Metagenomic Analyses of Glacier Ice

Dissertation

zur Erlangung des Doktorgrades der

Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

Carola Simon

aus Hannover

Göttingen 2008

Teile dieser Arbeit wurden zur Veröffentlichung eingereicht:

Simon, C., Herath, J., Rockstroh, S., and Daniel, R. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. Appl. Environ. Microbiol. (published ahead of print on 6 March 2009). DOI: 10.1128/AEM.02644-08.

Simon, C., Wiezer, A., Strittmatter, A., and Daniel, R. Phylogenetic and metabolic analyses of a glacial ice metagenome. Submitted for publication.

D7
Referent: PD Dr. R. Daniel
Korreferent: Prof. Dr. G. Gottschalk
Tag der mündlichen Prüfung: 21.01.2009

Table of Contents

Abbreviations

ACE	abundance-based coverage estimator
Ap^r	ampicillin resistance
ATP	adenosine-5'-triphosphat
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BLASTN	BLAST search using a nucleotide query
BLASTP	BLAST search using a protein query
BLASTX	BLAST search using a translated nucleotide query
bp	base pairs
C	cytosine
°C	degree celsius
ca.	circa
CDD	NCBI's Conserved Domain Database
cm^r	chloramphenicol resistance
CO	carbon monoxide
COG	Clusters of Orthologous Groups of proteins
C-terminal	carboxyterminal
dATP	deoxyadenosine
dd H$_2$O	double distilled water
DGGE	denaturing gradient gel electrophoresis
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
E.	*Escherichia*
EDTA	ethylenediamintetraacetatic acid
e. g.	*exempli gratia*, for example
et al.	*et alii/alia*, and others
FEN	flap endonuclease
Fig.	Figure
G/g	guanine or gram

Gap	genome assembly program
Gb	gigabase(s)
glb	glacier ice bacteria
x g	gravitation acceleration (9,81 m/s^2)
h	hour(s)
i. e.	*id est*, that is
IPTG	isopropyl beta-D-1-thiogalactopyranoside
k	kilo
Kan^r	kanamycin resistance
kb	kilobase(s)
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
l	liter
LB	Luria-Bertani
LMP	low melting point
ln	natural logarithm
μ	growth rate or micro
Mb	megabase(s)
MDA	multiple displacement amplification
M/m	molar (mol per liter) or meter
min	minute(s)
mm	millimeter
n	nano
NCBI	National Center for Biotechnology Information
nt	nucleotides
N-terminal	aminoterminal
OD	optical density
ORF	open reading frame
OTU	operational taxonomic unit
PCR	polymerase chain reaction
pH	power of hydrogen
PP$_i$	pyrophosphate anion ($P_2O_7^{4-}$)
PUFAs	polyunsaturated fatty acids
RDP	Ribosomal Database Project

rpm	rotations per minute
rRNA	ribosomal ribonucleic acid
RSR	random sequence reads
sp.	species
SSU RNA	small subunit RNA
T	thymine
TAE	tris-acetate-EDTA
TE	tris-EDTA
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane
TSA	tryptic soy agar
U	units (unit of enzyme activity)
UV	ultraviolet
V	Volt
v/v	volume per volume
WGA	whole genome amplification
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Zeo^R	zeocin resistance

Abbreviations for amino acids

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic acid	P	Pro	Proline
E	Glu	Glutamic acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

1   Introduction

In the present study, glacier ice samples were collected at the Northern Schneeferner (47° 25' N, 10° 59' E) on the Zugspitzplatt, Germany (Fig. 1). The Zugspitzplatt is a plateau on the Zugspitze, which is the highest mountain in Germany. The two glaciers on the Zugspitzplatt, the Northern and the Southern Schneeferner, are remains of the last ice age and were first described in 1820. The Northern Schneeferner is the largest and highest-located glacier of the five glaciers of the German Alps (http://www.zugspitze.de). It is located 2,820 to 2,560 m above sea level and covers an area of 340,000 m^2 with an inclination angle of 14°. The area is characterized by a precipitation of 400 mm per year, of which 80% is snow. On average the glacier is covered by 4.5 m of snow in a period of 8 to 10 months with temperatures below 0°C (Huettl, 1999).
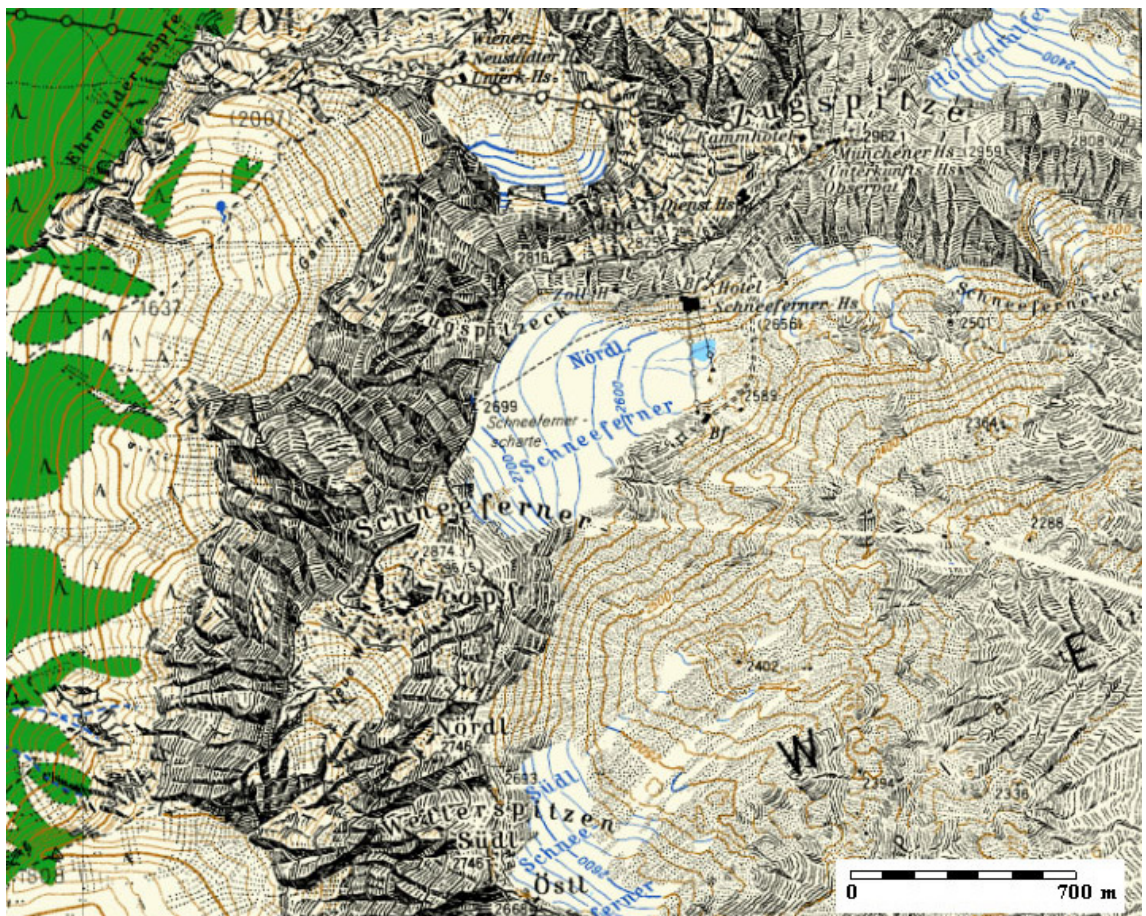
Figure 1. Location of the Northern Schneeferner, Zugspitzplatt, Germany (Courtesy of the Landesamt für Vermessung und Geoinformation, München).

1.1   Phylogenetic and functional analysis of glacier ice

More than 75% of the Earth's biosphere is constantly exposed to temperatures below 5°C. Permafrost and ice contain the lowest temperature settings of all habitats on Earth and exhibit high stability with respect to environmental conditions. These habitats represent a long-term, chronological archive of microorganisms. It has been postulated that permafrost and glacial ice harbor the oldest prokaryotes on Earth (Willerslev *et al.*, 2004b). Microorganisms have successfully colonized these environments by evolving unique characteristics of their proteins, membranes, and genetic responses to thermal shifts. The evidence for the presence of ice on Mars and Jupiter's moon Europa increased the interest in the investigation of microbial life in frozen habitats. Glacial ice is regarded as an environment, which is equivalent to extraterrestrial cold habitats (Miteva *et al.*, 2004). Recently, high-throughput pyrosequencing technology has been applied for metagenomic characterization of environmental microbial communities (Biddle *et al.*, 2008; Dinsdale *et al.*, 2008). The most important advantages of this cloning-independent approach are avoidance of cloning bias and bias introduced by PCR amplification. Metagenomic analyses of environmental samples have been proposed to be the most accurate quantitative approach for description of microbial communities (von Mering *et al.*, 2007). In addition to the assessment of the taxonomic composition, relative abundances of all genes and metabolic profiles of the communities can be determined. To date, a comprehensive metagenomic analysis of a permanently frozen habitat has not been conducted. In several studies, the prokaryotic diversity of glacial and subglacial habitats in America (Cheng and Foght, 2007; Skidmore *et al.*, 2005), Asia (Christner *et al.*, 2003; Zhang *et al.*, 2008). Antarctica (Priscu *et al.*, 1999), Greenland (Miteva *et al.*, 2004; Sheridan *et al.*, 2003), and New Zealand (Foght *et al.*, 2004) has been analyzed based on cultivation and analysis of 16S rRNA genes. However, studies on the microbial composition of European glaciers are rare. These glaciers have been mainly investigated with respect to the presence of yeasts (Turchetti *et al.*, 2008) and bacterial population sizes (Sharp, 1999).

Since the 1980s, glaciers in the European Alps are receding quickly. According to Haeberli (2007), these glaciers lost half of their total volume between 1850 and 1975. Recently, the speed of mass and volume losses of these glaciers accelerated. This is probably a result of the global climate change. It has been estimated that an almost complete deglaciation of the European Alps will happen within this century (Haeberli, 2007).

1.2   Exploitation of the microbial diversity present in glacier ice by metagenomic approaches

Currently, there is a global political initiative to support industrial or 'white' biotechnology. With respect to high energy costs, diminishing fossil resources, environmental pollution and globalized economy the large-scale use of biotechnology is regarded as essential. The development of novel enzymes, processes, and products is required for a wide range of applications and industries. Metagenomics promises to substantially impact industrial production by providing new molecules possessing novel functions (Lorenz and Eck, 2005). The metagenomic approach was first proposed in 1998 as a technique to exploit the entire genetic content of a habitat (Handelsman *et al.*, 1998; Handelsman, 2004). It comprises the application of modern genomic tools to analyze the collective genomes of whole microbial communities in any habitat (Beja *et al.*, 2000; Riesenfeld *et al.*, 2004b). For this approach, the whole genomic DNA of an environment is extracted and directly cloned into small-insert or large-insert vectors, which are then transformed into suitable hosts. Subsequently, the metagenomic libraries can be analyzed and screened for novel genes encoding biocatalysts. Thus, microbial genomes present in a given habitat can be exploited without prior cultivation of microorganisms. Cultivation is often difficult or impossible and can lead to an enrichment of dominant organisms under specific selective conditions (Daniel, 2004, 2005). However, the vast majority of the Earth's genetic and metabolic diversity is hidden in yet uncharacterized microbial genomes (Pace, 1997; Torsvik *et al.*, 2002). It has been estimated that approximately 99% of the microorganisms in most

environments cannot be cultured by standard methods. In addition, the uncultured majority of microorganisms comprise diverse organisms that are only distantly related to the cultured ones (Riesenfeld *et al.*, 2004b). Thus, culture-independent approaches are required to understand the genetic diversity, population structure, and ecological impact of microorganisms.

In an effort to isolate genes encoding novel biocatalysts metagenomic libraries derived from a large variety of different environments, such as soil (Henne *et al.*, 2000; Knietsch *et al.*, 2003; Rondon *et al.*, 2000), pondwater (Ranjan *et al.*, 2005), hot springs (Kim *et al.*, 2006; Rhee *et al.*, 2005), a deep sea hypersaline basin (Ferrer *et al.*, 2005), and Arctic sediment (Jeon *et al.*, 2008a) have recently been constructed by several researchers. To date, metagenomics has led to the discovery of a remarkable number of industrially novel biocatalysts, such as esterase/lipases (Hardeman and Sjöling, 2007; Jeon *et al.*, 2008a; Jeon *et al.*, 2008b; Rhee *et al.*, 2005), proteases (Gupta *et al.*, 2002), amylases (Yun *et al.*, 2004), amidases (Gabor *et al.*, 2004), nitrilases (Robertson *et al.*, 2004), and and cellulases (Voget *et al.*, 2006). However, despite the wealth of conducted metagenomic studies, the majority focused on encountering novel enzymes by simple screening strategies such as monitoring of enzymatic functions of individual clones by incorporation of chromophore-bearing derivatives of enzyme substrates (Jeon *et al.*, 2008a).

1.2.1 DNA polymerase I

DNA polymerases are vital for the maintenance of the genome. For DNA synthesis and transfer of genetic information several types of DNA polymerases in the cell are required. Many DNA polymerases possess the ability to select substrates with high accuracy and catalyze DNA synthesis with high efficiency and processivity. DNA polymerases exclude the incorporation of non-complementary bases with an accuracy of greater than 99.99%. The loss of this fidelity has been proposed to lead to premature aging and an increased risk of cancer (Loh and Loeb, 2005).

Based on sequence similarity and phylogenetic relationships DNA polymerases are grouped into the following six families: Family A, B, C, D, X, and Y (Ohmori *et al.*, 2001). DNA polymerase I belongs to family A and contains three different domains: a 5'-3' exonuclease domain at the N-terminus, a central proofreading 3'-5' exonuclease domain, and a polymerase domain at the C-terminus of the enzyme (Joyce and Steitz, 1994). The structure of family A DNA polymerases resembles a right human hand, exhibiting domains similar to the palm, thumb, and finger (Fig. 2).
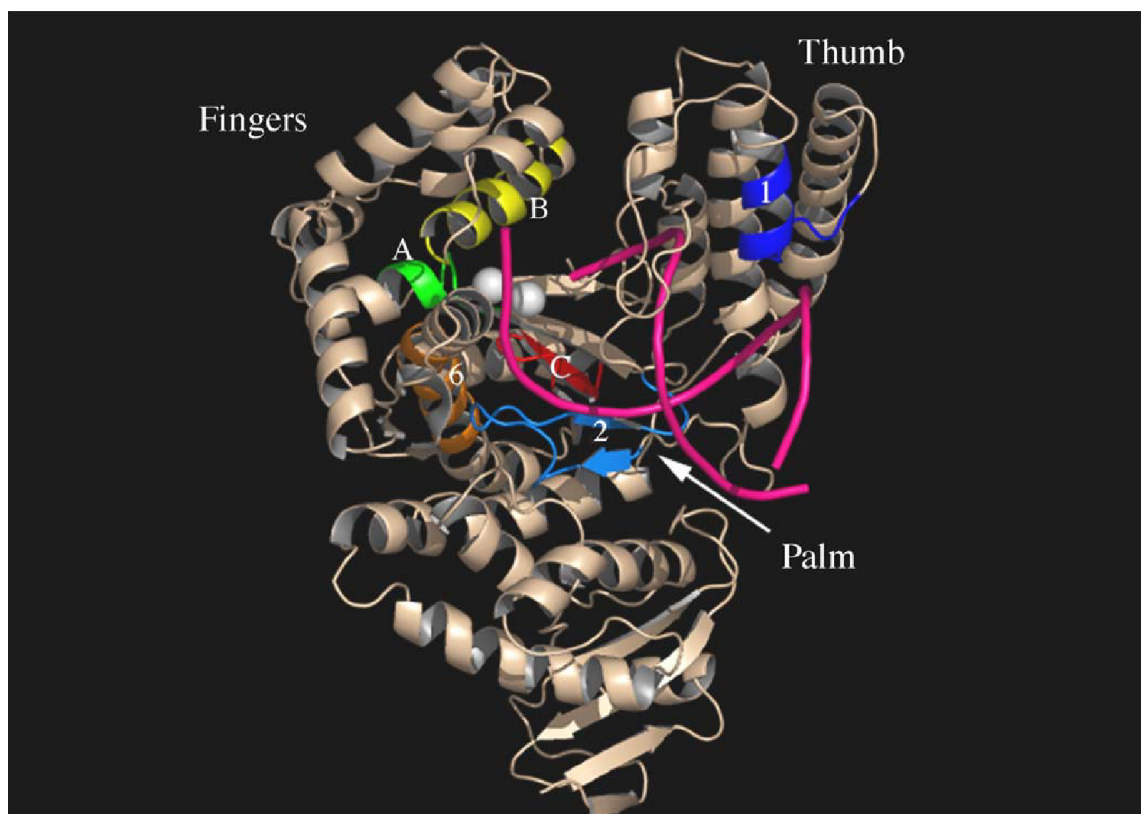
Figure 2. Crystal structure of *Taq* DNA polymerase I in closed conformation. Palm, thumb, fingers and exonuclease domains are labeled. Conserved Motifs 1 (purple), 2 (blue), A (green), B (yellow), 6 (orange), and C (red) are highlighted. Primer-template duplex DNA is colored in pink. The two catalytic magnesium ions bound to the palm are colored in white (Loh and Loeb, 2005).

Especially family A DNA polymerases are employed as tools in molecular biology including probe labeling, DNA sequencing, and mutagenic PCR (Loh and Loeb, 2005). Due to the large interest in these applications new polymerases with altered substrate specificities are created. To improve the

suitability for applications, various family A DNA polymerases have been modified, i. e., the Klenow fragment of the *Escherichia coli* DNA polymerase I has been designed by removal of the 5'–3' exonuclease domain (Klenow and Henningsen, 1970). Nevertheless, expanding the known DNA polymerase sequence space and discovery of polymerases with novel properties is required for the development of novel or improved molecular methods and tools (Loh and Loeb, 2005; Tvermyr *et al.*, 1998).

1.3   Aim of this study

The aim of this thesis was to analyze and exploit the microbial and metabolic diversity present in glacier ice of the Northern Schneeferner, which is located in the German Alps. Assessment of the taxonomic and metabolic diversity of a glacial microbial community was performed by analysis of a large pyrosequencing-derived dataset. To complement this approach, traditional methods for phylogenetic assessment such as PCR amplification of 16S rRNA genes and isolation of microorganisms were employed. In addition, DNA polymerases I were identified by function-driven screening of metagenomes. The screen for the targeted genes was based on complementation of a cold-sensitive lethal mutation in the *polA* gene of *E. coli* (Nagano *et al.*, 1999). For this purpose, small-insert and large-insert metagenomic libraries from DNA isolated from glacial ice were constructed. The employment of samples derived from this permanently frozen environment for library construction has not been reported by other researchers.

2 Materials and Methods

2.1 Sampling site and sampling of glacier ice

Glacier ice samples were collected at the Northern Schneeferner (47° 25' N, 10° 59' E) on the Zugspitzplatt, Germany (Fig. 1). The pH of the ice sample was slightly acidic (pH 5). Samples were taken in June 2005 when the glacier was not covered with snow. Glacier ice was collected by employing clean surface-sterilized spades and pick-axes. The overlying 30 cm of glacier ice were removed and discarded. Two hundred kg of ice up to a depth of approximately 0.5 m were collected and transferred into sterile polypropylene bags (Sarstedt, Nümbrecht), which were stored in sterile plastic containers. The samples for DNA extraction were returned frozen to the laboratory. Subsequently, the ice was melted at 4°C and filtered using a sterile cellulose acetate membrane (pore size 0.2 µm; Whatman, Dassel).

2.2 Bacterial strains and growth conditions

2.2.1 Organisms, plasmids, and oligonucleotides

Escherichia coli strains, plasmids, 16S rRNA gene primers, and primers for amplification of identified open reading frames (ORFs) employed in this thesis are listed in Tables 1 to 4.

Table 1. *E. coli* **strains used in this thesis.**

E. coli strain	Genotype	Source/reference
DH5α	F- φ80*lac* ZDM15 Δ(*lac* ZYA-*arg* F)U169 *rec* A1 *end* A1 *hsd* R17(rk-, mk+) *pho* A *sup* E44 *thi* -1 *gyr* A96 relA1 λ-	Invitrogen, Karlsruhe
TOP10	F- *mcr* A Δ(*mrr* -*hsd* RMS-*mcr* BC) φ80*lac* ZΔM15 Δ*lac* X74 *rec* A1 *ara* D139 Δ(*araleu*)7697 *gal* U *gal* K *rps* L (StrR) *end* A1 *nup* G	Invitrogen, Karlsruhe
EPI300-T1	[F- e14-(McrA-) D(mcr*C*- mrr) (TetR) *hsd* R514 *sup* E44 *sup* F58 *lac* Y1 or D(*lac* IZY)6 *gal* K2 *gal* T22 *met* B1 *trp* R55 I-]	Epicentre, Madison
cs2-29	F⁻ *fcsA29 thr leuB trp his thy thi ara lac gal xyl mtl str tonA*	Nagano *et al.* , 1999
CSH26 *fcsA29*	F⁻ *ara (lac-pro) thi fcsA29 met* ::Tn5	Nagano *et al.* , 1999

Table 2. Vectors used in this thesis.

Plasmid/fosmid	Relevant characteristic	Source
pCR2.1-TOPO	Ap^r, Kan^r, pMB1-Replikon	Invitrogen, Karlsruhe
pCR-XL-TOPO	Kan^r, Zeo^r; pMB1-Replikon	Invitrogen, Karlsruhe
pBAD	Ap^r, araBAD promoter	Invitrogen, Karlsruhe
pCC1FOS	Cm^r, *cos*	Epicentre, Madison

Table 3. Sequences and specificity of rDNA-targeted primers used in this thesis.

Primer	Sequence (5' - 3')	Specificity	Reference
8F	GGATCCAGACTTTGATYMTGGCTCAG	Bacteria	Ben-Dov et al., 2006
341F	CCTACGGGAGGCAGCAG	Bacteria	Muyzer et al., 1995
349F	AGGCAGCAGTGGGGAAT	Bacteria	Diez et al., 2001
907R	CCGTCAATTCCTTTRAGTTT	Bacteria	Muyzer et al., 1995
1114R	GGGTTGCGCTCGTTRC	Bacteria	modified from Reysenbach and Pace, 1995
Arch8F	TCCGGTTGATCCTGCCGG	Archaea	modified from Sorensen et al., 2004
Arch21F	TTCCGGTTGATCCYGCCGGA	Archaea	DeLong, 1992
Arch958R	YCCGGCGTTGAMTCCAATT	Archaea	DeLong, 1992
A1040F	GAGAGGWGGTGCATGGCC	Archaea	Reysenbach and Pace, 1995
1492R	TACCTTGTTACGACTT	Universal	Muyzer et al., 1995

Table 4. Primers used for amplification of genes encoding DNA polymerases I or domains typical for polymerases.

Primer	Sequence (5' - 3')
pCS1.1for	GAGCTCGATGCCTAAAAAATTG
pCS1.1rev	GAGCTCTCAAACGGGTAAATCC
pCS2.1for	GAGCTCGATGAAAAAACTATTTTTGC
pCS2.1rev	GAGCTCTTAGTGTGCTTCCAGCC
pCS3.1for	ATCTGTGAGCTCTATGGAACGCGCACTTTTAC
pCS3.1rev	CAACGGAGCTCCTATCTATCCTCAAGCATC
pCS4.1for	GAGCTCTATGACAAAAACATTTAA
pCS4.1rev	GAGCTCATCATAGATATAGTCCTAAC
pCS5.1for	GAGCTCAATGCCCCCGTTTATTTTG
pCS5.1rev	GAGCTCCTAATGCGCCACACC
pCS6.1for	GAGCTCGATGATAACAAAACAAAC
pCS6.1rev	GAGCTCTTAATGCGCCGCC
pCS7.1for	ATCTGTGAGCTCTATGCCACCATTTGTATTGGT
pCS7.1rev	CAACGGAGCTCCTAATGGGTATCGGGATTGGCTTG
pCS8.1for	GAGCTCAATGCCCCCGTTTATTTTG
pCS8.1rev	GAGCTCCTAATGCGCCACACC
fCS1.1for	ATCTGTGAGCTCTATGCAAAAAAAATTATTTCTCCT
fCS1.1rev	CAACGGAGCTCTTAGTGTGCCAGCAACCAAT

2.2.2   Growth media

All growth media listed below were sterilized in an autoclave at 121°C for 30 min. For preparation of agar plates, the media were supplemented with 1.5% (w/v) agar prior to sterilization.

LB (Luria-Bertani) medium (Sambrook, 1989)

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
dd H_2O	ad 1000 ml

SOB medium (Hanahan, 1983)

Tryptone	20 g
Yeast extract	5 g
NaCl	0.58 g
KCl	0.18 g
$MgCl_2$	2 g
$MgSO_4$ x 7 H_2O	2.5 g
dd H_2O	ad 1000 ml

Tryptic soy agar (TSA)

Peptone from caseine	17 g
Peptone from soymeal	3 g
D(+)-Glucose	2.5 g
NaCl	5 g
K_2HPO_4	2.5 g
dd H_2O	ad 1000 ml

The pH was adjusted to 7.3.

R2A (Reasoner and Geldreich, 1985)

Yeast extract	0.5 g
Proteose peptone (Difco no. 3)	0.5 g
Casamino acids	0.5 g
Glucose	0.5 g
Soluble starch	0.5 g
Na-pyruvate	0.3 g
K_2HPO_4	0.3 g
$MgSO_4$ x 7 H_2O	0.3 g
dd H_2O	ad 1000 ml

The pH was adjusted to 7.2 with crystalline K_2HPO_4 or KH_2PO_4.

L medium (Sambrook, 1989)

Bactopeptone	1 g
Yeast extract	0.5 g
NaCl	0.5 g
Glucose	0.5 g
dd H_2O	ad 1000 ml

The pH was adjusted to 7.2 with 1 M NaOH. For growth of *E. coli* cs2-29 (Nagano *et al.*, 1999) the medium was supplemented with 20 mg/l filter-sterilized thymine after cooling the autoclaved media to at least 60°C.

2.2.3 Antibiotics and media supplements

For growth of recombinant *E. coli* strains, the supplements listed in Table 5 were utilized.

Table 5. Media supplements and respective concentrations used in this study.

Supplement	Stock solution	Working concentration
Ampicillin	50 mg/ml 50% Ethanol	50-100 mg/ml
Chloramphenicol	25 mg/ml Ethanol	12.5 mg/ml
Kanamycin	25 mg/ml dd H_2O	50 mg/ml
IPTG	26 mg/ml dd H_2O	50 mg/ml
X-Gal	20 mg/ml Dimethylformamide	40 mg/ml
L-Arabinose	50 mg/ml dd H_2O	0.1 (w/v)
Thymine	4 mg/ml ddH_2O	20 µg/ml

All stock solutions were filter-sterilized. Addition of the supplements was carried out after cooling the autoclaved media to at least 60°C. Supplements were stored at -20°C.

2.2.4 Growth conditions and isolation of glacier ice bacteria

All recombinant strains of *E. coli* Top10, DH5α, and EPI300-T1 were grown overnight at 30°C or 37°C in Luria-Bertani (LB) medium (see 2.2.2) containing the required supplements (see 2.2.3). Recombinant strains of *E. coli* CSH26 *fcsA29* and cs2-29 were grown in L medium (see 2.2.2) containing the required supplements at 37 and 18°C for one to six days.

To isolate microorganisms from glacier ice the following media were employed: full strength R2A (Reasoner and Geldreich, 1985), 0.25 strength R2A, 0.5 tryptic soy agar (TSA) (Oxoid, Cambridge, UK), and 0.2 strength LB medium (see 2.2.2). The enrichment of microorganisms was initiated by spreading 100 µl of molten glacier ice directly onto agar plates consisting of the above-mentioned media. Initial enrichment experiments were carried out at 4°C. Isolates that formed colonies with different morphologies were purified by repeated transfer (four to five times) on agar plates. The ability of each isolate to form colonies at 4, 18, 30, and 37°C was determined.

2.2.5 Storage of strains

E. coli strains on agar plates were stored at 4°C for eight weeks at the most. For long-term storage of *E. coli* and isolates from glacier ice glycerol stocks were prepared. Cultures were grown in the respective growth media (see 2.2.2) at 30°C (*E. coli* strains) and 4°C (glacier ice isolates), respectively. In the logarithmic phase of growth, 0.75 ml of cell suspension was mixed with 0.25 ml glycerol [87% (v/v)]. Stocks prepared in this way were stored at -70°C.

Clone libraries were stored in 96-well microtiter plates. *E. coli* clones were grown in microtiter plates containing LB medium (see 2.2.2) supplemented with 10% (v/v) DSMO. Microtiter plates were stored at -70°C.

2.2.6   Determination of growth parameters

2.2.6.1   Determination of the optical density

Determination of the optical density (OD) was conducted by means of turbidity measurements against sterile medium at a wavelength of 600 nm using a spectrophotometer (UV-VIS Spektrophotometer V-550, Jasco, Jena). Semi-micro cuvettes, 1.6 ml (polystyrene, 10 x 4 mm) were used (Sarstedt, Nümbrecht). When cell densities OD_{600} > 0.3 were measured, samples were diluted accordingly.

2.2.6.2   Determination of the growth rate of recombinant *E. coli* CSH26 *fcsA29* strains

Growth rates of *E. coli* CSH26 *fcsA29* (Nagano *et al*., 1999) transformed with plasmids and the fosmid harboring the *polA* gene were determined. Of an overnight culture, 0.5 ml was used as inoculum for 10 ml LB medium (see 2.2.2) containing kanamycin (50 µg/ml). The cultures were incubated at 18°C with vigorous shaking. The OD_{600} was measured turbidimetrically at the start of the experiment and following in the logarithmic phase of growth regularly every hour. Experiments were carried out in duplicates with two replications. Growth rates were calculated as follows:

$$\frac{\ln x_t - \ln x_0}{(t - t_0)}$$

with x = cell mass and t = time.

