

RESEARCH

Open Access

# Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing

Jan Kjølhede Vester, Mikkel Andreas Glaring and Peter Stougaard\*

## Abstract

**Background:** The use of cold-active enzymes has many advantages, including reduced energy consumption and easy inactivation. The ikaite columns of SW Greenland are permanently cold (4–6°C) and alkaline (above pH 10), and the microorganisms living there and their enzymes are adapted to these conditions. Since only a small fraction of the total microbial diversity can be cultured in the laboratory, a combined approach involving functional screening of a strain collection and a metagenomic library was undertaken for discovery of novel enzymes from the ikaite columns.

**Results:** A strain collection with 322 cultured isolates was screened for enzymatic activities identifying a large number of enzyme producers, with a high re-discovery rate to previously characterized strains. A functional expression library established in *Escherichia coli* identified a number of novel cold-active enzymes. Both  $\alpha$ -amylases and  $\beta$ -galactosidases were characterized in more detail with respect to temperature and pH profiles and one of the  $\beta$ -galactosidases, BGal<sub>117E2</sub>, was able to hydrolyze lactose at 5°C. A metagenome sequence of the expression library indicated that the majority of enzymatic activities were not detected by functional expression. Phylogenetic analysis showed that different bacterial communities were targeted with the culture dependent and independent approaches and revealed the bias of multiple displacement amplification (MDA) of DNA isolated from complex microbial communities.

**Conclusions:** Many cold- and/or alkaline-active enzymes of industrial relevance were identified in the culture based approach and the majority of the enzyme-producing isolates were closely related to previously characterized strains. The function-based metagenomic approach, on the other hand, identified several enzymes ( $\beta$ -galactosidases,  $\alpha$ -amylases and a phosphatase) with low homology to known sequences that were easily expressed in the production host *E. coli*. The  $\beta$ -galactosidase BGal<sub>117E2</sub> was able to hydrolyze lactose at low temperature, suggesting a possibly use in the dairy industry for this enzyme. The two different approaches complemented each other by targeting different microbial communities, highlighting the usefulness of combining methods for bioprospecting. Finally, we document here that ikaite columns constitute an important source of cold- and/or alkaline-active enzymes with industrial application potential.

**Keywords:** Bioprospecting, 16S rRNA,  $\beta$ -Galactosidase,  $\alpha$ -Amylase, Metagenomics, MDA, Cold-active enzymes, Alkaline-active enzymes

\* Correspondence: [psg@plen.ku.dk](mailto:psg@plen.ku.dk)  
Department of Plant and Environmental Sciences, University of Copenhagen,  
Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

## Background

Many industrial and biotechnological applications make use of cold-active enzymes or could benefit from the use of such enzymes as they enable these processes to run at low temperature. Such processes may save energy and production costs, improve hygiene, maintain taste and other organoleptic properties, and reduce the risk of contaminations. In addition, cold-active enzymes are heat labile and can easily and selectively be inactivated by moderately elevated temperatures. Cold-active enzymes may be used in fine chemical synthesis, environmental biotechnology, production of biofuels and energy, and in the food and feed, detergent, pharmaceutical, medical and textile industries [1]. Approximately 75% of the Earth's biosphere is cold (less than 5°C) [2] and consequently, bacteria producing cold-active enzymes can be found in numerous habitats. Bioprospecting for cold-active enzymes has been conducted in many environments including Antarctic soil [3] and sediments [4], Arctic and Subarctic glaciers [5-8], the deep sea [9] and permafrost soils [10]. Also, enzymes that are active at high pH are of industrial interest, since these are used in food and feed, textile, waste management, medical and detergent industries [11]. Especially the detergent industry is of commercial interest, and alkaline proteases, amylases, cellulases and lipases are all used in detergents [12]. Thus, the combination of cold- and alkaline-active enzymes could be used in detergents for environment-friendly, low temperature washing.

The ikaite columns of SW Greenland are submarine tufa columns formed over alkaline springs by precipitation of the metastable hexahydrate of calcium carbonate, called ikaite [13]. The columns represent a permanently cold (4-6°C) and alkaline (above pH 10) ecological niche of moderate salinity (ca. 10‰) and together with the ice-covered Lake Untersee [14] and a series of small ponds [15] in the Antarctica, the ikaite columns constitute one of the very few permanently cold and alkaline environments on Earth. They harbor a rich microbial community adapted to these conditions and a significant part of the bacteria isolated from the ikaite columns represent previously uncharacterized species and genera [16,17]. So far, four new bacterial strains with cold-active enzymes have been described in detail; the  $\alpha$ -amylase- and protease-producing *Arsukibacterium ikkense* [18], the  $\alpha$ -amylase-,  $\alpha$ -,  $\beta$ -galactosidase- and  $\beta$ -glucuronidase-producing *Alkalilactibacillus ikkensis* [19], the phosphatase-, esterase-, protease- and  $\beta$ -galactosidase-producing *Rhodonellum psychrophilum* [20] as well as a lipase-producing  *$\gamma$ -Proteobacterium* [21]. Work conducted during the last decade on the ecology and microbial enzymes from the ikaite columns has recently been reviewed, highlighting the ikaite columns as a unique biological environment with good prospects for finding novel bacterial species and enzymes for industrial applications [22].

It is well established that only a fraction of the total bacterial diversity can be cultured in the laboratory [23] and that most bacterial phyla have no cultured representatives [24]. Previous work on the ikaite columns has focused on cultured isolates and it was recently reported that various attempts to optimize culturing conditions for ikaite column bacteria only improved the total diversity covered marginally [25]. In order to circumvent this problem, a combined culture dependent and independent approach was taken in a search for cold-active enzymes from the cold and alkaline ikaite columns. In this report, traditional screening of strain collections was coupled with screening of a functional expression library. Furthermore, a metagenomic sequence of the expression library was established and searched for putative enzyme-encoding sequences allowing comparisons between the discovery rate and value of the different approaches. Several cold- and alkaline-active enzymes were identified and a few selected  $\alpha$ -amylases and  $\beta$ -galactosidases were characterized in more detail.

## Results and discussion

### Culture dependent approach

Ikaite column material collected during expeditions over the last decade was used to establish a strain collection of 322 cultured isolates. The strain collection was screened for nine different enzymatic activities at pH 10, and 203 enzyme producing strains were identified (Table 1). The dominating activities were phosphatase,  $\alpha$ -galactosidase, protease and  $\alpha$ -amylase followed by  $\beta$ -galactosidase and  $\beta$ -glucanase. Only one positive was found for both cellulase and  $\beta$ -xylanase. Many of the activities were identified in strains showing more than one activity, with the combinations phosphatase/protease (32 isolates), and  $\alpha$ -galactosidase/ $\beta$ -galactosidase (20 isolates) being dominant.  $\beta$ -Galactosidase,  $\beta$ -glucanase, cellulase and  $\beta$ -xylanase were not found as single activities. Most of the activities were produced at all temperatures tested (10°C, 20°C and 28°C) (data not shown), indicating that the majority of strains are psychrotolerant and not true psychrophiles. More importantly, it also demonstrates that the ikaite column isolates produce industrially relevant extracellular enzymes active at both high pH and low temperature. Phylogeny determined by 16S rRNA gene analyses on 65 randomly picked strains representing all enzyme groups showed that most strains were either  $\alpha$ - (15%) or  $\gamma$ - (69%) *Proteobacteria*. A previous study of randomly picked cultured isolates from the ikaite columns identified *Firmicutes*, *CFB-group*,  $\alpha$ -*Proteobacteria* and  $\gamma$ -*Proteobacteria* [16]. The observation that  $\gamma$ -*Proteobacteria* dominate the cultured species with enzymatic activities is in agreement with two of the already characterized enzyme producing species, *A. ikkense* [18] and a lipase-producing strain [21], both being  $\gamma$ -*Proteobacteria*. In a

