity searches of higher sensitivity with consideration of the genomic context (Piao *et al.,* 2014). An emphasis, on developing more sensitive algorithms and incorporating additional information, such as auxillary domains or gene context, will help to improve sequence prediction based on homology and thereby to unravel “genomic dark matter”. I am confident that those improved bioinformatics tools will be available in the near future.

4. Conclusion

The results obtained in my thesis have demonstrated that the AMOR deep-sea hydrothermal vent systems are valuable sources for new microorganisms and biocatalysts to meet the needs in biotechnology. The described microbial isolates represent a valuable source of new biocatalysts, like in the case of the thermophilic *Geobacillus* sp. 12AMOR1, or divergent from other known microorganisms, such as the mesophilic *Lutibacter profundi* LP1^{T}. The combination of sequencing-based and functional-based metagenomics has complemented each other in terms of finding potential amylases. The high number of amylase activity in functional screening in combination with non-existing sequence similarities points towards the possibility of identifying novel ones. Furthermore, it allowed the discovery of other potential biocatalysts, such as β-polymer degrading enzymes, glycosyltransferases and proteases. Besides exploiting metagenomics for bioprospecting, the comprehensive data enabled an expansive insight to functional and metabolic traits in the deep-sea hydrothermal vent systems and linking them to taxonomical groups. The individual approaches provided different information, which combined contributed to our understanding of interactions in deep-sea hydrothermal food webs.

5. References

Adrio, J. L. and A. L. Demain (2014). "Microbial enzymes: tools for biotechnological processes." *Biomolecules* **4**(1): 117-139.

Albertsen, M., P. Hugenholtz, A. Skarshewski, K. L. Nielsen, G. W. Tyson and P. H. Nielsen (2013). "Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes." *Nat Biotechnol* **31**(6): 533-538.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." *J Mol Biol* **215**(3): 403-410.

Angiuoli, S. V., A. Gussman, W. Klimke, G. Cochrane, D. Field, G. Garrity, C. D. Kodira, N. Kyrpides, R. Madupu, V. Markowitz*, et al.* (2008). "Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation." *OMICS* **12**(2): 137-141.

Antranikian, G., P. L. Joergensen, M. Wuempelmann and S. T. Joergensen (1992). Novel thermostable pullulanases. Novo Nordisk As, Google Patents, WO1992002614 A1

Antranikian, G. and P. L. J. Rgensen (1996). Thermostable pullulanase. Novo Nordisk A/S, Google Patents, EP0578672 B1

Antranikian, G. and C. Sjoholm (1997). $i(*THERMOCOCCUS*) AMYLASE AND PULLULANASE. Novo Nordisk A/S, Google Patents, EP0793716 A1

Antranikian, G., C. E. Vorgias and C. Bertoldo (2005). "Extreme Environments as a Resource for Microorganisms and Novel Biocatalysts." *Adv in Biochem Engin/Biotechnol* **96**: 219-262.

Bayer, E. A., E. Morag and R. Lamed (1994). "The cellulosome--a treasure-trove for biotechnology." *Trends Biotechnol* **12**(9): 379-386.

Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich*, et al.* (2000a). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." *Science* **289**(5486): 1902-1906.

Beja, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich*, et al.* (2000b). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." *Environ Microbiol* **2**(5): 516-529.

Bell, G. S., R. J. Russell, H. Connaris, D. W. Hough, M. J. Danson and G. L. Taylor (2002). "Stepwise adaptations of citrate synthase to survival at life's extremes. From psychrophile to hyperthermophile." *Eur J Biochem* **269**(24): 6250-6260.

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

Biver, S., A. Stroobants, D. Portetelle and M. Vandenbol (2014). "Two promising alkaline beta-glucosidases isolated by functional metagenomics from agricultural soil, including one showing high

tolerance towards harsh detergents, oxidants and glucose." *J Ind Microbiol Biotechnol* **41**(3): 479-488.

Biwer, A., G. Antranikian and E. Heinzle (2002). "Enzymatic production of cyclodextrins." *Appl Microbiol Biotechnol* **59**(6): 609-617.

Brandelli, A. (2008). "Bacterial Keratinases: Useful Enzymes for Bioprocessing Agroindustrial Wastes and Beyond." *Food and Bioprocess Technology* **1**(2): 105-116.

Campbell, B. J., A. S. Engel, M. L. Porter and K. Takai (2006). "The versatile *epsilon-proteobacteria*: key players in sulphidic habitats." *Nat Rev Microbiol* **4**(6): 458-468.

Campbell, B. J., C. Jeanthon, J. E. Kostka, G. W. Luther, 3rd and S. C. Cary (2001). "Growth and phylogenetic properties of novel bacteria belonging to the epsilon subdivision of the *Proteobacteria* enriched from *Alvinella pompejana* and deep-sea hydrothermal vents." *Appl Environ Microbiol* **67**(10): 4566-4572.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." *Nucleic Acids Research* **37**: D233-D238.