2.2.6.3   Determination of the pH value

The pH value was measured utilizing a WTW pH meter pH 526 (WTW GmbH, Weilheim) and a glas-calomel electrode (Cahn-Ingold, Frankfurt). Prior to measuring, the pH meter was calibrated using calibration solutions (pH 4.0, pH 7.0, and pH 10.0) (Riedel de Haen AG, Hannover).

2.3   Genetic manipulations

2.3.1   Treatment of equipment and solutions

For inactivation of DNA-degrading enzymes all tools, vessels and solutions for work with DNA were autoclaved (20 min, 121°C). Tools that were not autoclavable were first rinsed with 70% (w/v) ethanol and subsequently with sterile dd H_2O. Nonautoclavable or heat-labile substances (e. g. lysozyme, proteinase K) were dissolved in sterile buffers or water and filter-sterilized.

2.3.2   Buffers and solutions

Following, the composition of buffers, which are not further explained in the subsequent text, is listed. Lower buffer concentrations were obtained by dilution of the buffer with dd H_2O.

10 x TE buffer

1 M Tris buffer (pH 8.0)	100 ml
0.5 M EDTA (pH 8.0)	10 ml
dd H_2O	ad 1000 ml

50 x TAE buffer

Tris base	242 g
Acetic acid	57 ml
0.5 M EDTA (pH 8.0)	100 ml
dd H_2O	ad 1000 ml

2.3.3   Precipitation, purification, and concentration of nucleic acids

2.3.3.1   Precipitation of nucleic acids

Concentration of nucleic acids was achieved by precipitation with ethanol or isopropanol. If necessary, the concentration of monovalent cations in the sample was adjusted by adding 3 M sodium acetate (pH 5.5) to a final concentration of 0.3 M. For precipitation of nucleic acids, 2 volumes of ethanol (96%) or 1 volume of isopropanol (96%) were added. Precipitation using ethanol was carried out by incubation at room temperature for at least 1 h, while precipitation with isopropanol was incubated overnight at room temperature. The nucleic acids were pelleted by centrifugation (13,000 to 16,000 x g, 4°C or room temperature, 30 min). The pellet was rinsed with 1 ml ice-cold ethanol (70% v/v), dried, and resuspended in dd H_2O to the desired concentration.

2.3.3.2   Determination of the DNA concentration

The DNA concentration and purity was determined using a NanoDrop ND-1000 spectrophotometer (Peqlab Biotechnologie, Erlangen).

2.3.3.3   Purification of DNA

DNA fragments were extracted from agarose gels using the QIAquick Gel Extraction kit (Qiagen, Hilden) or alternatively the peqGOLD Gel Extraction kit (Peqlab Biotechnologie, Erlangen) according to the manufacturer's instructions. The DNA was eluted with dd H_2O and stored at -20°C until further use.

PCR products and other linear DNA fragments were purified using the QIAquick PCR Purification kit (Qiagen, Hilden) or SureClean solution (Bioline, Luckenwalde) according to the manufacturer's instructions. The DNA was eluted with dd H_2O and stored at -20°C until further use.

2.3.4   Isolation of nucleic acids

2.3.4.1   Isolation of environmental DNA

Total genomic DNA from glacier ice was extracted from the membranes (see 2.1) by using the Nucleo Spin Tissue kit (Macherey-Nagel, Düren) according to the manufacturer's instructions with the following exception: The initial pre-lysis was performed with two cell-containing membranes that were cut into small pieces. Subsequently, the pieces were incubated in 2.5 ml pre-lysis buffer containing 1.8 mg/ml proteinase K at 56°C for 16 h.

2.3.4.2   Preparation of plasmid and fosmid DNA

For a rapid verification and characterization of cloning results, plasmid DNA was isolated as described previously (Holmes and Quigley, 1981). For obtaining high-purity plasmid or fosmid DNA, the NucleoSpin Plasmid kit (Macherey-Nagel, Düren) or the peqGOLD HP Plasmid Miniprep Kit II (Peqlab Biotechnologie, Erlangen) was used according to the manufacturer's instructions. Prior to preparation of fosmids, the copy control fosmid vectors

were induced to high copy number by supplementing the Copy Control Fosmid Autoinduction Solution (Epicentre Biotechnologies, Madison) into the cultures before inoculation as described by the manufacturer.

Metagenomic plasmid and fosmid libraries were isolated using the Qiagen Plasmid Midi Kit (Qiagen, Hilden) according to the manufacturer's instructions. Prior to isolation, fosmids were induced to high copy number as described above.

2.3.5   Polymerase chain reaction (PCR)

Polymerase chain reaction was employed to amplify specific DNA fragments for cloning or sequencing using the *Taq* polymerase (Fermentas, St. Leon-Rot). For amplification of putative *polA* genes encoding DNA polymerase I (see 2.3.17) the PCR Extender System (VWR International, Darmstadt) was used. Negative controls containing the entire reaction mixture except for the template DNA were included in all experiments. PCR reactions were carried out in a volume of 50 µl according to the manufacturer's instructions and included the following reagents:

Taq polymerase

Template DNA	max 200 ng
Oligonucleotides	each 2 µM
dNTP Mix	200 µM
DNA polymerase	1 U
Reaction buffer (10x)	5 µl
$MgCl_2$ or $MgSO_4$	1.75 mM
dd H_2O	ad 50 µl

PCR conditions

Initial denaturation	95°C	5 min
Three-step cycle		
Denaturation	95°C	1 min
Annealing	$(T_{ann.}-5)$°C	1 min
Elongation	72°C	1 min/kb
Number of cycles: 25 – 35		
Final elongation	72°C	10 min

By reducing the length of time for annealing and variation of concentration of $MgCl_2$ and the oligonucleotides, the specificity of the PCR was optimized. For amplification of the 16S rDNA of the glacial isolates (see 2.2.4), a cell suspension (1 µl) derived from single colonies was directly used as template for the PCR reaction.

PCR Extender System

Reaction Mix I

Oligonucleotides	each 300 nM
Template DNA	2 µl
dd H_2O	ad 10 µl

Reaction Mix II

10 x HighFidelity buffer	5 µl
dNTP Mix	280 µM
PCR Extender Polymerase Mix	1 U
dd H_2O	ad 40 µl

PCR conditions

Initial denaturation	94°C	2 min
Three-step cycle		
Denaturation	94°C	0.2 min
Annealing	$(T_{ann.}-5)$°C	0.2 min
Elongation	72°C	0.3 - 2 min/kb
Number of cycles: 30		
Final elongation	72°C	10 min

The PCR was carried out in a Mastercycler Gradient (Eppendorf, Hamburg). The obtained PCR products were analyzed by agarose gel electrophoresis (see 2.3.6).

To analyze the microbial community present in the glacier ice by denaturing gradient gel electrophoresis (DGGE) (see 2.3.7) 16S rRNA genes were amplified employing the primer pair 341F/907R (Table 3). To primer 341F a GC clamp (5'-CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGG-3') was attached at the 5' terminus. PCR reactions were performed with *Taq* polymerase using the respective reaction mixture additionally containing 10 µg bovine serum albumin. A touchdown PCR was carried out as described by Muyzer *et al.*, 1998.

2.3.6   Agarose gel electrophoresis

DNA fragments were analyzed by standard agarose gel electrophoresis (Sambrook, 1989). Gels with a volume of 10 to 25 ml and an agarose concentration of 0.8 or 1.5% were utilized. All samples were mixed with 0.2 volume of loading buffer prior to loading into the wells. Gels were run in 1 x TAE buffer (see 2.3.2) at a constant voltage of 120 V for 40 min. For visualization of the DNA, the gels were stained in ethidium bromide solution (1 µg/ml dd H_2O) for 10 to 20 min. The DNA was visualized under UV light and the DNA fragment

profiles were documented using a GelDoc system (Image Quant 100; GE Healthcare, Munich). Sizes of DNA fragments were determined using DNA markers (Fig. 3).

10 x Loading buffer

Bromphenol blue	125 mg
150 mM Tris (pH 7.6)	16.5 ml
Glycerol	30 ml
dd H_2O	3.5 ml
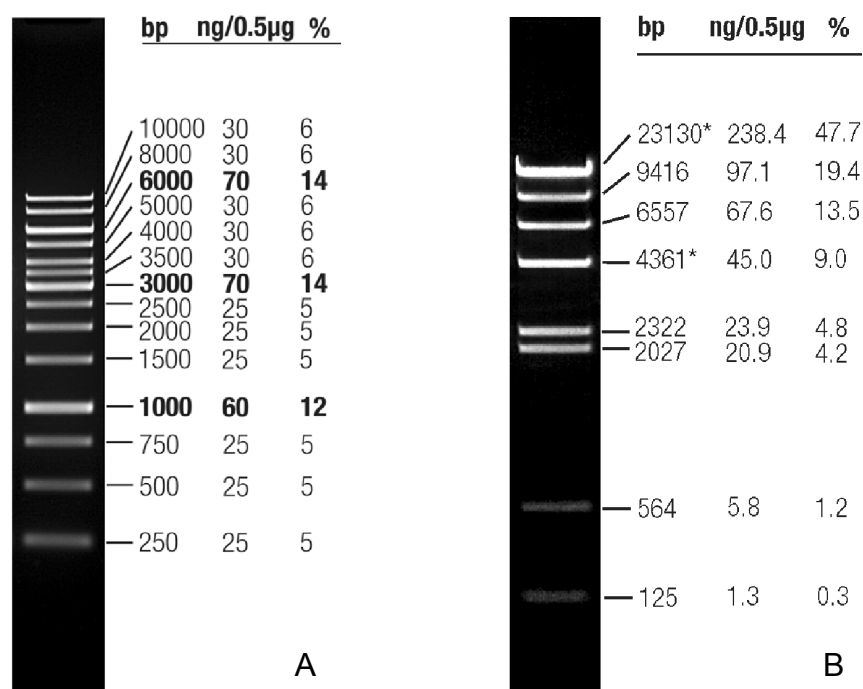
Figure 3. DNA ladders used: GeneRuler^{TM} 1 kb DNA ladder (A) and Lambda DNA/Hind III Marker (B) (Fermentas, St. Leon-Rot).

2.3.7   Denaturing gradient gel electrophoresis (DGGE)

DGGE analysis of 16S rRNA genes was performed using the Ingeny phorU system (Ingeny International BV, Goes). Purified bacterial amplicons (600 ng)

were applied to a 9% (wt/vol) polyacrylamide gel containing a 25% to 75% denaturing gradient (100% denaturant consisted of 7 M urea and 40% formamide). DGGE was performed at 60°C and 100 V for 16 h in TAE buffer (see 2.3.2). Subsequently, the gels were stained with SYBR Gold (Invitrogen, Karlsruhe). Dominant bands were excised from the gel and the DNA was eluted by incubation in 50 µl of TAE buffer at 4°C for 16 h. One microliter of the resulting DNA solution was employed as template for reamplification of the 16S rRNA genes. The above-described PCR reaction conditions and primers (see 2.3.5) were employed using *Taq* polymerase, but primer 341F was used without the GC clamp. Purification (see 2.3.3.3), cloning (see 2.3.12), and sequencing (see 2.3.16) of reamplified products were performed as described. In order to estimate the number of different 16S rRNA genes co-migrating within the same DGGE band the plasmids derived from three clones of each DGGE band were sequenced.

2.3.8   Mechanical modification of DNA

For subcloning of fosmids and plasmids harboring large inserts, the DNA was mechanically sheared to 4 and 1 kb, respectively, using a nebulizer (Invitrogen, Karlsruhe). On ice, 13 µg fosmid or plasmid DNA were added to 750 µl shearing buffer (TE buffer, pH 8, containing 10% glycerol) and pipetted into the nebulizer. The fosmid and plasmid DNA was sheared at 0.6 bar for 50 and 40 sec, respectively. Subsequently, the DNA was transferred to a sterile microcentrifuge tube and precipitated as described above (see 2.3.3.1). To ensure that only DNA fragments of the desired sizes were cloned the fragmented DNA was loaded onto a preparative gel and DNA fragments of 3 and 1 kb for the fosmid and plasmids, respectively, were recovered as described below (see 2.3.11.4) by using standard melting temperature agarose and subsequent DNA extraction from the gel by using the QIAquick Gel Extraction kit (see 2.3.3.3).

2.3.9   Enzymatic modification of DNA

Prior to sequencing all purified plasmids were subjected to restriction enzyme digestion to confirm the success of the cloning experiment. DNA was digested using enzyme and reaction mixture as follows:

Analytical digestion reaction

DNA solution	up to 1 μg
Reaction buffer (10 x)	1.5 μl
Restriction enzyme(s)	each 1 U
dd H_2O	ad 15 μl

The digestions were performed for one to four hours at the temperature and buffer optimal for the restriction enzyme's activity. The results of the digestion reaction were directly analyzed by agarose gel electrophoresis (see 2.3.6).
If one or more of the fragments from the digestion reactions were to be used in further cloning procedures, the reaction volumes were scaled up as follows:

Preparative digestion reaction

DNA solution	max 10 μg
Reaction buffer (10 x)	5 μl
Restriction enzyme	10 – 25 U
dd H_2O	ad 50 μl

The digestions were performed overnight at the temperature and buffer optimal for the restriction enzyme's activity. The results of the digestion reaction were directly analyzed by agarose gel electrophoresis (see 2.3.6).

2.3.10   Whole genome amplification (WGA) of glacier ice DNA

For construction of metagenomic plasmid libraries, prior to cloning into pCR-XL-TOPO whole genomic glacial ice DNA was amplified by multiple displacement amplification (MDA) using the GenomiPhi V2 DNA amplification kit (GE Healthcare, Munich). Amplification was carried out according to the manufacturer's instructions with the following exception: in order to minimize background DNA synthesis template DNA was added to the reaction in 10-fold excess (Hutchison *et al.*, 2005). The amplified DNA was purified using SureClean solution (Bioline, Luckenwalde).

2.3.11   Resolving hyperbranched DNA structures generated by WGA and preparation for cloning into pCR-XL-TOPO

To improve cloning efficiency and to avoid abnormal insert size distribution, hyperbranched structures generated during WGA (see 2.3.10) were resolved as suggested by Zhang *et al.* (2006) with modifications (Fig. 4). During the first step, the amplified DNA was incubated with phi29 DNA polymerase and dNTPs, but without primers (see 2.3.11.1). In this way, the density of branching junctions was reduced due to the strand-displacement activity of the phi29 polymerase. In the next step, 3' single-stranded overhangs were removed by digestion with S1 nuclease (see 2.3.11.2). The resulting nicks in the double-stranded DNA were removed by nick translation using DNA polymerase I. As the DNA was subsequently cloned into pCR-XL-TOPO (Invitrogen, Karlsruhe) requiring blunt-ended DNA with an adenosine overhang, incubation with S1 nuclease for nick translation and T4 polymerase for DNA polishing were carried out in one step (see 2.3.11.3). Following, the DNA was size-fractionated by low melting point (LMP) agarose gel electrophoresis (see 2.3.11.4). Finally, a deoxyadenosine was added to the size-fractionated blunt-ended DNA (see 2.3.11.5) followed by dephosphorylation using Antarctic phosphatase (2.3.11.6).
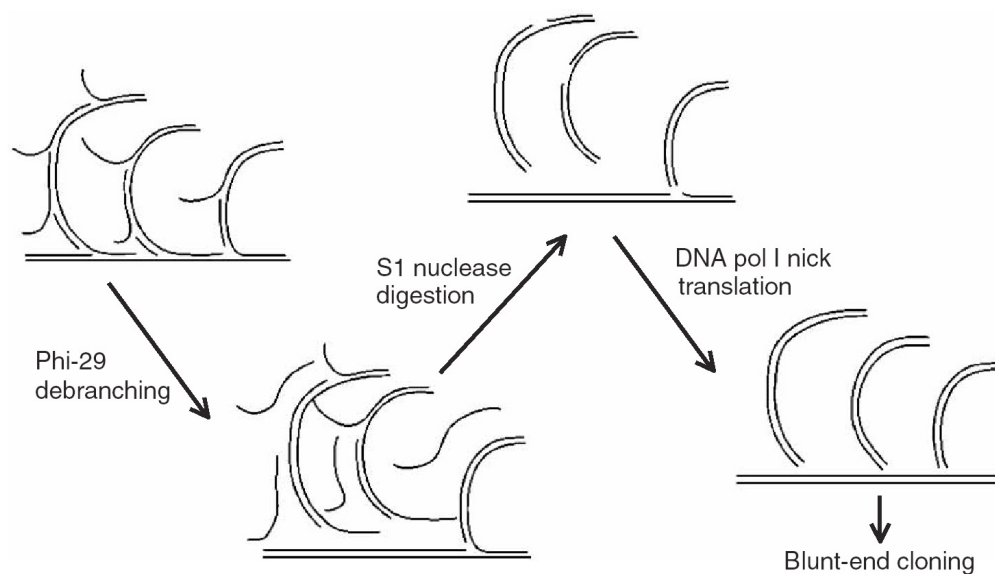
Figure 4. Scheme of required steps to resolve hyperbranched DNA structure derived by WGA as published previously (Zhang *et al.*, 2006).

2.3.11.1 Phi29 debranching

The reaction volumes were scaled up as follows:

DNA solution	up to 10 µg
Reaction buffer (10 x)	5 µl
dNTP Mix	1 mM
phi29 polymerase	10 U
dd H_2O	ad 50 µl

Phi29 debranching was carried out for 2 h at 30°C. The enzyme was heat-inactivated for 3 min at 65°C and subsequently the DNA was purified using SureClean solution (Bioline, Luckenwalde).

2.3.11.2 S1 nuclease digestion

The reaction volumes were scaled up as follows:

DNA solution	up to 10 μg
Reaction buffer (5 x)	10 μl
S1 nuclease	50 U
dd H_2O	ad 50 μl

S1 nuclease digestion was incubated for 30 min at 37°C and subsequently purified using SureClean solution (Bioline, Luckenwalde).

2.3.11.3 DNA polymerase I nick translation and end-repair

The reaction volumes were scaled up as follows:

DNA	up to 1 μg
T4 polymerase	3 U
DNA polymerase I	10 U
Taq polymerase buffer (10 x)	5 μl
dNTP Mix	0.5 mM
dd H_2O	ad 50 μl

The reaction mix was incubated for 1 h at room temperature and was subsequently heat inactivated for 10 min at 75°C.

2.3.11.4   Size fractionation of DNA by low melting point (LMP) agarose gel electrophoresis

The DNA from 2.3.11.3 was loaded onto a 1% LMP agarose gel in 1 x TAE buffer and gel electrophoresis was conducted as described in 2.3.6. In order to achieve a narrow separation of the DNA, the length of gel electrophoresis was shortened to 30 min. Following electrophoresis, the outer lanes of the gel containing the DNA marker were cut off and stained in ethidium bromide solution (see 2.3.6). The DNA was visualized under UV light and the position of marker bands at 3 and 10 kb were marked using a sterile scalpel. Subsequently, the gel was reassembled and a gel slice containing DNA that migrated between the marked positions of the DNA ladder was excised. The gel slice was transferred to a sterile 2 ml tube and DNA was recovered from the gel using GELase (Epicentre Biotechnologies, Madison) as recommended by the manufacturer.

2.3.11.5   Addition of dATP to blunt-ended DNA

By employing the *Taq* polymerase a deoxyadenosine was added to the blunt-ended DNA as follows:

Blunt-end DNA	up to 10 µg
dATP	2 mM
Taq polymerase	5 U
$MgCl_2$	2.5 mM
Reaction buffer (10 x)	7 µl
dd H_2O	ad 70 µl

The reaction mix was incubated for 20 min at 72°C with subsequent purification using SureClean solution (s. 2.3.3.3).

2.3.11.6   Dephosphorylation of DNA fragments

Prior to cloning into pCR-XL-TOPO, the 5'-phosphate residues at the end of the insert DNA fragments were removed by phosphatase treatment. To avoid religation of empty vectors the linearized pBAD vector, which was utilized for heterologous expression of putative *polA* genes (see 2.3.18) was dephosphorylated. To the linearized pBAD vector or the reaction mixture of 2.3.11.5, 5 U Antarctic phosphatase (NEB, Frankfurt) and 1 x reaction buffer were added. The mix was incubated for 15 min at 37°C and subsequently heat inactivated for 5 min at 65°C.

2.3.12   TOPO Cloning

For cloning of PCR products and amplified glacial DNA and for the construction of shotgun libraries of plasmids or fosmids, the vectors pCR2.1-TOPO or pCR-XL-TOPO were used (Invitrogen, Karlsruhe).

The linearized TOPO vectors have single, overhanging 3'-deoxythymidine residues. Thus, DNA fragments with a deoxyadenosine overhang can be ligated with the vector. The TOPO vectors contain a covalently bound topoisomerase I from the *Vaccinia* virus. The topoisomerase I binds to duplex DNA at certain sites and cleaves the phosphodiester backbone after 5'-CCCTT in one strand (Shuman, 1991). Subsequently, the phosphor-tyrosyl bond between the DNA and enzyme can be attacked by the 5'-hydroxyl of the original cleaved strand, thereby reversing the reaction and releasing topoisomerase (Shuman, 1994). The cloning reaction was carried out as recommended by the manufacturer.

2.3.13   Ligation using T4 DNA ligase

For ligation of DNA fragments into the pBAD expression vector (see 2.3.18) T4 DNA ligase was used, which catalyzes a phosphodiester linkage between 3'-

OH- and 5'-PO$_4$ ends of nucleic acids in the presence of ATP. An insert : vector molar ratio of 5 : 1 was chosen. The ligation mix was scaled up as follows:

Vector DNA	0.1 – 0.2 µg
Insert DNA	0.5 – 1 µg
T4 DNA ligase buffer (10 x)	1 U
dd H$_2$O	ad 20 µl

Ligation was carried out overnight at 16°C or for 2 to 4 h at room temperature.

2.3.14   Transformation of *E. coli*

Chemical transformation and electroporation was carried out as described by Hanahan (1983) and Dower *et al.* (1988), respectively. For preparation of chemically competent cells, *E. coli* Top10 and DH5α were grown at 18°C for 12 and 19 h, respectively. For preparation of electrocompetent cells, *E. coli* Top10, DH5α, and CSH26 *fcsA29* were grown at 30°C for 5 h and *E. coli* cs2-29 was grown for 22 h at 30°C. After the transformation insert-containing clones were selected by aid of α-complementation ('blue-white') screening (Sambrook, 1989).

2.3.15   Construction of a metagenomic fosmid library

Large glacial DNA fragments of 30 to 40 kb were cloned into fosmids using the CopyControl Fosmid Library Production kit (Epicentre Biotechnologies, Madison). Cloning was carried out as recommended by the manufacturer with the following exceptions: DNA was used for cloning without shearing, as the glacial DNA was already fragmented after DNA isolation. End-repair of fragmented DNA was carried out for 2 h after the size-selection step. The DNA was subsequently purified employing SureClean solution (Bioline,

Luckenwalde). For the ligation reaction a higher amount of DNA (600 ng of glacial DNA) than recommended was used. Ligation was carried out overnight at 16°C.

2.3.16   Sequence analysis

The Göttingen Genomics Laboratory (Göttingen, Germany) determined the sequences of plasmids and PCR products. The generated sequence data were edited by using the gap4 program of the Staden Package (Bonfield *et al.*, 1995). All 16S rRNA gene sequences were checked for chimeric artifacts using Mallard (Ashelford *et al.*, 2006) and the CHIMERA_CHECK program (Maidak *et al.*, 2001) of the Ribosomal Database Project (RDP) II database (Cole *et al.*, 2003). Phylogenetic classification was performed by using the RDP sequence match tool (http://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp) and the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI) database (Altschul *et al.*, 1990). Operational taxonomic units (OTUs) were determined at sequence similarity levels of 99, 97, 90, and 80% by employing the furthest-neighbor method of DOTUR (Schloss and Handelsman, 2005). To determine the number of observed unique OTUs as a function of the distance between sequences and the number of sequences sampled, rarefaction analysis was performed. In addition to the Shannon Weaver-Diversity Index, the Chao1 richness estimator and the abundance-based coverage estimator (ACE) were calculated using DOTUR. Only one sequence per OTU (>99% sequence identity cutoff) was used for construction of phylogenetic trees. Sequences of the nearest neighbors were retrieved from the NCBI and the RDP databases. The 16S rRNA gene sequences from clone libraries and the most similar neighbors were imported into the SSU Ref SILVA database (www.arb-silva/download/) of the ARB program package (Ludwig *et al.*, 2004). Multiple sequence alignments were checked manually and improved by the ARB editor tool. Phylogenetic trees were created by employing the maximum parsimony algorithm implemented in ARB. The robustness of

obtained tree topologies was evaluated by bootstrap analysis with 100 resamplings.

2.3.16.1 Gene prediction and annotation

ORFs putatively coding for proteins were predicted manually using the Artemis software (Rutherford *et al.*, 2000). Annotation was carried out manually by a BLASTP search (Altschul *et al.*, 1990) against the Swiss-Prot (Bairoch *et al.*, 2004), NCBI, Pfam (Finn *et al.*, 2008), and PROSITE (Hulo *et al.*, 2008) databases.

2.3.16.2 Analysis of metagenomic GS FLX data

The isolated glacial DNA was sequenced conducting two full runs (70 x 75 picotitre plates) on a Roche GS FLX sequencer (Roche, Mannheim). The pyrosequencing-derived dataset was analyzed with the phylogenetic algorithm CARMA (Krause *et al.*, 2008). Conserved Pfam domain and protein families were determined and classified into a higher-order taxonomy as described (Krause *et al.*, 2008). In addition, the dataset was compared with the NCBI-nr (as of July 2008), RDP II, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Groups (COG) databases using a BLAST search (Altschul *et al.*, 1990). Comparisons of the pyrosequencing-derived dataset to the NCBI-nr and COG databases were performed at a cutoff e-value of 10^{-5}. For comparison to the KEGG database a cutoff e-value of 10^{-3} was chosen. In order to identify taxonomic marker genes, sequences were analyzed by a BLASTN search against the RDP II database. Matches with an e-value $<10^{-3}$ and a match length >200 nt were accepted. For assignation to phylogenetic groups matches were re-analyzed by using the RDP Classifier (Wang *et al.*, 2007). Matches with a RDP confidence estimate below 60% were designated 'unclassified'.

2.3.17   Screening for genes encoding DNA polymerases

For the detection of plasmids and fosmids harboring the *polA* gene encoding the DNA polymerase I, complementation tests were carried out with the cold-sensitive *E. coli* mutant CSH26 *fcsA29* (Nagano *et al.*, 1999). *E. coli* CSH26 *fcsA29* carries a temperature-sensitive mutation in the *polA* gene causing lethality at temperatures below 20°C (Nagano *et al.*, 1999). Small-insert and large-insert metagenomic libraries from glacier ice (see 2.3.12 and 2.3.15) were transformed into CSH26 *fcsA29* and plated onto LB agar containing kanamycin (50 µg/ml) and chloramphenicol (12.5 µg/ml), respectively. Recombinant *E. coli* strains were incubated at 18°C for two (plasmids) and three (fosmids) days.

In order to confirm that complementation of the cold-sensitive mutation was fosmid-/plasmid-encoded, the recombinant fosmid and plasmids were isolated, retransformed into *E. coli* CSH26 *fcsA29* and *E. coli* cs2-29 (Nagano *et al.*, 1999), and subjected to the above-described screening for a second time. For *E. coli* cs2-29 L agar plates (see 2.2.2) containing the respective supplements were used and strains were incubated at 18°C for six days.

For all experiments the negative control *E. coli* CSH26 *fcsA29* or cs2-29 harboring an empty cloning vector was always included. Additionally, a control plate was always incubated at 37°C.

2.3.18   Heterologous expression of putative *polA* genes using pBAD *Myc*/His A

In order to verify that the identified putative *polA* genes (see Table 14) were responsible for complementation of the cold-sensitive phenotype of *E. coli* CSH26 *fcsA29* and cs2-29, sequences coding for the DNA polymerase I or domains typical for DNA polymerases were cloned into the expression vector pBAD *Myc*/His A (Invitrogen, Karlsruhe). This vector is a derivate of pBR322 designed for regulated, dose-dependent recombinant protein expression in *E. coli*. The *araBAD* promoter (P_{BAD}) from *E. coli* is used to produce optimum levels of soluble, recombinant protein. The regulatory protein, AraC, is supplied

on the pBAD *Myc*/His A expression vector allowing regulation of P_{BAD}. In the presence of L-arabinose, expression from P_{BAD} is turned on, whereas very low levels of transcription are induced from P_{BAD} in the absence of L-arabinose (Lee *et al.*, 1987). These uninduced transcription levels can be further repressed by growth in the presence of glucose. Glucose lowers expression of the catabolite-repressed P_{BAD} promoter by reducing the levels of 3', 5'-cyclic AMP (Miyada *et al.*, 1984). The predicted genes coding for the DNA polymerase I or typical domains of the DNA polymerase I were amplified by PCR (see 2.3.5). To the 5' end of the primers (Table 4), recognition sites for the restriction enzyme *Sac* I were attached. In between the recognition site and the start codon of the predicted ORF one additional nucleotide was inserted to ensure that the coding region was in frame with the P_{BAD} promoter. To ensure correct digestion of the DNA at the recognition sites, a deoxyadenosine was added to the PCR products (see 2.3.11.5) and subsequently these were cloned into pCR2.1-TOPO (see 2.3.12). In order to generate complementary DNA strands, the PCR products and the pBAD vector were both digested with *Sac* I (see 2.3.9). The vector was dephosphorylated (see 2.3.11.6) and following, the DNA fragments were ligated into pBAD as described in 2.3.13. The correct orientation of the cloned PCR products was verified by restriction analysis (see 2.3.9) and sequencing.