**Table 1 Enzymatic activities identified by screening the ikaite strain collection**

Activity	Isolates	No. of 16S sequences	Closest relative (16S rDNA % identity)	Class
Phosphatase	102			
$\alpha$ -Galactosidase	75			
Protease	73			
$\alpha$ -Amylase	57			
$\beta$ -Galactosidase	25			
$\beta$ -Glucanase	24			
Cellulase	1			
$\beta$ -Xylanase	1			
$\beta$ -Mannanase	0			
<i>Total</i>	358			
Groups of activity	Total no.			
$\alpha$ -Galactosidase	45	3	<i>Loktanella vestfoldensis</i> (98-100%)	$\alpha$ -Proteobacteria
Phosphatase + protease	32	5	<i>Arsukibacterium ikkense</i> (98-100%)	$\gamma$ -Proteobacteria
Phosphatase	29	4	<i>Pseudomonas sabulinigri</i> (99%)	$\gamma$ -Proteobacteria
		1	<i>Loktanella vestfoldensis</i> (100%)	$\alpha$ -Proteobacteria
		1	<i>Stenotrophomonas rhizophila</i> (100%)	$\gamma$ -Proteobacteria
		1	<i>Arsukibacterium ikkense</i> (100%)	$\gamma$ -Proteobacteria
		1	<i>Idiomarina fontislapidosi</i> (95%)	$\gamma$ -Proteobacteria
$\alpha$ -Galactosidase + $\beta$ -galactosidase	20	2	<i>Natronobacillus azotifigens</i> (98%)	Bacilli
		1	<i>Rhodonellum psychrophilum</i> (100%)	Bacteroidetes
		1	<i>Klebsiella pneumoniae</i> (95%)	$\gamma$ -Proteobacteria
$\alpha$ -Amylase	17	1	<i>Pseudomonas poae</i> (100%)	$\gamma$ -Proteobacteria
		1	<i>Marinovum algicola</i> (100%)	$\alpha$ -Proteobacteria
		1	<i>Thermoleophilum minutum</i> (100%)	Actinobacteria
		1	<i>Rhodobacter veldkampii</i> (96%)	$\alpha$ -Proteobacteria
Phosphatase + $\alpha$ -amylase + protease + $\beta$ -glucanase	17	9	<i>Arsukibacterium ikkense</i> (98-100%)	$\gamma$ -Proteobacteria
Phosphatase + $\alpha$ -amylase + protease	10	7	<i>Arsukibacterium ikkense</i> (100%)	$\gamma$ -Proteobacteria
		1	<i>Pseudomonas sabulinigri</i> (99%)	$\gamma$ -Proteobacteria
Protease	8	2	<i>Pseudomonas sabulinigri</i> (96-97%)	$\gamma$ -Proteobacteria
		1	<i>Aquimonas voraii</i> (93%)	$\gamma$ -Proteobacteria
Phosphatase + $\alpha$ -amylase	7	5	<i>Arsukibacterium ikkense</i> (98-100%)	$\gamma$ -Proteobacteria
Phosphatase + protease + $\beta$ -glucanase	7	4	<i>Arsukibacterium ikkense</i> (98-100%)	$\gamma$ -Proteobacteria
		1	<i>Klebsiella variicola</i> (97%)	$\gamma$ -Proteobacteria
$\alpha$ -Galactosidase + $\beta$ -galactosidase + protease	3	3	<i>Rhodonellum psychrophilum</i> (96-99%)	Bacteroidetes
$\alpha$ -Galactosidase + $\beta$ -galactosidase + $\alpha$ -amylase	2	2	<i>Rhodonellum psychrophilum</i> (95-99%)	Bacteroidetes
$\alpha$ -Galactosidase + $\alpha$ -amylase	2	2	<i>Loktanella vestfoldensis</i> (100%)	$\alpha$ -Proteobacteria
$\alpha$ -Galactosidase + $\alpha$ -amylase + protease	1	1	<i>Loktanella vestfoldensis</i> (100%)	$\alpha$ -Proteobacteria
$\alpha$ -Galactosidase + phosphatase	1	1	<i>Loktanella vestfoldensis</i> (100%)	$\alpha$ -Proteobacteria
$\alpha$ -Galactosidase + cellulase + $\beta$ -xylanase	1	1	<i>Demequina aestuarii</i> (98%)	Actinobacteria
$\alpha$ -Amylase + protease	1	1	<i>Arsukibacterium ikkense</i> (100%)	$\gamma$ -Proteobacteria
<i>Total</i>	203	65		

A total of 322 cultured isolates were screened for 9 different enzyme activities at pH 10 and 10, 20 and 28°C. The closest relative was determined for selected isolates by partial sequencing of the 16S rRNA gene.

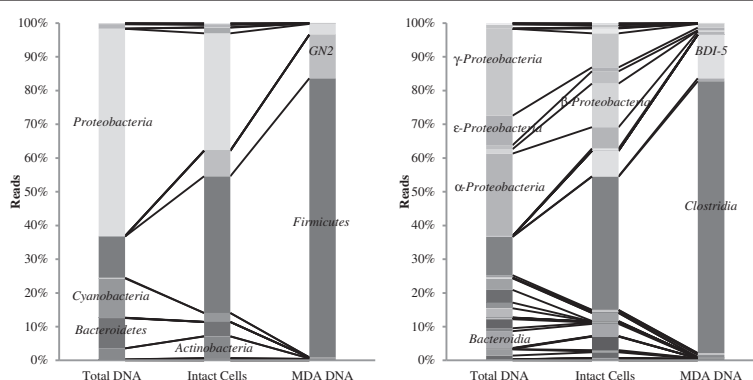
similar study of a strain collection from Arctic sea ice, 70% of the isolates were  $\gamma$ -*Proteobacteria* [26], highlighting the easy cultivation of this class of bacteria. Almost half of the selected enzyme producing isolates were closely related to *A. ikkense* (32 of 65 sequenced isolates) and six showed sequence similarity to *R. psychrophilum*, indicating that the re-discovery rate was high as most strains were closely related to previously characterized strains. The one strain with activity on cellulose was related to *Demequina aestuarii*, a mesophilic *Actinobacterium* isolated from a tidal sediment in South Korea [27] and to the cellulolytic *Cellulomonas fermentans* [28]. The isolate could be involved in degradation of algae in ikaite columns, since it showed activity on cellulose and xylan, which are known components of algal cell walls.

### Functional expression approach

#### Diversity analysis

Although a large number of enzyme-producing strains were identified in the culture based approach, their phylogenetic affiliations were similar, highlighting the bias introduced by cultivation. To obtain cold-active enzymes from other groups of bacteria, a culture independent approach based on functional screening of an expression library was included. This would also allow for a more direct route to recombinant expression of enzymes in a relevant host. Intact bacterial cells were extracted from fresh ikaite material from 10 different columns. DNA from these cells was isolated and because the amount of extracted DNA was too low for direct library construction, the DNA was amplified using multiple displacement amplification (MDA) prior to generation of the functional expression library. Both the cell extraction and MDA was expected to introduce bias in the final DNA pool. In order to gain insight into this bias, pyrosequencing of the 16S rRNA gene was performed on (i) DNA extracted directly from the columns, (ii) DNA extracted after separation of intact cells,

and (iii) DNA after MDA treatment (Figure 1). The initial diversity was dominated by *Proteobacteria*, *Firmicutes*, *Cyanobacteria* and *Bacteroidetes*, in agreement with a previous analysis of ikaite columns [22], and as expected, significant bias was introduced in each step together with a concomitant loss of diversity. The resulting DNA used for the expression library consisted mainly of *Firmicutes* followed by candidate division *GN02* and *Proteobacteria*. *Cyanobacteria*, *Bacteroidetes* and *Actinobacteria* were lost in the MDA step. At the class level, *BD1-5* from the phylum *GN02* and  $\beta$ -*Proteobacteria* were selected during the cell extraction. The MDA step is known to be heavily biased on complex communities [29] and in this case the MDA reaction favored *BD1-5* and *Clostridia*. Diversity down to the genus level of the different DNA samples is given in Table 2. The most abundant operational taxonomic units (OTUs) in the total DNA were related to the genera *Rhodobaca* and *Thioalkalivibrio*, while OTUs related to *Fusibacter*, *Proteiniclasticum*, *Tindallia/Anoxyatronum* and *Alkaliphilus* dominated the MDA DNA used for the expression library. *Proteiniclasticum* is a proteolytic genus [30], the genera *Alkaliphilus* [31] and *Tindallia/Anoxyatronum* [32] are alkaliphiles and the presence of OTUs related to these is most likely a result of the alkaline conditions inside the ikaite columns. The four genera dominating the MDA DNA (including *Fusibacter* [33]) are anaerobic and are consequently not expected to be found in the strain collection. The OTUs related to *Rhodobaca*, dominating the total DNA extraction, are known to be capable of both phototrophic and chemotrophic growth [34]. They were still present after cell extraction, but were almost completely lost at the MDA step (data not shown). The  $\gamma$ -*Proteobacteria* dominating the enzyme producers in the strain collection were not identified in the MDA DNA used for the library (data not shown), and the anaerobic genera found in the library were not found in the strain collection, confirming that the culture independent



**Figure 1 Changes in microbial diversity during library preparation.** Diversity shifts at phylum (left) and class (right) level of samples introduced during preparation of DNA for the expression library. Total DNA represents DNA extracted directly before manipulation, Intact Cells represents DNA from extracted cells, and MDA DNA represents DNA after MDA treatment.