Cavicchioli, R., R. Amils, D. Wagner and T. McGenity (2011a). "Life and applications of extremophiles." *Environ Microbiol* **13**(8): 1903-1907.

Cavicchioli, R., T. Charlton, H. Ertan, S. Mohd Omar, K. S. Siddiqui and T. J. Williams (2011b). "Biotechnological uses of enzymes from psychrophiles." *Microb Biotechnol* **4**(4): 449-460.

Cherry, J. R. and A. L. Fidantsef (2003). "Directed evolution of industrial enzymes: an update." *Curr Opin Biotechnol* **14**(4): 438-443.

Chevreux, B., T. Pfisterer, B. Drescher, A. J. Driesel, W. E. Muller, T. Wetter and S. Suhai (2004). "Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs." *Genome Res* **14**(6): 1147-1159.

Choi, D. H. and B. C. Cho (2006). "*Lutibacter litoralis* gen. nov., sp. nov., a marine bacterium of the family *Flavobacteriaceae* isolated from tidal flat sediment." *International journal of systematic and evolutionary microbiology* **56**(Pt 4): 771-776.

Coorevits, A., A. E. Dinsdale, G. Halket, L. Lebbe, P. De Vos, A. Van Landschoot and N. A. Logan (2012). "Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly '*thermoglucosidasius*'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov." *Int J Syst Evol Microbiol* **62**(7): 1470-1485.

Corliss, J. B., J. Dymond, L. I. Gordon, J. M. Edmond, R. P. V. Herzen, R. D. Ballard, K. Green, D. Williams, A. Bainbridge, K. Crane*, et al.* (1979). "Submarine Thermal Springs on the Galapagos Rift." *Science* **203**(4385): 1073-1083.

Crabb, W. D. and J. K. Shetty (1999). "Commodity scale production of sugars from starches." *Curr Opin Microbiol* **2**(3): 252-256.

Craig, J. W., F. Y. Chang, J. H. Kim, S. C. Obiajulu and S. F. Brady (2010). "Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse *Proteobacteria*." *Appl Environ Microbiol* **76**(5): 1633-1641.

D'Amico, S., C. Gerday and G. Feller (2003). "Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase." *J Mol Biol* **332**(5): 981-988.

Dahle, H., I. Okland, I. H. Thorseth, R. B. Pederesen and I. H. Steen (2015). "Energy landscapes shape microbial communities in hydrothermal systems on the Arctic Mid-Ocean Ridge." *ISME J* **9**(7): 1593-1606.

Dahle, H., I. Roalkvam, I. H. Thorseth, R. B. Pedersen and I. H. Steen (2013). "The versatile in situ gene expression of an *Epsilonproteobacteria*-dominated biofilm from a hydrothermal chimney." *Environ Microbiol Rep* **5**(2): 282-290.

Dalmaso, G. Z., D. Ferreira and A. B. Vermelho (2015). "Marine extremophiles: a source of hydrolases for biotechnological applications." *Mar Drugs* **13**(4): 1925-1965.

De Santi, C., B. Altermark, D. de Pascale and N. P. Willassen (2016a). "Bioprospecting around Arctic islands: Marine bacteria as rich source of biocatalysts." *J Basic Microbiol* **56**(3): 238-253.

De Santi, C., B. Altermark, M. M. Pierechod, L. Ambrosino, D. de Pascale and N. P. Willassen (2016b). "Characterization of a cold-active and salt tolerant esterase identified by functional screening of Arctic metagenomic libraries." *BMC Biochem* **17**(1): 1.

Delavat, F., V. Phalip, A. Forster, M. C. Lett and D. Lievremont (2012). "Deciphering the role of *Paenibacillus* strain Q8 in the organic matter recycling in the acid mine drainage of Carnoules." *Microb Cell Fact* **11**: 16.

Demain, A. L. and J. L. Adrio (2008). "Contributions of microorganisms to industrial biology." *Mol Biotechnol* **38**(1): 41-55.

Donk, P. J. (1920). "A Highly Resistant Thermophilic Organism." *J Bacteriol* **5**(4): 373-374.

Dubilier, N., C. Bergin and C. Lott (2008). "Symbiotic diversity in marine animals: the art of harnessing chemosynthesis." *Nat Rev Microbiol* **6**(10): 725-740.

Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman*, et al.* (2009). "Real-time DNA sequencing from single polymerase molecules." *Science* **323**(5910): 133-138.

Ekkers, D. M., M. S. Cretoiu, A. M. Kielak and J. D. Elsas (2012). "The great screen anomaly--a new frontier in product discovery through functional metagenomics." *Appl Microbiol Biotechnol* **93**(3): 1005-1020.