The resulting constructs were transformed into *E. coli* cs2-29 (Nagano *et al.*, 1999) and plated onto L agar containing ampicillin (100 µg/ml). Recombinant *E. coli* strains were incubated at 18°C for five to six days. For each experiment, one control plate and the negative control *E. coli* cs2-29 harboring the empty expression vector were incubated at 37 and 18°C, respectively.

2.4   Nucleotide sequence accession numbers

The 16S rDNA sequences were deposited in the GenBank database under accession numbers EU978474–EU978633, EU978636-EU978652, and EU978654-EU978854. Pyrosequencing data has been deposited in the NCBI Short Read Archive under accession number SRA001163. The nucleotide

sequences of the inserts of pCS1 to pCS8, and fCS1 harboring genes encoding DNA polymerases I or domains typical for DNA polymerase I have been deposited in the GenBank database under accession numbers FJ384787 to FJ384794, and FJ384795, respectively.

3   Results

In order to characterize the taxonomic and metabolic diversity of glacial ice and to exploit this subzero temperature habitat with respect to novel biocatalysts, ice was sampled at the Northern Schneeferner glacier, which is located at the Zugspitze, Germany. To extract the DNA from such a low-biomass environment, the samples were melted at 4°C and subsequently, cells were concentrated by filtration. The cell-containing membrane filters were used as starting material for DNA isolation. Several DNA isolation methods and kits were tested. Application of the Nucleo Spin Tissue kit was best with respect to yield and purity of the isolated DNA. Approximately 5 µg of DNA per litre of melted glacier ice were recovered. The isolated DNA was subjected to PCR amplification of the 16S rRNA gene sequences and pyrosequencing, which yielded 1,076,539 reads (239.7 Mb of sequence) with an average read length of 223 bp. Moreover, the isolated glacial DNA was employed for whole genome amplification (WGA) and construction of small-insert and large-insert metagenomic libraries. Subsequently, the metagenomic libraries were screened for genes encoding a DNA polymerase I.

3.1   Phylogenetic analysis of the microbial glacier ice community

The bacterial and archaeal diversity of the glacial ice metagenome was assessed by the following three approaches: analysis of a constructed 16S rRNA gene library, identification and classification of 16S rRNA gene sequences in the pyrosequencing-derived dataset, and evaluation of the pyrosequencing-derived dataset employing the CARMA algorithm. Additionally, the composition of the microbial community was assessed by denaturing gradient gel electrophoresis (DGGE) analysis. In order to complement the results established by these molecular approaches, microorganisms were isolated and phylogenetically characterized.

3.1.1   Analysis of a 16S rDNA clone library derived from glacier ice

The 16S rRNA genes present in the glacial DNA were amplified employing different sets of primers that were specific for Bacteria and Archaea. The products were then cloned and sequenced. All attempts to amplify archaeal 16S rRNA genes, including variation of primer sets and reaction conditions or employment of different DNA polymerases and $MgCl_2$ concentrations, were unsuccessful.

A total of 419 bacterial 16S rDNA sequences were amplified from glacial DNA and analyzed. Nearest neighbors were determined by searching both the NCBI and RDP II databases. Identification of chimeric sequences was performed using Mallard and the CHIMERA_CHECK program. Potential chimeras (66 sequences) and duplicates (15 sequences) were identified and removed. The remaining database contained 338 16S rRNA gene sequences that ranged in length from 770 to 1,500 bp. To identify unique phylotypes and to estimate the bacterial richness, the sequences were analyzed using DOTUR (Schloss and Handelsman, 2005). This revealed that the 338 16S rDNA sequences represented 108 operational taxonomic units (OTUs) based on a >99% sequence identity cutoff (Table 6).

Table 6. Richness and diversity estimates of the bacterial 16S rRNA gene clone libraries constructed from glacier ice of the Northern Schneeferner. The estimates were calculated by employing DOTUR (Schloss and Handelsman, 2005).

Distance	Richness[*]	ACE[†]	Chao1[‡]	Shannon[§]
0.01	108	173	153	4.09
0.03	72	132	107	3.44
0.1	28	38	35	2.19
0.2	13	15	14	1.66

[*]richness is expressed as number of observed unique operational taxonomic units (OTUs).

[†]abundance-based coverage estimator (ACE), nonparametric richness estimator based on distribution of abundant (>10) and rare (≤ 10) OTUs.

[‡]nonparametric richness estimator based on distribution of singletons and doubletons.

[§]Shannon-Weaver Index of diversity. A higher number indicates more diversity.

Fifty-eight OTUs contained two (doubleton) or more clones and the remaining 50 OTUs harbored a single clone (singleton). The richness of the bacterial lineages in the microbial community was analyzed by rarefaction analysis, calculation of the Chao1 and ACE richness estimates, and the Shannon-Weaver Index of diversity implemented in DOTUR (Table 6). Rarefaction curves reached saturation at a distance level of 20% (phylum level), but not at a distance level of 3% (species level) (Fig. 5). This indicated that a substantial fraction of the bacterial diversity was recovered during this study. In addition, the sampling effort gives an almost complete picture of the phyla present in the glacier ice.
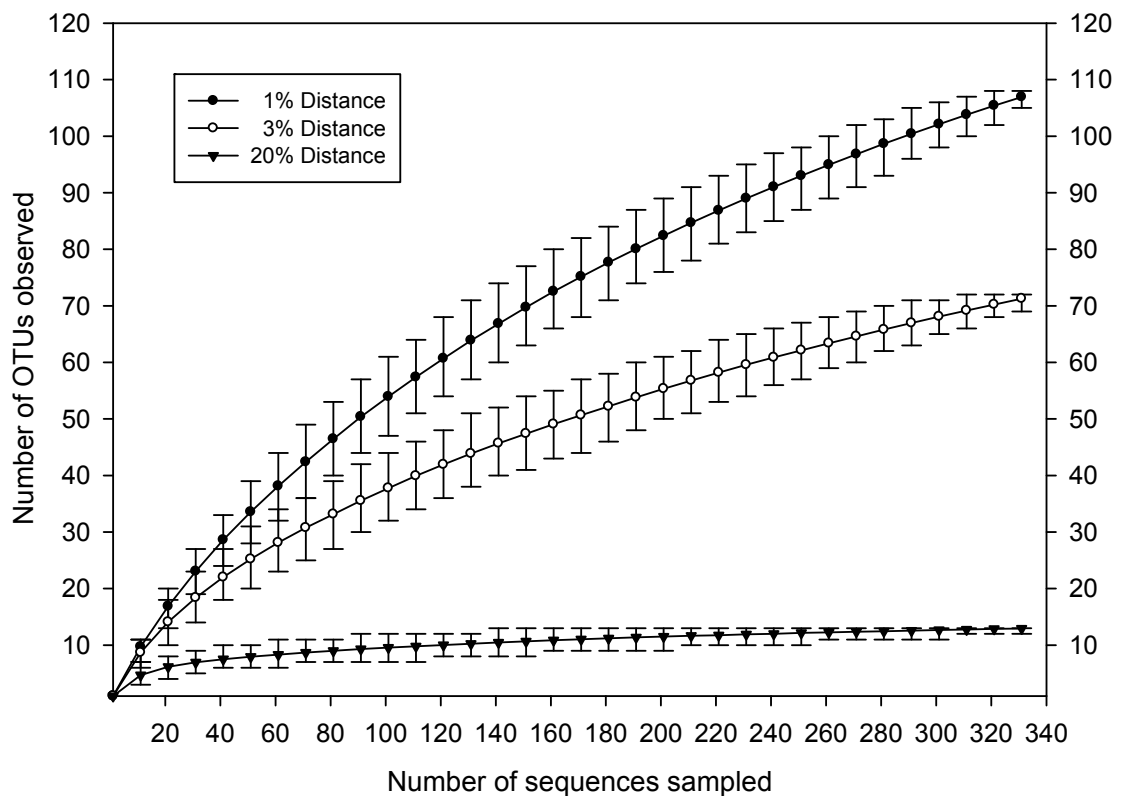
Figure 5. Rarefaction curves calculated with DOTUR (Schloss and Handelsman, 2005) indicating the observed number of OTUs within the 16S rRNA gene libraries derived from ice of the Northern Schneeferner glacier. OTUs are shown at the 1, 3, and 20% distance level. Error bars represent the 95% confidence interval.

This was confirmed by calculation of the Chao1 and ACE richness estimates: At a distance level of 3% 72 OTUs were observed, but the number of expected OTUs was 132 and 107, respectively (Table 6). At a distance level of 20% the

number of observed OTUs was 13, which is in agreement with the calculated number of expected OTUs. In addition, a high bacterial species diversity of the analyzed glacier ice was indicated by calculation of the Shannon-Weaver Index of diversity at a genetic distance of 3% (3.44) (Table 6). For most other environmental samples, this index ranges from 1.5 to 3.5 (Magurran, 1988).

The analyzed sequences represented a wide bacterial diversity and were affiliated to the following eleven phylogenetic groups: Proteobacteria (190 sequences, 55 OTUs), Bacteroidetes (101 sequences, 31 OTUs), Actinobacteria (33 sequences, 11 OTUs), Cyanobacteria (4 sequences, 2 OTUs), Gemmatimonadetes (3 sequences, 2 OTUs), Chloroflexi (1 sequence, 1 OTU), Verrucomicrobia (1 sequence, 1 OTU), Acidobacteria (1 sequence, 1 OTU), Deinococcus-Thermus (2 sequences, 2 OTUs), and candidate division TM7 (1 sequence, 1 OTU). One sequence showed no significant similarity to any known phylum and was assigned to the unclassified Bacteria (Fig. 6 and Fig. 7A-D).
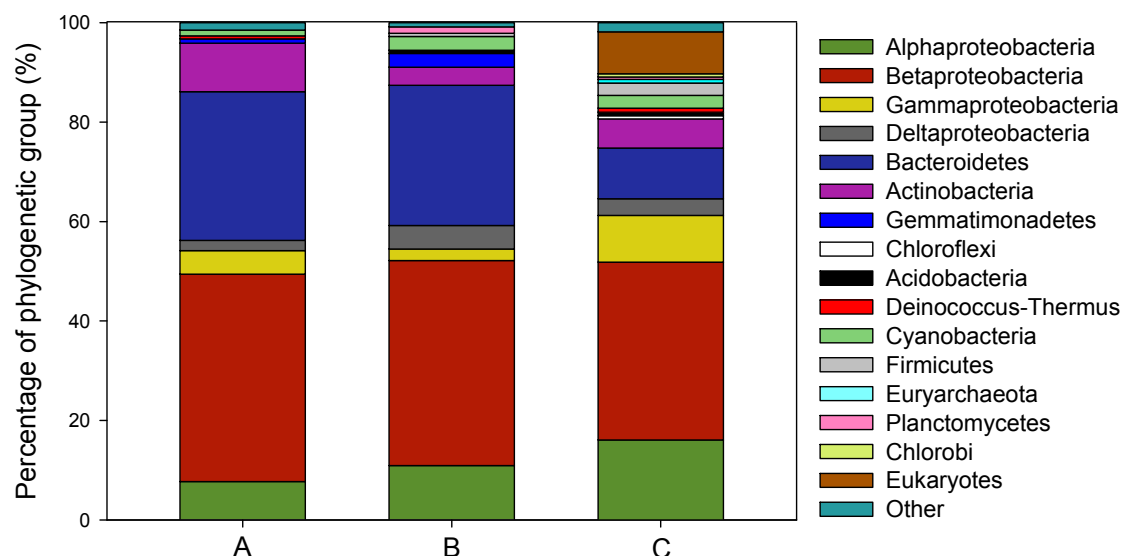
Figure 6. Distribution of phylogenetic groups present in glacier ice. Three different approaches were applied: analysis of amplified 16S rRNA gene sequences (A), comparison of pyrosequencing-derived dataset with the RDP II database (B) [cutoffs: match length >200 nt and RDP Classifier (Wang *et al.*, 2007) confidence estimate >60%], and taxonomic assignment of the pyrosequencing-derived dataset using CARMA (C) (Krause *et al.*, 2008). Shown are the percentages of the phylogenetically classified sequences. Phylogenetic groups accounting for less than 0.5% of the sequences are summarized in the artificial group 'Other'.

3.1.1.1   Construction of phylogenetic trees

Phylogenetic trees were constructed using one 16S rRNA gene sequence per OTU at a >99% sequence identity cutoff (Fig. 7A-D). Sequences were designated glb1a to glb353c. The maximum parsimony algorithm implemented in the ARB software package (Ludwig *et al.*, 2004) was employed for the analysis. Robustness of all trees was tested by bootstrap analysis with 100 resamplings.

The majority of the OTUs (55) belonged to the phylum Proteobacteria. Most of these OTUs (31) were part of the class Betaproteobacteria. The Alpha-, Gamma-, and Deltaproteobacteria were represented by 13, 8, and 2 OTUs, respectively. In addition, one OTU was assigned to the unclassified Proteobacteria. All betaproteobacterial sequences, except for sequences glb342c and glb281b, were affiliated to the order Burkholderiales. This included the families Comamonadaceae, Oxalobacteraceae, and the unclassified Burkholderiales (Fig. 7A). Sequences glb342c and glb281b were grouped into the unclassified Betaproteobacteria. Sequence glb342c revealed the lowest similarity (85%) of all betaproteobacterial sequences to its closest relative in the GenBank database, which is an uncultured bacterium isolated from a trembling aspen (EF019329). All other sequences exhibited sequence identities from 94 to 100% to their next neighbors. Most of the betaproteobacterial sequences were closely related to 16S rRNA gene sequences of uncultured organisms from cold habitats such as glacier bacterium FJS31 (AY315178), glacier bacterium FXI10 (AY315181), or Antarctic bacterium R-7724 (AJ440986). The sequences of the closest cultured relatives were derived from psychrophilic organisms such as *Polaromonas rhizosphaerae* (EF127651) or organisms typically isolated from water samples, e. g., *Aquabacterium* sp. (AF523022).

The 13 phylotypes within the Alphaproteobacteria displayed a high diversity and were related to the Rhodobacterales, Rhizobiales, Caulobacterales, Sphingomonadales, Rhodospirillales, and the unclassified Alphaproteobacteria (Fig. 7B). Identities of the selected sequences to their next neighbors ranged from 89 to 99%. In all cases, the nearest neighbors were uncultured organisms, which are only represented by 16S rRNA gene sequences. Many of these

sequences were derived from cold habitats, such as the ice cover of Antarctica (DQ521501), glacial ice (AY315166, DQ628964), or the snow cover at Spitzberg, Svalbard (DQ497241, DQ497241). The eight OTUs related to the Gammaproteobacteria were affiliated to the orders Xanthomonadales, Enterobacteriales, and Pseudomonadales (Fig. 7B). All sequences exhibited high identities (96 to 100%) to their closest relatives. The Deltaproteobacteria were represented by two OTUs, which belonged to the Bdellovibrionales or the Myxococcales (Fig. 7B). One OTU (glb216b) was assigned to the unclassified Proteobacteria, since it exhibited lowest identity (82%) of all proteobacterial sequences to its nearest neighbor in the database, the 16S rRNA gene sequence of an uncultured bacterium (EU287205) isolated from a Pacific Arctic surface sediment. The majority of the 31 OTUs belonging to the Bacteroidetes were related to the order Sphingobacteriales, including the Sphingobacteriaceae, Flexibacteraceae, Crenotrichaceae, and the unclassified Sphingobacteriales (Fig. 7C). The Flavobacteriaceae and the unclassified Bacteroidetes were represented by one OTU and two OTUs, respectively. Within the Bacteroidetes, sequences glb308b and glb314b exhibited the lowest similarity (84 and 85%, respectively) to their nearest neighbor, which was the 16S rRNA gene sequence of an uncultured Sphingobacteriales bacterium (AJ697700).

Figure 7. Maximum parsimony phylogenetic trees based on bacterial 16S rRNA gene sequences assigned to the Betaproteobacteria derived from ice of the Northern Schneeferner glacier, and their nearest relatives in the GenBank database. The tree was calculated using the ARB software package (Ludwig *et al.*, 2004). Sequences are characterized by sample designation (glb, glacier ice bacteria), clone number, length of sequence (a, 770 bp, b, 1100 bp, c, 1500 bp), and GenBank accession number. Sequences derived from glacier ice are shown in red. Numbers at branch nodes are bootstrap values (only values ≥50% are shown). The scale bar represents 0.1 nucleotide substitution per sequence position. Trees are rooted with the 16S rRNA gene sequence of *Thermotoga maritima*. Phylogenetic groups depicted: (A) Betaproteobacteria; (B) Alpha-, Gamma-, and Deltaproteobacteria; (C) Bacteroidetes; (D) Actinobacteria, candidate division TM7, Chloroflexi, Cyanobacteria, Verrucomicrobia, Acidobacteria, Gemmatimonadetes, Deinococcus-Thermus, and unclassified Bacteria (see following pages).

A

Betaproteobacteria

0.10

B

C

glb179b (EU978669)
glb263b (EU978751)
uncultured bacterium JEG.c5 (DQ228396)
55
glb16a (EU978515)
98
uncultured bacterium GKS2-106 (AJ290025)
70
glb103a (EU978594)
67
glb278b (EU978766)
uncultured Bacteroidetes bacterium AKYG1727 (AY921801)
glb175b (EU978665)
glb239b (EU978728)
82
glb112a (EU978603)
68
glb287b (EU978775)
uncultured Bacteroidetes bacterium CB_07 (EF56255)
uncultured Bacteroidetes bacterium IRD18F11 (AY947961)
89
glb70a (EU978563)
96
100
glb229b (EU978718)
57
Sphingobacterium sp. P-38 (AM411962)
96
glb279b (EU978767)
100
uncultured Bacteroidetes bacterium clone Amb_16S_9 (EF018676)
100
glb318b (EU978806)
100
uncultured Bacteroidetes bacterium SW11 (AJ575719)
100
Chryseobacterium sp. TB4-8-II (AY599655)
59
100
Flavobacterium sp. Ep300 (AF493662)
100
Antarctic bacterium R-7579 (AJ440981)
glb110a (EU978601)
100
glb208b (EU978697)
76
uncultured bacterium clone ANTLV2_H03 (DQ521530)
uncultured bacterium (DQ404663)
89
glb297b (EU978785)
100
uncultured Bacteroidetes bacterium IRD18G07 (AY947961)
100
uncultured Bacteroides sp. (AY510262)
84
glb240b (EU978729)
100
uncultured bacterium clone FCPP542 (EF516493)
94
Sphingoterrabacterium pocheensis (AB267718)
95
glb300b (EU978788)
65
Pochenobacter ginsengarvi TR6-06 (AB267720)
100
glb304b (EU978792)
65
uncultured bacterium Chun-w-71 (EF632777)
100
100
glb215b (EU978704)
Ginsengisolibacter pocheensis (AB267722)
98
glb283b (EU978771)
100
uncultured bacterium GKS2-216 (AJ290033)
Flectobacillus speluncae (AY065626)
uncultured Sphingobacteriales bacterium SF33 (AJ697700)
100
glb308b (EU978796)
glb314b (EU978802)
uncultured Bacteroidetes bacterium F15_8C_FL (EF683049)
100
glb187b (EU978677)
Cytophaga hutchinsonii (T) M58768 (CP000383)
70
uncultured bacterium clone KD6-118 (AY218749)
glb291b (EU978779)
88
glb40a (EU978537)
glb159a (EU978649)
37
glb115a (EU978606)
75
glb352c (EU978840)
Hymenobacter sp. 1004 (EF423320)
77
Hymenobacter sp. Pb17 (AB251884)
90
glb123a (EU978614)
99
glb150a (EU978640)
uncultured *Hymenobacter* group bacterium KL-2-5-13 (AF408273)
98
Hymenobacter sp. X2-1g (DQ888329)
Thermotoga maritima (AJ401017)

50

Bacteroidetes

0.10

D

Taxon / clade labels
Actinobacteria
TM7
Chloroflexi
Cyanobacteria
Verrucomicrobia
Acidobacteria
Unclassified Bacteria
Gemmatimonadetes
Deinococcus - Thermus

Cryocola sp. KAR3 (EF451633)
Frigoribacterium sp. MSL 08 (EF466126)
uncultured bacterium ARKDMS-27 (AF468262)
Arctic sea ice bacterium ARK10173 (AF468440)
glb275b (EU978763)
Cryobacterium sp. 1021 (EF423331)
glb290b (EU978778)
Cryobacterium psychrophilum (AM410676)
Cryobacterium aff. *psychrophilum* A1/C-aer/OI (AJ297438)
glb221b (EU978710)
glacier bacterium FJS11 (AY315169)
actinobacterium MH3-1 (EF187352)
Arctic sea ice bacterium ARK10186 (AF468441)
glb202b (EU978691)
uncultured bacterium ANTLV2_A11 (DQ521506)
glb265b (EU978753)
glacial ice bacterium G200-C11 (AF479342)
glb219b (EU978708)
uncultured bacterium clone ANTLV2_G12 (DQ521529)
glb248b (EU978737)
Micrococcineae str. Ellin184 (AF409026)
glb316b (EU978804)
uncultured actinobacterium TR1D12 (EU117950)
glb245b (EU978734)
uncultured bacterium BF0001B029 (AM697003)
glb269b (EU978757)
uncultured actinobacterium E1B-A4-114 (EF016798)
glb284b (EU978772)
uncultured bacterium FFCH11714 (EU132728)
glb21a (EU978520)
uncultured candidate division TM7 bacterium E1B-B (EF016808)
uncultured soil bacterium (AF507689)
uncultured soil bacterium (AF507687)
uncultured candidate division TM7 bacterium (DQ451478)
glb315b (EU978803)
uncultured bacterium BF0002B007 (AM697057)
uncultured bacterium (AY921524)
uncultured soil bacterium (AF507694)
glb171b (EU978661)
Chamaesiphon subglobosus PCC 7430 (AY170472)
uncultured cyanobacterium clone SepB-17 (EF032663)
glb296b (EU978784)
uncultured cyanobacterium RBE1CI-123 (EF111085)
glb210b (EU978699)
uncultured bacterium SLB624 (DQ787718)
uncultured verrucomicrobium DEV005 (AJ401105)
glb264b (EU978752)
uncultured Acidobacteria bacterium AKYG1078 (AY921950)
uncultured soil bacterium (AY289382)
glb320b (EU978808)
uncultured bacterium FFCH7771 (EU135026)
glb247b (EU978736)
uncultured bacterium ANTLV1_E09 (DQ521491)
glb306b (EU978794)
uncultured Gemmatimonadetes bacterium AKYG753 (AY922060)
uncultured bacterium (AF524020)
glb246b (EU978735)
bacterium Ellin510 (AY960773)
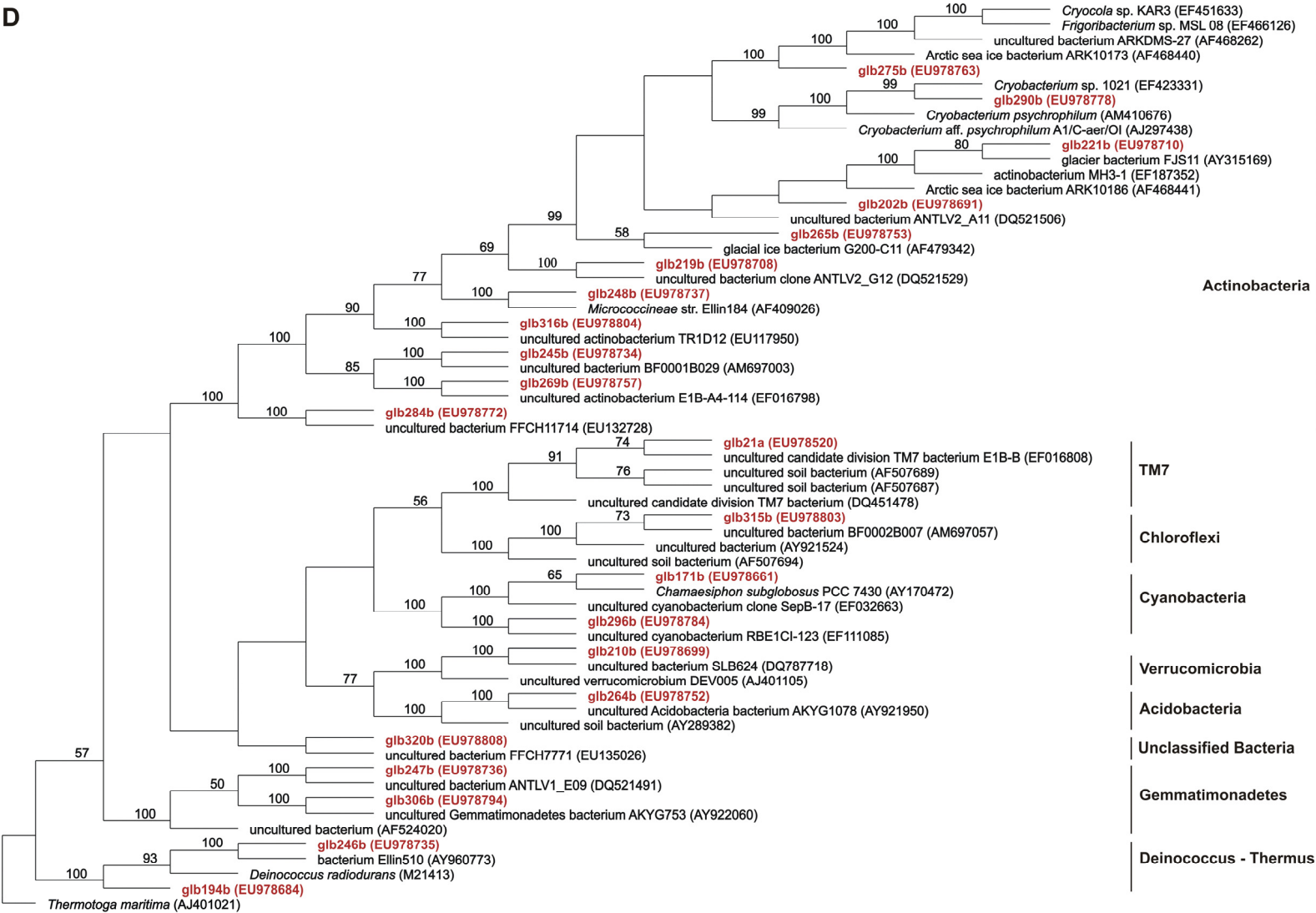*Deinococcus radiodurans* (M21413)
glb194b (EU978684)
Thermotoga maritima (AJ401021)

0.10

All other sequences belonging to the Bacteroidetes revealed identities ranging from 92 to 99% to their nearest neighbors. Almost all of the closest relatives of these sequences were 16S rRNA gene sequences amplified from temperate and cold environments such as a glacier of the Tianshan Mountain, China (EF423320), or Antarctic environments (AY218749, AJ440981, DQ521530).

Eleven OTUs were affiliated to the Actinobacteria, representing the Actinomycetales and the unclassified Actinobacteria (Fig. 7D). Five of these OTUs were related to the family Microbacteriaceae and exhibited high sequence identities (97 to 99%) to their next neighbors in the database, which were mainly clones derived from cold habitats such as glacial and subglacial environments, and ice samples from Antarctica (AF468440, EF423331, AY315169, DQ521506, AF479342). Other clades of the Actinobacteria were represented by phylotypes that belonged to the Intrasporangiaceae and the Frankiaceae. These sequences showed identities from 94 to 98% to their closest known relatives, which were all clones derived from extreme habitats such as the ice cover of Lake Vida in Antarctica (DQ521529), or desert soil (EF016798).

The remaining eleven OTUs were affiliated to the Chloroflexi (1 OTU), Cyanobacteria (2 OTUs), Verrucomicrobia (1 OTU), Acidobacteria (1 OTU), Gemmatimonadetes (2 OTUs), Deinococcus-Thermus (2 OTUs), the unclassified Bacteria (1 OTU), and candidate division TM7 (1 OTU) (Fig. 7D). All of these sequences exhibited similarities of 96 to 99% to their nearest neighbors, except for two sequences belonging to the phylum Deinococcus-Thermus and the unclassified Bacteria, respectively. These singletons glb194b and glb320b displayed low sequence identities (87 and 89%, respectively) to their next neighbors; the 16S rRNA gene sequences of *Deinococcus radiodurans* (M21413) and an uncultured bacterium (EU135026) isolated from tallgrass prairie soil, respectively.

3.1.2   Phylogenetic assessment of glacier ice by evaluation of a pyrosequencing-derived dataset

The glacial DNA was subjected to pyrosequencing using a Roche GS FLX sequencer. Pyrosequencing yielded 1,076,539 reads (239.7 Mb of sequence) with an average read length of 223 bp. In order to validate the results obtained by the 16S rDNA clone library and to identify 16S rDNA gene fragments, the pyrosequencing-derived dataset was compared to the RDP II database. Sequences with a match length of <200 nt and an e-value >10^{-3} were removed. In this way, 1,116 16S rRNA gene fragments were determined, of which 468 were affiliated to distinct phylogenetic groups (Fig. 6 and see Dataset S1 on supplementary CD). The following twelve bacterial phyla were established: Proteobacteria (59%), Bacteroidetes (28%), Actinobacteria (4%), Gemmatimonadetes (3%), Cyanobacteria (3%), Firmicutes (1%), Planctomycetes (1%), candidate division TM7 (<1%), Verrucomicrobia (<1%), Acidobacteria (<1%), Aquificae (<1%), and candidate division OD1 (<1%). The phyla Planctomycetes, Firmicutes, Aquificae, and candidate division OD1 were identified additionally to those determined in the 16S rDNA clone library. However, of all classified pyrosequencing-derived sequences these phyla were represented by 1% at the most. Thus, overall classification of the hereby identified 16S rRNA gene sequences confirmed the results of the 16S rDNA clone library with respect to the number of detected phyla and the dominant phylogenetic groups.