**Table 2 Top ten most abundant taxonomic groups during library preparation**

Sample	% reads	Phylum	Class	Order	Family	Genus
Total DNA	11.35%	<i>Proteobacteria</i>	<i>α-Proteobacteria</i>	<i>Rhodobacterales</i>	<i>Rhodobacteraceae</i>	<i>Rhodobaca</i>
	6.84%	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Alteromonadales</i>	<i>Chromatiaceae</i>	
	5.20%	<i>Proteobacteria</i>	<i>ε-Proteobacteria</i>	<i>Campylobacterales</i>	<i>Helicobacteraceae</i>	
	3.99%	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Chromatiales</i>	<i>Ectothiorhodospiraceae</i>	<i>Thioalkalivibrio</i>
	3.82%	<i>Bacteroidetes</i>	<i>Bacteroidia</i>	<i>Bacteroidales</i>	<i>ML635J-40</i>	
	3.46%	<i>Proteobacteria</i>	<i>ε-Proteobacteria</i>	<i>Campylobacterales</i>	<i>Helicobacteraceae</i>	<i>Sulfurimonas</i>
	3.01%	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Oceanospirillales</i>	<i>Oceanospirillaceae</i>	<i>Marinomonas</i>
	2.42%	<i>Cyanobacteria</i>	<i>Chloroplast</i>	<i>Stramenopiles</i>		
	2.38%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Tindallia_Anoxynatronum</i>
	2.31%	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Alteromonadales</i>	<i>Chromatiaceae</i>	<i>Other</i>
44.77%						
Intact cells	14.71%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Proteiniclasticum</i>
	11.65%	<i>Proteobacteria</i>	<i>β-Proteobacteria</i>	<i>Rhodocyclales</i>	<i>Rhodocyclaceae</i>	<i>Azoarcus</i>
	10.63%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Acidaminobacteraceae</i>	<i>Fusibacter</i>
	7.54%	<i>GN02</i>	<i>BD1-5</i>			
	4.14%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Tindallia_Anoxynatronum</i>
	4.01%	<i>Actinobacteria</i>	<i>OPB41</i>			
	2.77%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Alkaliphilus</i>
	2.45%	<i>Bacteroidetes</i>	<i>Bacteroidia</i>	<i>Bacteroidales</i>	<i>ML635J-40</i>	
	2.38%	<i>Cyanobacteria</i>	<i>Synechococcophycideae</i>	<i>Synechococcales</i>	<i>Synechococcaceae</i>	<i>Paulinella</i>
	2.31%	<i>Proteobacteria</i>	<i>α-Proteobacteria</i>	<i>Rhodobacterales</i>	<i>Rhodobacteraceae</i>	<i>Rhodobaca</i>
62.59%						
MDA DNA	46.19%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Acidaminobacteraceae</i>	<i>Fusibacter</i>
	12.93%	<i>GN02</i>	<i>BD1-5</i>			
	9.16%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Proteiniclasticum</i>
	5.62%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Tindallia_Anoxynatronum</i>
	5.29%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	<i>Alkaliphilus</i>
	3.29%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Acidaminobacteraceae</i>	<i>WH1-8</i>
	2.52%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>		
	2.26%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Clostridiaceae</i>	
	1.60%	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiales</i>	<i>Acidaminobacteraceae</i>	
	0.73%	<i>Firmicutes</i>				
89.59%						

Phylogenetic affiliation of the top ten most abundant taxonomic groups identified in the total DNA, DNA from intact cells, and MDA DNA preparations.

approach is likely to identify a different set of enzymes than the culture dependent approach. Even though the DNA extraction applied in this study is biased, functional diversity is not necessarily severely affected due to functional redundancy in a community, as was demonstrated by Delmont et al. [35].

#### Functional metagenomics

The MDA DNA was partially digested, inserted into the bacterial artificial chromosome (BAC) shuttle vector, mod.pGNS-BAC, and transformed into *E. coli* to yield a

functional expression library of 2,843 clones. The average insert size was around 15 kb and 14% had no insert, giving a total size of the cloned metagenome of approximately 36 Mbp (data not shown). The library was screened for various enzymatic activities as presented in Table 3. Three  $\alpha$ -amylase, two  $\beta$ -galactosidase and one phosphatase producing clones were identified. The  $\alpha$ -amylase clones and one of the  $\beta$ -galactosidase clones showed activity at 15°C, but not at 37°C. This indicated that these enzymes were only active at low temperature, although it cannot be ruled out that the lack of activity was an effect of decreased

**Table 3 Enzymatic activities identified in the functional metagenomic library of 2,843 *E. coli* clones**

Activity	Total no.	15°C	37°C
$\alpha$ -Amylase	3*	3	0
$\beta$ -Galactosidase	2	1	1
Phosphatase	1		1

No protease, cellulase,  $\beta$ -glucanase,  $\beta$ -xylanase,  $\beta$ -mannanase or lipase producing isolates were identified. \*: The three  $\alpha$ -amylase positive clones carried inserts covering the same gene.

enzyme production in *E. coli* at higher temperatures. One of the main advantages of functional expression is that once positive clones are identified, a suitable host for expression of the enzymes has already been established. Several factors however, have to work in concert for enzymatic activities to be picked up in a functional expression screening and there are many limiting factors including codon usage, promoter recognition, presence of chaperones and successful secretion [36]. Gabor et al. [37] calculated that an estimated 40% of genes from 32 selected different genomes had expression signals that would be recognized in *E. coli*. *Firmicutes* were best with approximately 70% and *Actinobacteria* worst with around 10%, suggesting that *E. coli* could be a suitable host for the ikaite library given the high fraction of *Firmicutes*, although the study did not consider other limiting factors. The library constructed had relatively few clones, but the hit-rate was comparable to those of similar studies: one  $\alpha$ -amylase was found when screening a cosmid-library of 35,000 clones [38], six esterases were identified in a screening of 60,000 fosmid clones [8], 11 cellulases were found when screening a small insert BAC-library of 10,000 clones [39], three  $\beta$ -galactosidases in a screening of 2,100 plasmid clones [40], one amyolytic enzyme in a screening of 30,000 plasmid clones [41], one cellulase in a plasmid library of 8,500 clones [42], and 38 amylases, 13 phosphatases but no proteases in a screening of 32,000 plasmid clones [43].

#### Metagenomic sequencing

In order to get an overview of the enzyme potential in the functional expression library, high throughput Illumina sequencing was carried out on a pool of all BAC clones (see Table 4 for sequence statistics). The final assembly consisted of 4,621 contigs with an average size of 2,215 bp giving a total metagenome size of 10.2 Mbp. This is considerably lower than the estimated size of 36 Mbp for the cloned metagenome (see above), suggesting either an unequal coverage in the sequencing or a significant redundancy among the BAC clone inserts. The latter is consistent with the results on the microbial diversity in the starting material, which was dominated by a few related phylogenetic groups (Figure 1).

**Table 4 Metagenome sequencing and assembly statistics**

Sequencing	
Filtered reads	8,042,722
Filtered reads, nucleotide count	1,097,898,162 bp
Average read length	136.5 bp
Metagenome assembly	
Total size	10,236,415 bp
Contigs	4,621
Average contig size	2,215 bp
Minimum/maximum contig size	200/28,743 bp
N50	4,501 bp
GC %	41.6%

The predicted full-length and partial coding sequences in the resulting metagenome contigs were annotated by comparison to the Pfam protein family database, searched for relevant enzyme domains, and compared to the data obtained from the functional expression library (Table 5). A total of 60 domains with similarity to GH families known to contain  $\alpha$ -amylase,  $\alpha$ - and  $\beta$ -galactosidase, cellulase and  $\beta$ -xylanase activities were identified in 47 unique coding sequences (Additional file 1: Table S1). Since there is considerable functional overlap between, and variety within, GH families, this is likely to be an overestimate of the abundance of these specific enzyme activities. Some GH domains could be assigned to more than one activity and the data given in Table 5 represents the number of non-redundant domain matches for each activity. Putative proteases were identified from protease domain-containing Pfam families, as specified by the MEROPS peptidase database, resulting in 313 protease domains in 289 unique coding sequences (Additional file 1: Table S1). The majority of these are likely to be intracellular house-keeping proteases involved in normal cellular metabolism. A non-exhaustive search for phosphatases identified at least 68 phosphatase domains in 17 Pfam families (data not shown).

It is clear that the number of relevant enzymes identified in the metagenome is considerably higher than those obtained from the functional screening. Apart from the limiting factors discussed above, the AZCL-linked substrates used for detecting  $\alpha$ -amylase, cellulase,  $\beta$ -glucanase,  $\beta$ -xylanase and protease activity are extracellular and requires secretion of the active enzymes from *E. coli* for optimal detection. The subcellular localization of the identified enzyme-encoding sequences was predicted using PSORTb. Seven GH sequences were predicted to be extracellular in either gram-positive or gram-negative bacteria and six of these were from family GH13, which include the majority of  $\alpha$ -amylases (Additional file 1: Table S1). The predicted proteases included 15 potentially extracellular enzymes and another 13 possibly localized to the

**Table 5 Summary of enzymes identified by the different approaches**

Activity	Metagenome <sup>a</sup>	Functional screening	Strain collection
$\alpha$ -Amylase	17	3*	59
Protease	-	0	73
Cellulase	4	0	1
$\beta$ -Glucanase	0	0	20
$\beta$ -Xylanase	7	0	1
$\beta$ -Mannanase	3	0	0
$\beta$ -Galactosidase	15	2	25
Phosphatase	-	1	102
$\alpha$ -Galactosidase	10	N/A	75

Number of putative enzyme encoding open reading frames (ORFs) in the BAC clone metagenome (metagenome), number of enzyme producing *E. coli* clones carrying the BAC library (functional screening), and number of enzyme producing native bacteria in the strain collection (strain collection).

