World Journal of Microbiology & Biotechnology **20**: 801–809, 2004.

Use of low nutrient enrichments to access novel amylase genes in silent diversity of thermophiles

Cédric F.V. Hobel[1], Viggó T. Marteinsson[1,*], Sigurbjörg Hauksdóttir[1], Ólafur H. Fridjónsson[1], Sigurlaug Skírnisdóttir[1], Gudmundur Ó. Hreggvidsson[1,2] and Jakob K. Kristjánsson[1,2]
[1]*Prokaria Ltd., Gylfaflöt 5, sturlugata 7, IS-101 Reykjavik, Iceland*
[2]*Institute of Biology, University of Iceland, Sturlugatu 7, IS-101 Reykjavik, Iceland*
**Author for correspondence: Tel.: +354-570-7900, Fax: +354-570-7901, E-mail: viggo@prokaria.com*

Received 20 October 2003; accepted 4 March 2004

Keywords: 16S rRNA, amylase genes, enrichments, hot spring, PCR, thermophiles

Summary

By combining low nutrient enrichments and molecular methods, a high diversity of new amylase genes was detected in a neutral sulphide-rich hot spring in Iceland. Enrichments based on hot spring water and low concentrations of starch were used to select slow-growing, starch-degrading microorganisms. Six enrichments had in total 17 bacterial types detected by 16S rRNA analysis, mostly related to the *Thermus-Deinococcus* group, green non-sulphur bacteria, gram positives, and uncultivated new candidate divisions. No *Archaea* were found. The apparent 16S rRNA species composition of the enrichments was very different from that of the microbial mat in the same hot spring. DNA samples obtained from 4 enrichments and from hot spring biomass were screened by PCR for amylase genes in glycoside-hydrolase family 13. Degenerate primers, based on conserved amino acid sequences from multiple alignments of family 13, enabled the detection of 18 amylase sequence types in the enrichments, including α-amylases, α-glucosidases, 1,4-α-glucan branching enzymes, cyclomaltodextrin hydrolases, maltogenic amylases and neopullulanases, and unspecified family 13 glycoside-hydrolases. Only one unique neopullulanase sequence, also found in most of the enrichments, was detected in the hot spring biomass DNA. The results suggest that the enrichment method combined with sequence-based screening is an efficient way to access the silent, i.e. not detectable, gene diversity in natural environments.

Introduction

Developments in the starch-processing industry require continued discovery and development of new enzymes. Until now all commercial enzymes have been derived from cultivated bacteria or fungi. Notably, intensive research has been performed aiming at the isolation of unique thermostable and thermoactive amylases from thermophilic and hyperthermophilic microorganisms, therefore allowing more industrial processes to run at higher temperatures (Niehaus *et al.* 1999).

The realization that less than 1% of naturally occurring microorganisms can be isolated and grown in pure culture has created great interest in developing methods to get access to uncultivable microbes. New methods would allow the exploitation of a larger fraction of microbial diversity than what is available with the present technology (Amann *et al.* 1995). Direct microscopic count and diversity studies based on the analysis of 16S rRNA genes have shown that both the number and types of the microbes present in environmental samples exceed by several orders of magnitude the viable cell count done on plates (Amann *et al.* 1995; Skirnisdottir *et al.* 2000). The results have expanded our view of microbial diversity in different environments (Amann *et al.* 1995) and many novel microorganisms, yet uncultivated, have been identified by these techniques in extreme environments, such as in hot springs (Hugenholtz *et al.* 1998; Ward *et al.* 1998; Skirnisdottir *et al.* 2000; Hjorleifsdottir *et al.* 2001; Marteinsson *et al.* 2001). Although the use of molecular phylogenetic techniques has alleviated to some extent our reliance on culture techniques for the analysis of microbial diversity, there are many intrinsic limitations (Ward *et al.* 1998). The quality and amount of DNA are clear methodological limitations of the standard 16S rRNA approach and rare organisms (e.g. abundance below 0.1%) will hardly be detected. Moreover, the use of PCR for amplifying the 16S rRNA genes out of an environmental biomass sample seems to introduce a noticeable bias (Suzuki & Giovannoni 1996). This eventually leads to a distorted view of the "real world" (von Witzingerode *et al.* 1997).

The discrepancy observed between populations characterized by culture methods and those described by culture-independent diversity analysis (Jackson *et al.* 1996; Santegoeds *et al.* 1996; Ward *et al.* 1998; Hjorle-

ifsdottir *et al.* 2001) prompted us to use techniques that could bridge the gap between culture and non-culture methods. We describe here a technique inspired from oligotrophic cultivation methods that aims at the enrichment of low-abundancy microorganisms, and henceforth provide access to new amylase genes that otherwise would remain undetectable in the original hot spring biomass.