Additionally, the phylogenetic composition of the glacier ice metagenome was determined by analysis of the pyrosequencing-derived dataset using CARMA (Krause *et al.*, 2008). This algorithm assigns putative taxonomic origins based on conserved protein family (Pfam) domains and protein families, which are detected in unassembled sequence reads. A taxonomic classification was achieved at the superkingdom and phylum level for 14% and 11% of the reads, respectively. Most of the classified reads were assigned to the Bacteria (83% of the sequences). Other detected phylogenetic groups were the Eukaryota (9% of the sequences), Archaea (<1% of the sequences), and Viruses (<1% of the

sequences) (Fig. 8). Strikingly, in contrast to the analyses of 16S rDNA sequences a few reads (<1%) were assigned to the Archaea by CARMA analysis.

The following 17 bacterial phyla were detected employing this approach: Proteobacteria (65%), Bacteroidetes (10%), Actinobacteria (6%), Chloroflexi (<1%), Acidobacteria (<1%), Deinococcus-Thermus (<1%), Cyanobacteria (3%), Firmicutes (3%), Spirochaetes (<1%), Planctomycetes (<1%), Aquificae (<1%), Chlamydiae (<1%), Nitrospira (<1%), Fusobacteria (<1%), Chlorobi (<1%), Thermotogae (<1%), and Fibrobacteres (<1%). The bacterial composition with respect to the dominant phyla was highly similar to that established by the 16S rRNA-based approaches. In addition to nine bacterial phyla, which were also detected by the above-reported analyses the presence of eight further bacterial phyla was indicated. However, these phyla were represented by less than 1% of all classified sequences.
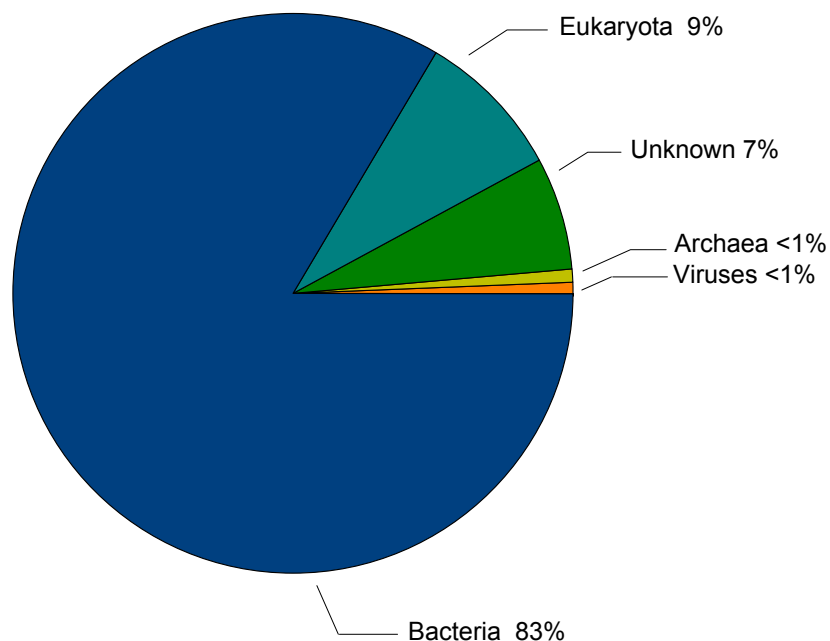
Figure 8. Distribution of phylogenetic groups in glacier ice determined by employing the CARMA algorithm (Krause *et al.*, 2008). Fourteen percent (150,478 sequences) of the pyrosequencing-derived dataset were phylogenetically classified at the superkingdom level. Shown are the percentages of all classified sequences.

3.1.3   Comparison of different approaches for taxonomic assessment

Comparison of the three methods for taxonomic assessment revealed a high degree of similarity; especially the results obtained by the two 16S rDNA analyses were in accordance (Fig. 6). The dominant phylogenetic group determined by analysis of the 16S rRNA gene library (A), comparison of the pyrosequencing-derived dataset to the RDP II database (B), and CARMA analysis (C) was the phylum Proteobacteria, representing 56% (A), 59% (B), and 65% (C) of all classified sequences. Most proteobacterial sequences were assigned to the class Betaproteobacteria [42% (A), 41% (B), and 36% (C)]. The class Alphaproteobacteria included 8% (A), 11% (B), and 16% (C) of all classified sequences. The second most frequently determined phylum Bacteroidetes comprised 30% (A), 28% (B), and 10% (C) of the sequences. In addition, a large part of the sequences grouped into the Gram-positive Bacteria, i. e., Actinobacteria. All other bacterial phyla were represented by less than 4% of the classified sequences.

3.1.4   Analysis of the bacterial community by DGGE

In order to support the results of the above-reported approaches for phylogenetic assessment DGGE analysis of bacterial 16S rRNA genes was performed. Thirteen discrete bands, designated A to M, were excised from the gel, re-amplified, and cloned (Fig. 9 and Table 7). To establish the number of phylotypes co-migrating at the same position on the gel, three clones derived from each band were sequenced. In four cases, the same 16S rRNA gene sequences were recovered from the three clones (bands C, F, G, J). In four cases two different clones were present (A, B, I, M) and in five cases all three analyzed clones were related to different organisms (D, E, H, K, L) (Fig. 9). A total of 27 different 16S rRNA gene sequences were recovered. The majority of the clones was related to uncultured organisms (92 to 100% sequence identity) isolated from cold habitats, such as glaciers of the Tianshan Mountains, China

(EU263765, EF423339, EU263763, EU263701, EU263711, EU263783), or Antarctica (EU636023, AM934651, EU636065, EU636026).

The Beta-, Gamma-, and Alphaproteobacteria were represented by 16, 3, and 2 sequences, respectively. The remaining six sequences were affiliated to the Bacteroidetes (Table 7). Phylotypes within the Betaproteobacteria were related to the Burkholderiales and Methylophilales. Two sequences did not show significant similarity to sequences derived from known betaproteobacterial orders and were therefore assigned to the unclassified Betaproteobacteria. Sequences belonging to the Gammaproteobacteria were affiliated to the Xanthomonadales and Pseudomonadales and alphaproteobacterial sequences grouped into the orders Rickettsiales and Sphingomonadales. The Bacteroidetes were represented by the orders Sphingobacteriales and the Flavobacteriales.

Overall, the results obtained by DGGE analysis are consistent with the results of the above-reported taxonomic approaches. The Betaproteobacteria and Bacteroidetes were in all analyses the main phylogenetic groups. However, the Actinobacteria, which represented 4 to 10% of the classified sequences established by the approaches described above, were not identified in the analyzed DGGE bands.
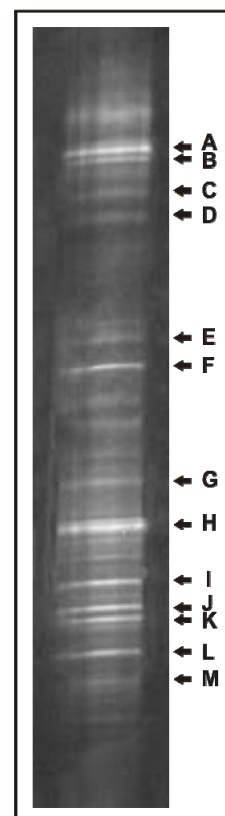
Figure 9. DGGE fingerprint of the microbial community present in glacier ice of the Northern Schneeferner. Bands designated A to M were excised from the gel, amplified, cloned, and sequenced. A 9% acrylamide gel with a denaturing gradient of urea and formamide from 25% to 75% was used. Electrophoresis was carried out at 60°C and 100 V for 16 h in TAE buffer.

Table 7. Similarity of glacial 16S rRNA gene sequences derived from DGGE analysis.

Band	Designation of the 16S rRNA gene (accession number)	Closest relative (accession number)	Identity	Phylum
A	glbD1 (EU978474)	Antarctic bacterium strain CA1 (EU636023)	99 (%)	Betaproteobacteria
	glbD2 (EU978475)	Uncultured bacterium clone N05Dec-106 (EU442913)	98 (%)	Bacteroidetes
B	glbD3 (EU978476)	Flavobacterium sp. WB2.3-68 (AM934651)	96 (%)	Bacteroidetes
	glbD4 (EU978477)	Uncultured bacterium clone Ni_B10_DMDN (EU150282)	99 (%)	Betaproteobacteria
C	glbD5 (EU978478)	Flavobacterium sp. 3008 (EF423339)	97 (%)	Bacteroidetes
D	glbD6 (EU978479)	Uncultured Flavobacteriaceae bacterium clone Amb_16S_1373 (EF018885)	89 (%)	Bacteroidetes
	glbD7 (EU978480)	Antarctic bacterium strain 3C2 (EU636065)	99 (%)	Alphaproteobacteria
	glbD8 (EU978481)	Uncultured bacterium clone KuyT-water-20 (EU263783)	99 (%)	Betaproteobacteria
E	glbD9 (EU978482)	Uncultured beta proteobacterium clone PIB-31 (AM849442)	99 (%)	Betaproteobacteria
	glbD10 (EU978483)	Uncultured bacterium clone P53-120 (EU375440)	99 (%)	Betaproteobacteria
	glbD11 (EU978484)	Uncultured bacterium clone KuyT-ice-89 (EU263763)	99 (%)	Betaproteobacteria
F	glbD12 (EU978485)	Uncultured bacterium clone KNA6-EB12 (AB179689)	98 (%)	Betaproteobacteria
G	glbD13 (EU978486)	Uncultured bacterium clone FCPP735 (EF516037)	98 (%)	Gammaproteobacteria
H	glbD14 (EU978487)	Uncultured Variovorax sp. clone 82I (AY466780)	94 (%)	Betaproteobacteria
	glbD15 (EU978488)	Uncultured bacterium clone 3.3.5 (EU528236)	99 (%)	Gammaproteobacteria
	glbD16 (EU978489)	Uncultured Comamonadaceae bacterium clone 469C10 (EU127420)	97 (%)	Betaproteobacteria
I	glbD17 (EU978490)	Uncultured bacterium clone KuyT-ice-18 (EU263756)	98 (%)	Bacteroidetes
	glbD18 (EU978491)	Antarctic bacterium strain G024 (EU636026)	99 (%)	Betaproteobacteria

Table 7 continued

Band	Designation of the 16S rRNA gene (accession number)	Closest relative (accession number)	Identity	Phylum
J	glbD19 (EU978492)	Uncultured betaproteobacterium TTGE band A3 (AJ867851)	99 (%)	Betaproteobacteria
K	glbD20 (EU978493)	Uncultured bacterium partial 16S rRNA gene, clone M6A-646 (AM991246)	92 (%)	Alphaproteobacteria
	glbD21 (EU978494)	Uncultured bacterium clone 2\SC\12 (EU340165)	98 (%)	Betaproteobacteria
	glbD22 (EU978495)	Uncultured bacterium clone 110ds10 (AY212562)	94 (%)	Betaproteobacteria
L	glbD23 (EU978496)	Aquabacterium sp. Aqua2 (AF089858)	98 (%)	Betaproteobacteria
	glbD24 (EU978497)	Uncultured bacterium clone Kuy-SL-34 (EU263701)	99 (%)	Betaproteobacteria
	glbD25 (EU978498)	Uncultured bacterium clone 202 (DQ158102)	99 (%)	Gammaproteobacteria
M	glbD26 (EU978499)	Uncultured bacterium clone Kuy-SL-50 (EU263711)	100 (%)	Bacteroidetes
	glbD27 (EU978500)	Uncultured bacterium clone KuyT-ice-51 (EU263765)	99 (%)	Betaproteobacteria

3.1.5 Isolation of microorganisms from glacier ice

To isolate microorganisms, samples of melted glacier ice were used to inoculate full strength R2A, 0.25 strength R2A, 0.5 TSA, and 0.2 strength LB agar plates at 4°C. Appearance of the first colonies was visible after five days of incubation. Over a time period of twelve weeks approximately 90 colonies with different morphologies were selected for 16S rRNA gene analysis. A total of 13 different isolates, designated glbl1 to glbl13, were obtained and purified by repeated transfer (four to five times) to agar plates (Table 8 and 9). All analyzed organisms were able to grow at temperatures of 4 and 18°C. Isolates glbl1, glbl2, glbl3, glbl7, glbl10, and glbl12 showed growth at 30°C, but none of the isolates was able to grow at a temperature of 37°C. This implied that all of the isolates are either psychrophilic or psychrotolerant species. Most of the purified organisms were rod-shaped and the majority of the isolates formed pigmented colonies (pink, orange, yellow, cream) (Table 8).

Seven of the bacterial isolates were related to members of the Proteobacteria (glbl1, glbl2, glbl3, glbl7, glbl9, glbl11, glbl12). The four betaproteobacterial isolates (glbl2, glbl7, glbl9, glbl11) grouped into the orders Burkholderiales and Rhodocyclales, and the three gammaproteobacterial isolates (glbl1, glbl3, glbl12) grouped into the order Pseudomonadales. Three isolates were related to the Sphingobacteriales within the phylum Bacteroidetes (glbl4, glbl8, glbl10), and three isolates grouped into the order Actinomycetales within the Actinobacteria (glbl5, glbl6, glbl13). All of the isolates exhibited sequence identities of 98 and 99% to their nearest described relatives, except for isolate glbl4, which displayed a sequence similarity of 95% to its next described neighbor, *Sphingoterrabacterium pocheensis* (AB267718). Thus, the latter isolate putatively represented a new species. Psychrophilic organisms, such as *Cryobacterium psychrophilum* (EF467640), *Pedobacter cryoconitis* (AJ438170), and *Polaromonas naphtalenivorans* (CP000529) and psychrotrophic organisms, such as *Janthinobacterium lividum* (DQ473538) were the closest relatives of the isolates (Table 9).

Table 8. Physiological characteristics of isolates recovered from glacier ice.

Isolate (accession number)	Colony color	Cell morphology	Growth temperature[*]			
			4°C	18°C	30°C	37°C
glbl1 (EU978842)	white	short rods	+	+	+	-
glbl2 (EU978843)	white	short rods	+	+	+	-
glbl3 (EU978844)	white	rods	+	+	+	-
glbl4 (EU978845)	pink	cocci	+	+	-	-
glbl5 (EU978846)	orange	rods	+	+	-	-
glbl6 (EU978847)	yellow	cocci	+	+	-	-
glbl7 (EU978848)	white	short rods	+	+	+	-
glbl8 (EU978849)	cream	long thin rods	+	+	-	-
glbl9 (EU978850)	translucent yellow	long thin rods	+	+	-	-
glbl10 (EU978851)	translucent yellow	thin rods	+	+	+	-
glbl11 (EU978852)	translucent white	short rods	+	+	-	-
glbl12 (EU978853)	translucent cream	short rods	+	+	+	-
glbl13 (EU978854)	orange	cocci	+	+	-	-

[*] + and – indicate growth and no growth, respectively.

Table 9. Nearest neighbors of isolates recovered from glacier ice.

Isolate (accession number)	Organism (accession number)	Identity	Phylum
glbI1 (EU978842)	*Pseudomonas veronii* strain A1XB2-5 (AY512619)	99 (%)	Gammaproteobacteria
glbI2 (EU978843)	*Janthinobacterium lividum* GA01 (DQ473538)	99 (%)	Betaproteobacteria
glbI3 (EU978844)	*Pseudomonas mephitica* strain CCUG 2513T (AM748811)	99 (%)	Gammaproteobacteria
glbI4 (EU978845)	*Sphingoterrabacterium pocheensis* (AB267718)	95 (%)	Bacteroidetes
glbI5 (EU978846)	*Cryobacterium psychrophilum* strain RuGl_7 (EF467640)	99 (%)	Actinobacteria
glbI6 (EU978847)	*Cryobacterium aff. psychrophilum* A1/C-aer/OI (AJ297438)	99 (%)	Actinobacteria
glbI7 (EU978848)	*Janthinobacterium lividum* strain XT1 (EU642885)	98 (%)	Betaproteobacteria
glbI8 (EU978849)	*Sphingobacterium siyangensis* strain SY1 (EU046272)	98 (%)	Bacteroidetes
glbI9 (EU978850)	*Zoogloea ramigera* (X74914)	98 (%)	Betaproteobacteria
glbI10 (EU978851)	*Pedobacter cryoconitis* (AJ438170)	99 (%)	Bacteroidetes
glbI11 (EU978852)	*Polaromonas naphthalenivorans* CJ2 (CP000529)	99 (%)	Betaproteobacteria
glbI12 (EU978853)	*Pseudomonas lini* (AY035996)	98 (%)	Gammaproteobacteria
glbI13 (EU978854)	*Cryobacterium psychrophilum* strain RuGl_7 (EF467640)	98 (%)	Actinobacteria

The results of the culture-dependent analysis of the glacier ice agree with the results of the molecular approaches described above. All of the isolated organisms belonged to the Proteobacteria, Bacteroidetes, and Actinobacteria, which were the dominant phyla established in the molecular analyses.

3.2   Functional analysis of the glacier ice metagenome

3.2.1   Assessment of the functional diversity

To establish the functional diversity present in the Northern Schneeferner, the pyrosequencing-derived dataset was compared to the COG and KEGG databases. Of the dataset, 308,546 sequences (29% of the sequences) were assigned to 4,876 unique KO (KEGG Orthology) groups at a cutoff e-value of 10^{-3}. In addition, 355,512 sequences (33% of all sequences) showed matches to the COG database, which correspond to 3,709 unique COGs at a cutoff e-value of 10^{-5}.

The functional category metabolism in the COG and KEGG databases was represented by 45% and 57% of all sequences, respectively. Genetic information processing (KEGG) and cellular processes and signaling (COG) yielded 22% and 20% of all matches, respectively. The category environmental information processing (KEGG) and information storage and processing (COG) included 16% and 19% of all classified sequences, respectively (Fig. 10).

Within the KEGG categories, matches were separated into different subcategories and several sequences occupied more than one subcategory (Table 10 and see supplementary Dataset S2). Most of the glacial sequences in the subcategory carbohydrate metabolism shared homologies to genes involved in pyruvate (8,523 sequences), butanoate (7,975 sequences), propanoate (7,990 sequences), starch and sucrose (4,218 sequences), and glyoxylate and dicarboxylate (4,041 sequences) metabolism.

Figure 10. COG (A) and KEGG (B) category distributions of the pyrosequencing-derived dataset deduced from glacier ice. Percentages of all classified sequences in each category are shown. Analyses were performed using BLASTX (Altschul et al., 1990) with a cutoff e-value of 10^{-5} and 10^{-3} for COG (Tatusov et al., 2001) and KEGG (Kanehisa et al., 2008) databases, respectively.
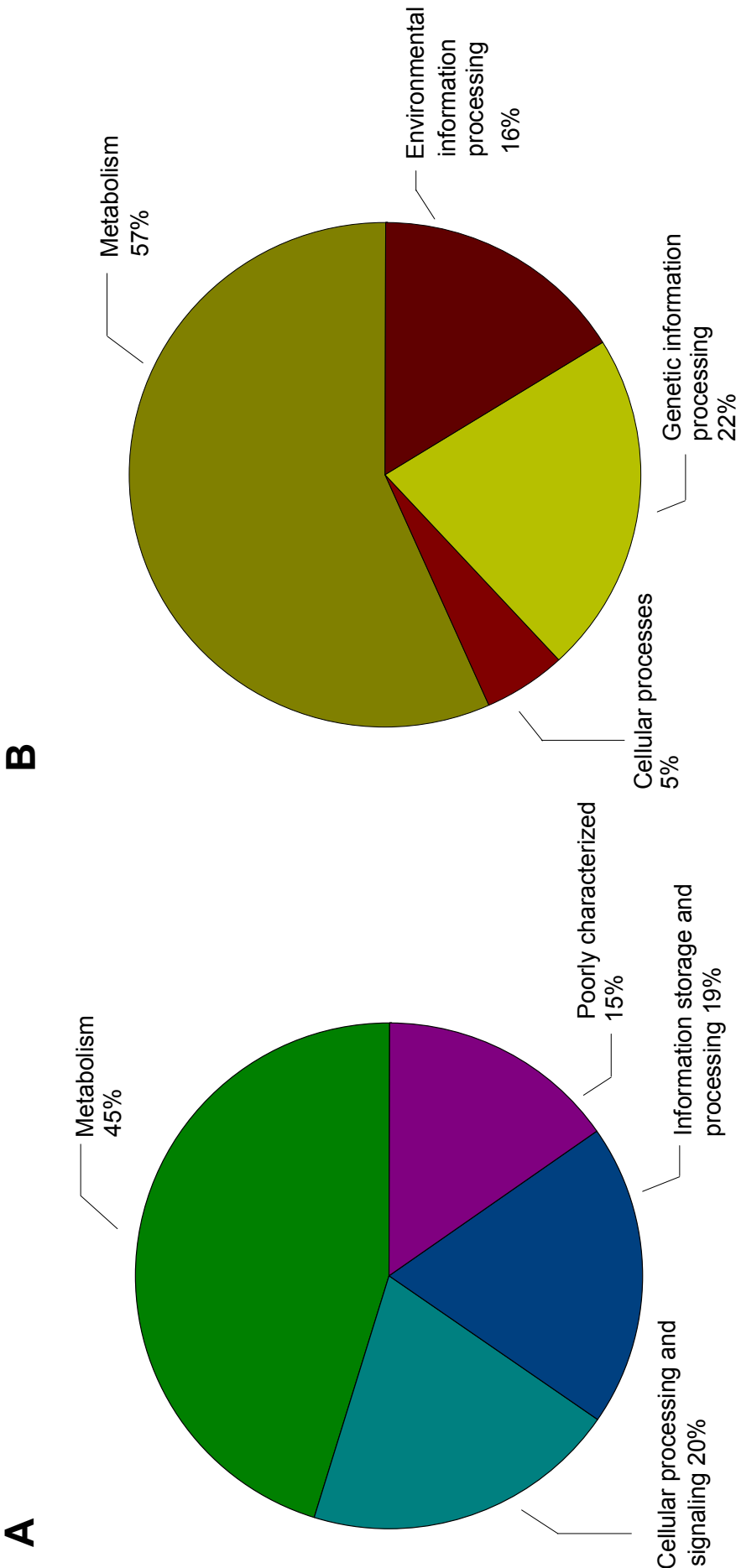
In addition, sequences were homologous to genes responsible for glycolysis/gluconeogenesis (6,450 sequences), the citrate cycle (5,677 sequences), and the pentose-phosphate pathway (4,077 sequences).

The majority of the glacial sequences in the category energy metabolism were related to genes that participate in oxidative phosphorylation (8,954 sequences). Most sequences related to photosynthesis (1,511 sequences) were associated with ATPase-like genes and only few sequences showed similarities to genes encoding components of photosystems. Thus, a low degree of photosynthesis was indicated. Genes involved in nitrogen metabolism (5,539 sequences) were highly represented. These included assimilatory and respiratory nitrate/nitrite reductase-like genes. The presence of these genes implied that assimilation and respiration of anorganic nitrogen sources is very important in glacial ice. The detection of many glutamate/glutamine synthase-like genes provided evidence for the production of glutamate, which is the most common response of bacteria to increased osmolarity (Medigue *et al.*, 2005) in addition to its function as a central nitrogen carrier. Most sequences associated with sulfur metabolism (1,866 sequences) showed similarities to cysteine synthases, sulfate adenylyltransferases, and sulfite reductases, which participate in sulfur assimilation (Table 10).

The importance of an autotrophic lifestyle in glacial ice was indicated by the presence of a significant number of genes involved in carbon fixation via the Calvin cycle and the reductive carboxylate cycle. Within the category methane metabolism (2,694 sequences) glacial sequences were mainly related to catalases/peroxidases, formate dehydrogenases, CO dehydrogenases, and glycine hydroxymethyltransferases. The latter enzyme is involved in assimilation of C_1 substrates via the serine pathway. In addition, catalases/peroxidases represent important antioxidative enzymes. In general, unsaturated fatty acids are more susceptible to oxidative damage than saturated fatty acids. Nevertheless, unsaturated fatty acids are an important characteristic of bacteria living at low temperatures, as they maintain the fluidity of the membrane (Methe *et al.*, 2005). Correspondingly, 1,075 sequences were homologous to genes associated with the biosynthesis of unsaturated fatty acids (supplementary Dataset S2, and see Table 11)

.

Table 10. Number of pyrosequencing-derived sequences showing homologies to genes associated with KEGG pathways in the categories 'carbohydrate metabolism' and 'energy metabolism'. The values were determined by comparison of pyrosequencing-derived sequences derived from the glacial DNA to the KEGG database (Kanehisa *et al.*, 2008) at a cutoff e-value of 10^{-5}. Dataset S2 on supplementary CD shows a detailed list of all matches to KEGG pathways and associated KO and COG groups.

KEGG class ID	KEGG category [ko number]	Number of matches
1110	*Carbohydrate Metabolism*	*41723*
10	Glycolysis / Gluconeogenesis [PATH:ko00010]	6450
20	Citrate cycle (TCA cycle) [PATH:ko00020]	5677
30	Pentose phosphate pathway [PATH:ko00030]	4077
31	Inositol metabolism [PATH:ko00031]	518
40	Pentose and glucuronate interconversions [PATH:ko00040]	1323
51	Fructose and mannose metabolism [PATH:ko00051]	3340
52	Galactose metabolism [PATH:ko00052]	1761
53	Ascorbate and aldarate metabolism [PATH:ko00053]	1181
500	Starch and sucrose metabolism [PATH:ko00500]	4218
520	Nucleotide sugars metabolism [PATH:ko00520]	1982
530	Aminosugars metabolism [PATH:ko00530]	2949
562	Inositol phosphate metabolism [PATH:ko00562]	785
620	Pyruvate metabolism [PATH:ko00620]	8523
630	Glyoxylate and dicarboxylate metabolism [PATH:ko00630]	4041
640	Propanoate metabolism [PATH:ko00640]	7990
650	Butanoate metabolism [PATH:ko00650]	7975
660	C_5-Branched dibasic acid metabolism [PATH:ko00660]	1072
1120	*Energy Metabolism*	*28859*
190	Oxidative phosphorylation [PATH:ko00190]	8954
191	Pyruvate/Oxoglutarate oxidoreductases	1068
192	ATPases	1976
195	Photosynthesis [PATH:ko00195]	1511
196	Photosynthesis - antenna proteins [PATH:ko00196]	92
680	Methane metabolism [PATH:ko00680]	2694
710	Carbon fixation [PATH:ko00710]	3676
720	Reductive carboxylate cycle (CO_2 fixation) [PATH:ko00720]	4676
910	Nitrogen metabolism [PATH:ko00910]	5539
920	Sulfur metabolism [PATH:ko00920]	1866

Interestingly, only few sequences (286 sequences) were associated with ether lipid metabolism (supplementary Dataset S2). Since ether lipids are typical for Archaea, this supports the above-reported detection of only few Archaea during CARMA analysis of the pyrosequencing-derived data.

The metabolism of secondary metabolites was mostly associated with limonene and pinene degradation (2,924 sequences) (supplementary Dataset S2). In addition, the synthesis of carotenoids (599 sequences) was indicated. This is consistent with the above-reported pigmentation of the organisms isolated from the Northern Schneeferner. A high degradative potential of the analyzed microbial glacier ice community was indicated by the large number of sequences that showed similarities to genes involved in degradation and metabolism of xenobiotics (14,523 sequences) (supplementary Dataset S2). Strikingly, the KEGG category environmental information processing contained 4,043 sequences exhibiting similarities to genes associated with the type II secretion system but only 580 and 671 sequences for type III and IV secretion system, respectively (supplementary Dataset S2). Recently, it has been shown that type II secretion systems can possess important functions with respect to survival in environmental niches. It has been proposed that type II exoproteins promote growth at low temperatures (Soderberg *et al.*, 2004).

3.2.2   Determination of characteristics for a psychrophilic lifestyle

To identify features related to a psychrophilic lifestyle the pyrosequencing-derived dataset was compared to the NCBI-nr database employing a cutoff e-value of 10^{-5}. The thereby recovered matches were searched for keywords associated with a psychrophilic lifestyle (Table 11).

Since the solubility of gasses increases rapidly at low temperatures, the ability to respond to reactive oxygen species is an essential function for organisms living at low temperatures. To prevent oxidative harm genes similar to antioxidative enzymes such as catalases/peroxidases, dioxygenases and superoxide dismutases were present (Table 10 and Table 11). The avoidance of cell damage by formation of ice crystals is also a prerequisite for living at

subzero temperatures. Correspondingly, many genes involved in the synthesis of well-known cryo- and osmoprotectants, such as glycine, betaine, choline, sarcosine, and glutamate were detected (Table 11). Maintenance of protein folding is indicated by the presence of genes encoding peptidyl-prolyl cis-trans isomerases, which are typical for adaptation to a cold-adapted lifestyle (D'Amico *et al.*, 2006). Additionally, a large number of desaturases that are essential for the biosynthesis of unsaturated fatty acids were found, thereby supporting the above-reported results of the functional analysis of glacier ice (Table 11).

In summary, a distinct functional profile of the microbial glacier ice community was established by comparison of the pyrosequencing-derived dataset to the KEGG, COG, and NCBI-nr databases. Genes associated with several metabolic pathways and traits required for life at low temperatures were highly represented.

Table 11. Enzymes, amino acids, and compounds associated with a psychrophilic lifestyle and their functions.

Keyword	Number of sequence annotations containing keyword[*]	Selected function
Desaturase	1012	Catalyzes synthesis of unsaturated fatty acids (Bolter, 2004)
Peptidyl-prolyl cis-trans isomerase	1021	Maintains protein-folding rates at low temperatures (Thomas and Dieckmann, 2002)
Glycine	2644	Osmoprotectant (Thomas and Dieckmann, 2002)
Betaine	471	Osmoprotectant and cryoprotectant (Methe *et al*., 2005)
Choline	827	Precursor of glycine betaine (Methe *et al*., 2005)
Sarcosine	233	Intermediate in the metabolism of choline to glycine (Methe *et al*., 2005)
Glutamate	5782	Osmoprotectant (Medigue *et al*., 2005)
Dioxygenase	2706	Antioxidative enzyme (Medigue *et al*., 2005)
Superoxide dismutase	257	Antioxidative enzyme (Methe *et al*., 2005)

[*]NCBI-nr database matches of the 454 sequence annotations (cutoff e-value 10^{-5}) containing the keyword.