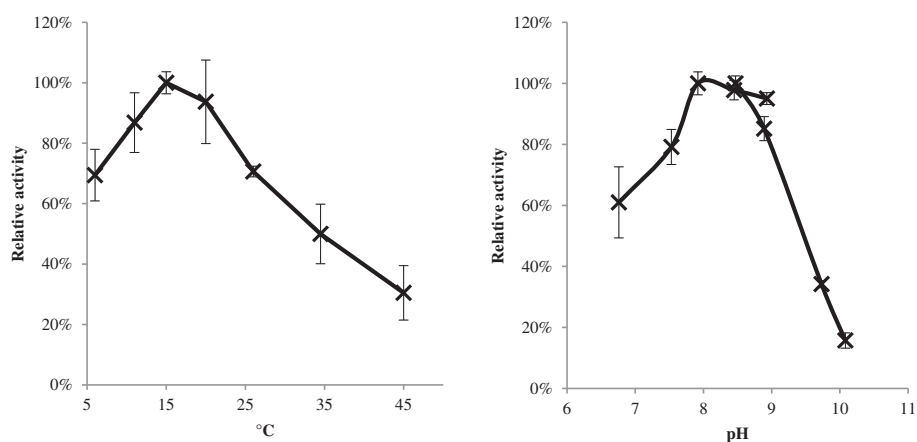
<sup>a</sup>: Identified enzymes in the metagenome belonged to the following GH families:  $\alpha$ -amylase: GH13 + 57;  $\alpha$ -galactosidase: GH4 + 27 + 57;  $\beta$ -galactosidase: GH1 + 2 + 42; cellulase: GH5 + 8;  $\beta$ -xylanase: GH5 + 8 + 30;  $\beta$ -mannanase: GH5. \*: The three  $\alpha$ -amylase positive clones carried inserts from the same genomic region.

outer membrane/periplasm or cell wall, indicating that the absence of protease-producing clones in the functional library is not caused by a lack of protease-encoding sequences (Additional file 1: Table S1).

#### $\alpha$ -Amylases (IKA3C6, IKA16D10 and IKA28E6)

The three  $\alpha$ -amylases and two  $\beta$ -galactosidases identified in the expression library were analyzed in more detail. End-sequencing of the three *E. coli* clones with  $\alpha$ -amylase activity (IKA3C6, IKA28E6 and IKA16D10) and comparison to the obtained metagenomic sequence of the library, showed that the three inserts were overlapping and covering the same genomic region, indicating that they were most likely expressing the same gene. Of the three clones, IKA3C6 was the most active, indicating that the location of the gene in the insert in this clone was more favorable for expression (data not shown).

The  $\alpha$ -amylase encoded by IKA3C6 was successfully expressed in an optimized *E. coli* expression system and the activity of crude extract was characterized with regard to temperature and pH (Figure 2). The IKA3C6  $\alpha$ -amylase showed a temperature optimum around 15°C and a pH optimum around pH 8–9, and retained more than 60% activity at 10°C. These profiles are somewhat similar to the well characterized  $\alpha$ -amylase from the Antarctic bacterium *Pseudoalteromonas haloplanctis*, which has a temperature optimum at 30°C and pH 7 [44], as well as the  $\alpha$ -amylase from an Arctic sea-ice isolate related to *Brachybacterium faecium*, with an optimum at 30°C and pH 7 [26]. The  $\alpha$ -amylase activity of the *E. coli* clone was primarily intracellular, which together with the pH profile showing optimal activity at near neutral pH could indicate that the IKA3C6  $\alpha$ -amylase is an intracellular enzyme in the natural host,



**Figure 2 Temperature and pH profiles of the  $\alpha$ -amylase.** Temperature (left) and pH (right) profile of crude extract of the IKA3C6  $\alpha$ -amylase using amylopectin as substrate. Error bars indicate standard deviations from triplicate experiments.

although it could also be an effect of the heterologous expression in *E. coli*. The  $\alpha$ -amylase from the IKA3C6 clone, Amy<sub>I3C6</sub>, is currently the focus of further studies.

#### ***$\beta$ -galactosidases (IKA3H5 and IKA17E2)***

Sequence information was obtained for the two  $\beta$ -galactosidases identified in the expression library (IKA3H5 and IKA17E2). The two  $\beta$ -galactosidase protein sequences from IKA3H5 (BGal<sub>I3H5</sub>) and IKA17E2 (BGal<sub>I17E2</sub>) were compared to the LacZ sequence from mesophilic *E. coli* as well as to the cold-active  $\beta$ -galactosidase gene from *A. ikkensis* previously isolated from the ikaite columns, which retains more than 60% activity at 0°C [45]. Both sequences were shorter than LacZ and the *A. ikkensis* enzyme, especially BGal<sub>I17E2</sub> with only 455 amino acids (Table 6). BGal<sub>I3H5</sub> belonged to glycosyl hydrolase (GH) family 2 like LacZ and the  $\beta$ -galactosidase from *A. ikkensis*, whereas BGal<sub>I17E2</sub> represented GH family 1. Both BGal<sub>I17E2</sub> and BGal<sub>I3H5</sub> displayed typical adaptations to low temperature compared to LacZ (lower arginine, proline and arginine/ (arginine + lysine ratio)) [46], and for BGal<sub>I17E2</sub> this was even more pronounced than the *A. ikkensis*  $\beta$ -galactosidase, indicating that the enzymes are indeed cold-adapted. The active site glutamic acid (E) was conserved in all four enzymes as was the adjacent methionine (M) (Table 6). The neighboring tyrosine (Y) was conserved in all except *A. ikkensis*. The closest homologues of BGal<sub>I17E2</sub> and BGal<sub>I3H5</sub> were  $\beta$ -galactosidases from *Clostridium hathewayi* (53% identity) and *Roseburia hominis* (51% identity), respectively. These are both obligate anaerobic *Firmicutes* belonging to the class *Clostridia* [47] indicating that the natural hosts of the enzymes are associated with *Clostridia*. This is in agreement with *Clostridia* being the dominating class in the DNA used for the library (Figure 1), highlighting the complementary nature of the culture dependent and independent approaches used in this study.

The two enzymes were successfully expressed in an optimized *E. coli* expression system and the activities of

crude extracts were characterized with regards to temperature and pH (Figure 3). Both  $\beta$ -galactosidases showed similar profiles with a temperature optimum around 37°C and a pH optimum around pH 6. Both enzymes retained 20-30% activity at 10°C. These profiles are similar to a  $\beta$ -galactosidase from the Antarctic soil bacterium *Paracoccus* sp. 32d [49], a  $\beta$ -galactosidase from an Arctic sea-ice relative of *Psychromonas antarctica* [26], as well as a  $\beta$ -galactosidase obtained from a metagenome from the Baltic Sea [50].

One possible application for  $\beta$ -galactosidases is the hydrolysis of lactose to galactose and glucose in order to generate lactose free milk for lactose intolerant people. An optimal enzyme for this purpose would be active at pH 6.7-6.8 and at 4-8°C [51], which correlates with the pH profiles of the two identified  $\beta$ -galactosidases. Several cold-active  $\beta$ -galactosidases with the ability to hydrolyze lactose have been reported [45,49,50,52-54] and interestingly, recombinantly produced  $\beta$ -galactosidase from the Antarctic bacterium *Pseudoalteromonas haloplanctis* was able to outperform a commercial yeast enzyme at 4°C, where it retained around 20% of its activity even though the pH optimum was 8.5 [54]. The two enzymes from this study also show approximately 20% activity at 5°C and their pH optimum is around the pH of milk, suggesting that they could be good candidates for enzymatic hydrolysis of lactose. Therefore, their ability to hydrolyze lactose at 37°C and 5°C was analyzed (Figure 4). BGal<sub>I3H5</sub> was unable to hydrolyze lactose, whereas BGal<sub>I17E2</sub> showed activity at both 37°C and 5°C. Further studies on e.g. inhibitory effects of glucose and galactose on the hydrolytic activity, performance in milk and potential transglycosylating activity are needed in order to determine the potential use of BGal<sub>I17E2</sub> for generation of lactose free milk.

#### **Conclusions**

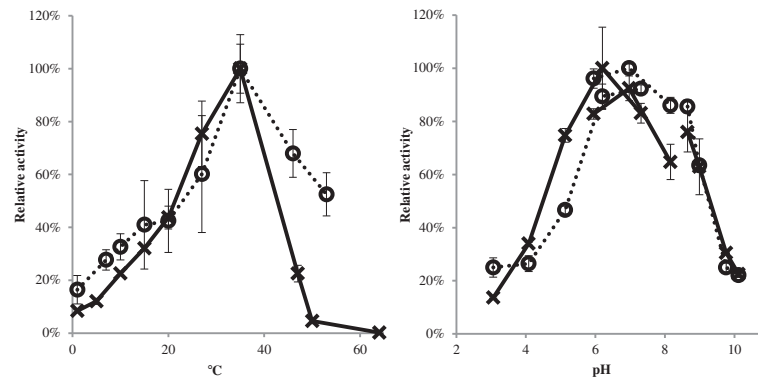
Using a combination of culture dependent and independent approaches we have demonstrated that the ikaite

**Table 6 Comparison of  $\beta$ -galactosidases**

	BGal <sub>I17E2</sub>	BGal <sub>I3H5</sub>	<i>A. ikkensis</i> $\beta$ -galactosidase	<i>E. coli</i> LacZ
Monomeric size	455aa	820aa	1,041aa	1,029aa
	52.75 kDa	92.25 kDa	119.13 kDa	116.97 kDa
GH family	1	2	2	2
Closest relatives	<i>Clostridium hathewayi</i> (53% identity) [GenBank:WP_006771328]	<i>Roseburia hominis</i> (51% identity) [GenBank:YP_004838333]	-	-
Arg	3.3%	4.5%	4.0%	6.4%
Arg/Arg + Lys ratio	0.31	0.44	0.42	0.77
Pro	3.7%	3.3%	4.8%	6.1%
Active site	WFA <b>E</b> YTKVM	LVTE <b>Y</b> NGHM	ILCE <b>F</b> SHAM	ILCE <b>Y</b> AHAM

Comparison of the  $\beta$ -galactosidases from the two library clones, IKA17E2 (BGal<sub>I17E2</sub>) and IKA3H5 (BGal<sub>I3H5</sub>), to  $\beta$ -galactosidases from the psychrotrophic *A. ikkensis* [45] and the mesophilic *E. coli* (LacZ) [48]. Active site glutamic acid (E) highlighted in bold.