Feller, G. (2010). "Protein stability and enzyme activity at extreme biological temperatures." *J Phys Condens Matter* **22**(32): 323101.

Ferrer, M., A. Beloqui, K. N. Timmis and P. N. Golyshin (2009). "Metagenomics for mining new genetic resources of microbial communities." *J Mol Microbiol Biotechnol* **16**(1-2): 109-123.

Finn, R. D., J. Clements and S. R. Eddy (2011). "HMMER web server: interactive sequence similarity searching." *Nucleic Acids Res* **39**(Web Server issue): W29-37.

Flores, G. E., J. H. Campbell, J. D. Kirshtein, J. Meneghin, M. Podar, J. I. Steinberg, J. S. Seewald, M. K. Tivey, M. A. Voytek, Z. K. Yang, *et al.* (2011). "Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge." *Environ Microbiol* **13**(8): 2158-2171.

Flores, G. E. and A. L. Reysenbach (2011). Hydrothermal Environments, Marine. Encyclopedia of Geobiology. Netherlands, Springer**:** 456-467.

Fullwood, M. J., C. L. Wei, E. T. Liu and Y. Ruan (2009). "Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses." *Genome Res* **19**(4): 521-532.

German, C. R., E. Ramirez-Llodra, M. C. Baker, P. A. Tyler and C. Committee (2011). "Deep-water chemosynthetic ecosystem research during the census of marine life decade and beyond: a proposed deep-ocean road map." *PloS one* **6**(8).

Giardina, T., A. P. Gunning, N. Juge, C. B. Faulds, C. S. Furniss, B. Svensson, V. J. Morris and G. Williamson (2001). "Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose." *J Mol Biol* **313**(5): 1149-1159.

Gonzalez, J. M., B. Fernandez-Gomez, A. Fernandez-Guerra, L. Gomez-Consarnau, O. Sanchez, M. Coll-Llado, J. Del Campo, L. Escudero, R. Rodriguez-Martinez, L. Alonso-Saez, *et al.* (2008). "Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (*Flavobacteria*)." *Proc Natl Acad Sci U S A* **105**(25): 8724-8729.

Guillen, D., S. Sanchez and R. Rodriguez-Sanoja (2010). "Carbohydrate-binding domains: multiplicity of biological roles." *Appl Microbiol Biotechnol* **85**(5): 1241-1249.

Gupta, R., R. Rajput, R. Sharma and N. Gupta (2013). "Biotechnological applications and prospective market of microbial keratinases." *Appl Microbiol Biotechnol* **97**(23): 9931-9940.

Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy and R. M. Goodman (1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products." *Chem Biol* **5**(10): R245-249.

Healy, F. G., R. M. Ray, H. C. Aldrich, A. C. Wilkie, L. O. Ingram and K. T. Shanmugam (1995). "Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose." *Appl Microbiol Biotechnol* **43**(4): 667-674.

Hendriksen, H. V. and B. E. Norman (2002). Method for producing maltose syrup by using a hexosyltransferase. Novozymes A/S, Google Patents, WO2002010427 A1

Henrissat, B. (1991). "A classification of glycosyl hydrolases based on amino acid sequence similarities." *Biochem J* **280 (Pt 2)**: 309-316.

Hii, S. L., J. S. Tan, T. C. Ling and A. B. Ariff (2012). "Pullulanase: role in starch hydrolysis and potential industrial applications." *Enzyme Res* **2012**: 921362.

Homaei, A., M. Ghanbarzadeh and F. Monsef (2016). "Biochemical features and kinetic properties of alpha-amylases from marine organisms." *Int J Biol Macromol* **83**: 306-314.

Huber, J. A., D. A. Butterfield and J. A. Baross (2003). "Bacterial diversity in a subseafloor habitat following a deep-sea volcanic eruption." *FEMS Microbiol Ecol* **43**(3): 393-409.

Hussein, A. H., B. K. Lisowska and D. J. Leak (2015). "The genus *Geobacillus* and their biotechnological potential." *Adv Appl Microbiol* **92**: 1-48.

Hyatt, D., G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer and L. J. Hauser (2010). "Prodigal: prokaryotic gene recognition and translation initiation site identification." *BMC Bioinformatics* **11**: 119.

Izquierdo, J. A., L. Goodwin, K. W. Davenport, H. Teshima, D. Bruce, C. Detter, R. Tapia, S. Han, M. Land, L. Hauser*, et al.* (2012). "Complete Genome Sequence of *Clostridium clariflavum* DSM 19732." *Stand Genomic Sci* **6**(1): 104-115.

Jaeschke, A., S. L. Jorgensen, S. M. Bernasconi, R. B. Pedersen, I. H. Thorseth and G. L. Fruh-Green (2012). "Microbial diversity of Loki's Castle black smokers at the Arctic Mid-Ocean Ridge." *Geobiology* **10**(6): 548-561.