3.3   Metagenomic bioprospecting

3.3.1   Construction of metagenomic libraries

To exploit the microbial diversity present in glacier ice, small-insert and large-insert metagenomic libraries were constructed in plasmids and fosmids, respectively. Since the glacial DNA yield is much lower than that of high-biomass environments such as soils and sediments, starting material for the construction of small-insert metagenomic libraries was generated by multiple displacement amplification (MDA). To improve cloning efficiency and to avoid abnormal insert size distribution of the amplified DNA, hyperbranched structures that were introduced during MDA were resolved as published previously (Zhang *et al.*, 2006). The prepared DNA was cloned into the plasmid vector pCR-XL-TOPO. In this way, a small-insert library was constructed, which comprised 230,000 *E. coli* clones with an average insert size of 4 kb. The percentage of plasmids containing inserts was approximately 97%. Additionally, a large-insert fosmid library was constructed without a prior amplification step of glacial DNA by using the fosmid pCC1FOS as vector. The fosmid library consisted of about 4,000 fosmids with an average insert size of 36 kb. In summary, both constructed metagenomic libraries harbored ca. 1.07 Gb of cloned glacial DNA.

3.3.2   Screening for genes encoding DNA polymerases

In order to identify novel DNA polymerases a function-driven screening approach was chosen, which was based on complementation of a cold-sensitive lethal mutation in the *polA* gene of *Escherichia coli* CSH26 *fcsA29* (Nagano *et al.*, 1999). This strain carries a temperature-sensitive lethal mutation in the 5'–3' exonuclease domain of DNA polymerase I, which causes filamentation of the cells with dispersed nuclei. This mutation is lethal at temperatures below 20°C (Nagano *et al.*, 1999). The glacial DNA-containing

recombinant plasmids and fosmids were transformed into the mutant and the resulting *E. coli* CSH26 *fcsA29* clones were incubated at 18°C. Thus, only recombinant *E. coli* strains harboring a gene conferring polymerase activity could grow under the employed conditions. Positive clones with a colony diameter of >3 mm were visible after 48 to 72 h of incubation. The negative control *E. coli* CSH26 *fcsA29* harboring the cloning vector without insert showed no growth under the employed conditions.

To identify one positive clone during the initial screen approximately 1,000 clones (plasmid library) and 100 clones (fosmid library) needed to be tested. Seventeen plasmid-containing clones and one fosmid-containing clone were randomly chosen for further analysis. In order to confirm that complementation of the cold-sensitive mutation was encoded by the inserts of the vectors, the fosmid and plasmids were isolated from the positive clones, retransformed into *E. coli* CSH26 *fcsA29*, and screened again under selective conditions. The fosmid (fCS1) and fifteen of the plasmids (pCS1 to pCS15) conferred a stable phenotype at 18°C and were subjected to further analyses.

3.3.3   Identification and characterization of putative *polA* genes

In order to determine the complete sequence of the fosmid (fCS1), the insert-containing fosmid was mechanically sheared and subsequently subcloned into a plasmid vector. The inserts of all recombinant plasmids and the fosmid were sequenced and analyzed with the gap4 program of the Staden Package. Putative open reading frames (ORFs) were identified using the Artemis software and were subsequently compared to the database of the NCBI.

Analysis of the fosmid fCS1 insert (32 kb) revealed the presence of 26 ORFs, including one putative *polA* gene (Table 12). Fifteen of the predicted ORFs were similar to genes from *Rhodoferax ferrireducens* or *Polaromonas* sp., which both belong to the Comamonadaceae (Betaproteobacteria) (Table 12).

Table 12. ORFs and genes predicted on the fosmid fCS1 and observed similarities to known gene products from other organisms [*]

ORF	No. of amino acids	Closest similar protein, no. of amino acids (accession no. of similar protein)	Organism	Region of similar amino acids (% identity)	E-value
orf1	644[†]	Beta-glycosidase-like protein (YP_679442.1)	Cytophaga hutchinsonii ATCC 33406	391-699/1152 (46)	1E-51
orf2	377	Ppx/GppA phosphatase (YP_523288)	Polaromonas sp. JS666	126-498/499 (64)	3E-138
orf3	606	Diguanylate cyclase/phosphodiesterase (YP_523286.1)	Rhodoferax ferrireducens T118	8-590/598 (65)	0
orf4	165	Hypothetical protein Rfer_2019 (YP_523275)	Rhodoferax ferrireducens T119	1-136/136 (51)	5E-25
sixA	153	Phosphohistidine phosphatase, SixA (YP_523274)	Rhodoferax ferrireducens T118	1-153/153 (71)	4E-55
ppk	700	Polyphosphate kinase (YP_523273.1)	Rhodoferax ferrireducens T118	2-692/700 (77)	0
orf7	360	Glycosyl transferase, group 1 (YP_549068.1)	Polaromonas sp. JS666	1-366/367 (70)	3E-148
orf8	287	Metallophosphoesterase (YP_549067.1)	Polaromonas sp. JS666	1-286/286 (72)	7E-121
orf9	271	Ornithine-acyl(acyl carrier protein) N-acyltransferase (ABD69753)	Rhodoferax ferrireducens T118	30-282/282 (79)	6E-117
orf10	301	Putative transmembrane protein (YP_547856)	Polaromonas sp. JS666	6-293/301 (70)	3E-111
orf11	305	Hypothetical protein Pnap_1274 (YP_981510.1)	Polaromonas naphthalenivorans CJ2	1-305/305 (83)	5E-147
orf12	292	Adenylate cyclase (YP_980335.1)	Polaromonas naphthalenivorans CJ2	200-498/505 (42)	6E-52
orf13	361	Putative carboxypeptidase (YP_523281.1)	Rhodoferax ferrireducens T118	28-373/373 (68)	8E-143
orf14	171	GCN5-related N-acetyltransferase (YP_982814.1)	Polaromonas naphthalenivorans CJ2	10 - 172/173 (77)	3E-68
glM	443	Phosphoglucosamine mutase (YP_523269.1)	Rhodoferax ferrireducens T118	1 - 442/443 (88)	0
orf16	153	Dihydropteroate synthase (YP_523268.1)	Rhodoferax ferrireducens T118	167 - 292/292 (75)	2E-42
amaB	272	N-carbamoyl-L-amino acid amidohydrolase (ZP_01721100)	Algoriphagus sp. PR1	157 - 427/430 (57)	1E-81
orf18	248	Putative hydrolase (YP_552647)	Burkholderia xenovorans LB400	14 - 248/387 (31)	5E-31
orf19	290	Oxidoreductase (NP_924951.1)	Gloeobacter violaceus PCC 7421	5 - 289/290 (69)	2E-114

Table 12 continued

ORF	No. of amino acids	Closest similar protein, no. of amino acids (accession no. of similar protein)	Organism	Region of similar amino acids (% identity)	E-value
orf20	205	Hypothetical protein Bphy_7244 (YP_001863288)	Burkholderia phymatum STM815	10 - 202/218 (28)	2E-19
orf21	513	Oxidoreductase (NP_924950)	Gloeobacter violaceus PCC 7421	23 - 529/535 (66)	0
orf22	447	Hypothetical protein gll2003 (NP_924949.1)	Gloeobacter violaceus PCC 7421	2 - 442/445 (69)	0
polA	962	Putative DNA polymerase I (ZP_01718371)	Algoriphagus sp. PR1	8 - 941/941 (57)	0
orf24	317	5,10-methylenetetrahydrofolate reductase (ZP_01692153.1)	Microscilla marina ATCC 23134	3 - 317/318 (57)	3E-112
orf25	274	ABC transporter, ATP-binding protein (ZP_01718881)	Algoriphagus sp. PR1	1 - 244/245 (75)	4E-102
orf26	149†	Auxin-regulated protein (ZP_01885154)	Pedobacter sp. BAL39	1 - 149/508 (46)	4E-34

*The identification of proteins homologous to the predicted gene products encoded by fCS1 and the determination of percent identity, region of similar amino acids and e-value was performed by using the BLAST programs (Altschul et al., 1997). The DNA sequence was submitted to GenBank under accession number FJ384795. Identified ORFs were designated a gene name, if the observed e-value was $<10^{-50}$.

†Incomplete gene.

Partial sequencing of the flanking regions of the inserts from plasmids pCS9 to pCS15 revealed that they were identical to the inserts of pCS1, pCS4, pCS7, or pCS8. Therefore, pCS9 to pCS15 were not studied further. The high number of duplicates is probably a result of the amplification of the DNA by MDA. The insert sizes of pCS1 to pCS8 ranged from 3.5 to 15 kb (Table 13).

Table 13. Insert sizes and detected Pfam domains within the predicted *polA* genes of plasmids pCS1 to pCS8 and fosmid fCS1. The domains were identified by searches against the Conserved Domain Database (CDD) database (Marchler-Bauer *et al.*, 2007).

Plasmid/fosmid (accession no.)	Insert size (kb)	Detected Pfam domains within predicted gene[*]
pCS1 (FJ384787)	3.5	53 exo, 35 exo
pCS2 (FJ384788)	15.0	53 exo, 35 exo, DNA_polA
pCS3 (FJ384789)	6.3	53 exo, DNA_polA
pCS4 (FJ384790)	4.8	53 exo
pCS5 (FJ384791)	6.4	53 exo, 35 exo, DNA_polA
pCS6 (FJ384792)	4.6	53 exo, 35 exo, DNA_polA
pCS7 (FJ384793)	6.2	53 exo, 35 exo, DNA_polA
pCS8 (FJ384794)	9.0	53 exo, 35 exo, DNA_polA
fCS1 (FJ384795)	32.0	53 exo, 35 exo, DNA_polA

[*]The e-values are given in Fig. 11.

Sequencing of the complete plasmid inserts by primer walking was possible for pCS1, pCS5, and pCS6 but not for the five remaining plasmids. The latter was caused by the presence of repeat structures. The formation of these chimeric artifacts is a well-known drawback of MDA (Zhang *et al.*, 2006). To circumvent this problem, shotgun libraries of the plasmids with insert sizes of approximately 1 kb were constructed and sequenced.

Sequence analyses of pCS1 to pCS8 and fCS1 revealed that all inserts contained ORFs, which exhibited similarities to known PolA-encoding genes. Four of the plasmids (pCS2, pCS5, pCS6, pCS8) and the fosmid (fCS1) contained a putative *polA* gene that encodes all three domains typical for DNA polymerase I (Table 13 and Fig. 11). The number of amino acids deduced from

corresponding proteins (927 to 962 amino acids) is similar to that of DNA polymerase I from *E. coli* (928 amino acids), which is the prototype for these kinds of enzymes (Riley *et al.*, 2006) (Table 14 and Fig. 11). In addition, the plasmid pCS7 contained an almost complete version of the *polA* gene, which lacked part of the C-terminal polymerase domain. The amino acid sequence of the putative *polA* gene product encoded by pCS3 is slightly shorter (803 amino acids) than that of *E. coli*. The central region of the deduced enzyme showed no significant similarities to central 3'-5' exonuclease domains of other DNA polymerases. The amino acid sequence showed highest identity (35%) to the DNA polymerase I of *Thermus thermophilus* (YP_144320) (Table 14). The remaining two plasmids (pCS1 and pCS4) harbored complete ORFs, which encode shorter versions of PolA (Fig. 11). The gene product encoded by pCS1 (557 amino acids) contained a putative 5'-3' exonuclease domain and a 3'-5' exonuclease domain. The protein encoded by pCS4 (282 amino acids) was the smallest of all and contained solely a 5'-3' exonuclease domain.

The mutation of the complemented *E. coli* host strain is located in the 5'–3' exonuclease domain of DNA polymerase I (Nagano *et al.*, 1999). Correspondingly, the identified genes located on the inserts of pCS1 to pCS8 and fCS1 encoded at least this domain. Furthermore, all amino acid sequences of 5'–3' exonuclease domains of DNA polymerases I derived from glacier ice harbored regions characteristic for these domains. Based on the sequence alignment of ten bacterial and bacteriophage nucleases six conserved sequence motifs containing 14 invariant amino acids were identified, of which nine were carboxylate residues (Fig. 12) (Gutman and Minton, 1993). It was suggested that some of these highly conserved carboxylate amino acids are involved in binding metal ligands that are indispensable for the nuclease activity (Amblar *et al.*, 2001; Joyce and Steitz, 1994). Strikingly, the 5'-3' exonuclease domain encoded by pCS4 differed in four of the highly conserved residues as compared to all other 5'-3' exonuclease domains derived from glacier ice.

Table 14. Sequence similarities of the presumptive DNA polymerase activity-conferring gene products encoded by pCS1 to pCS8 and fCS1 to gene products from other organisms. The determination of the values for the percentage of identity, the region of similar amino acids, and the e-value were performed by using the BLAST programs (Altschul et al., 1990).

Plasmid/ fosmid	No. of amino acids	Closest similar protein, no. of amino acids (accession no. of similar protein)	Organism	Region of similar amino acids (% identity)	E-value
pCS1	557	putative DNA polymerase I, 941 (ZP_01718371)	Algoriphagus sp. PR1	8 - 521 (52)	1E-155
pCS2	944	DNA polymerase I, 955 (ZP_01689558)	Microscilla marina ATCC 23134	1 - 955 (58)	0
pCS3	803	DNA polymerase I, 833 (YP_144320)	Thermus thermophilus HB8	7 - 827 (35)	5E-124
pCS4	282	Exodeoxyribonuclease, 276 (AAX12058)	Bacteriophage T5	2 - 257 (43)	8E-52
pCS5	927	DNA polymerase I, 923 (YP_001845275)	Acinetobacter baumannii ACICU	1 - 923 (65)	0
pCS6	942	DNA polymerase I, 937 (ZP_01884419)	Pedobacter sp. BAL39	2 - 937 (57)	0
pCS7	646	DNA polymerase I, 920 (AAG43148)	Rhodococcus erythropolis	1 - 647 (82)	0
pCS8	927	DNA polymerase I, 923 (YP_001845275)	Acinetobacter baumannii ACICU	1 - 923 (66)	0
fCS1	962	putative DNA polymerase I, 941 (ZP_01718371)	Algoriphagus sp. PR1	8 - 941 (57)	0

Figure 11. Domain structures of the DNA polymerase activity-conferring gene products of pCS1 to pCS8 and fCS1. The localization and similarities of the domains were determined by searches against the CDD database (Marchler-Bauer *et al.*, 2007). The domain structure of DNA polymerase I of *E. coli* was taken from Riley *et al.* (2006). The amino acid length of the gene products is given below the boxes. The localization of the detected domains is indicated above the boxes. The recorded e-values for the domains 53 exo (cd00008), 35 exo (cd00007), and DNA_polA (pfam00476) are given in parenthesis.
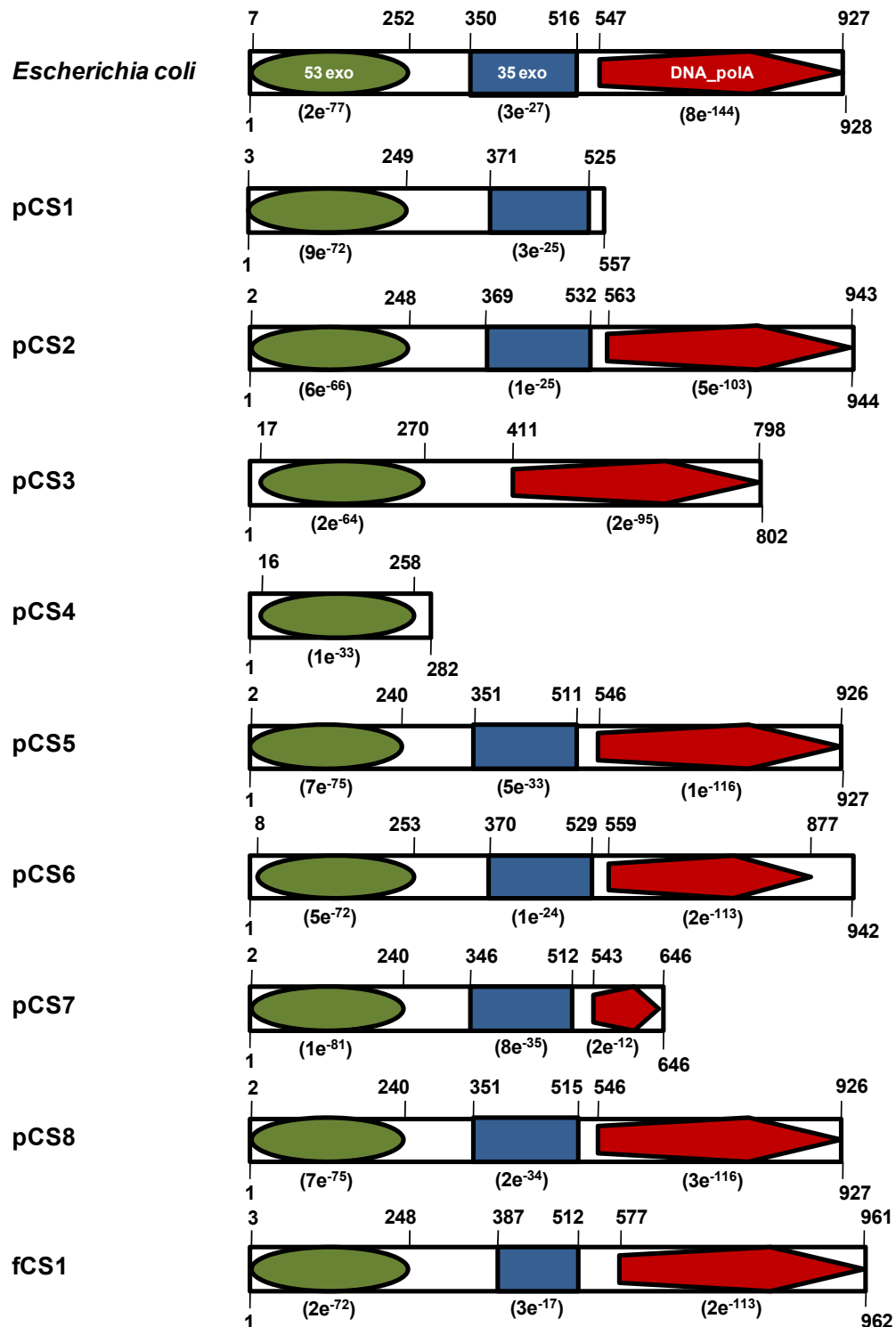
Figure 12. Alignment of the deduced amino acid sequences of 5'-3' exonuclease domains encoded by the putative *polA* genes of pCS1 to pCS8 and fCS1. The amino acid sequence of the 5'-3' exonuclease domain of the *E. coli* DNA polymerase I (Riley *et al.*, 2006) was used as reference. The six conserved 5'-3' exonuclease motifs containing 14 invariant amino acids (shown in bold) as described by Gutman and Minton (1993) are marked with letters A to F and are highlighted in yellow. The consensus sequences (Gutman and Minton, 1993) are indicated below conserved motifs. The amino acid position (Asp116) responsible for the cold-sensitive mutation *fcsA29* (Nagano *et al.*, 1999) is highlighted in red.

```
                              A
E.coli  ----------MVQIPQNPLILVDGSSYLYRAYHAFP--PLTNSAGEPTGAMYGVLNMLRS  48
pCS5    ------------MP-PFILVDGSYYLFRAFHALP--PLTTSQGQTTNAIKGALSALQK  43
pCS7    ------------MP--PFVLVDGSYFLFRAFHALP--PLTTSTGLQTNAIRGAISAIQK  43
pCS1    -------------MPKKLFLLDAMALIYRAYYALIRSPRITSKGRNTNAQFGFTSTLID  46
pCS2    --------------MKKLFLLDSMALIYRAHFAFINNPRQTSKGLDTSAVLGFINVLLE  45
fCS1    -------------MQKKLFLLDAFALVFRAYYALIRNPRITSKGRNTNAQFGFTNTLVD  46
pCS6    ---------MITKQTDKKLFLLDAFALIYRAYFAFSSNPRINSKGFNTSAIFGFTNTLLE  51
pCS3    MERALLLMATKIKSDKKTLVLLDSHAILHRAYHALP--DFASPSGEPTGALYGVVAMLLK  58
pCS4    ---MTKTFKTLSASSSDALMVVDSLNLAFRYKHSRAVDFADDYMDTVQSLRRSYHCDKLV  57
                              LLLVDG----F

              B           C
E.coli  LIMQYKPTHAAVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIE-PLHAMVKAMGLPLL 107
pCS5    LMRRMQPTHMAVIFDTPEPTFRHMLSPEYKAHRPAMPSELSTQIP-YLHAIIKGLGIPLL 102
pCS7    LMRRVQPTHMAVIFDTPEPTFRHVLSPIYKGDRPSMPEELSQQIP-YLHALIRALGIPLH 102
pCS1    LINKEKPTHLAVCFDTEAATERHTDFAEYKANRQAAPEDLISALP-DIKKIITAFNIPVV 105
pCS2    ILKKEKPTHIAAAFDLSAPTFRHIEYTAYKANRPTQPEGITAGIP-YIKMLLKAMNIPIL 104
fCS1    LINNQKPTHMAVCFDTHALTERHTDFADYKANRQETPEDILAAVP-DIKKIIEALNIPVI 105
pCS6    ILTKEKPSHIAVVFDMEGPTQRHIEFETYKANREEMPEDLRKSIP-IIIEMIKGFNIEVL 110
pCS3    IIEDFKPDYIAACFDLPEPTYRHEAFAAYKGTRSKTDDALVQQII-RSRDIFAAFGIPIY 117
pCS4    IAGDMGSSSYRKALSPIYKQNRKDKFAEQTEQEAAEFEAFFAEVQGILVRYEEEATYPLL 117
            IVLFD         FR-E----YK--R

              D                   E
E.coli  AVSGVEADDVIGTLAREAEKAG-RPVLISTGDKDMAQLVTPNITLINT-----MTNTILG 161
pCS5    LLPGAEADDLIGTLACRAVREG-HHVLISTGDKDMAQLVNDHIKLEDS-----FRDQVLD 156
pCS7    TLPGAEADDIIGTLAKRAEKAG-HQVLISTGDKDMAQLVTDKVTLEDS-----FKDKPMD 156
pCS1    ELDGYEADDVIGTLAWQAADKG-YEVFMVTPDKDYGQLLIHNNVFIYKPPYQGGKEDILD 164
pCS2    QLEGYEADDVIGTIAKKLSAPD-LEIYMMTSDKDYCQLLEENRIFMFRPASKFSPNEVWG 163
fCS1    AIDGYEADDVIGALARQAELAG-YDVYMVTPDKDYGQLVTEK-IKIYKPPYQGGSIEIMG 163
pCS6    GLEGYEADDVIGTLAIKAEKAG-FTTYMMTPDKDYGQLVDEN-TFIYKPARLGNGAEILG 168
pCS3    EHPGFEADDMLGTIAYQTRDDKDLTVIIASGDMDTMQCVEKKRVQVYTLKKGIKDTILYD 177
pCS4    RFPGVEADDIAAYIVSKRKQFNLSQIWLISSDRDWDLLVAQGVSRFSY-----VTRKEVT 172
            EADDV-A-L           IIS-DKD

                              F
E.coli  PEEVVNKYGVP-PELIIDFLALMGDSSDNIPGVPGVGEKTAQALLQGLGGLDTLYAEPEK 220
pCS5    HDGVIAKFGVR-PDQIIDYLTLMGDASDGIAGVPGVGKVTAAKLLNQYQTIGGILENAHL 215
pCS7    VDGVIEKFGVR-PDQIIDYLTLMGDASDGIRGVPGVGAKTAAKLLNEYGTIAGILENVDN 215
pCS1    AKKICEKWGIERVDQVVDMLGLMGDAVDNIPGIAGIGEKTACKLLKEFDNLENILANADN 224
pCS2    VSKALEKFGIKRVEQVIDMLGLQGDAVDNIPGLPGVGEKTAQKLLEEYDNIENIIANVQN 223
fCS1    PKEVCEKWGIKNVSQVIDILGLMGDAVDNIPGIRGVGEKTAAKLLAEYETLENILANADT 223
pCS6    TEEICKKWEIRNVAELIDILGLMGDKVDNIPGIPGVGEKTAIQLIKDFGSIENLLENTDK 228
pCS3    EEAVKERFGFG-PKLVPDYKGLRGDTSDNIPGIVGIGEKTATDLITQFGSIDDIYKKLKK 236
pCS4    LDNWSTHYDFT-QEEYISIKCLMGDSGDNVIGVPGIGPKKAAQLVLDYGSAYDIIDALPI 231
                              L-GD—SDNI-GV-G-G
```

Growth experiments revealed that the growth rates of all recombinant strains containing pCS1 to pCS8 were in the same range (0.18 h^{-1} to 0.2 h^{-1}) and the recombinant strains containing fCS1 exhibited a slower growth (0.07 h^{-1}) (Table 15). This indicated that the complementation of the mutant is independent from the presence of a 3'-5' exonuclease domain or a polymerase domain.

Table 15. Growth rates (µ) of *E. coli* CSH26 *fcsA29* transformed with pCS1 to pCS8, and fCS1, respectively.

Plasmid/fosmid	Growth rate µ (h^{-1})
pCS1	0.21
pCS2	0.18
pCS3	0.18
pCS4	0.20
pCS5	0.20
pCS6	0.19
pCS7	0.20
pCS8	0.20
fCS1	0.07

All predicted *polA* gene products exhibited amino acid sequence identities of 35% (pCS3) to 82% (pCS7) to DNA polymerases from other organisms. These proteins were derived from a variety of different genera, such as *Algoriphagus*, *Pedobacter*, *Microscilla*, *Thermus, Acinetobacter,* and *Rhodococcus* (Table 14). In order to verify that the identified putative *polA* genes were responsible for complementation of the cold-sensitive *E. coli* mutant, the genes were amplified by PCR and cloned into the expression vector pBAD *Myc*/His A, thereby placing the genes under control of the arabinose-inducible *araBAD* promoter. Since arabinose is toxic for *E. coli* CSH26 *fcsA29*, this strain was not a suitable host for these experiments. Alternatively, the *E. coli* strain cs2-29 (Nagano *et al.*, 1999) was used as host. This strain carries the same cold-sensitive mutation of *polA* as *E. coli* CSH26 *fcsA29*, but is able to grow in the presence of arabinose (Nagano *et al.*, 1999). Recombinant *E. coli* cs2-29 clones containing the original recombinant plasmids (pCS1 to pCS8) or the fosmid (fCS1) were

indistinguishable from the corresponding *E. coli* CSH26 *fcsA29* clones with respect to growth at 18°C. The pBAD *Myc*/His A constructs harboring the different identified *polA* genes were transformed into *E. coli* cs2-29. Subsequently, the resulting recombinant strains were used to inoculate agar plates supplemented with arabinose. Growth of all strains was detected after five to six days of incubation at 18°C. The negative control containing the expression vector without an insert showed no growth under the employed conditions. Thus, these results confirmed that the identified genes were responsible for complementation of the cold-sensitive *E. coli* mutants.

Nine novel genes coding for entire DNA polymerases I or individual polymerase domains were identified in metagenomic small-insert and large-insert libraries derived from glacial DNA by complementation of a cold-sensitive mutation in the *polA* gene of *E. coli*.

4   Discussion

Molecular ecology and metagenomics applied to the analysis of microbial biodiversity in various environments has altered our understanding of the biosphere. The perspective of environmental microbiology has improved enormously over the last two decades with respect to knowledge and methodology (Lau and Liu, 2007). A vast diversity of Archaea and Bacteria has been unraveled by application of molecular tools in various habitats. Attempts to link function to the phylogenetic diversity of ecosystems have lead to the discovery of novel metabolisms and to a different view of the ecological implications (Lopez-Garcia and Moreira, 2008). Thus, metagenomics has proven to be a powerful tool to estimate the phylogenetic content and the metabolic potential of microbial communities in various environments. In addition, a variety of novel biocatalysts have been discovered (Manichanh *et al.*, 2008).