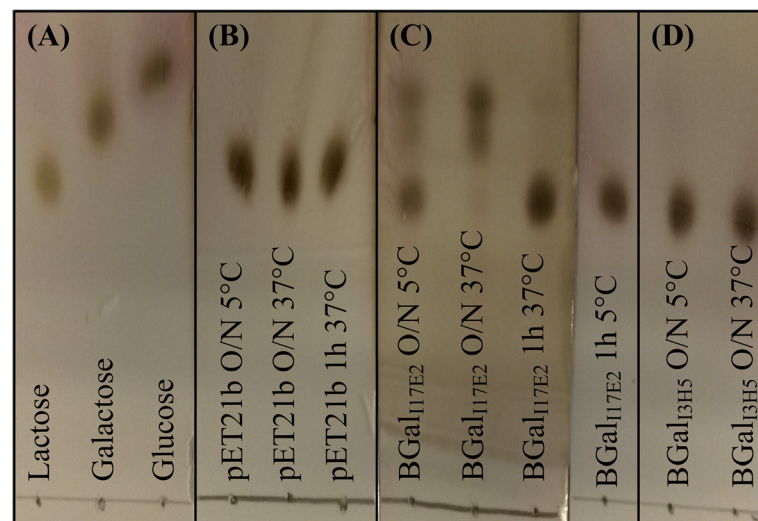




**Figure 3** Temperature and pH profiles of  $\beta$ -galactosidases. Temperature (left) and pH (right) profile of crude extracts of the two  $\beta$ -galactosidases BGal<sub>13H5</sub> (dotted line) and BGal<sub>117E2</sub> (solid line) with ortho-nitrophenyl- $\beta$ -galactoside (ONPG) as substrate. Error bars show standard deviations from triplicate experiments.

columns are a rich source of cold and/or alkaline-active enzymes. The two approaches complemented each other by targeting different phylogenetic groups of bacteria. The strength of the cultured strain collection was the high hit rate, although there was a significant degree of phylogenetic redundancy in the collection and most of the native bacterial isolates were closely related to previously characterized strains. Functional expression had a very low hit-rate, but the identified sequences were immediately available in a relevant production organism. The identified  $\alpha$ -amylase and  $\beta$ -galactosidases showed characteristics of cold-adapted enzymes, and BGal<sub>117E2</sub> was able to hydrolyze lactose at low temperature. Metagenomic sequencing of the library revealed a much higher degree of

novelty among the phylogenetic groups covered, but also that most of the potential activities were not expressed. Interesting candidate genes could therefore be chosen for heterologous expression in future studies. A useful combination of culture dependent and independent approaches is to use genomic DNA from natural isolates for functional expression, in order to pick up easily produced enzymes in a relevant host. Such an approach has been applied to identify  $\alpha$ -amylases from *Paenibacillus* sp. [55] and *Halothermothrix orenii* [56] and could be similarly applied to isolates from the ikaite strain collection. This study shows that the combination of approaches for bioprospecting can be highly beneficial in the discovery of novel enzyme activities, and in addition to the immediate



**Figure 4** TLC analysis of lactose hydrolysis at 5°C and 37°C. (A) Controls of Lactose, Galactose and Glucose (each 5 mg/ml). (B) Lactose (5 mg/ml) incubated with crude extract of *E. coli* carrying the empty pET21b vector. (C) Lactose (5 mg/ml) incubated with crude extract of BGal<sub>117E2</sub>. (D) Lactose (5 mg/ml) incubated with crude extract of BGal<sub>13H5</sub>.

relevance of cold- and alkaline-active enzymes in industrial applications, it also underlines the significance of the ikaite columns as a unique biological resource for bioprospecting.

## Materials and methods

### Bacterial strains and vectors

The bacterial strains and vectors used in this study are presented in Table 7. Media compositions are reported for the specific experiments.

### Establishment and screening of a strain collection

Material from ikaite columns collected from 2001 to 2011 and stored at  $-18^{\circ}\text{C}$  was spread on R2 plates buffered to pH 10.4 [16] supplemented with NaCl (10 g/L) and incubated at  $10^{\circ}\text{C}$ . Single colonies were picked, streaked to purity, grown in liquid R2 in 96 well format, supplemented with glycerol (15%) and stored at  $-80^{\circ}\text{C}$ . Screening of the strain collection was performed on 1/10 R2 plates buffered to pH 10.4 with NaCl (1 g/L) and supplemented with 0.05% starch (w/v) and the following substrates, all at a final concentration of 20  $\mu\text{g/ml}$ : 5-bromo-4-chloro-3-indolyl phosphate disodium salt (BCIP), 5-bromo-4-chloro-3-indolyl- $\beta$ -d-galactopyranoside (X-gal), and 5-bromo-4-chloro-3-indolyl- $\alpha$ -d-galactopyranoside (X- $\alpha$ -gal) for the detection of intracellular phosphatase,  $\beta$ -galactosidase and  $\alpha$ -galactosidase, respectively. Extracellular protease,  $\alpha$ -amylase,  $\beta$ -glucanase, cellulase,  $\beta$ -xy lanase and  $\beta$ -mannanase were identified with the azurine-cross linked (AZCL)-coupled substrates (Megazyme, Wicklow, Ireland): AZCL-casein, AZCL-amylose, AZCL-curdlan and -pachyman, AZCL-cellulose, AZCL-xy lan and AZCL-galactomannan, respectively (final concentration

of 0.05% w/v). Strains were transferred using a hand held 96 well pin replicator, and plates were incubated at  $10^{\circ}\text{C}$ ,  $20^{\circ}\text{C}$  and  $28^{\circ}\text{C}$  for 30 days and scored continuously for color development.

### Phylogenetic analysis of the strain collection

Template for 16S rRNA gene PCR was obtained by transferring colony material to 50  $\mu\text{l}$  demineralized water, boiling for 10 min followed by cooling and centrifugation and using 2  $\mu\text{l}$  of the supernatant as template. The PCR reaction (30  $\mu\text{l}$ ) was performed with 200  $\mu\text{M}$  of each dNTP, 0.3  $\mu\text{M}$  of each primer (27 F: AGAGTTTGATCM TGGCTCAG and BAC805R: GACTACCAGGGTATC TAATCC), 1x Phusion HF buffer and 0.02 U/ $\mu\text{l}$  Phusion HotStart DNA polymerase (Finnzymes, Vantaa, Finland). The PCR program consisted of an initial denaturation at  $98^{\circ}\text{C}$  for 3 min, followed by 30 cycles of  $98^{\circ}\text{C}$  for 15 s,  $55^{\circ}\text{C}$  for 45 s,  $72^{\circ}\text{C}$  for 60 s, and a final elongation at  $72^{\circ}\text{C}$  for 10 min. Sequencing was performed using the primer BAC338F (ACTCCTACGGGAGGCAG) sequencing the V3 and V4 regions of the 16S rRNA gene. Sequences were end-trimmed with a phred quality score limit of 20 and manually inspected to remove low quality sequences. Phylogenetic affiliation was determined by Blast analysis against the GenBank 16S Microbial database in CLC Main Workbench ([www.clcbio.com](http://www.clcbio.com)) with default settings.