Janecek, S., B. Svensson and E. A. MacGregor (2011). "Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals." *Enzyme Microb Technol* **49**(5): 429-440.

Janecek, S., B. Svensson and E. A. MacGregor (2014). "alpha-Amylase: an enzyme specificity found in various families of glycoside hydrolases." *Cell Mol Life Sci* **71**(7): 1149-1170.

Jorgensen, S. L., B. Hannisdal, A. Lanzen, T. Baumberger, K. Flesland, R. Fonseca, L. Ovreas, I. H. Steen, I. H. Thorseth, R. B. Pedersen*, et al.* (2012). "Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge." *Proc Natl Acad Sci U S A* **109**(42): E2846-2855.

Kakirde, K. S., L. C. Parsley and M. R. Liles (2010). "Size does matter: Application-driven approaches for soil metagenomics." *Soil Biology & Biochemistry* **42**(11): 1911-1923.

Kasana, R. C., R. Salwan, H. Dhar, S. Dutt and A. Gulati (2008). "A Rapid and Easy Method for the Detection of Microbial Cellulases on Agar Plates Using Gram's Iodine." *Current Microbiology* **57**(5): 503-507.

Kelley, D. S., J. A. Baross and J. R. Delaney (2002). "Volcanoes, fluids, and life at mid-ocean ridge spreading centers." *Annual Review of Earth and Planetary Sciences* **30**: 385-491.

Kelly, R. M., L. Dijkhuizen and H. Leemhuis (2009). "Starch and alpha-glucan acting enzymes, modulating their properties by directed evolution." *J Biotechnol* **140**(3-4): 184-193.

Kennedy, J., B. Flemer, S. A. Jackson, D. P. Lejon, J. P. Morrissey, F. O'Gara and A. D. Dobson (2010). "Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism." *Mar Drugs* **8**(3): 608-628.

Kindle, K. L., S. E. Mainzer, D. L. Marlatt and C. B. Sawyer (1986). Thermostable alpha amylase having a low requirement for calcium ions, derived from a *Bacillus* microorganism. Corning Glass Works. US, Google Patents, US4600693 A

Knapp, S., S. Kardinahl, N. Hellgren, G. Tibbelin, G. Schafer and R. Ladenstein (1999). "Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon *Sulfolobus acidocaldarius* at 2.2 A resolution." *J Mol Biol* **285**(2): 689-702.

Knittel, K. and A. Boetius (2009). "Anaerobic oxidation of methane: progress with an unknown process." *Annu Rev Microbiol* **63**: 311-334.

Kodzius, R. and T. Gojobori (2015). "Marine metagenomics as a source for bioprospecting." *Mar Genomics* **24 Pt 1**: 21-30.

Kracun, S. K., J. Schuckel, B. Westereng, L. G. Thygesen, R. N. Monrad, V. G. Eijsink and W. G. Willats (2015). "A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes." *Biotechnol Biofuels* **8**: 70.

Kragh, K. M., H. Leemhuis, L. Dijkhuizen and B. W. Dijkstra (2010). Cyclodextrin glycosyltransferase (CGTase) polypeptides with modified hydrolysis activity. Danisco A/S, Google Patents, US7803598 B2

Lammle, K., H. Zipper, M. Breuer, B. Hauer, C. Buta, H. Brunner and S. Rupp (2007). "Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning." *J Biotechnol* **127**(4): 575-592.

Lange, L., Y. Huang and P. K. Busk (2016). "Microbial decomposition of keratin in nature-a new hypothesis of industrial relevance." *Appl Microbiol Biotechnol* **100**(5): 2083-2096.

Lanzen, A., S. L. Jorgensen, M. M. Bengtsson, I. Jonassen, L. Ovreas and T. Urich (2011). "Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA." *FEMS Microbiol Ecol* **77**(3): 577-589.

Le Moine Bauer, S., I. Roalkvam, I. H. Steen and H. Dahle (in review). "*Lutibacter profundi* sp. nov., isolated from a deep-sea hydrothermal system on the Arctic mid-ocean ridge." *International Journal of Systematic and Evolutionary Microbiology* **in Revision**.

Lee, D. H., S. L. Choi, E. Rha, S. J. Kim, S. J. Yeom, J. H. Moon and S. G. Lee (2015). "A novel psychrophilic alkaline phosphatase from the metagenome of tidal flat sediments." *BMC Biotechnol* **15**: 1.

Lee, H. S., M. S. Kim, H. S. Cho, J. I. Kim, T. J. Kim, J. H. Choi, C. Park, H. S. Lee, B. H. Oh and K. H. Park (2002). "Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other." *J Biol Chem* **277**(24): 21891-21897.

Leiros, H. K., A. E. Fedoy, I. Leiros and I. H. Steen (2012). "The complex structures of isocitrate dehydrogenase from *Clostridium thermocellum* and *Desulfotalea psychrophila* suggest a new active site locking mechanism." *FEBS Open Bio* **2**: 159-172.