Materials and methods

Sample collection

Samples were collected in a sulphide-rich hot spring located on a 2–3 m high riverbank in Grensdalur, Iceland, previously studied by Skirnisdottir *et al.* (2000). Most of the biomass was growing in the basin overflow (20–30 cm from the source), which was going directly to the river 2 m below. The mat was several cm long, formed by long white and yellow filaments. The water temperature was 74 °C in the spring outlet and 65 °C at the biomass location in the overflow. The whole biomass from the basin and from the mat in the overflow was collected in a sterile flask. Some 30 l of hot spring water were also collected into sterile containers. Analysis of the sulphide was done in the field by mercury-acetate titration as described previously (Skirnisdottir *et al.* 2000). The spring water contained 20.0 mg of sulphide l^{-1}, the total dissolved salt concentration was 0.285 g l^{-1} and the pH was 7.0.

Low nutrient enrichments

Six static enrichments were prepared within 4 h after sampling using 500 ml of untreated spring water (Table 1). Erlenmeyer flasks were used for aerobic conditions, and bottles closed with a rubber stopper for anaerobic conditions. Effects of the pH and salinity were studied respectively by adjustments to pH 4 with sterile 1 M HCl, and by adjustments to 0.5% (w/v) NaCl with a sterile solution of 10% NaCl, when appropriate. Anaerobic conditions were achieved by replacing the headspace with nitrogen gas (N_2) and reducing the media with drops of a sterile solution of 0.025% $Na_2S \cdot 9H_2O$. The biomass sample was homogenized by vigorous shaking and the resultant slurry was diluted in series with spring water up to 10^{-4}. Enrichments were inoculated with 5 ml of the 10^{-4} dilution of the biomass mix. After 16 days of incubation at 65 °C yeast extract was added to all enrichments to give 0.005% final concentration, and the incubation continued. The enrichments were monitored with a Leica DM LB light microscope equipped with a phase-contrast oil immersion objective and were counted by using a Petroff–Hausser chamber (depth 0.02 mm) and stopped when the cell concentration had reached about 10^6–10^7 cells ml^{-1}.

Table 1. Media components added to untreated hot spring fluid for low nutrient enrichment.

Enrichment code	Starch (%) [a]	$(NH_4)_2SO_4$ (%)	Headspace [b]	pH	NaCl (%)	Incubation time (days)
HS2	0.1	1.0	Air			18
HS3	0.1	1.0	N_2			23
HS5	0.1	1.0	Air		0.5	18
HS6	0.1	1.0	Air	4.0		60
LS7	0.02	0.2	N_2			50
LS10	0.02	0.2	Air		0.5	60

Yeast extract was added to final concentration of 0.005% (wt/vol) after 16 days.
[a] wt/vol.
[b] N_2: anaerobic conditions.

Molecular analysis of the microbial diversity in the enrichments

Enrichments were centrifuged at 10,000 g for 30 min at 4 °C, the pellets were homogenized with a sterile glass mortar and DNA extracted as described by Marteinsson *et al.* (2001). The PCR amplifications of the 16S rRNA genes were performed as described by Skirnisdottir *et al.* (2000). Oligonucleotide primers 23FPL and 1391R were used for detection of *Archaea,* and F9, R805 and R1544 for *Bacteria.* The cloning of 16S rRNA genes was performed with the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Plasmid DNA from single colonies was isolated and screening of the clone library was performed by sequencing with R805 primer on an ABI 377 DNA sequencer by using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA). The sequences were analysed with Sequencher v4.0 (GeneCodes, Ann Arbor, MI). Sequences from the enrichments were selected and identified with BLAST searches (Altschul *et al.* 1997), then manually aligned with closely related sequences obtained from the Ribosomal Database Project (RDP) (Cole *et al.* 2003). Whole sequencing of selected clones was performed with the following sets of primers: reverse M13, R357, R805, R1195, R1544, forward M13, F9, F338 and F515 (Skirnisdottir *et al.* 2000). Phylogenetic analysis was performed with the ARB program (W. Ludwig *et al.* submitted for publication) by constructing maximum-likelihood and maximum-parsimony trees with the respective softwares included in the ARB package. Bacterial group filters were used to selectively exclude alignment portions for treeing such as uninformative or ambiguous columns, as well as to compute similarity percentages between clones and closest relative. The CHECK-CHIMERA program of the RDP server was used for searches of chimera artifacts.