### Extraction and preparation of DNA and establishment of functional expression library

Material from the interior of 10 different ikaite columns was collected on site directly after harvest of the columns in August 2011. The material was homogenized and kept at  $5^{\circ}\text{C}$ . To determine the total diversity, DNA was extracted from the pooled material with the PowerLyzer

**Table 7 Bacterial strains and vectors used in this study**

Strain or vector	Description	Source or reference
<b>Strains</b>		
<i>E. coli</i> MegaX DH10B T1 <sup>R</sup>	F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ), $\Phi$ 80/ <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 recA1 endA1 araD139 <math>\Delta</math>(<i>ara, leu</i>)7697 <i>galU galK</i> <math>\lambda^{-}</math> <i>rpsL nupG tonA</i></i>	Life Technologies
<i>E. coli</i> Tuner <sup>TM</sup>	F- <i>ompT hsdSB</i> ( <i>r<sub>B</sub>- m<sub>B</sub>-</i> ) <i>gal dcm lacY1</i>	Novagen
IKA3C6	MegaX DH10B T1 <sup>R</sup> <i>E. coli</i> carrying metagenomic library clone with $\alpha$ -amylase activity	This study
IKA16D10	MegaX DH10B T1 <sup>R</sup> <i>E. coli</i> carrying metagenomic library clone with $\alpha$ -amylase activity	This study
IKA28E6	MegaX DH10B T1 <sup>R</sup> <i>E. coli</i> carrying metagenomic library clone with $\alpha$ -amylase activity	This study
IKA3H5	MegaX DH10B T1 <sup>R</sup> <i>E. coli</i> carrying metagenomic library clone with $\beta$ -galactosidase activity	This study
IKA17E2	MegaX DH10B T1 <sup>R</sup> <i>E. coli</i> carrying metagenomic library clone with $\beta$ -galactosidase activity	This study
<b>Plasmids</b>		
pGNS-BAC	Gram negative BAC shuttle vector	[59]
mod.pGNS-BAC	pGNS-BAC modified to include a multiple cloning site with four unique restriction sites ( <i>Apa</i> I, <i>Bsp</i> DI, <i>Bbv</i> CI and <i>Nsi</i> I) into the <i>Hind</i> III site	This study
pET21b	<i>E. coli</i> expression vector	Novagen

PowerSoil DNA Isolation Kit (MoBio, Carlsbad, CA, USA) modified with G1-blocker (Carsten Suhr Jacobsen, GEUS, Denmark). Intact cells were extracted after 110 days using a modified version of the method developed by Kallmeyer et al. [57]: In brief, a slurry was made by mixing 130 g ikaite material with 450 ml 0.9% NaCl in a Warring blender at low speed. The slurry was centrifuged at  $3,000 \times g$  for 5 min and pellets were resuspended in 100 ml 0.9% NaCl and harvested at  $3,000 \times g$  for 5 min; this was done twice. Pellets were resuspended in 100 ml 0.9% NaCl with 0.1% Na<sub>3</sub>N and 17 ml methanol and 17 ml detergent mix (100 mM EDTA-Na<sub>2</sub>, 100 mM Na<sub>4</sub>O<sub>7</sub>P<sub>2</sub>, 1% (v/v) Tween 80) were added and cells were detached by vortexing at 1400 rpm for 60 min. Cells were separated from particles by centrifugation at  $500 \times g$  for 2 min and collected from the supernatant by centrifugation at  $10,000 \times g$  for 10 min. All steps were performed at 4°C. DNA from extracted cells was obtained by resuspending pellets in 1.5 ml STET-buffer (8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM EDTA, 50 mM Tris-HCl) with lysozyme (2 mg/ml) and incubating at 37°C for 30 min before adding SDS (2% final concentration) and continuing incubation for 30 min at 37°C followed by 30 min at 65°C. Finally, DNA was extracted using traditional phenol/chloroform extraction [58]. High molecular weight (HMW) DNA (>8 kb) was gel-purified using a QIAquick Gel Extraction Kit (Qiagen) without dyes. Purified DNA was used as template for MDA with Repli-g Mini Kit (Qiagen) following the standard protocol.

A modified version of the pGNS-BAC vector [59], mod.pGNS-BAC, was produced by introducing a new multiple cloning site with four unique restriction sites (*Apa*LI, *Bsp*DI, *Bbv*CI and *Nsi*I) into the *Hind*III site of the original vector to reduce the self-ligation rate. The vector was digested with *Apa*LI and *Nsi*I (New England Biolabs, Ipswich, MA, USA), gel-purified with a QIAquick Gel Extraction Kit and phosphatase treated with Shrimp Alkaline Phosphatase (New England Biolabs, Ipswich, MA, USA). MDA DNA was partially digested with *Apa*LI and *Nsi*I and HMW DNA (>8 kb) from the digestion was gel-purified using GELase (Epicentre, Chicago, IL, USA). The purified DNA was ligated into the mod.pGNS-BAC vector using T4 DNA ligase (New England Biolabs, Ipswich, MA, USA), transformed into MegaX DH10B T1<sup>R</sup> electrocompetent *E. coli* cells (Life Technologies), and spread onto LB library plates supplemented with 12.5 µg/ml chloramphenicol. Colonies were picked into 96 well format in LB with 10% glycerol and 12.5 µg/ml chloramphenicol using a QPix colony picker (Genetix - Molecular devices, Workingham, UK), and grown over night at 37°C with shaking before being stored at -80°C. Randomly picked clones (29) were analyzed for insert size by purification of the BAC-vector and digestion with *Apa*LI and *Nsi*I, followed by gel electrophoresis.

### Pyrosequencing analysis of bacterial diversity

A fragment covering the V3 and V4 hypervariable regions of the 16S rRNA gene from bacteria and archaea was amplified from DNA extracted after sampling, cell extraction, and MDA using the primers 341 F (CCTAYGGGRBGCASCAG) and 806R (GGACTACNNGGGTATCTAAT). Amplification, pyrosequencing and phylogenetic analysis was performed as previously described [60]. Briefly, pyrosequencing was performed at The National High-Throughput Sequencing Centre at University of Copenhagen on a Genome Sequencer FLX pyrosequencing system (454 Life Sciences, Roche, Branford, CT, USA). Trimming and quality-filtering of the resulting sequences was performed using Biopieces ([www.biopieces.org](http://www.biopieces.org)) using a minimum average phred quality score of 25. Sequences shorter than 250 bases and sequences containing more than one ambiguous nucleotide were discarded. Phylogenetic analysis was performed using the QIIME pipeline ([www.qiime.org](http://www.qiime.org)) [61]. Operational taxonomic units (OTUs) were clustered at 97% identity using the USEARCH [62] quality filter pipeline in QIIME, which included reference-based detection of chimeric sequences and removal of clusters containing only one sequence (singletons). The taxonomy of the resulting cleaned set of OTUs was determined using the RDP classifier at a confidence threshold of 50% and the Greengenes taxonomy database version 13\_05 (<http://greengenes.lbl.gov/>) [63].

### Screening of functional expression library

The functional expression library was screened on LB plates supplemented with 12.5 µg/ml chloramphenicol, 0.01% (w/v) arabinose and the appropriate substrates as described above for the strain collection. In addition, lipolytic activity was screened on plates containing 1% tributyrin. Strains were transferred to plates using a hand held 96 well pin replicator, and the clones were grown over night at 37°C and then transferred to 20°C for two days followed by transfer to 15°C. Enzyme activities were scored continuously.

### Characterization of α-amylase

Open reading frames (ORFs) in the three BAC library clones encoding α-amylase activity (IKA3C6, IKA16D10 and IKA28E6) were identified by BlastX analysis on a combination of BAC and metagenome sequences. The open reading frame (ORF) encoding α-amylase in IKA3C6 (GenBank: KJ790257) was cloned into the expression vector pET21b with a C-terminal 6x His-tag and transformed into *E. coli* Tuner cells (Merck Millipore, Darmstadt, Germany). Positive clones were identified on LB plates supplemented with 100 µg/ml ampicillin, 1 mM IPTG and AZCL-amylose (0.05% w/v). The α-amylase enzyme was produced from 50 ml liquid culture grown at

37°C at 150 rpm. Expression was induced at an OD<sub>600</sub> of 0.8 by addition of 1 mM IPTG and incubation was continued at 20°C for 16 h, before harvesting at 10,000 × g for 10 min. Cell pellets were resuspended in 2 ml 100 mM phosphate-buffer pH 7.6 and intracellular proteins were extracted by bead beating in a FastPrep homogenizer (Thermo Scientific) with 3 × 25 s at setting 5.5 with cooling on ice in between. The lysed cells were centrifuged at 10,000 × g for 5 min at 4°C and the supernatant was collected. Temperature and pH profiles were produced on the crude extract using an assay for reducing-end sugars [64] after incubation in 100 mM buffer (Tris–HCl buffer pH 8.6 for the temperature profile and pH 6–9, and glycine–NaOH buffer for pH 8–10) with 5 mg/ml amylopectin as substrate. Assays for the pH profile were performed at 20°C.

#### Characterization of β-galactosidases

ORFs in the two BAC library clones encoding β-galactosidase activity, IKA17E2 (GenBank: KJ790256) and IKA3H5 (GenBank: KJ790255) were identified by BlastX analysis on a combination of BAC and metagenome sequences. Genes were cloned into the expression vector pET21b with a C-terminal 6x His-tag and transformed into *E. coli* Tuner cells (Merck Millipore, Darmstadt, Germany). Positive clones were identified on LB plates supplemented with 100 μg/ml ampicillin, 1 mM IPTG and 10 mg/ml X-gal. The β-galactosidase enzymes were produced from 50 ml liquid cultures grown overnight at 37°C at 150 rpm, and intracellular extracts were obtained as described above. Temperature and pH profiles were produced on the crude extracts using an ONPG assay [65] in 100 mM buffer (phosphate buffer pH 7 for the temperature profile, citrate-phosphate for pH 3–7, phosphate buffer for pH 6–8, and glycine–NaOH buffer for pH 8–10). Assays for the pH profile were performed at 20°C. Lactose assays were performed by adding 5 μl crude enzyme extract to 1 ml lactose (5 mg/ml) and incubating at 5°C or 37°C for 1 h or overnight before analysis on TLC aluminum sheets (Merck Millipore, Darmstadt, Germany) running in 1-butanol:2-propanol:water (3:12:4).