Levasseur, A., E. Drula, V. Lombard, P. M. Coutinho and B. Henrissat (2013). "Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes." *Biotechnol Biofuels* **6**(1): 41.

Littlechild, J. A. (2015). "Archaeal Enzymes and Applications in Industrial Biocatalysts." *Archaea* **2015**: 147671.

Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu and M. Law (2012a). "Comparison of next-generation sequencing systems." *J Biomed Biotechnol* **2012**: 251364.

Liu, Y., Y. Lei, X. Zhang, Y. Gao, Y. Xiao and H. Peng (2012b). "Identification and phylogenetic characterization of a new subfamily of alpha-amylase enzymes from marine microorganisms." *Mar Biotechnol (NY)* **14**(3): 253-260.

MacGregor, E. A. and B. Svensson (1989). "A super-secondary structure predicted to be common to several alpha-1,4-D-glucan-cleaving enzymes." *Biochem J* **259**(1): 145-152.

Mallick, P., D. R. Boutz, D. Eisenberg and T. O. Yeates (2002). "Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds." *Proc Natl Acad Sci U S A* **99**(15): 9679-9684.

Mann, A. J., R. L. Hahnke, S. Huang, J. Werner, P. Xing, T. Barbeyron, B. Huettel, K. Stuber, R. Reinhardt, J. Harder, *et al.* (2013). "The genome of the alga-associated marine *flavobacterium Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides." *Appl Environ Microbiol* **79**(21): 6813-6822.

Marcy, Y., C. Ouverney, E. M. Bik, T. Losekann, N. Ivanova, H. G. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. A. Relman, *et al.* (2007). "Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth." *Proc Natl Acad Sci U S A* **104**(29): 11889-11894.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, *et al.* (2005). "Genome sequencing in microfabricated high-density picolitre reactors." *Nature* **437**(7057): 376-380.

Maruthamuthu, M., D. J. Jimenez, P. Stevens and J. D. van Elsas (2016). "A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes." *BMC Genomics* **17**(1): 86.

Mathur, E. J. (1996). Purified thermostable *Pyrococcus furiosus* DNA polymerase I. Stratagene, Google Patents, US5545552 A

Matsunaga, A., H. Amano and S. Yamaguchi (2011). Beta-amylase, gene coding therefor and manufacturing method thereof. Amano Enzyme Inc., Google Patents, US20110059491 A1

McLean, B. W., A. B. Boraston, D. Brouwer, N. Sanaie, C. A. Fyfe, R. A. Warren, D. G. Kilburn and C. A. Haynes (2002). "Carbohydrate-binding modules recognize fine substructures of cellulose." *J Biol Chem* **277**(52): 50245-50254.

Michalska, K., A. D. Steen, G. Chhor, M. Endres, A. T. Webber, J. Bird, K. G. Lloyd and A. Joachimiak (2015). "New aminopeptidase from "microbial dark matter" archaeon." *FASEB J* **29**(9): 4071-4079.

Michel, G., T. Barbeyron, B. Kloareg and M. Czjzek (2009). "The family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity." *Glycobiology* **19**(6): 615-623.

Mirete, S., V. Morgante and J. E. Gonzalez-Pastor (2016). "Functional metagenomics of extreme environments." *Curr Opin Biotechnol* **38**: 143-149.

Moyer, C. L., F. C. Dobbs and D. M. Karl (1995). "Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii." *Appl Environ Microbiol* **61**(4): 1555-1562.

Nakagawa, S., Y. Takaki, S. Shimamura, A. L. Reysenbach, K. Takai and K. Horikoshi (2007). "Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens." *Proc Natl Acad Sci U S A* **104**(29): 12146-12150.

Nakamura, K. and K. Takai (2014). "Theoretical constraints of physical and chemical properties of hydrothermal fluids on variations in chemolithotrophic microbial communities in seafloor hydrothermal systems." *Prog. Earth Planet Sci.* **1**: 5.

Nedashkovskaya, O. I., S. Van Trappen, N. V. Zhukova and P. De Vos (2015). "*Lutibacter holmesii* sp. nov., a novel marine bacterium of the family *Flavobacteriaceae* isolated from the sea urchin *Strongylocentrotus intermedius* and emended description of the genus *Lutibacter*." *Int J Syst Evol Microbiol*.

Nedwin, G. E., V. Sharma and J. K. Shetty (2013). Alpha-amylase blend for starch processing and method of use thereof. Danisco Us Inc., Google Patents, US8545907 B2

Neveu, J., C. Regeard and M. S. DuBow (2011). "Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts." *Appl Microbiol Biotechnol* **91**(3): 635-644.