In the present study, different approaches for phylogenetic assessment of the microbial diversity in glacier ice were combined. In addition to analysis of a 16S rDNA clone library and assessment of the community structure by denaturing gradient gel electrophoresis (DGGE) a large pyrosequencing-derived dataset was evaluated. The pyrosequencing-derived dataset was searched for 16S rDNA fragments and was evaluated by using the algorithm CARMA (Krause *et al.*, 2008). CARMA assigns phylogenetic groups to sequence fragments based on similarities to conserved Pfam domains and protein families (Krause *et al.*, 2008). To complement the results derived from these molecular approaches, microorganisms were isolated and phylogenetically classified. In addition, metagenomic small-insert and large-insert libraries were constructed from glacial ice DNA. Subsequently, the constructed libraries were screened for genes encoding DNA polymerases I by using a novel function-based screening approach. Fig. 13 shows an schematic overview of the metagenomic analyses conducted in this study.
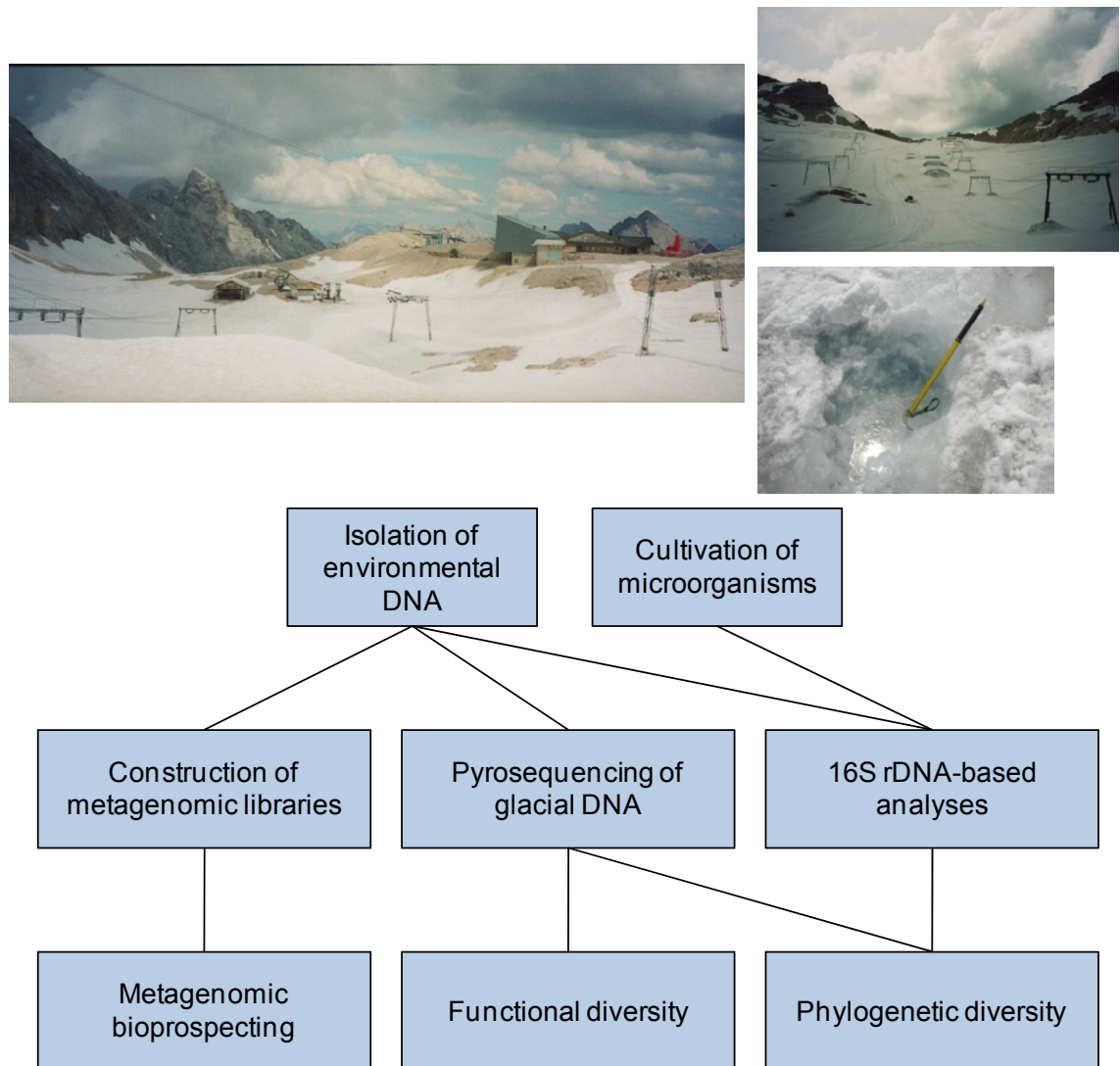
Figure 13. Schematic representation of the approaches included in the metagenomic analyses of glacier ice. The pictures at the top show the sampling site at the Northern Schneeferner, Germany.

4.1 Phylogenetic analysis of the bacterial glacier ice community

4.1.1 Approaches employed for phylogenetic assessment

One of the first applications of molecular phylogeny was the recognition of the Archaea and the construction of a tripartite tree of life by Woese (1987).

Evolutionary relationships between organisms were determined by comparative sequence analysis of the small subunit ribosomal RNA (16S rRNA). Based on analyses of cultivated organisms, the bacterial domain comprised approximately twelve phylogenetic groups. However, organisms that can be isolated by standard techniques are estimated to constitute less than 1% of all microbial species (Staley and Konopka, 1985). Pace *et al.* proposed in 1985 a molecular technique that bypassed the need to cultivate microorganisms for identification of their 16S rDNA sequences. The authors employed direct analysis of environmental 5S and 16S rRNA gene sequences for the description of phylogenetic diversity (Pace *et al.*, 1985). The current number of recognized bacterial phyla ranges from 50 to 88 (Lopez-Garcia and Moreira, 2008). The difference in the number of acknowledged phyla reflects both phylogenetic artifacts and contradictory taxonomic practices. Many 16S rDNA-based analyses are characterized by a lack of resolution of the deepest nodes. A phylogenetic group can be misplaced in trees because of insufficient phylogenetic signal and other tree construction artifacts, such as incomplete taxonomic sampling or mutational saturation. These obstacles can lead to a false description of novel phyla. Fig. 14 shows the 53 bacterial phyla (Lopez-Garcia and Moreira, 2008) that are acknowledged by at least three of the five taxonomic frameworks used in greengenes, which is a recently developed tool for online analysis of user-generated sequences (DeSantis *et al.*, 2006). This application allows cross-comparison of different taxonomic nomenclatures from different curators, i. e., the NCBI, RDP, Ludwig *et al.* (2004), Hugenholtz (2002), or Pace (1997).

Many researchers have employed the rRNA approach for analysis of a wide variety of environmental samples over the past decades. This is the most common approach to determine microbial diversity. However, PCR-based analyses are inherently biased. An additional disadvantage of 16S rRNA gene analysis is the varying number of gene copies between taxa. This results in an overestimation of microorganisms containing a high number and an underestimation of those containing a low number of 16S rRNA genes (Sheridan *et al.*, 2003).
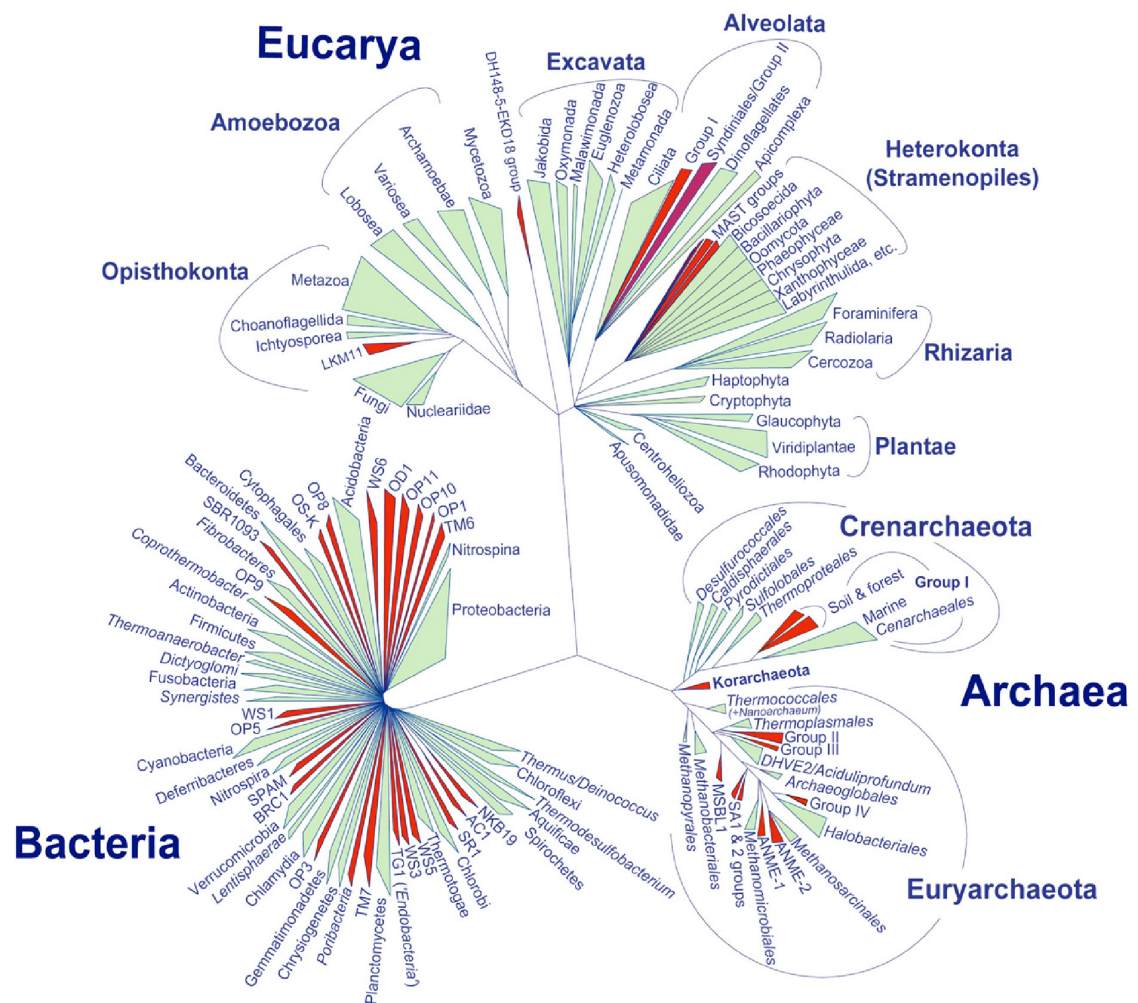
Figure 14. Schematic tree of life. Green triangles represent phyla for which at least one member has been cultivated and/or properly described; red triangles represent candidate divisions or highly divergent lineages without cultivated/described species. The tree is simplified, only a fraction of known eukaryotic phyla are depicted, and, in the case of bacteria, only phyla and candidate divisions found by at least three classification systems, as acknowledged in greengenes (http://greengenes.lbl.gov) (Lopez-Garcia and Moreira, 2008).

As 16S rRNA gene analysis relies on a single gene in the genome, it represents an incomplete system. For example, in case rRNA genes were frequently transferred horizontally between organisms this phylogenetic marker would not be characteristic for the evolutionary roots of the organism. Since the 16S rRNA gene is highly conserved, important phenotypic and ecological differentiation may be linked with identical 16S rRNA genes (Konstantinidis and Tiedje, 2007).

Venter *et al.* (2004) attempted to circumvent the disadvantages attached to the traditional 16S rDNA analysis by evaluating different phylogenetic markers, such as RecA/RadA, heat shock protein 70, elongation factor Tu, and elongation factor G in their shotgun sequencing-derived dataset of the Sargasso Sea. The authors concluded that the description deduced from additional phylogenetic markers was qualitatively similar to that based on analysis of 16S rDNA. However, the quantitative description was significantly different for certain taxonomic groups (Venter *et al.*, 2004). Manichanh *et al.* (2008) compared results of a 16S rDNA clone library with those derived from analysis of random sequence reads (RSR). The authors proposed that phylogenetic evaluation of a RSR-dataset is a cost-effective alternative to the traditional 16S rRNA-based approach for characterization of the microbial diversity. Although evaluation of the 16S rDNA clone library lead to more precise results, RSRs provide a reliable estimate of the diversity present in a given habitat (Manichanh *et al.*, 2008). Recently, this metagenomic approach for establishing the phylogenetic diversity of various environments was employed in several projects, such as characterization of an acid mine biofilm (Tyson *et al.*, 2004), seawater samples (DeLong *et al.*, 2006; Venter *et al.*, 2004), deep-sea sediment (Hallam *et al.*, 2004), soil and whale falls (Tringe and Rubin, 2005). All of these studies employed Sanger sequencing, which depends on cloning, fluorescent dideoxynucleotides, and capillary electrophoresis (Sanger *et al.*, 1977). In 2005, a novel parallel sequencing system based on an emulsion method for DNA amplification was introduced. Margulies *et al.* (2005) developed an instrument for sequencing by DNA synthesis using a pyrophosphate-based sequencing ('pyrosequencing') protocol for picolitre-scale volumes. The principle of pyrosequencing is depicted in Fig. 15. According to the authors, 25 million bases at an accuracy of at least 99% can be sequenced in a single four-hour run (Margulies *et al.*, 2005). The present drawbacks of this approach are a short read length and a slightly higher error rate than Sanger sequencing. This is due to inaccurate sequencing in homopolymer regions. In October 2008, a new GS FLX Titanium series kit and the respective software have been launched. According to the manufacturer, an improved read length of 400 bp and a five-fold increase in throughput of up to 600 million bp per run is provided (www.roche-applied-science.com). The most important advantages of

pyrosequencing for metagenomic analyses is a 10-fold cost reduction per base pair compared to Sanger sequencing and the elimination of time-consuming cloning. It is expected that pyrosequencing will replace traditional Sanger sequencing for microbial community analyses (Liu *et al.*, 2007).

In the present study, pyrosequencing was employed for assessment of the phylogenetic diversity of the glacier ice. In addition, a 16S rDNA clone library was constructed and analyzed, a DGGE analysis was carried out and microorganisms from glacier ice were cultivated and phylogenetically classified.
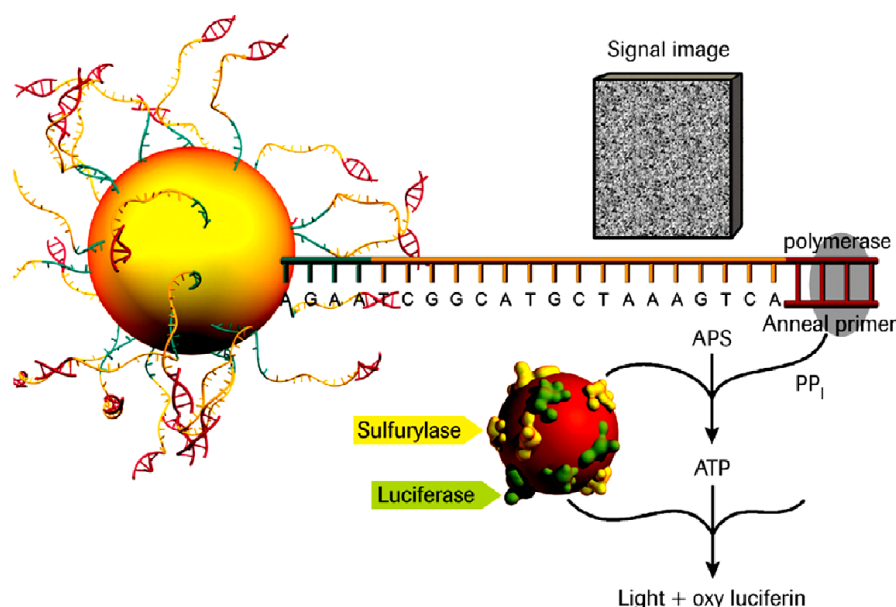
Figure 15. Scheme of the pyrosequencing technique. A DNA capture bead (left) containing millions of copies of a single clonal fragment is shown. Simultaneous sequencing is processed in hundreds of thousands of picoliter-size wells. The four nucleotides are added stepwise to the immobilized template hybridized to a primer. The pyrophosphate released in the DNA polymerase-catalyzed reaction is detected by ATP-sulfurylase- and luciferase-catalyzed reactions. The height of the signal is proportional to the number of bases, which have been incorporated (www.roche-applied-science.com).

4.1.2   Phylogenetic diversity

The general comparison of the bacterial community derived from the Northern Schneeferner with the results of previous studies of glacier ice and other permanently cold habitats from geographically different regions revealed similarities (Foght *et al.*, 2004; Mosier *et al.*, 2007; Segawa *et al.*, 2005; Skidmore *et al.*, 2005). The major phylogenetic groups established in all studies were identical, but the abundance of the respective groups differed. The Proteobacteria, Bacteroidetes, and Actinobacteria represented the dominant phyla in glacier ice from the Northern Schneeferner (Fig. 6; see 3.1). These groups were also abundant in ice derived from several other glaciers located in Canada (Bhatia *et al.*, 2006; Dancer *et al.*, 1997), Tibet (Christner *et al.*, 2003; Zhang *et al.*, 2008), or Greenland (Miteva *et al.*, 2004; Sheridan *et al.*, 2003). In addition to glacial ice, these dominant phylogenetic groups were also found in other permanently cold habitats, such as subglacial habitats (Cheng and Foght, 2007; Foght *et al.*, 2004; Skidmore *et al.*, 2005), snow (Segawa *et al.*, 2005), or an Antarctic lake (Mosier *et al.*, 2007). However, only in few of these studies, the composition of glacial ice microbial communities has been analyzed by molecular methods, e. g., Xiang *et al.* (2004) analyzed a sample of the Guliya ice core (Tibet) by 16S rRNA PCR. The majority of the studies focused on isolation of organisms from glacier ice. Most of the previously isolated glacial bacteria were assigned to the Gram-positive bacteria followed by the Proteobacteria (Miteva *et al.*, 2004; Xiang *et al.*, 2005; Zhang *et al.*, 2008). In this study, bacteria from glacial ice of the Northern Schneeferner were also cultivated and phylogenetically characterized (see 3.1.5). Analysis of the 16S rDNA revealed that all isolated organisms belonged to the three most abundant phyla established by the molecular approaches (Table 9). A molecular analysis by Mosier *et al.* (2007) revealed that Actinobacteria or Proteobacteria were the dominant phyla in Lake Vida, Antarctica, depending on the depth of the sampling site. The second most frequently recovered phylogenetic group was the phylum Bacteroidetes. This is in accordance with the microbial community established in this study (see 3.1). The results of the 16S rDNA analysis and evaluation of the pyrosequencing-derived dataset were supported by DGGE

analysis (Fig. 9, see 3.1.4). Only representatives of the phyla Proteobacteria (mainly, Betaproteobacteria) and Bacteroidetes were detected by DGGE (Table 7).

In this study, the class Betaproteobacteria was the most abundant phylogenetic group (Fig. 6). This corresponds to studies of glacier ice (Foght *et al.*, 2004), subglacial habitats (Cheng and Foght, 2007; Skidmore *et al.*, 2005), and mountain snow (Segawa *et al.*, 2005). Within this class, the family Comamonadaceae includes many genera, which have been found in polar and alpine environments. This indicated that these organisms are important members of glacial bacterial communities (Foght *et al.*, 2004; Skidmore *et al.*, 2005). All dominant phyla reported here are known to be widely spread in cold environments. Many of the nearest neighbors of 16S rRNA gene sequences derived from the Northern Schneeferner glacier were detected in other cold environments, such as glaciers, Antarctica, and snow. In addition, many sequences were related to genera such as *Polaromonas, Sphingomonas, Cryobacterium,* and *Stenotrophomonas* (Fig. 7), which have been found in other permanently cold environments (Christner, 2000; Dancer *et al.*, 1997; Foght *et al.*, 2004; Mikucki and Priscu, 2007; Miteva *et al.*, 2004; Xiang *et al.*, 2005; Zhang *et al.*, 2008). These results support the perception that related phylotypes exist in geographically diverse cold environments because of similar strategies for survival and remaining active at low temperatures (Priscu, 2004). Culture-dependent methods revealed that Gram-positive bacteria also represent a large part of the prokaryotes in glacial ice (Miteva *et al.*, 2004; Xiang *et al.*, 2005; Zhang *et al.*, 2008). In this study, Actinobacteria were found by analysis of the 16S rDNA clone library derived from the Northern Schneeferner (Fig. 6; see 3.1.1). Phylogenetic analysis of the pyrosequencing-derived dataset dataset revealed the additional presence of Firmicutes (1 to 2% of all classified sequences) (Fig. 6; see 3.1.2). Ancient ice cores are often dominated by sporulating Gram-positive bacteria (Mosier *et al.*, 2007). However, in previous studies of a Greenland glacier and the perennial ice cover of an Antarctic lake Actinobacteria accounted for a large part of the determined phylotypes, but only few representatives of endospore-forming bacteria were reported (Miteva *et al.*, 2004; Mosier *et al.*, 2007; Sheridan *et al.*, 2003). It has been shown that Actinobacteria have a higher ability to survive in old permafrost than low GC

Gram-positive and Gram-negative bacteria (Mosier *et al.*, 2007; Willerslev *et al.*, 2004a). Willerslev *et al.* (2004a) suggested that Actinobacteria develop resting forms with low metabolic activity.

Surprisingly, no 16S rRNA gene amplification product was obtained by using several archaeal primer pairs and/or various reaction conditions. Archaea are ubiquitous extremophile microorganisms, present in most terrestrial and aquatic environments. According to Cavicchioli (2006), the largest proportion and greatest diversity of Archaea occurs in cold environments, but cold-adapted Archaea have been underestimated and have been the subject of only few studies. In general, Archaea are present in cold environments, but were not found in any other published study of glacial or subglacial ice. Sheridan *et al.* (2003), Christner *et al.* (2001), and Skidmore *et al.* (2005) employed archaeal primers in their 16S rRNA analyses of glacial or subglacial ice, but no or no unambiguous PCR was recovered. In contrast, Archaea were detected in young, seasonal Antarctic and Arctic sea ice (Junge *et al.*, 2004). In polar and other cold environments Bacteria dominate and are present in a larger diversity than Archaea (Deming, 2002). Exceptions are the cold interior and deep waters of the ocean. Mosier *et al.* (2007) suggested that the lack of Archaea in ancient glacial ice is caused by outcompetition by Bacteria or by a high susceptibility to death and decay of Archaea in ice.

In conclusion, this study gives a comprehensive description of the bacterial diversity of a glacier in the German Alps for the first time. The Proteobacteria (mainly, Betaproteobacteria), Bacteroidetes, and Actinobacteria were established to be the dominating phylogenetic groups in the Northern Schneeferner. In contrast to previous investigations of glacier ice, a broad variety of different approaches for phylogenetic assessment including analysis of a pyrosequencing-derived dataset was employed. Thus, this study characterizes the microbial community structure of a low-temperature environment thereby providing insight into a habitat, which can be regarded as analogous to frozen extraterrestrial habitats.

4.1.3   Comparison of different approaches for phylogenetic assessment

In general, the different methods employed for the assessment of the microbial diversity in ice from the Northern Schneeferner yielded identical results with respect to the dominant phyla (Fig. 6; see 3.1.3). Eleven bacterial phylogenetic groups were detected by analysis of the 16S rDNA clone library (see 3.1.1). The rarefaction analysis, which was conducted for estimation of the bacterial richness, confirmed that almost all of the phyla present in the Northern Schneeferner were recovered (Fig. 5 and Table 6). Nevertheless, taxonomic analyses of the pyrosequencing-derived dataset revealed the existence of at least eleven additional phyla, but these phyla were only represented by less than 1% of all classified sequences (see 3.1.2.). The concept of calculating operational taxonomic units (OTUs) to measure the microbial richness of an environmental sample emerged from the attempt to link traditional taxonomy with modern 16S rRNA gene sequencing (Schloss and Handelsman, 2005). OTUs have become essential estimates for diversity of microbial species. Defining species based on sequence similarity is the accepted standard to analyze uncultured microorganisms, although sequence similarity does not necessarily correspond to functional conservation (Venter *et al.*, 2004).

The comparison of the results obtained by CARMA analysis with those of the 16S rDNA sequence analyses exhibited differences in the abundance of phylogenetic groups, i. e., CARMA analysis revealed a high representation of Alphaproteobacteria and lower representation of Bacteroidetes than 16S rRNA analysis (Fig. 6; see 3.1.3). These differences can be explained by the method used for taxonomic assignment. Traditionally, the phylogenetic profile of an environmental habitat is explored by analysis of conserved 16S rRNA genes, whereas CARMA assigns phylogenetic groups based on conserved Pfam domain and protein families. According to Krause *et al.* (2008), conserved gene fragments as short as 27 amino acids can be precisely classified with an average specificity of 97% (superkingdom level) to 93% (order level). The average sensitivity ranges from 84% (superkingdom level) to 61% (order level). However, the accuracy depends on the grade of representation of taxonomic groups in the Pfam database. In this study, 14% of the metagenomic

pyrosequencing-derived sequences were assigned to a phylogenetic group (Fig. 8; see 3.1.2). This result is consistent with the percentage of classified sequences of a synthetic metagenome (Krause *et al.*, 2008), but higher than that of other metagenomes (Biddle *et al.*, 2008; Edwards *et al.*, 2006; Gill *et al.*, 2006). Based on the high similarity of the results derived by the different employed taxonomic approaches, it can be concluded that pyrosequencing-derived sequences are equally suitable as traditional methods based on 16S rRNA for exploring the phylogenetic profile of an environment.

4.2   Functional analysis of the glacier ice metagenome

In order to understand metabolic processes and interactions of microbial populations, components of the community and their associated functions have to be identified. Traditionally, microbial functions and relationships were determined by observation of microbial growth and physiological traits of individual microorganisms. This approach underestimates the functional capability of microorganisms, since the complex biotic and abiotic environment cannot be exactly reproduced. The most effective way to understand the functional potential of a given habitat is the identification of functional genes. Metagenomics provides tools for unraveling the interactions between microorganisms and their environment and for assigning ecological functions and roles to microbial communities (Sjöling, 2008). The high yield and low cost of pyrosequencing allows a fast identification of key processes of a given environment (Edwards *et al.*, 2006). In this way, specific traits of the microbial community, such as the primary energy sources used and the dominant biosynthetic pathways can be defined.

To date, assessment of the functional diversity of few habitats has been analyzed by pyrosequencing. These habitats include an obesity-associated gut microbiome (Turnbaugh *et al.*, 2006), the Soudan mine (Edwards *et al.*, 2006), a coral-associated microbial community (Wegley *et al.*, 2007), a comparison of nine biomes (Dinsdale *et al.*, 2008), ocean surface waters (Frias-Lopez *et al.*, 2008), and the Peru Margin subseafloor (Biddle *et al.*, 2008). These studies

evaluated datasets derived by pyrosequencing using a GS 20 sequencer, which yields read lengths of approximately 100-120 bases. In the present study, the second generation of the pyrosequencing platform, GS FLX, was used, which yields average read lengths of 250 bp. The employment of a longer read length resulted in a classification which is more reliable than that of the above-mentioned studies.

4.2.1   Metabolic potential of the Northern Schneeferner

The metabolic potential of glacier ice was determined by comparing the pyrosequencing-derived dataset to higher-order cellular processes (KEGG pathways) and functional categories (COG) (Fig. 10, see 3.2.1). The classification of metabolic genes in KEGG categories showed that genes involved in nitrogen metabolism and utilization of C_1 compounds were highly represented in the glacial microbial community (Table 10). A high percentage of sequences was associated with degradation of C_1 substrates via the serine pathway. Together with the numerous genes related to biodegradation of xenobiotics and different biopolymers (supplementary Dataset S2), a high metabolic versatility of the microorganisms present in glacial ice was indicated. This might be a result of survival in a habitat with low nutrient concentrations. Correspondingly, genes characteristic for autotrophic bacteria such as genes involved in carbon fixation via the Calvin cycle, the reductive carboxylate cycle, or the serine pathway were also detected (Table 10).

To date, little is known about the dominant metabolic functions in permanently frozen ecosystems. Stibal *et al.* (2008) proposed that most of the organic matter present in glacial ice originates from allochthonous airborne material and atmospheric deposition on the glacier surface. An additional source of nutrients is glacial flour, which is generated by glacial erosion and abrasion of solid material (Stibal *et al.*, 2008). Bottrell and Tranter (2002) provided geochemical evidence for microbial oxidation of metal sulfides in glacier flour of oxic and anoxic glacier bed environments. In glacial ice of the Northern Schneeferner, no evidence for the presence of sulfide/sulfite oxidizing bacteria was detected

(Table 10). Hodson *et al.* (2005) reported the presence of inorganic and organic forms of nitrogen and phosphorus for two glaciers in Arctic Svalbard. The authors proposed the occurrence of NH_4^+ assimilation and nitrification on the glacier surface and denitrification and sulfate reduction in subglacial environments. In this study, no evidence for nitrification or respiratory sulfate reduction was detected, but dissimilatory and assimilatory nitrate/nitrite reduction was found (Table 10). In addition, genes involved in assimilatory sulfate reduction were detected, indicating the presence of anorganic sulfur sources in glacier ice as suggested by Bottrell and Tranter (2002).

Based on the central metabolic pathways, it can be concluded that the microorganisms present in the upper layer of the Northern Schneeferner glacier (up to 0.5 m) are predominantly aerobic or facultative aerobic bacteria. In addition, strictly anaerobic bacteria such as acetogenic, methanogenic, or sulfate/sulfur reducing bacteria were not established by phylogenetic analyses. The presence of genes homologous to denitrifying enzymes revealed the alternative use of nitrate/nitrite as electron acceptor in the absence of oxygen (Table 10). This is in accordance with the vast majority of Proteobacteria detected in the ice of the Northern Schneeferner (Fig 6). Most denitrifying bacteria belong to the Proteobacteria, which are known to be metabolically versatile facultative aerobic bacteria. In accordance with the low number of photosynthetic bacteria determined by phylogenetic analyses, the amount of genes related to photosynthesis was low.

In this study, the genetic diversity and metabolic capacity of a microbial community derived from a permanently frozen habitat has been elucidated for the first time by evaluation of a large pyrosequencing-derived dataset. A wide metabolic diversity was established, including a large degradative capacity and the ability to assimilate inorganic and organic nitrogen and sulfur sources. The ability to grow under anaerobic conditions by utilizing nitrate/nitrite as electron acceptor was also shown.

4.2.2   Characteristics for a psychrophilic lifestyle

Survival of bacteria at low temperatures arouses scientific interest because of several reasons. Apart from unraveling life in icy habitats on Earth and extraterrestrial environments knowledge about the psychrophilic lifestyle is useful in controlling pathogenic bacteria in cold-stored food material. Cold adaptation of bacteria involves the modification of several specific traits of the cellular biochemistry, such as enzyme functionality and stability and cell membrane fluidity (Bowman, 2008). Genomic data provides essential information of the genetic characteristics of microorganisms. Insight in cellular metabolism and functionality can be gained based on annotation data. In the pyrosequencing-derived dataset of glacial ice several enzymes and compounds, which are associated with adaptation to low temperatures were identified (see 3.2.2). Examples for these are antioxidative enzymes, cryoprotectants, desaturases, and peptidyl-prolyl cis-trans isomerases (Tables 10 and 11, and supplementary Dataset S2).