#### Metagenomic sequencing and analysis

DNA for sequencing of the functional expression library was extracted from a pool of liquid cultures of all *E. coli* BAC clones using the BACMAX DNA Purification Kit (Epicentre). The metagenome sequence was obtained by 2 × 250 bp paired-end sequencing of a short-insert library on an Illumina MiSeq system at The National High-Throughput DNA Sequencing Centre at University of Copenhagen. The resulting sequences were cleaned using Biopieces (<http://www.biopieces.org>) by trimming of adaptors and poor-quality sequence from sequence

ends and removal of sequences containing ambiguous nucleotides or with an average quality score of less than 30. Before assembly, all sequences showing at least 95% identity to the mod.pGNS-BAC vector or to the *E. coli* K12 genome were discarded. Assembly was performed in CLC Assembly Cell (<http://www.clcbio.com>) (see Table 4 for sequence statistics). Initial analysis and gene-calling of the assembled contigs was carried out on the MG-RAST server (<http://metagenomics.anl.gov/>). The predicted full-length and partial coding sequences were annotated by a batch search against the Pfam protein family database (<http://pfam.sanger.ac.uk/>) with an e-value cut-off of 1e-5. The resulting identified protein domains were searched for relevant enzyme targets using information from the Pfam database, the carbohydrate-active enzymes database (CAZy; [www.cazy.org](http://www.cazy.org)) and the MEROPS peptidase database (<http://merops.sanger.ac.uk/>). Protein localization was predicted using PSORTb version 3.0 for both gram-positive and gram-negative bacteria (<http://www.psорт.org/psорт/>).

#### Additional file

**Additional file 1: Table S1.** Predicted proteases and glycosyl hydrolases from sequencing of the functional expression library. Predicted Pfam domains from the metagenomic sequence of the functional expression library with relevant glycosyl hydrolase (GH) or protease function.

#### Abbreviations

AZCL: Azurine cross-linked; BAC: Bacterial artificial chromosome; BCIP: 5-bromo-4-chloroindoxyl phosphate; GH: Glycosyl hydrolase; HMW: High molecular weight; MDA: Multiple displacement amplification; ONPG: Ortho-nitrophenyl-β-galactoside; ORF: Open reading frame; OTUs: Operational taxonomic units; X-gal: 5-Bromo-4-chloroindoxyl-β-d-galactopyranoside; X-α-gal: 5-Bromo-4-chloroindoxyl-α-d-galactopyranoside.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

JKV carried out the study, analyzed data and results, participated in the design and drafted the manuscript. MIG performed the bioinformatics analysis, participated in the design and helped to draft the manuscript. PSG conceived the study, and participated in its design and coordination and helped to draft the manuscript. All authors have read and approved the final manuscript.

#### Acknowledgements

The Government of Greenland is thanked for permission to sample ikaite material. Mark Lever is acknowledged for help and suggestions on establishing a protocol for extracting intact cells from ikaite material. Claus Böttcher Jørgensen is thanked for access to the colony picker, and Mark Liles is thanked for the original pGNS-BAC vector. Agnès Garcías López is acknowledged for assisting with screening of the strain collection, and Aviaja Lyberth Hauptmann, Eva Selsner-Rasmussen, and Jeanette Eva Lylloff are acknowledged for establishing the strain collection. This work was supported by a grant to JKV from University of Copenhagen and by a grant to PS from the Danish Environmental Protection Agency.

Received: 12 March 2014 Accepted: 13 May 2014  
Published: 20 May 2014

## References

- Cavicchioli R, Charlton T, Ertan H, Mohd OS, Siddiqui KS, Williams TJ: **Biotechnological uses of enzymes from psychrophiles.** *Microb Biotechnol* 2011, **4**:449–460.
- Huston AL: **Biotechnological Aspects of Cold-Adapted Enzymes.** In *Psychrophiles: From Biodiversity to Biotechnology.* New York: Springer; 2008.
- Cieslinski H, Bialkowska A, Tkaczuk K, Dlugolecka A, Kur J, Turkiewicz M: **Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library.** *Pol J Microbiol* 2009, **58**:199–204.
- Yu Y, Li HR, Zeng YX, Chen B: **Bacterial diversity and bioprospecting for cold-active hydrolytic enzymes from culturable bacteria associated with sediment from Nella Fjord, Eastern Antarctica.** *Mar Drugs* 2011, **9**:184–195.
- Reddy PVV, Rao SSSN, Pratibha MS, Sailaja M, Kavya B, Manorama RR, Singh SM, Srinivas TNR, Shivaji S: **Bacterial diversity and bioprospecting for cold-active enzymes from culturable bacteria associated with sediment from a melt water stream of Midtre Lovenbreen glacier, an Arctic glacier.** *Research in Microbiology* 2009, **160**:538–546.
- Prasad S, Manasa P, Buddhi S, Tirunagari P, Begum Z, Rajan S, Shivaji S: **Diversity and Bioprospective Potential (Cold-Active Enzymes) of Cultivable Marine Bacteria from the Subarctic Glacial Fjord, Kongsfjorden.** *Curr Microbiol* 2013, **68**:233–238.
- Simon C, Herath J, Rockstroh S, Daniel R: **Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice.** *Appl Environ Microbiol* 2009, **75**:2964–2968.
- Jeon JH, Kim JT, Kang SG, Lee JH, Kim SJ: **Characterization and its potential application of two esterases derived from the arctic sediment metagenome.** *Mar Biotechnol (NY)* 2009, **11**:307–316.
- Goodfellow M, Fiedler HP: **A guide to successful bioprospecting: informed by actinobacterial systematics.** *Antonie Van Leeuwenhoek* 2010, **98**:119–142.
- Petrovskaya LE, Novototskaya-Vlasova KA, Spirina EV, Khokhlova GV, Rivkina EM, Gilichinsky DA, Dolgikh DA, Kirpichnikov MP: **Lipolytic enzymes of microorganisms from permafrost cryopegs.** *Dokl Biol Sci* 2012, **445**:279–282.
- Gupta R, Beg QK, Lorenz P: **Bacterial alkaline proteases: molecular approaches and industrial applications.** *Appl Microbiol Biotechnol* 2002, **59**:15–32.
- Horikoshi K: **Alkaliphiles: some applications of their products for biotechnology.** *Microbiol Mol Biol Rev* 1999, **63**:735–750. table.
- Buchardt B, Seaman P, Stockmann G, Vous M, Wilken U, Duwel L, Kristiansen A, Jenner C, Whitticar MJ: **Submarine columns of ikaite tufa.** *Nature* 1997, **390**:129–130.
- Andersen DT, Sumner DY, Hawes I, Webster-Brown J, McKay CP: **Discovery of large conical stromatolites in Lake Untersee, Antarctica.** *Geobiology* 2011, **9**:280–293.
- Jungblut AD, Wood SA, Hawes I, Webster-Brown J, Harris C: **The Pyramid Trough Wetland: environmental and biological diversity in a newly created Antarctic protected area.** *FEMS Microbiol Ecol* 2012, **82**:356–366.
- Stougaard P, Jorgensen F, Johnsen MG, Hansen OC: **Microbial diversity in ikaite tufa columns: an alkaline, cold ecological niche in Greenland.** *Environ Microbiol* 2002, **4**:487–493.
- Schmidt M, Prieme A, Stougaard P: **Bacterial diversity in permanently cold and alkaline ikaite columns from Greenland.** *Extremophiles* 2006, **10**:551–562.
- Schmidt M, Prieme A, Stougaard P: **Arsukibacterium ikkense gen. nov., sp. nov., a novel alkaliphilic, enzyme-producing gamma-Proteobacterium isolated from a cold and alkaline environment in Greenland.** *Syst Appl Microbiol* 2007, **30**:197–201.
- Schmidt M, Prieme A, Johansen A, Stougaard P: **Alkalilactibacillus ikkensis, gen. nov., sp. nov., a novel enzyme-producing bacterium from a cold and alkaline environment in Greenland.** *Extremophiles* 2012, **16**:297–305.
- Schmidt M, Prieme A, Stougaard P: **Rhodonellum psychrophilum gen. nov., sp. nov., a novel psychrophilic and alkaliphilic bacterium of the phylum Bacteroidetes isolated from Greenland.** *Int J Syst Evol Microbiol* 2006, **56**:2887–2892.
- Schmidt M, Larsen DM, Stougaard P: **A lipase with broad temperature range from an alkaliphilic gamma-proteobacterium isolated in Greenland.** *Environ Technol* 2010, **31**:1091–1100.
- Vester JK, Lylloff JE, Glaring MA, Stougaard P: **Microbial Diversity and Enzymes in Ikaite Columns; A Cold and Alkaline Environment in Greenland.** In *Cellular Origin, Life in Extreme Habitats and Astrobiology.* Edited by Seckbach J, Oren A, Stan-Lotter H. Netherlands: Springer; 2013:365–380. ISBN: 978-94-007-6487-3.
- Amann RL, Ludwig W, Schleifer KH: **Phylogenetic identification and in situ detection of individual microbial cells without cultivation.** *Microbiol Rev* 1995, **59**:143–169.
- Achtman M, Wagner M: **Microbial diversity and the genetic nature of microbial species.** *Nat Rev Microbiol* 2008, **6**:431–440.
- Vester JK, Glaring MA, Stougaard P: **Improving diversity in cultures of bacteria from an extreme environment.** *Can J Microbiol* 2013, **59**:581–586.
- Groudieva T, Kambourova M, Yusef H, Royer M, Grote R, Trinks H, Antranikian G: **Diversity and cold-active hydrolytic enzymes of culturable bacteria associated with Arctic sea ice, Spitzbergen.** *Extremophiles* 2004, **8**:475–488.
- Yi H, Schumann P, Chun J: **Demequina aestuarii gen. nov., sp. nov., a novel actinomycete of the suborder Micrococccineae, and reclassification of Cellulomonas fermentans Bagnara et al. 1985 as Actinotalea fermentans gen. nov., comb. nov.** *Int J Syst Evol Microbiol* 2007, **57**:151–156.
- Bagnara C, Gaudin C, Belaich JP: **Purification and partial characterization of two extracellular endoglucanases from Cellulomonas fermentans.** *Biochem Biophys Res Commun* 1986, **140**:219–229.
- Yilmaz S, Allgaier M, Hugenholtz P: **Multiple displacement amplification compromises quantitative analysis of metagenomes.** *Nat Methods* 2010, **7**:943–944.
- Zhang K, Song L, Dong X: **Proteiniclasticum ruminis gen. nov., sp. nov., a strictly anaerobic proteolytic bacterium isolated from yak rumen.** *Int J Syst Evol Microbiol* 2010, **60**:2221–2225.
- Takai K, Moser DP, Onstott TC, Spoelstra N, Pflfner SM, Dohnalkova A, Fredrickson JK: **Alkaliphilus transvaalensis gen. nov., sp. nov., an extremely alkaliphilic bacterium isolated from a deep South African gold mine.** *Int J Syst Evol Microbiol* 2001, **51**:1245–1256.
- Garnova ES, Zhilina TN, Tourova TP, Lysenko AM: **Anoxynatronum sibiricum gen. nov., sp. nov. alkaliphilic saccharolytic anaerobe from cellulolytic community of Nizhnee Beloe (Transbaikal region).** *Extremophiles* 2003, **7**:213–220.
- Ravot G, Magot M, Fardeau ML, Patel BK, Thomas P, Garcia JL, Ollivier B: **Fusibacter paucivorans gen. nov., sp. nov., an anaerobic, thiosulfate-reducing bacterium from an oil-producing well.** *Int J Syst Bacteriol* 1999, **49 Pt 3**:1141–1147.
- Milford AD, Achenbach LA, Jung DO, Madigan MT: **Rhodobaca bogoriensis gen. nov. and sp. nov., an alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes.** *Arch Microbiol* 2000, **174**:18–27.
- Delmont TO, Robe P, Clark I, Simonet P, Vogel TM: **Metagenomic comparison of direct and indirect soil DNA extraction approaches.** *J Microbiol Methods* 2011, **86**:397–400.
- Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD: **The great screen anomaly—a new frontier in product discovery through functional metagenomics.** *Appl Microbiol Biotechnol* 2012, **93**:1005–1020.
- Gabor EM, Alkema WB, Janssen DB: **Quantifying the accessibility of the metagenome by random expression cloning techniques.** *Environ Microbiol* 2004, **6**:879–886.
- Sharma S, Khan FG, Qazi GN: **Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas.** *Appl Microbiol Biotechnol* 2010, **86**:1821–1828.
- Berlemont R, Delsaute M, Pipers D, D'Amico S, Feller G, Galleni M, Power P: **Insights into bacterial cellulose biosynthesis by functional metagenomics on Antarctic soil samples.** *ISME J* 2009, **3**:1070–1081.
- Wang K, Li G, Yu SQ, Zhang CT, Liu YH: **A novel metagenome-derived beta-galactosidase: gene cloning, overexpression, purification and characterization.** *Appl Microbiol Biotechnol* 2010, **88**:155–165.
- Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, Ryu S: **Characterization of a novel amyolytic enzyme encoded by a gene from a soil-derived metagenomic library.** *Appl Environ Microbiol* 2004, **70**:7229–7235.
- Bhat A, Riyaz-Ul-Hassan S, Ahmad N, Srivastava N, Johri S: **Isolation of cold-active, acidic endocellulase from Ladakh soil by functional metagenomics.** *Extremophiles* 2013, **17**:229–239.
- Lammle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, Rupp S: **Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning.** *J Biotechnol* 2007, **127**:575–592.
- Feller G, Lonhienne T, Deroanne C, Libioule C, Van BJ, Gerday C: **Purification, characterization, and nucleotide sequence of**