Nyyssonen, M., H. M. Tran, U. Karaoz, C. Weihe, M. Z. Hadi, J. B. Martiny, A. C. Martiny and E. L. Brodie (2013). "Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries." *Front Microbiol* **4**: 282.

Orcutt, B. N., J. B. Sylvan, N. J. Knab and K. J. Edwards (2011). "Microbial ecology of the dark ocean above, at, and below the seafloor." *Microbiol Mol Biol Rev* **75**(2): 361-422.

Park, B. H., T. V. Karpinets, M. H. Syed, M. R. Leuze and E. C. Uberbacher (2010). "CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database." *Glycobiology* **20**(12): 1574-1584.

Pedersen, R. B., H. T. Rapp, I. H. Thorseth, M. D. Lilley, F. J. Barriga, T. Baumberger, K. Flesland, R. Fonseca, G. L. Fruh-Green and S. L. Jorgensen (2010a). "Discovery of a black smoker vent field and vent fauna at the Arctic Mid-Ocean Ridge." *Nat Commun* **1**: 126.

Pedersen, R. B., I. H. Thorseth, T. E. Nygaard, M. D. Lilley and D. S. Kelley (2010b). "Hydrothermal Activity at the arctic Mid-Ocean Ridges." *Geophysical Monograph* **188**: 67-90.

Peng, H., Y. Zheng, M. Chen, Y. Wang, Y. Xiao and Y. Gao (2014). "A starch-binding domain identified in alpha-amylase (AmyP) represents a new family of carbohydrate-binding modules that contribute to enzymatic hydrolysis of soluble starch." *FEBS Lett* **588**(7): 1161-1167.

Piao, H., J. Froula, C. Du, T. W. Kim, E. R. Hawley, S. Bauer, Z. Wang, N. Ivanova, D. S. Clark, H. P. Klenk*, et al.* (2014). "Identification of novel biomass-degrading enzymes from genomic dark matter: Populating genomic sequence space with functional annotation." *Biotechnol Bioeng* **111**(8): 1550-1565.

Podar, M. and A. L. Reysenbach (2006). "New opportunities revealed by biotechnological explorations of extremophiles." *Curr Opin Biotechnol* **17**(3): 250-255.

Pope, P. B., A. K. Mackenzie, I. Gregor, W. Smith, M. A. Sundset, A. C. McHardy, M. Morrison and V. G. Eijsink (2012). "Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci." *PLoS One* **7**(6): e38571.

Prakash, O. and N. Jaiswal (2010). "alpha-Amylase: an ideal representative of thermostable enzymes." *Appl Biochem Biotechnol* **160**(8): 2401-2414.

Prasad, S., P. Manasa, S. Buddhi, P. Tirunagari, Z. Begum, S. Rajan and S. Shivaji (2014). "Diversity and bioprospective potential (cold-active enzymes) of cultivable marine bacteria from the subarctic glacial Fjord, Kongsfjorden." *Curr Microbiol* **68**(2): 233-238.

Qin, Q. L., X. Y. Zhang, X. M. Wang, G. M. Liu, X. L. Chen, B. B. Xie, H. Y. Dang, B. C. Zhou, J. Yu and Y. Z. Zhang (2010). "The complete genome of *Zunongwangia profunda* SM-A87 reveals its adaptation to the deep-sea environment and ecological role in sedimentary organic nitrogen degradation." *BMC Genomics* **11**: 247.

Radu, C. D., O. Parteni and L. Ochiuz (2016). "Applications of cyclodextrins in medical textiles - review." *J Control Release* **224**: 146-157.

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

Rashid, M. and U. Stingl (2015). "Contemporary molecular tools in microbial ecology and their application to advancing biotechnology." *Biotechnol Adv*.

Rawlings, N. D. and F. R. Morton (2008). "The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome." *Biochimie* **90**(2): 243-259.

Rinke, C., P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J. F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies*, et al.* (2013). "Insights into the phylogeny and coding potential of microbial dark matter." *Nature* **499**(7459): 431-437.

Russell, R. J., U. Gerike, M. J. Danson, D. W. Hough and G. L. Taylor (1998). "Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium." *Structure* **6**(3): 351-361.

Sarmiento, F., R. Peralta and J. M. Blamey (2015). "Cold and Hot Extremozymes: Industrial Relevance and Current Trends." *Front Bioeng Biotechnol* **3**: 148.

Schildkraut, I. and E. Schildkraut (2008). Discovery, cloning and purification of *Thermococcus* sp. (strain 9°n-7) dna ligase. Inc. New England Biolabs, Google Patents, EP1929000 A1

Schipper, C., C. Hornung, P. Bijtenhoorn, M. Quitschau, S. Grond and W. R. Streit (2009). "Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*." *Appl Environ Microbiol* **75**(1): 224-233.

Schloss, P. D. and J. Handelsman (2003). "Biotechnological prospects from metagenomics." *Curr Opin Biotechnol* **14**(3): 303-310.