In order to maintain homeoviscosity of the cell membrane, adaptation to low temperatures includes modification of the membrane composition. An increased membrane viscosity reduces permeability, which inhibits the capacity of a cell to absorb and transport nutrients for growth (Bowman, 2008). This is critical under low temperatures due to reduction of enzymatic catalytic efficiency. In general, temperature-dependent membrane modification comprises replacement of membrane lipids by lipids with reduced melting temperatures (D'Amico *et al.*, 2006). Depending on the organism these modifications include reduction of fatty acid acyl chain length, increased unsaturation of fatty acid chains, and an increased shift from iso-branched to anteiso-branched fatty acids. The production of monounsaturated and polyunsaturated fatty acids (PUFAs) has been considered a classic feature of psychrophily (Bowman, 2008). Therefore, desaturases which convert saturated fatty acids into unsaturated fatty acids are an essential characteristic of adaptation to a low-temperature environment (Bowman, 2008; D'Amico *et al.*, 2006). Evidence for the presence of a large number of desaturases in glacier ice has been provided in this study (Table 11).

Many of the isolated bacteria from the ice of the Northern Schneeferner were pigmented (Table 8). Moreover, sequences homologous to genes associated with carotenoid production were identified in the pyrosequencing-derived dataset (supplementary Dataset S2). This is in accordance with characteristics of previously isolated organisms from glacial, subglacial, and Arctic habitats (Amato *et al.*, 2007; Cheng and Foght, 2007; Foght *et al.*, 2004; Miteva *et al.*, 2004; Zhang *et al.*, 2008). Carotenoid pigments play an important role in adaptation to low temperature (Chattopadhyay, 2006). In Antarctic prokaryotes carotenoids were found to be associated with the cell membranes. In response to increased synthesis of membrane-fluidizing fatty acids, polar carotenoids seem to aid in rigidifying the membrane to counterbalance the effects of fatty acids (Chattopadhyay, 2006). Thus, the modulation of membrane fluidity in psychrophilic prokaryotes is achieved by change of fatty acid composition and by changing the levels of polar and non-polar carotenoids (Bolter, 2004). In addition, pigmentation protects against UV radiation and is involved in detoxification of superoxide and free radicals (Ponder *et al.*, 2005; Zhang *et al.*, 2008).

Cellular osmoprotection and cryoprotection can be achieved by uptake of polyols, sugars, amino acids, amino acid derivatives, betaine, and carnithine depending on the organism (Bowman, 2008). Evidence for the presence of these compounds was found in glacier ice (Table 11). The compounds can be accumulated within the cytoplasm to high levels without disturbing cellular functions. By interacting with the hydration sphere around macromolecules these compounds stabilize ionic interactions when the osmotic pressure changes. Although betaine is a well-known bacterial cryoprotectant its cryoprotective mechanism is not completely understood. It is indicated that betaine prevents cold-induced aggregation of cellular proteins (Chattopadhyay, 2002). It possesses chaperoning effects, which inhibits denaturation of cellular proteins at low temperatures. Furthermore, betaine seems to be involved in maintenance of membrane fluidity by promoting the synthesis of specific fatty acids (Chattopadhyay, 2002). Compatible solutes, such as glycine and glutamate are important osmoprotectants and balance the osmotic disturbance. Although their functions at low temperatures are largely unknown, it has been proposed that they serve similar functions as betaine (Wemekamp-Kamphuis *et*

al., 2004). Moreover, Panoff *et al.* (2000) suggested, that adaptation to cold stress could be obtained by cross-adaptation with osmolytes that are known to be cryoprotectants.

As expected, an enhanced antioxidant capacity was detected in glacial ice (Table 10 and 11). In aerobic metabolism, the ability to respond to reactive oxygen species is an important function. In addition, PUFAs which are present in the membranes of psychrophiles are the cellular molecules with the highest susceptibility to oxidative damage (Nishida *et al.*, 2006). Reactive oxygen species can directly attack PUFAs in membranes and initiate lipid peroxidation. One of the main effects of lipid peroxidation is a reduction in membrane fluidity, which changes membrane properties and can disrupt membrane-bound proteins significantly (Cabiscol *et al.*, 2000).

Additional indications for a psychrophilic lifestyle were unraveled in ice from the Northern Schneeferner. A high number of sequences of the glacial pyrosequencing-derived dataset were similar to peptidyl-prolyl cis-trans isomerases (Table 11), which were recently found to facilitate protein folding at low temperatures (D'Amico *et al.*, 2006). Moreover, several sequences were homologous to genes involved in the type II secretion system (see supplementary Dataset S2). Type II exoproteins were demonstrated to promote growth at low temperatures, but the mechanism is still not solved (Soderberg *et al.*, 2004).

Several enzymes and compounds essential for a psychrophilic lifestyle in glacial ice were detected in the present study. Thus, further evidence for the presence of a viable psychrophilic and/or psychrotolerant microbial community was provided. The reported results will contribute to elucidate microbial activity and strategies allowing cell survival in subzero temperature ecosystems. This might also contribute to unravel prerequisites for life in frozen extraterrestrial habitats.

4.3   Metagenomic bioprospecting

The genetic diversity present in different ecological niches is a valuable source of novel agricultural, pharmaceutical, and industrial products. In the majority of metagenomic studies soil samples have been analyzed and exploited (Sjöling, 2008). Only few enzymes, which are active at low temperatures have been isolated and characterized by function-driven screening of metagenomes. These include nitrilases from the deep sea and polar regions (Robertson *et al.*, 2004), beta-lactamases from cold-seep sediments (Song *et al.*, 2005), a cold-active xylanase of waste water (Lee *et al.*, 2006a), and a lipase of Baltic Sea marine sediment (Hardeman and Sjöling, 2007). Due to their high diversity and the low level of current exploitation cold habitats are predicted to be a rich resource for future discovery of commercially interesting cold-active enzymes (Sjöling, 2008). In the present study, glacier ice has been utilized for the construction of metagenomic libraries.

The search for novel biocatalysts by screening of metagenomic libraries is usually initiated by isolation of environmental DNA. A sufficient amount of high-molecular weight DNA of high quality is required for library construction. This DNA should be representative of the microbial community in the investigated habitat (Daniel, 2004, 2005; Schmeisser *et al.*, 2007; Streit and Schmitz, 2004). In general, two strategies for DNA extraction can be employed: direct lysis of cells in the environmental sample with subsequent separation of the DNA from the matrix and cell debris, or separation of the cells from the matrix followed by cell lysis (Daniel, 2005). The extracted environmental DNA is subsequently cloned into metagenomic libraries. The resulting libraries can be divided into two groups regarding average insert sizes: small-insert libraries in plasmid vectors (<15 kb) and large-insert libraries in cosmid, fosmid (20 to 40 kb) or BAC vectors (>40 kb). In most cases, *Escherichia coli* is chosen as host for construction and maintenance of metagenomic libraries (Daniel, 2004, 2005). The choice of the vector system depends on the aim of the study. Considerations have to embrace the quality of the isolated DNA, the intended average insert size, the copy number of the vector, and finally the screening

strategy. Important advantages of small-insert libraries include the detection of weakly-expressed foreign genes due to high copy number and the expression of foreign genes from vector promoters. Furthermore, the construction of small-insert libraries is simpler than that of large-insert libraries. Entire operons and large gene clusters can only be detected in large-insert clone libraries. In addition, a smaller number of clones have to be screened to obtain positives (Daniel, 2005).

In order to combine the advantages of small-insert and large-insert libraries, metagenomic plasmid libraries and fosmid libraries were constructed from glacier ice (see 3.3.1). To generate starting material for cloning into plasmids the glacial DNA was amplified by rolling circle amplification employing phi29 polymerase (Fig. 16). Phi29 DNA polymerase is a sensitive proofreading enzyme, which amplifies DNA up to 70 kb by strand-displacement. The enzyme increases the amount of DNA 10^4 to 10^6-fold (Dean *et al.*, 2001; Yokouchi *et al.*, 2006). Whole genome amplification (WGA) allows the amplification of very low amounts of genomic DNA.
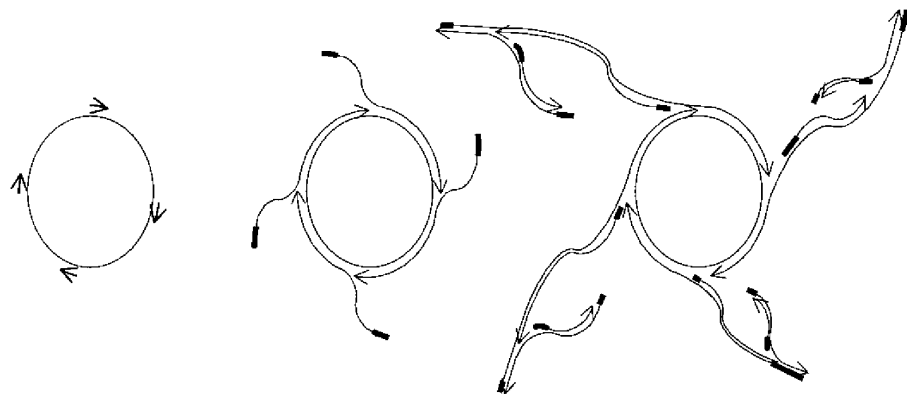
Figure 16. Scheme for whole genome amplification using phi29 polymerase. Oligonucleotide primers are hybridized to the template DNA. The 3' ends of the DNA strands are indicated by arrowheads to show the polarity of polymerization. Thickened lines indicate the location of the original primer sequences within the product strands. Extension of the primers results in displacement of already synthesized DNA strands. Secondary priming events can subsequently occur on the displaced product strands of the initial rolling circle amplification step (Dean *et al.*, 2001).

Recently, this technique has been employed for metagenomic studies of diverse environments, such as the Soudan mine (Edwards *et al.*, 2006), microbial populations in contaminated sediments (Abulencia *et al.*, 2006), methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain Aquifer (Erwin *et al.*, 2005), and a microbial community associated with scleratinian coral (Yokouchi *et al.*, 2006). However, WGA may result in an amplification bias depending on template accessibility and priming efficiency (Abulencia *et al.*, 2006). A major pitfall when amplifying small amounts of template DNA is 'background' DNA synthesis derived from contaminating traces of DNA or primer dimers (Zhang *et al.*, 2006). In order to minimize background DNA synthesis template glacial DNA was added to the WGA reaction in excess (Hutchison *et al.*, 2005). In contrast to plasmid libraries, glacial fosmid libraries were constructed without a prior amplification step.

The small-insert plasmid library comprised 230,000 clones with an average insert size of 4 kb and the large-insert fosmid library consisted of approximately 4,000 clones with an average insert size of 36 kb. Metagenomic fosmid/cosmid libraries constructed by other researchers contained from 267 cosmids (drinking water biofilm; Elend *et al.*, 2006) to 64,333 fosmids (tidal flat sediments; Lee *et al.*, 2006b). Other published plasmid libraries ranged from 2,727 plasmids (pondwater; Ranjan *et al.*, 2005) to 730,000 plasmids (soil; Henne *et al.*, 2000). In this study, the constructed libraries harbored approximately 1.07 Gb of cloned glacial DNA. Although larger metagenomic libraries have been reported, such as a fosmid library from tidal flat sediments (13.5 Gb), the metagenomic libraries constructed in this study contain the highest amount of DNA isolated from an extreme environment. Therefore, it can be concluded that the techniques for metagenomic library construction applied in the present study were suitable for DNA derived from this subzero temperature environment.

4.3.1   Identification of putative *polA* genes

Several high-throughput and sensitive screening methods have been used to recover novel biocatalysts from metagenomic libraries. In general, two different types can be differentiated: function-based and sequence-based screenings. For the sequence-driven screening, PCR or hybridization using target-specific probes is usually employed. However, the applicability of this approach is limited to the identification of new members of known gene families, since primers and probes are derived from conserved regions of known genes (Daniel, 2005; Handelsman, 2004). A function-based screening, which was employed in this study (see 3.3.2), relies on the heterologous expression of genes and the detection of gene products by identification of a specific function, such as enzymatic or antibiotic activity. In this way, it is ensured that full-length genes or gene clusters are identified and entirely novel genes encoding new biocatalysts can be detected. Furthermore, genes that code for enzymes with very low or no identities to known gene products can be isolated (Daniel, 2004, 2005; Handelsman, 2004; Schmeisser *et al.*, 2007). In order to detect enzymatic activities of individual clones, screening techniques often include the incorporation of chemical dyes and insoluble or chromophore-bearing derivatives of enzyme substrates in the growth medium. For example, recombinant *E. coli* clones with proteolytic or lipolytic activity can be identified on agar plates containing skimmed milk or triolein, respectively (Daniel, 2005). A different screening approach requires heterologous complementation of host strains or mutants of host strains for growth under selective conditions. This approach has been successfully used for the isolation of genes encoding Na^+/H^+ antiporters (Majernik *et al.*, 2001), antibiotic resistance (Riesenfeld *et al.*, 2004a), enzymes involved in poly-3-hydroxybutyrate metabolism (Wang *et al.*, 2006), or naphthalene dioxygenases (Ono *et al.*, 2007). In the present study, a function-based screening approach based on complementation of the cold-sensitive lethal mutation *fcsA29,* which is located in the *polA* gene of *E. coli* was chosen (Nagano *et al.*, 1999). Usually, screening of metagenomic libraries containing a large number of clones is laborious due to a low gene-detection frequency during function-driven screening (Daniel, 2005). However, the

screening technique employed in this study allowed a rapid identification of genes conferring DNA polymerase activity. In order to identify one positive clone 1,000 clones (plasmid library) and 100 clones (fosmid library) needed to be screened. This corresponds to approximately one positive clone per 3.6 – 4.0 Mb of cloned glacial DNA. Furthermore, 16 out of the 18 initially positive clones conferred a stable phenotype under selectice growth conditions and harbored a *polA* gene or derivative of this gene. Thus, almost no false positive clones were recorded. This is caused by the requirement of heterologous complementation for growth under selective conditions. The high percentage of positive clones compared to other metagenomic studies (Jeon *et al.*, 2008a; Riesenfeld *et al.*, 2004a; Rondon *et al.*, 2000) revealed that the chosen approach is a powerful tool for rapid and reliable screening of metagenomic libraries.

4.3.2   Molecular characterization of putative *polA* genes

Replication of double-stranded DNA proceeds semi-discontinuous with being continuous on the leading strand and discontinuous on the lagging strand. DNA synthesis on the lagging strand is initiated from RNA primers of 8 to 12 nucleotides in length. It proceeds for approximately 100 to 2,000 nucleotides, thereby constituting the Okazaki fragment. DNA synthesis on the lagging strand stops in response to signals, such as the 5' end of an RNA primer (prokaryotes) or DNA flaps (eukaryotes). DNA polymerase I extends the 5' ends of the upstream Okazaki fragment and creates a substrate for DNA ligase (Bayliss *et al.*, 2005). DNA polymerases are grouped into the six families A, B, C, D, X, and Y (Ohmori *et al.*, 2001). DNA polymerase I encoded by *polA* belongs to family A and possesses three enzymatic activities: polymerase, 3'-5' exonuclease, and 5'-3' exonuclease. Each activity is located on a separate structural domain (Loh and Loeb, 2005). The polymerase activity of the enzyme drives nick translation, thereby producing displaced 5' ends, which serve as substrate for the 5'-3' exonuclease. The 5'-3' exonuclease is a structure-specific nuclease with specificity for the junction between a 5' overhang and a duplex region (Xu *et al.*,

1997). The 3'-5' exonuclease, or proofreading domain, removes polymerase errors. The incorporation of a mismatched nucleotide on a DNA strand interrupts further DNA synthesis. Thereby, the affinity of the polymerase for the template-primer is reduced and the delay allows editing of the mismatch by the 3'-5' exonuclease domain (Kukreti *et al.*, 2008).

For detection of putative *polA* genes in the glacial metagenomic libraries a complementation-based screening approach was employed (see 3.3.2). The cold-sensitive *E. coli* mutant, which was used as host for the screening carries a mutation in the 5'-3' exonuclease domain of the *polA* gene. All of the analyzed plasmids and the fosmid contained putative *polA* genes or derivatives of this gene (Tables 12 and 14; see 3.3.3). Five identified putative *polA* genes encoded all three typical domains for DNA polymerase I (pCS2, pCS5, pCS6, pCS8, fCS1). One plasmid harbored a putative *polA* gene, which lacked part of the C-terminal polymerase domain (pCS7) and three recovered plasmids contained shorter versions of the *polA* gene (pCS1, pCS3, pCS4). All of the analyzed plasmids and the fosmid contained the 5'-3' exonuclease domain (Table 13 and Fig. 11). No differences in the phenotype of all recombinant strains containing the *polA* harboring plasmids were evident. A similar growth rate was established for all recombinant strains containing pCS1 to pCS8 (Table 15). This indicated that the 5'-3' exonuclease domain alone was sufficient for complementation of the *E. coli* mutant as not all plasmids encoded all three domains of the *polA* gene (Fig. 11). Correspondingly, Xu *et al.* (1997) suggested that the active site of the 5'-3' exonuclease domain operates fairly autonomously. It does not require assistance in binding or catalysis from the remaining polymerase molecule (Xu *et al.*, 1997).

The deduced amino acid sequences of the identified putative *polA* genes showed a low degree of similarity (35-82%) to DNA polymerases from other organisms (Table 12 and 14). In general, *polA* genes are conserved and can be used as a phylogenetic marker gene (Croan *et al.*, 1997). Thus, the low degree of similarity to known DNA polymerases in most cases indicated that the *polA* genes were recovered from uncharacterized and novel microorganisms. A high diversity of the DNA stored in the constructed metagenomic libraries was evident. In most cases, the most similar DNA polymerases were derived from organisms, which were affiliated to the Proteobacteria, Bacteroidetes, or

Actinobacteria (Table 14). These groups were the dominating phyla in ice from the Northern Schneeferner (Fig. 6). Most of the identified ORFs on the fosmid fCS1 exhibited highest identity to *Rhodoferax* or *Polaromonas* (Table 12), which belong to the most commonly identified genera from glacier ice. Members of this family are well-known to occur in cold environments (Miteva, 2008).

Family A DNA polymerases share the six conserved motifs A, B, C, 1, 2, and 6, which are all located in the polymerase domain of the enzyme (Loh and Loeb, 2005). These motifs constitute the active site and contact the substrate. Motifs A, B, and C are most highly conserved, and motif A and C are present in all known DNA and RNA polymerases. Motifs 1, 2, and 6 vary in amino acid sequence, but are conserved structurally (Loh and Loeb, 2005). An overview of the six conserved motifs is given in Table 17.

Table 16. Conserved motifs of family A DNA polymerases (Loh and Loeb, 2005).

Motif	Conservation	Contacts to substrate
A	Family A, B, X, Y, and reverse transcriptase	DNA primer and Mg^{2+}
B	Family A, B, and X	Incoming nucleotide and template base
C	Family A, B, X, Y, and reverse transcriptase	DNA primer terminus and Mg^{2+}
1	Family A	DNA template and primer backbone
2	Family A	DNA minor groove and template backbone
6	Family A	DNA minor groove and template backbone

It can be concluded from the available structural and sequence data that the polymerase domain is the most conserved part of all polymerases (Joyce and Steitz, 1994). The conserved motifs A, B, C, 1, 2, and 6 were present in all identified polymerase domains encoded on the plasmids and the fosmid from the metagenomic glacier ice libraries (Fig. 17). Slight modifications in the deduced amino acid sequences were evident. However, the amino acid residues, which are conserved in all members of family A polymerases (Loh and Loeb, 2005) (highlighted in yellow in Fig. 17) were identical in all putative polymerase sequences identified in the glacial metagenomic libraries.

```
                    Motif 1                                    Motif 2
E.coli   578 FNLSSTKQLQTIL 590      660 HQAVTATGRLSSTDPNLQ 677
pCS5     577 FNVSSPKQLGEIL 589      657 HQAITATGRLSSTDPNLQ 674
pCS2     594 FNIASPQQLGKVL 606      677 NQAVVATGRLSSTNPNLQ 694
fCS1     608 FNLASPKQLGEVL 620      692 AQAVAVTGRLSSTNPNLQ 709
pCS6     590 FNVSSPKQVGEIL 602      674 NQVVAVTGRLSSDNPNLQ 691
pCS3     442 FNINSPKQMGEVL 454      525 VQTGAATGRMASQNPNLQ 542

                    Motif A                                    Motif B
E.coli   702 VSADYSQIELR 712        754 RRSAKAINFGLIY 766
pCS5     699 LAADYSQIELR 709        751 RRQAKAINFGLLY 763
pCS2     719 LSADYSQIELR 729        771 RRMAKTANFAILY 783
fCS1     734 VSADYSQIELR 744        886 RYKAKSVNFGIIY 898
pCS6     716 LSADYSQIELR 726        768 RYKSKSVNFGIIY 780
pCS3     567 VSIDYSQIELR 577        619 RRKAKVINFGILY 631

                    Motif 6                                    Motif C
E.coli   843 AINAPMQGTAADIIK 857    877 IMQVHDELVFE 887
pCS5     838 AINAPLQGSAADIIK 852    872 LLQVHDELVFE 882
pCS2     859 AVNSPIQGTAADMIK 873    893 ILQVHDELVFD 903
fCS1     875 AINSPIQGTAADMIK 889    909 LLQVHDELVFD 919
pCS6     857 AINAPIQGSAADMIK 871    891 LLQVHDELVFD 901
pCS3     712 AINAPVQGTAADAMR 726    746 LLQVHDELLFE 756
```

Figure 17. Alignment of the conserved motifs of DNA polymerase family A: motifs 1, 2, A, B, 6, and C encoded by the putative *polA* genes of pCS2, pCS3, pCS5, pCS6, and fCS1. The amino acid sequence of the conserved motifs of the *E. coli* DNA polymerase I (Riley *et al.*, 2006) was used as reference. Conserved residues of known members of family A are highlighted in yellow (Loh and Loeb, 2005).

The identified putative *polA* gene encoded by pCS3 lacked the central 3'-5' exonuclease domain of DNA polymerase I (Fig. 11 and Table 13). Correspondingly, the enzyme was 125 amino acids shorter than the DNA polymerase I of *E. coli*, which is the prototype for family A (Diaz *et al.*, 1992). Interestingly, this putative *polA* gene was most similar to a DNA polymerase of *Thermus thermophilus* (Table 14). The *Streptococcus pneumonia* and *Thermus thermophilus* polymerases are 51 and 94 amino acids shorter than the *E. coli* DNA polymerase I (Diaz *et al.*, 1992). These DNA polymerases also lack the proofreading activity encoded by the central domain (Diaz *et al.*, 1992). The absence of the proofreading activity and the shorter size of the protein are presumably due to several deletions in the 3'-5' exonuclease domain.

An alignment of all deduced amino acid sequences with the sequence of the *E. coli* DNA polymerase I revealed that the six conserved motifs A to F, which are characteristic for 5'-3' exonuclease domains of DNA polymerases I (Gutman and Minton, 1993) were present in all deduced gene products (Fig. 12). Based on the amino acid alignment of ten bacterial DNA polymerases and bacteriophage 5' nucleases Gutman and Minton (1993) identified 14 invariant amino acids, of which nine were carboxylate residues. Sequence comparisons revealed that prokaryotic *polA*-associated 5' nucleases, or flap endonucleases (FEN), share significant sequence similarities to the polymerase-independent 5' nucleases from several bacteriophages (Gutman and Minton, 1993). The prokaryotic 5' nucleases are also related to mammalian flap endonuclease-1 (FEN-1) proteins and structure-specific nucleases from yeast. FENs catalyze the hydrolysis of nucleic acids during DNA repair and replication. In double-stranded DNA these enzymes exhibit 5'-3' exonuclease activity but they can also cleave bifurcated nucleic acids, such as flap substrates (Feng *et al.*, 2004). Therefore, sequences of polymerase-dependent and –independent 5' nucleases from different bacteria and the eukaryotic FEN-1 enzyme were additionally included in the above-mentioned alignment (Amblar *et al.*, 2001). The alignment revealed that only six of the conserved 14 invariant residues are present in all prokaryotic and eukaryotic 5' nucleases (Amblar *et al.*, 2001). Except for the 5' nuclease encoded by pCS4, none of the 5'-3' exonuclease domains derived from the glacial metagenomic libraries exhibited differences in the 14 invariant amino acids present in prokaryotes. Surprisingly, the gene product encoded by pCS4 differed in four of these conserved amino acids, including one of the six residues, which have been established to be invariant in all prokaryotic and eukaryotic 5' nucleases (Fig. 12). The 5'-3' exonucleases must contain a very flexible active site in order to recognize different DNA substrates with different structures and to adopt different conformations. Thus, sequence differences of the active site reflect the various functional conformations that these sites can adopt (Amblar *et al.*, 2001).

It was proposed by several researchers that some of the highly conserved aspartate and glutamate residues are involved in metal binding at the active site of FENs (Amblar *et al.*, 2001; Feng *et al.*, 2004; Xu *et al.*, 1997). Divalent metal cofactors are required by many of the enzymes involved in key cellular

processes. Phosphatases, nucleases, and polymerases are well-characterized examples of essential metal ion-dependent phosphoryl transferases (Feng *et al.*, 2004). For prokaryotic 5'-3' exonucleases Amblar *et al.* (2001) suggested the presence of two metal binding sites: one divalent metal ion stabilizes the generated transition state and the other one positions the DNA correctly at the active site for cleavage. Although structural and mutational data have emerged from studies on various FENs, conflicting conclusions with respect to the function of the different residues were drawn (Feng *et al.*, 2004). Amblar *et al.* (2001) conducted analyses of the 5'-3' exonuclease domain of DNA polymerase I of *Streptococcus pneumonia* by introducing mutations at the carboxylate residues Asp^{10}, Glu^{88}, and Glu^{114}. According to the authors, Asp^{10} has a critical role in the exonuclease reaction. This residue is essential by activating a water molecule and producing the hydroxyl group required for nucleophilic attack. The residue Glu^{114} is less essential for the exonucleolytic reaction and seems to be involved in substrate binding. It was concluded that Glu^{88} possesses a nonessential function for the exonuclease activity. Furthermore, the results of the mutational studies indicated that none of these three residues is a metal-ligand at the active site (Amblar *et al.*, 2001).

In conclusion, the chosen function-driven screening strategy was an efficient way to identify the targeted DNA polymerase-encoding genes. Nine novel genes coding for complete DNA polymerases I or for DNA polymerase domains were identified. The complementation of the cold-sensitive *E. coli* mutant allowed a simple and rapid screening of both metagenomic libraries derived from glacial ice. Since almost no false positives were encountered, a high selectivity of the approach was evident. Sequence analysis of first metagenome-derived DNA polymerase-encoding genes revealed that all encode domains typical for DNA polymerases belonging to family A. Most of the protein sequences exhibited low similarities to DNA polymerases from a variety of different microorganisms. This demonstrated that a permanently frozen environment, such as glacier ice harbors a variety of unknown genes from yet uncharacterized organisms.

5   Summary

The largest part of the Earth's microbial biomass is stored in cold environments, which represent almost untapped reservoirs of novel species, processes, and genes. In this study, the first metagenomic survey of the metabolic and phylogenetic diversity of a glacial microbial community was conducted. In addition, the so far unexplored habitat glacier ice was exploited with respect to novel genes encoding DNA-modifying enzymes.

DNA (5 µg per kg) was isolated from glacial ice of the Northern Schneeferner, Germany. Pyrosequencing of this DNA yielded 1,076,539 reads (239.7 Mbp). The phylogenetic composition of the microbial community was assessed by evaluation of a pyrosequencing-derived dataset by using the CARMA algorithm and a comparison to the RDP II database. In addition, a 16S rDNA clone library was constructed, sequenced and analyzed. Furthermore, bacteria from the Northern Schneeferner were cultivated and phylogenetically analyzed. Phylogenetic trees were constructed and the bacterial richness and diversity of glacier ice was estimated. A total of 338 16S rRNA gene sequences represented 108 operational taxonomic units based on a >99% sequence identity cutoff. It was indicated that all phyla predicted to be present in glacial ice were recovered and a substantial fraction of the high bacterial species diversity was assessed. The Proteobacteria (mainly, Betaproteobacteria), Bacteroidetes, and Actinobacteria were the dominant phylogenetic groups. As expected for microorganisms living in a low-nutrient environment, a high metabolic versatility with respect to degradation of organic substrates was detected by comparison of the pyrosequencing-derived dataset to the KEGG and COG databases. The presence of autotrophic microorganisms was indicated by identification of genes typical for different ways of carbon fixation. In accordance with the results of the phylogenetic studies, in which mainly aerobic and facultative aerobic bacteria were detected, genes typical for central metabolism of aerobes were found. Nevertheless, the capability of growth under anaerobic conditions was indicated by genes involved in dissimilatory nitrate/nitrite reduction. Numerous characteristics for physiological and metabolic adaptations associated with a psychrophilic lifestyle, such as the

formation of cryoprotectants and maintenance of membrane fluidity by incorporation of unsaturated fatty acids were found.

Large-insert and small-insert metagenomic libraries in plasmids and fosmids, respectively, were constructed from the isolated glacial DNA. To generate starting material for cloning into plasmids the glacial DNA was amplified by employing whole genome amplification. The small-insert and large-insert libraries comprised 230,000 and 4,000 clones, respectively. The average insert size was 4 and 36 kb, respectively. Thus, both types of libraries harbored approximately 1.07 Gbp of glacial DNA. Subsequently, the libraries were screened for DNA polymerase-encoding genes by complementation of an *Escherichia coli polA* mutant. Nine novel genes encoding complete DNA polymerases I or typical domains of these proteins were recovered by employing this rapid screening approach. Although *polA* genes are highly conserved, the identified putative DNA polymerases displayed low sequence identities to known polymerases. Thus, a high diversity of the DNA stored in the constructed metagenomic libraries was evident. This indicated that the libraries derived from a permanently frozen habitat are a rich resource for the discovery of genes that originate from uncharacterized organisms. Furthermore, the phylogenetic and metabolic analysis of the glacial metagenome provided insight into microbial live in frozen habitats on Earth, thereby possibly shedding light onto microbial life in analogous extraterrestrial environments.