- thermolabile alpha-amylase from the antarctic psychrotroph *Alteromonas haloplanctis* A23. *J Biol Chem* 1992, **267**:5217–5221.
45. Schmidt M, Stougaard P: Identification, cloning and expression of a cold-active beta-galactosidase from a novel Arctic bacterium, *Alkalilactibacillus ikkense*. *Environ Technol* 2010, **31**:1107–1114.
  46. Feller G, Gerday C: Psychrophilic enzymes: molecular basis of cold adaptation. *Cell Mol Life Sci* 1997, **53**:830–841.
  47. Rainey FA: Class II. Clostridia class nov. In *Bergey's Manual of Systematic Bacteriology. Volume 3*. 2nd edition. New York: Springer; 2009:736.
  48. Gebler JC, Aebersold R, Withers SG: Glu-537, not Glu-461, is the nucleophile in the active site of (lac Z) beta-galactosidase from *Escherichia coli*. *J Biol Chem* 1992, **267**:11126–11130.
  49. Wierzbicka-Wos A, Cieslinski H, Wanarska M, Kozłowska-Tylingo K, Hildebrandt P, Kur J: A novel cold-active beta-D-galactosidase from the *Paracoccus* sp. 32d—gene cloning, purification and characterization. *Microb Cell Fact* 2011, **10**:108.
  50. Wierzbicka-Wos A, Bartasun P, Cieslinski H, Kur J: Cloning and characterization of a novel cold-active glycoside hydrolase family 1 enzyme with beta-glucosidase, beta-fucosidase and beta-galactosidase activities. *BMC Biotechnol* 2013, **13**:22.
  51. Trimbur DE, Gutshall KR, Prema P, Brenchley JE: Characterization of a psychrotrophic *Arthrobacter* gene and its cold-active beta-galactosidase. *Appl Environ Microbiol* 1994, **60**:4544–4552.
  52. Bialkowska AM, Cieslinski H, Nowakowska KM, Kur J, Turkiewicz M: A new beta-galactosidase with a low temperature optimum isolated from the Antarctic *Arthrobacter* sp. 20B: gene cloning, purification and characterization. *Arch Microbiol* 2009, **191**:825–835.
  53. Nakagawa T, Fujimoto Y, Ikehata R, Miyaji T, Tomizuka N: Purification and molecular characterization of cold-active beta-galactosidase from *Arthrobacter psychrolactophilus* strain F2. *Appl Microbiol Biotechnol* 2006, **72**:720–725.
  54. Hoyoux A, Jennes I, Dubois P, Genicot S, Dubail F, Francois JM, Baise E, Feller G, Gerday C: Cold-adapted beta-galactosidase from the Antarctic psychrophile *Pseudoalteromonas haloplanktis*. *Appl Environ Microbiol* 2001, **67**:1529–1535.
  55. Rajesh T, Kim YH, Choi YK, Jeon JM, Kim HJ, Park SH, Park HY, Choi KY, Kim H, Kim HJ, Lee SH, Yang YH: Identification and functional characterization of an alpha-amylase with broad temperature and pH stability from *Paenibacillus* sp. *Appl Biochem Biotechnol* 2013, **170**:359–369.
  56. Mijts BN, Patel BK: Cloning, sequencing and expression of an alpha-amylase gene, *amyA*, from the thermophilic halophile *Halothermothrix orenii* and purification and biochemical characterization of the recombinant enzyme. *Microbiology* 2002, **148**:2343–2349.
  57. Kallmeyer J, Smith DC, Spivack AJ, D'Hondt S: New cell extraction procedure applied to deep subsurface sediments. *Limnology and Oceanography-Methods* 2008, **6**:236–245.
  58. Sambrook J, Russell DW: *Molecular cloning: A Laboratory Manual*, Volume 1. 3rd edition. New York: Cold Spring Harbor Laboratory Press; 2001.
  59. Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W, Liles MR: Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene* 2010.
  60. Michelsen CF, Pedas P, Glaring MA, Schjoerring JK, Stougaard P: Bacterial diversity in Greenlandic soils as affected by potato cropping and inorganic versus organic fertilization. *Polar Biology* 2014, **37**:61–71.
  61. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JL, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunencko T, Zaneveld J, Knight R: QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010, **7**:335–336.
  62. Edgar RC: Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010, **26**:2460–2461.
  63. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006, **72**:5069–5072.
  64. Anthon GE, Barrett DM: Determination of reducing sugars with 3-methyl-2-benzothiazolinonehydrazone. *Anal Biochem* 2002, **305**:287–289.
  65. Miller JH: In *Experiments in Molecular Genetics*. 3rd edition. Edited by Miller JH. Cold Spring Harbor: Cold Spring Harbor Laboratory; 1972:466. ISBN 9780879691066.

doi:10.1186/1475-2859-13-72

Cite this article as: Vester *et al.*: Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. *Microbial Cell Factories* 2014 **13**:72.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