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

Schouw, A., T. L. Eide, R. Stokke, R. B. Pedersen, I. H. Steen and G. Bodtker (2016). "*Abyssivirga alkaniphila*, gen. nov. sp. nov., an alkane degrading anaerobic bacterium from a deep-sea hydrothermal vent system. Emended description of *Natranaerovirga pectinivora* and *Natranaerovirga hydrolytica*." *Int J Syst Evol Microbiol*.

Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." *Bioinformatics* **30**(14): 2068-2069.

Shendure, J., G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra and G. M. Church (2005). "Accurate multiplex polony sequencing of an evolved bacterial genome." *Science* **309**(5741): 1728-1732.

Shinohara, M. L., M. Ihara, M. Abo, M. Hashida, S. Takagi and T. C. Beck (2001). "A novel thermostable branching enzyme from an extremely thermophilic bacterial species, *Rhodothermus obamensis*." *Appl Microbiol Biotechnol* **57**(5-6): 653-659.

Shiosaka, M. (1976). Heat stable cyclodextrin glycosyltransferase. Incorporated Hayashibara Biochemical Laboratories, Google Patents, US3988206 A

Sievert, S. M. and C. Vetriani (2012). "Chemoautotrophy at Deep-Sea Vents Past, Present, and Future." *Oceanography* **25**(1): 218-233.

Simon, C. and R. Daniel (2011). "Metagenomic analyses: past and future trends." *Appl Environ Microbiol* **77**(4): 1153-1161.

Simon, C., J. Herath, S. Rockstroh and R. Daniel (2009). "Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice." *Appl Environ Microbiol* **75**(9): 2964-2968.

Singh, B. K. (2010). "Exploring microbial diversity for biotechnology: the way forward." *Trends Biotechnol* **28**(3): 111-116.

Smith, S. P. and E. A. Bayer (2013). "Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex." *Curr Opin Struct Biol* **23**(5): 686-694.

Starnes, R. L., P. C. Trackman and D. M. Katkocin (1996). Thermostable cyclodextrin glycosyl transferase and processes using it. Novo Nordisk A/S, Google Patents, US5501968 A

Steen, I. H., H. Dahle, R. Stokke, I. Roalkvam, F.-L. Daae, H.-T. Rapp, R.-B. Pedersen and I. H. Thorseth (2016). "Novel barite chimneys at the Loki´s Castle Vent Field shed light on key factors shaping microbial communities and functions in hydrothermal systems." *Frontiers in Microbiology* **6**.

Stokke, R., H. Dahle, I. Roalkvam, J. Wissuwa, F. L. Daae, A. Tooming-Klunderud, I. H. Thorseth, R. B. Pedersen and I. H. Steen (2015). "Functional interactions among filamentous *Epsilonproteobacteria* and *Bacteroidetes* in a deep-sea hydrothermal vent biofilm." *Environ Microbiol*.

Takai, K., S. Nakagawa, A. L. Reysenbach and J. Hoek (2006). Microbial Ecology of Mid-Ocean Ridges and Bac-Arc Basins. In *Back-Arc Spreading Systems: Geological, Biological, Chemical, and Physical Interactions*. Edited by Charles R. Fisher David M. Christie, Sang-Mook Lee andSharon Givens. American Geophysical Union. **Geophysical Monograph Series.** pp. 185-213.

Takai, K. and K. Nakamura (2011). "Archaeal diversity and community development in deep-sea hydrothermal vents." *Current Opinion in Microbiology* **14**(3): 282-291.

Tamuri, M., M. Kanno and Y. Ishii (1981). Heat and acid-stable alpha-amylase enzymes and processes for producing the same. Cpc International Inc., Google Patents, US4284722 A

Tasse, L., J. Bercovici, S. Pizzut-Serin, P. Robe, J. Tap, C. Klopp, B. L. Cantarel, P. M. Coutinho, B. Henrissat, M. Leclerc, *et al.* (2010). "Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes." *Genome Res* **20**(11): 1605-1612.

Teather, R. M. and P. J. Wood (1982). "Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen." *Appl Environ Microbiol* **43**(4): 777-780.

Thornburg, C. C., T. M. Zabriskie and K. L. McPhail (2010). "Deep-sea hydrothermal vents: potential hot spots for natural products discovery?" *J Nat Prod* **73**(3): 489-499.

Tivey, M. K. (2007). "Generation of seafloor hydrothermal vent fluids and associated mineral deposits." *Oceanography* **20**(1): 50-65.

Tomimura, E. (1991). Novel thermoduric and aciduric pullulanase enzyme and method for its production. Enzyme Bio-Systems Ltd., Google Patents, EP0405283 A2

Torsvik, V. and L. Ovreas (2002). "Microbial diversity and function in soil: from genes to ecosystems." *Curr Opin Microbiol* **5**(3): 240-245.