6   Zusammenfassung

Der größte Teil der mikrobiellen Biomasse der Erde befindet sich in kalten Gebieten, die ein fast unberührtes Reservoir für die Entdeckung neuer mikrobieller Arten, biologischer Prozesse und Gene darstellen. In dieser Arbeit wurde die erste Metagenomanalyse der metabolischen und phylogenetischen Diversität einer mikrobiellen Gemeinschaft aus Gletschereis durchgeführt. Darüber hinaus wurden Metagenombanken konstruiert und auf das Vorhandensein von Genen für DNA-modifizierende Enzyme durchmustert.

DNA (5 µg pro kg) wurde aus Gletschereis des Nördlichen Schneeferner, Deutschland, isoliert. Pyrosequenzierung dieser DNA erbrachte 1.076.539 Sequenzen (239,7 Mbp). Die phylogenetische Diversität der mikrobiellen Gemeinschaft wurde durch Analyse der Pyrosequenzierungsdaten mit der CARMA Software und einen Vergleich mit der RDP II Datenbank analysiert. Zusätzlich wurden 16S rRNA Gene amplifiziert, sequenziert und mit den Datenbanken der NCBI und RDP II verglichen. Phylogenetische Stammbäume wurden erstellt und die bakterielle Diversität in Gletschereis wurde statistisch berechnet. Zusätzlich wurden Bakterien aus Gletschereis isoliert und phylogenetisch analysiert. Insgesamt 338 16S rRNA-Gensequenzen wurden in 108 operational taxonomic units eingeteilt, wobei eine genetische Distanz von <1% zugrunde gelegt wurde. Es wurden alle Phyla und ein Großteil der bakteriellen Arten, die statistisch zu erwarten waren, bestimmt. Als dominante Phyla in Eis des Nördlichen Schneeferner wurden Proteobacteria (hauptsächlich Betaproteobacteria), Bacteroidetes und Actinobacteria nachgewiesen. Durch einen Vergleich der Sequenzdaten mit der KEGG Datenbank wurde eine hohe metabolische Vielseitigkeit der mikrobiellen Gemeinschaft im Hinblick auf den Abbau von organischen Substraten nachgewiesen. Dies ist typisch für Mikroorganismen, die in einem Habitat mit niedrigem Nährstoffgehalt leben. Einen Hinweis auf das Vorhandensein von autotrophen Mikroorganismen ergab die Identifikation von Genen, die charakteristisch für verschiedene Wege der Kohlenstofffixierung sind. In Einklang mit den Ergebnissen der phylogenetischen Untersuchungen, in denen hauptsächlich aerobe und fakultativ anaerobe Bakterien nachgewiesen wurden,

konnten durch die funktionelle Analyse typische Gene für den zentralen Metabolismus von aeroben Mikroorganismen identifiziert werden. Dennoch war die Fähigkeit des Wachstums unter anaeroben Bedingungen durch die Identifikation von Genen für eine dissimilatorische Nitrat/Nitritreduktion angezeigt. Zahlreiche Besonderheiten für physiologische und metabolische Anpassungen im Zusammenhang mit einem psychrophilen Lebenstil wurden nachgewiesen. Dazu gehören z. B. Gene für die Synthese von Frost- und Osmoseschutzmitteln (Betain, Glycin, Glutamat) und für Enzyme, die an der Aufrechterhaltung der Membranfluidität durch Synthese von ungesättigten Fettsäuren beteiligt sind.

Zusätzlich wurden Plasmid- und Fosmidgenbanken aus der isolierten Gletschereis-DNA konstruiert. Um eine größere Menge an klonierbarer DNA zu gewinnen, wurde die Gletschereis-DNA vor dem Klonieren in Plasmide durch 'whole genome amplification' vermehrt. Die Plasmid- und Fosmidgenbanken enthielten 230.000 bzw. 4.000 Klone mit durchschnittlichen Insertgrößen von 4 bzw. 36 kb. Insgesamt wurden ca. 1,07 Gbp Gletschereis-DNA kloniert. Anschließend wurden die Genbanken auf das Vorhandensein von Genen für DNA-Polymerasen durch Komplementation einer *Escherichia coli polA* Mutante durchmustert. Neun neuartige Gene, die für eine komplette DNA-Polymerase I oder charakteristische Domänen dieses Proteins kodieren, wurden durch Anwendung dieses schnellen Screeningverfahrens identifiziert. Obwohl *polA* Gene hochkonserviert sind, wiesen die identifizierten putativen DNA-Polymerasen eine geringe Sequenzidentität zu bekannten Polymerasen auf. Hierdurch war eine hohe Diversität der konstruierten Metagenombanken angezeigt.

Es wurde belegt, dass die Genbanken aus einem dauerhaft gefrorenem Standort eine reichhaltige Ressource für Gene aus bisher uncharakterisierten Organismen sind. Darüber hinaus lieferte die phylogenetische und metabolische Analyse des Gletschereismetagenoms Einblicke in die Lebensweise von Mikroorganismen an Tieftemperaturstandorten. Hierdurch könnten Hinweise zur Aufklärung von mikrobiellem Leben in extraterrestrischen Räumen gegeben werden.

7   References

Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W., Chang, S.H., Chang, H.W., Watson, D., Brodie, E.L., Hazen, T.C., and Keller, M. (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* **72**: 3291-3301.

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

Amato, P., Hennebelle, R., Magand, O., Sancelme, M., Delort, A.M., Barbante, C., Boutron, C., and Ferrari, C. (2007) Bacterial characterization of the snow cover at Spitzberg, Svalbard. *FEMS Microbiol Ecol* **59**: 255-264.

Amblar, M., de Lacoba, M.G., Corrales, M.A., and Lopez, P. (2001) Biochemical analysis of point mutations in the 5'-3' exonuclease of DNA polymerase I of *Streptococcus pneumoniae.* Functional and structural implications. *J Biol Chem* **276**: 19172-19181.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72**: 5734-5741.

Bairoch, A., Boeckmann, B., Ferro, S., and Gasteiger, E. (2004) Swiss-Prot: juggling between evolution and stability. *Brief Bioinform* **5**: 39-55.

Bayliss, C.D., Sweetman, W.A., and Moxon, E.R. (2005) Destabilization of tetranucleotide repeats in *Haemophilus influenzae* mutants lacking RnaseHI or the Klenow domain of PolI. *Nucleic Acids Res* **33**: 400-408.

Ben-Dov, E., Shapiro, O.H., Siboni, N., and Kushmaro, A. (2006) Advantage of using inosine at the 3' termini of 16S rRNA gene universal primers for the study of microbial diversity. *Appl Environ Microbiol* **72**: 6902-6906.

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

Bhatia, M., Sharp, M., and Foght, J. (2006) Distinct bacterial communities exist beneath a high Arctic polythermal glacier. *Appl Environ Microbiol* **72**: 5838-5845.

Biddle, J.F., Fitz-Gibbon, S., Schuster, S.C., Brenchley, J.E., and House, C.H. (2008) Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc Natl Acad Sci U S A* **105**: 10583-10588.

Bolter, M. (2004) Ecophysiology of psychrophilic and psychrotolerant microorganisms. *Cell Mol Biol (Noisy-le-grand)* **50**: 563-573.

Bottrell, S., and Tranter, M. (2002) Sulphide oxidation under partially anoxic conditions at the bed of Haut Glacier d'Arolla, Switzerland. *Hydrological Processes* **16**: 2363-3468.

Bonfield, J.K., Smith, K., and Staden, R. (1995) A new DNA sequence assembly program. *Nucleic Acids Res* **23**: 4992-4999.

Bowman, J.P. (2008) Genomic analysis of psychrophilic prokaryotes. In *Psychrophiles: from biodiversity to biotechnology*. Margesin, R., Schinner, F., Marx, J.-C., Gerday, C. (ed). Heidelberg: Springer-Verlag, pp. 265-284.

Cabiscol, E., Tamarit, J., and Ros, J. (2000) Oxidative stress in bacteria and protein damage by reactive oxygen species. *Int Microbiol* **3**: 3-8.

Cavicchioli, R. (2006) Cold-adapted archaea. *Nat Rev Microbiol* **4**: 331-343.

Chattopadhyay, M.K. (2002) The cryoprotective effects of glycine betaine on bacteria *Trends in Microbiology* **10**: 311.

Chattopadhyay, M.K. (2006) Mechanism of bacterial adaptation to low temperature. *J Biosci* **31**: 157-165.

Cheng, S.M., and Foght, J.M. (2007) Cultivation-independent and -dependent characterization of Bacteria resident beneath John Evans Glacier. *FEMS Microbiol Ecol* **59**: 318-330.

Christner, B.C., Mosley-Thompson, E., Thompson, L.G., and Reeve, J.N. (2003) Bacterial recovery from ancient glacial ice. *Environ Microbiol* **5**: 433-436.

Christner, B.C., Mosley-Thompson, E., Thompson, L. G., Zagorodnov, V., Sandman, K., Reeve, J. N. (2000) Recovery and identification of viable bacteria immured in glacial ice. *Icarus* **144**: 479-485.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M., and Tiedje, J.M. (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* **31**: 442-443.

Croan, D.G., Morrison, D.A., and Ellis, J.T. (1997) Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. *Mol Biochem Parasitol* **89**: 149-159.

D'Amico, S., Collins, T., Marx, J.C., Feller, G., and Gerday, C. (2006) Psychrophilic microorganisms: challenges for life. *EMBO Rep* **7**: 385-389.

Dancer, S.J., Shears, P., and Platt, D.J. (1997) Isolation and characterization of coliforms from glacial ice and water in Canada's High Arctic. *J Appl Microbiol* **82**: 597-609.

Daniel, R. (2004) The soil metagenome--a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* **15**: 199-204.

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

Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11**: 1095-1099.

DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci U S A* **89**: 5685-5689.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., Chisholm, S.W., and Karl, D.M. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496-503.

Deming, J.W. (2002) Psychrophiles and polar regions. *Curr Opin Microbiol* **5**: 301-309.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006) Greengenes, a chimera-

checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069-5072.

Diaz, A., Pons, M.E., Lacks, S.A., and Lopez, P. (1992) *Streptococcus pneumoniae* DNA polymerase I lacks 3'-to-5' exonuclease activity: localization of the 5'-to-3' exonucleolytic domain. *J Bacteriol* **174**: 2014-2024.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M.A., Nelson, K.E., Nilsson, C., Olson, R., Paul, J., Brito, B.R., Ruan, Y., Swan, B.K., Stevens, R., Valentine, D.L., Thurber, R.V., Wegley, L., White, B.A., and Rohwer, F. (2008) Functional metagenomic profiling of nine biomes. *Nature* **452**: 629-632.

Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**: 6127-6145.

Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M., Saar, M.O., Alexander, S., Alexander, E.C., Jr., and Rohwer, F. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7**: 57.

Elend, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J.D., Steele, H.L., Reymond, J.L., Jaeger, K.E., and Streit, W.R. (2006) Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl Environ Microbiol* **72**: 3637-3645.

Erwin, D.P., Erickson, I.K., Delwiche, M.E., Colwell, F.S., Strap, J.L., and Crawford, R.L. (2005) Diversity of oxygenase genes from methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain aquifer. *Appl Environ Microbiol* **71**: 2016-2025.

Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I., and Sayers, J.R. (2004) Roles of divalent metal ions in flap endonuclease-substrate interactions. *Nat Struct Mol Biol* **11**: 450-456.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins Dos Santos, V.A., Yakimov, M.M., Timmis, K.N., and Golyshin, P.N. (2005) Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* **12**: 895-904.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., and Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Res* **36**: D281-288.

Foght, J., Aislabie, J., Turner, S., Brown, C.E., Ryburn, J., Saul, D.J., and Lawson, W. (2004) Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb Ecol* **47**: 329-340.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**: 3805-3810.

Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* **6**: 948-958.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355-1359.

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

Gutman, P.D., and Minton, K.W. (1993) Conserved sites in the 5'-3' exonuclease domain of *Escherichia coli* DNA polymerase. *Nucleic Acids Res* **21**: 4406-4407.

Haeberli, W., Hoelzle, M., Paul, F. and Zemp, M. (2007) Integrated monitoring of mountain glaciers as key indicators of global climate change: the European Alps. *Annals of Glaciology* **46**: 150-160.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**: 1457-1462.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557-580.

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

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

Hardeman, F., and Sjöling, S. (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol Ecol* **59**: 524-534.

Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66**: 3113-3116.

Hodson, A.J., Mumford, P. N., Kohler, J. and Wynn, P. M. (2005) The high arctic glacial ecosystem: new insights from nutrient budgets. *Biogeochemistry* **72**: 233-256.

Holmes, D.S., and Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114**: 193-197.

Huettl, C. (1999) Steuerungsfaktoren und Quantifizierung der chemischen Verwitterung auf dem Zugspitzplatt (Wettersteingebirge, Deutschland). In *Faculty of Geosciences*. PhD Munich: Ludwig-Maximilians-University, pp. 171.

Hugenholtz, P. (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**: REVIEWS0003.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P.S., and Sigrist, C.J. (2008) The 20 years of PROSITE. *Nucleic Acids Res* **36**: D245-249.

Hutchison, C.A., 3rd, Smith, H.O., Pfannkoch, C., and Venter, J.C. (2005) Cell-free cloning using phi29 DNA polymerase. *Proc Natl Acad Sci U S A* **102**: 17332-17336.

Jeon, J.H., Kim, J.T., Kang, S.G., Lee, J.H., and Kim, S.J. (2008a) Characterization and its potential application of two esterases derived from the Arctic sediment metagenome. *Mar Biotechnol (NY)*.

Jeon, J.H., Kim, J.T., Kim, Y.J., Kim, H.K., Lee, H.S., Kang, S.G., Kim, S.J., and Lee, J.H. (2008b) Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biotechnol*.

Joyce, C.M., and Steitz, T.A. (1994) Function and structure relationships in DNA polymerases. *Annu Rev Biochem* **63**: 777-822.

Junge, K., Eicken, H., and Deming, J.W. (2004) Bacterial Activity at -2 to -20 degrees C in Arctic wintertime sea ice. *Appl Environ Microbiol* **70**: 550-557.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36**: D480-484.

Kim, Y.J., Choi, G.S., Kim, S.B., Yoon, G.S., Kim, Y.S., and Ryu, Y.W. (2006) Screening and characterization of a novel esterase from a metagenomic library. *Protein Expr Purif* **45**: 315-323.

Klenow, H., and Henningsen, I. (1970) Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc Natl Acad Sci U S A* **65**: 168-175.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and Daniel, R. (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* **69**: 1408-1416.

Konstantinidis, K.T., and Tiedje, J.M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10**: 504-509.

Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F., Edwards, R.A., and Stoye, J. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res* **36**: 2230-2239.

Kukreti, P., Singh, K., Ketkar, A., and Modak, M.J. (2008) Identification of a new motif required for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I (Klenow fragment): the RRRY motif is necessary for the binding of single-stranded DNA substrate and the template strand of the mismatched duplex. *J Biol Chem* **283**: 17979-17990.

Lau, S.C., and Liu, W.T. (2007) Recent advances in molecular techniques for the detection of phylogenetic markers and functional genes in microbial communities. *FEMS Microbiol Lett* **275**: 183-190.

Lee, C.C., Kibblewhite-Accinelli, R.E., Wagschal, K., Robertson, G.H., and Wong, D.W. (2006a) Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. *Extremophiles* **10**: 295-300.

Lee, M.H., Lee, C.H., Oh, T.K., Song, J.K., and Yoon, J.H. (2006b) Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. *Appl Environ Microbiol* **72**: 7406-7409.

Lee, N., Francklyn, C., and Hamilton, E.P. (1987) Arabinose-induced binding of AraC protein to $araI_2$ activates the *araBAD* operon promoter. *Proc Natl Acad Sci U S A* **84**: 8814-8818.

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

Loh, E., and Loeb, L.A. (2005) Mutability of DNA polymerase I: implications for the creation of mutant DNA polymerases. *DNA Repair (Amst)* **4**: 1390-1398.

Lopez-Garcia, P., and Moreira, D. (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Res Microbiol* **159**: 67-73.

Lorenz, P., and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* **3**: 510-516.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., and Schleifer, K.H. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.

Magurran, A.E. (1988) *Ecological diversity and its measurement.* Princeton, N.J.: Princeton University Press.

Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Jr., Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M., and Tiedje, J.M. (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**: 173-174.

Majernik, A., Gottschalk, G., and Daniel, R. (2001) Screening of environmental DNA libraries for the presence of genes conferring Na^+ $(Li^+)/H^+$ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* **183**: 6645-6653.

Manichanh, C., Chapple, C.E., Frangeul, L., Gloux, K., Guigo, R., and Dore, J. (2008) A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. *Nucleic Acids Res*.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Krylov, D., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Thanki, N., Yamashita, R.A., Yin, J.J., Zhang, D., and Bryant, S.H. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**: D237-240.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., and Rothberg, J.M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.

Medigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P.N., Cheung, F., Cruveiller, S., D'Amico, S., Duilio, A., Fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E.P., Rouy, Z., Sekowska, A., Tutino, M.L., Vallenet, D., von Heijne, G., and Danchin, A. (2005) Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* **15**: 1325-1335.

Methe, B.A., Nelson, K.E., Deming, J.W., Momen, B., Melamud, E., Zhang, X., Moult, J., Madupu, R., Nelson, W.C., Dodson, R.J., Brinkac, L.M., Daugherty, S.C., Durkin, A.S., DeBoy, R.T., Kolonay, J.F., Sullivan, S.A., Zhou, L., Davidsen, T.M., Wu, M., Huston, A.L., Lewis, M., Weaver, B., Weidman, J.F., Khouri, H., Utterback, T.R., Feldblyum, T.V., and Fraser, C.M. (2005) The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proc Natl Acad Sci U S A* **102**: 10913-10918.

Mikucki, J.A., and Priscu, J.C. (2007) Bacterial diversity associated with Blood Falls, a subglacial outflow from the Taylor Glacier, Antarctica. *Appl Environ Microbiol* **73**: 4029-4039.

Miteva, V.I., Sheridan, P.P., and Brenchley, J.E. (2004) Phylogenetic and physiological diversity of microorganisms isolated from a deep greenland glacier ice core. *Appl Environ Microbiol* **70**: 202-213.

Miteva, V.I. (2008) Bacteria in snow and glacier ice. In *Psychrophiles: from biodiversity to biotechnology*. Margesin, R., Schinner, F., Marx, J.-C., Gerday, C. (ed). Heidelberg: Springer-Verlag, pp. 462.

Miyada, C.G., Stoltzfus, L., and Wilcox, G. (1984) Regulation of the araC gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on *araBAD* expression. *Proc Natl Acad Sci U S A* **81**: 4120-4124.

Mosier, A.C., Murray, A.E., and Fritsen, C.H. (2007) Microbiota within the perennial ice cover of Lake Vida, Antarctica. *FEMS Microbiol Ecol* **59**: 274-288.

Muyzer, G., Teske, A., Wirsen, C.O., and Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira species* and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**: 165-172.

Muyzer, G., Brinkhoff, T., Nübel, U., Santegoeds, C., Schäfer, H., Wawer, C. (1998) Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In *Molecular Microbial Ecology Manual*. Vol. 3.4.4. Akkermans, A.D.L., Van Elsas, J. D., De Bruijn, F. D. (ed.) Dordrecht: Kluwer Academic Publishers, pp. 1-27.

Nagano, K., Wachi, M., Takada, A., Takaku, F., Hirasawa, T., and Nagai, K. (1999) *fcsA29* mutation is an allele of *polA* gene of *Escherichia coli*. *Biosci Biotechnol Biochem* **63**: 427-429.

Nishida, T., Orikasa, Y., Ito, Y., Yu, R., Yamada, A., Watanabe, K., and Okuyama, H. (2006) *Escherichia coli* engineered to produce eicosapentaenoic acid becomes resistant against oxidative damages. *FEBS Lett* **580**: 2731-2735.

Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G.C., Wang, Z., and Woodgate, R. (2001) The Y-family of DNA polymerases. *Mol Cell* **8**: 7-8.

Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2007) Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl Microbiol Biotechnol* **74**: 501-510.

Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**: 734-740.

Pace, N.R., Stahl, D. A., Lane, D. J., and Olsen, G. J (1985) Analyzing natural microbial populations by rRNA sequences. *ASM News* **51**: 4-12.

Panoff, J.M., Thammavongs, B., and Gueguen, M. (2000) Cryoprotectants lead to phenotypic adaptation to freeze-thaw stress in *Lactobacillus delbrueckii* ssp. *bulgaricus* CIP 101027T. *Cryobiology* **40**: 264-269.

Ponder, M.A., Gilmour, S.J., Bergholz, P.W., Mindock, C.A., Hollingsworth, R., Thomashow, M.F., and Tiedje, J.M. (2005) Characterization of potential stress responses in ancient Siberian permafrost psychroactive bacteria. *FEMS Microbiol Ecol* **53**: 103-115.

Priscu, J.C., Adams, E.E., Lyons, W.B., Voytek, M.A., Mogk, D.W., Brown, R.L., McKay, C.P., Takacs, C.D., Welch, K.A., Wolf, C.F., Kirshtein, J.D., and Avci, R. (1999) Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science* **286**: 2141-2144.

Priscu, J.C., Christner, B. C. (2004) Earth's icy biosphere. In *Microbial diversity and bioprospecting*. Bull, A.T. (ed). Washington, D. C.: ASM Press, pp. 130-145.

Ranjan, R., Grover, A., Kapardar, R.K., and Sharma, R. (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochem Biophys Res Commun* **335**: 57-65.

Reasoner, D.J., and Geldreich, E.E. (1985) A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol* **49**: 1-7.

Reysenbach, A.-L., and Pace, N.R. (1995) In *Archaea: A Laboratory Manual - Thermophiles*. Robb, F.T., Place, A.R. (ed). New York: Cold Spring Harbour Laboratory Press, pp. 101-107.

Rhee, J.K., Ahn, D.G., Kim, Y.G., and Oh, J.W. (2005) New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl Environ Microbiol* **71**: 817-825.

Riesenfeld, C.S., Goodman, R.M., and Handelsman, J. (2004a) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* **6**: 981-989.

Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004b) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**: 525-552.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., Mori, H., Perna, N.T., Plunkett, G., 3rd, Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D., and Wanner, B.L. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res* **34**: 1-9.

Robertson, D.E., Chaplin, J.A., DeSantis, G., Podar, M., Madden, M., Chi, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K., McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J., and Short, J.M. (2004) Exploring nitrilase sequence space for enantioselective catalysis. *Appl Environ Microbiol* **70**: 2429-2436.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., and Goodman, R.M. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**: 2541-2547.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor: Cold Spring Harbor Lab Press.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**: 5463-5467.

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

Schmeisser, C., Steele, H., and Streit, W.R. (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* **75**: 955-962.

Segawa, T., Miyamoto, K., Ushida, K., Agata, K., Okada, N., and Kohshima, S. (2005) Seasonal change in bacterial flora and biomass in mountain snow from the Tateyama Mountains, Japan, analyzed by 16S rRNA gene sequencing and real-time PCR. *Appl Environ Microbiol* **71**: 123-130.

Sharp, M., Parkes, J., Cragg, B., Fairchild, I. J., Lamb, H., and Tranter, M. (1999) Widespread bacterial populations at glacier beds and their relationship to rock weathering and carbon cycling. *Geology* **27**: 107-110.

Sheridan, P.P., Miteva, V.I., and Brenchley, J.E. (2003) Phylogenetic analysis of anaerobic psychrophilic enrichment cultures obtained from a greenland glacier ice core. *Appl Environ Microbiol* **69**: 2153-2160.

Shuman, S. (1991) Site-specific DNA cleavage by vaccinia virus DNA topoisomerase I. Role of nucleotide sequence and DNA secondary structure. *J Biol Chem* **266**: 1796-1803.

Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J Biol Chem* **269**: 32678-32684.

Sjöling, S., and Cowan, D. A. (2008) Metagenomics: microbial community genomes revealed In *Psychrophiles: from biodiversity to biotechnology*. Margesin, R., Schinner, F., Marx, J.-C., Gerday, C. (ed). Heidelberg: Springer-Verlag, pp. 462.

Skidmore, M., Anderson, S.P., Sharp, M., Foght, J., and Lanoil, B.D. (2005) Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. *Appl Environ Microbiol* **71**: 6986-6997.

Soderberg, M.A., Rossier, O., and Cianciotto, N.P. (2004) The type II protein secretion system of *Legionella pneumophila* promotes growth at low temperatures. *J Bacteriol* **186**: 3712-3720.

Song, J.S., Jeon, J.H., Lee, J.H., Jeong, S.H., Jeong, B.C., Kim, S.J., Lee, J.H., and Lee, S.H. (2005) Molecular characterization of TEM-type beta-lactamases identified in cold-seep sediments of Edison Seamount (south of Lihir Island, Papua New Guinea). *J Microbiol* **43**: 172-178.

Sorensen, K.B., Lauer, A., Teske, A. (2004) Archaeal phylotypes in a metal-rich and low-activity deep subsurface sediment of the Peru Basin, ODP leg 201, Site 1231. *Geobiology* **2**: 151-161.

Staley, J.T., and Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**: 321-346.

Stibal, M., Tranter, M., Benning, L.G., and Rehak, J. (2008) Microbial primary production on an Arctic glacier is insignificant in comparison with allochthonous organic carbon input. *Environ Microbiol* **10**: 2172-2178.

Streit, W.R., and Schmitz, R.A. (2004) Metagenomics--the key to the uncultured microbes. *Curr Opin Microbiol* **7**: 492-498.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22-28.

Thomas, D.N., and Dieckmann, G.S. (2002) Antarctic Sea ice--a habitat for extremophiles. *Science* **295**: 641-644.

Torsvik, V., Ovreas, L., and Thingstad, T.F. (2002) Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296**: 1064-1066.

Tringe, S.G., and Rubin, E.M. (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* **6**: 805-814.

Turchetti, B., Buzzini, P., Goretti, M., Branda, E., Diolaiuti, G., D'Agata, C., Smiraglia, C., and Vaughan-Martini, A. (2008) Psychrophilic yeasts in glacial environments of Alpine glaciers. *FEMS Microbiol Ecol* **63**: 73-83.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027-1031.

Tvermyr, M., Kristiansen, B.E., and Kristensen, T. (1998) Cloning, sequence analysis and expression in *E. coli* of the DNA polymerase I gene from *Chloroflexus aurantiacus*, a green nonsulfur eubacterium. *Genet Anal* **14**: 75-83.

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

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., and Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.

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

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

Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., and Charles, T.C. (2006) Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol* **72**: 384-391.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261-5267.

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

Wemekamp-Kamphuis, H.H., Sleator, R.D., Wouters, J.A., Hill, C., and Abee, T. (2004) Molecular and physiological analysis of the role of osmolyte transporters BetL, Gbu, and OpuC in growth of *Listeria monocytogenes* at low temperatures. *Appl Environ Microbiol* **70**: 2912-2918.

Willerslev, E., Hansen, A.J., Ronn, R., Brand, T.B., Barnes, I., Wiuf, C., Gilichinsky, D., Mitchell, D., and Cooper, A. (2004a) Long-term persistence of bacterial DNA. *Curr Biol* **14**: R9-10.

Willerslev, E., Hansen, A. J., and Poinar, H. N. (2004b) Isolation of nucleic acids and cultures from fossil ice and permafrost. *TRENDS in Ecology and Evolution* **19**: 141-147.

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

Xiang, S., Yao, T., An, L., Xu, B., and Wang, J. (2005) 16S rRNA sequences and differences in bacteria isolated from the Muztag Ata glacier at increasing depths. *Appl Environ Microbiol* **71**: 4619-4627.

Xu, Y., Derbyshire, V., Ng, K., Sun, X.C., Grindley, N.D., and Joyce, C.M. (1997) Biochemical and mutational studies of the 5'-3' exonuclease of DNA polymerase I of *Escherichia coli*. *J Mol Biol* **268**: 284-302.

Yokouchi, H., Fukuoka, Y., Mukoyama, D., Calugay, R., Takeyama, H., and Matsunaga, T. (2006) Whole-metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using phi29 polymerase. *Environ Microbiol* **8**: 1155-1163.

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

Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24**: 680-686.

Zhang, X.F., Yao, T.D., Tian, L.D., Xu, S.J., and An, L.Z. (2008) Phylogenetic and physiological diversity of bacteria isolated from Puruogangri ice core. *Microb Ecol* **55**: 476-488.

Acknowledgements

Curriculum vitae

Carola Simon, Dipl.-Ing. agr.
Born on August, 25th 1975 in Hanover, Germany

Scientific Background

Since 01/05	Scientific assistant at the Institute of Microbiology and Genetics, Georg-August University Göttingen
10/97 – 06/04	University of Hannover, Diploma program Horticulture, Diploma thesis: 'Genetic and pathogenic diversity of strains of *Ralstonia solanacearum* originating from ginger (*Zingiber officinale* Roscoe)'
01/03 – 05/03	Practical part of the diploma thesis at the Cooperative Research Centre for Tropical Plant Pathology, Department of Microbiology and Parasitology, University of Queensland, Australia
08/01 – 10/01	Internship at the Institut für Pflanzenvirologie, Mikrobiologie und biologische Sicherheit, Biologische Bundesanstalt für Land- und Forstwirtschaft (BBA), Braunschweig
09/00 – 07/01	Scottish Agricultural College (SAC), Ayr, Scotland; Erasmus program
04/99 – 09/00	Scientific assistant at the Fraunhofer Institute for Toxicology and Aerosol Research, Hannover

Practical Experience

07/95 – 07/97	Apprenticeship as gardener; Gärtnerei Behre, Benthe

Education

1988 – 1995	Secondary education at the Matthias-Claudius-Gymnasium Gehrden