Trincone, A. (2011). "Marine biocatalysts: enzymatic features and applications." *Mar Drugs* **9**(4): 478-499.

Tyler, P. A., C. R. German, E. Ramirez-Llodra and C. L. Van Dover (2002). "Understanding the biogeography of chemosynthetic ecosystems." *Oceanologica Acta* **25**(5): 227-241.

Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar and J. F. Banfield (2004). "Community structure and metabolism through reconstruction of microbial genomes from the environment." *Nature* **428**(6978): 37-43.

Uchiyama, T., T. Abe, T. Ikemura and K. Watanabe (2005). "Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes." *Nat Biotechnol* **23**(1): 88-93.

Uchiyama, T. and K. Miyazaki (2009). "Functional metagenomics for enzyme discovery: challenges to efficient screening." *Curr Opin Biotechnol* **20**(6): 616-622.

Uchiyama, T. and K. Miyazaki (2010). "Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes." *Appl Environ Microbiol* **76**(21): 7029-7035.

Ufarte, L., E. Laville, S. Duquesne and G. Potocki-Veronese (2015a). "Metagenomics for the discovery of pollutant degrading enzymes." *Biotechnol Adv* **33**(8): 1845-1854.

Ufarte, L., G. Potocki-Veronese and E. Laville (2015b). "Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology." *Front Microbiol* **6**: 563.

Urich, T., A. Lanzen, R. Stokke, R. B. Pedersen, C. Bayer, I. H. Thorseth, C. Schleper, I. H. Steen and L. Ovreas (2014). "Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics." *Environ Microbiol* **16**(9): 2699-2710.

van der Maarel, M. J., B. van der Veen, J. C. Uitdehaag, H. Leemhuis and L. Dijkhuizen (2002). "Properties and applications of starch-converting enzymes of the alpha-amylase family." *J Biotechnol* **94**(2): 137-155.

Vester, J. K., M. A. Glaring and P. Stougaard (2014). "Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing." *Microb Cell Fact* **13**: 72.

Vidal-Melgosa, S., H. L. Pedersen, J. Schuckel, G. Arnal, C. Dumon, D. B. Amby, R. N. Monrad, B. Westereng and W. G. Willats (2015). "A new versatile microarray-based method for high throughput screening of carbohydrate-active enzymes." *J Biol Chem* **290**(14): 9020-9036.

Vieille, C. and G. J. Zeikus (2001). "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability." *Microbiol Mol Biol Rev* **65**(1): 1-43.

von Mering, C., P. Hugenholtz, J. Raes, S. G. Tringe, T. Doerks, L. J. Jensen, N. Ward and P. Bork (2007). "Quantitative phylogenetic assessment of microbial communities in diverse environments." *Science* **315**(5815): 1126-1130.

Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper and J. I. Gordon (2003). "A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis." *Science* **299**(5615): 2074-2076.

Xu, J., F. Ren, C. H. Huang, Y. Zheng, J. Zhen, H. Sun, T. P. Ko, M. He, C. C. Chen, H. C. Chan, *et al.* (2014). "Functional and structural studies of pullulanase from *Anoxybacillus* sp. LM18-11." *Proteins* **82**(9): 1685-1693.

Yan, X., A. Geng, J. Zhang, Y. Wei, L. Zhang, C. Qian, Q. Wang, S. Wang and Z. Zhou (2013). "Discovery of (hemi-) cellulase genes in a metagenomic library from a biogas digester using 454 pyrosequencing." *Appl Microbiol Biotechnol* **97**(18): 8173-8182.

Yang, X., J. H. Lee, H. Y. Yoo, H. Y. Shin, L. P. Thapa, C. Park and S. W. Kim (2014). "Production of bioethanol and biodiesel using instant noodle waste." *Bioprocess Biosyst Eng* **37**(8): 1627-1635.

Yin, Y., X. Mao, J. Yang, X. Chen, F. Mao and Y. Xu (2012). "dbCAN: a web resource for automated carbohydrate-active enzyme annotation." *Nucleic Acids Res* **40**(Web Server issue): 445-451.

Zeigler, D. R. (2001). The Genus *Geobacillus* - Introduction and Strain Catalog. *Bacillus* Genetic Stock Center - Catalog of Strains, 7th Edition, Volume 3. Department of Biochemistry Ohio State University, The *Bacillus* Genetic Stock Center.

Zeigler, D. R. (2014). "The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet?" *Microbiology-Sgm* **160**: 1-11.

Zeikus, J. G. and H. H. Hyun (1987). Thermostable beta-amylase. Michigan Biotechnology Institute, Google Patents, US4647538 A

Zhou, J., M. A. Bruns and J. M. Tiedje (1996). "DNA recovery from soils of diverse composition." *Appl Environ Microbiol* **62**(2): 316-322.