Molecular detection of amylase genes

About 150 amino acid sequences of various amylolytic enzymes from the glycoside-hydrolase family 13 (Coutinho *et al.* 1999a, b) were retrieved from Pfam (Sonnhammer *et al.* 1997) and GenBank. Multiple

alignments were constructed with ClustalX v1.8 (Thompson *et al.* 1997) and the alignments were processed with Blockmaker (Henikoff *et al.* 1995) to generate blocks of conserved amino acid sequences. Primer design was performed with the Codehop method (Rose *et al.* 1998) using the Blockmaker resulting blocks as input. The primers were designed using either 'Equal' or '*Bacillus subtilis*' as codon usage biases and with a maximum degeneracy score of 32. The oligonuclotides were degenerate at the 3′ core region of length 11 bp across four codons of highly conserved amino acids. In contrast, they were non-degenerate at the 5′ region (consensus clamp region) of 13–29 bp with the most probable nucleotide predicted for each position.

All forward and reverse α-amylase screening primers were tested against each other in a matrix using DNA samples obtained from the enrichments and from the hot spring microbial mat as templates. The PCR was carried out in 50 μl reaction mixture containing 10–100 ng of DNA, 1.0 μM of both reverse and forward primers, 200 μM of each dNTPs in 1X DyNAzyme polymerase buffer and 2.0 U DyNAzyme DNA polymerase (Finnzymes, Espoo, Finland) with a MJ Research thermal cycler PTC-0225. The reaction mixture was first denatured at 95 °C for 3 min, followed by 35 cycles of denaturing at 95 °C for 45 s, annealing at 42 °C for 45 s and extension at 72 °C for 90 s. This was followed by a final extension at 72 °C for 5 min to obtain deoxyadenosine overhangs. PCR products were loaded on 1% agarose gels, the bands obtained were excized from the agarose and purified by using GFX PCR DNA and Gel Band Purification kit (Amersham Biosciences, Hørsholm, Denmark) according to the manufacturer. Cloning was done by using the TA cloning method (Invitrogen), according to the manufacturer's instructions. Sequencing was performed by amplification of the insert with forward and reverse M13 primers, followed by enzymatic cleansing of the PCR products using the ExoSapit kit (Amersham Biosciences) and sequencing with forward M13 primer and the Big Dye Terminator Cycle Sequencing kit. The sequences obtained were identified with BLASTX searches (Altschul *et al.* 1997). Sequence groups (amylase sequence type) were constructed using a pairwise similarity algorithm included in BioEdit Sequence Alignment Editor software (Hall 1999).

Nucleotide sequence accession numbers

All partial and whole 16S rRNA sequences (designated as LNE) were deposited in the GenBank database under accession numbers AY531638 to AY531654.

Results

Enrichments

Negligible growth was observed in all six enrichments during the first 16 days of incubation on starch only. The addition of 0.005% of yeast extract resulted in increase in density and cultures reached 10^6–10^7 cells ml^{-1} after 2–42 days (Tables 1 and 2). Growth was observed as homogeneous turbidity, small floating agglutinated bodies or a film adhering at the bottom of the glass. Microscope analyses showed that enrichments contained mixed population of thick and thin rods of various size, coccoid cells, filament and sporeformers.

Table 2. Identification of cloned 16S rRNA sequences from low nutrient enrichments.

Enrichments	No. of clones	Closest database match	Similarity (% [a])	Phylotype
HS2	38	*Thermus* sp. SRI-96	99.7	LNE-1
	11	*Thermus oshimai*	99.9	LNE-2
	8	*Anoxybacillus flavithermus*	99.4	LNE-3
Total HS2	57			
HS3	22	*Anoxybacillus flavithermus*	99.4	LNE-3
	1	*Caloramator indicus*	99.8	LNE-4
	12	*Clostridium* sp. IrT-R5M2-31	89.8	LNE-5
	8	*Caloramator fervidus*	98.0	LNE-6
	3	*Thermoterrabacterium ferrireducens*	98.1	LNE-7
Total HS3	46			
HS5	16	*Anoxybacillus flavithermus*	99.4	LNE-3
Total HS5	16			
HS6	15	*Thermus* sp. SRI-96	98.9	LNE-8
	1	*Thermus kawarayuensis*	98.7	LNE-9
	1	*Thermus* sp. SRI248	99.4	LNE-10
Total HS6	17			
LS7	6	*Caloramator fervidus*	98.0	LNE-6
	8	Candidate OP9 clone OPB47	99.7	LNE-11
	14	Unidentified GNSB[b] clone OPB34	98.6	LNE-12
	9	*Moorella glycerini*	97.2	LNE-13
	2	*Thermoterrabacterium ferrireducens*	92.7	LNE-14
	1	*Thermoterrabacterium ferrireducens*	91.3	LNE-15
	2	Uncultured Low G+C GPB [c] clone OPB54	91.0	LNE-16
Total LS7	42			
LS10	2	*Thermus* sp. SRI-96	99.7	LNE-1
	18	*Thermus oshimai*	99.9	LNE-2
	10	*Meiothermus cerbereus*	97.1	LNE-17
Total LS10	30			

[a] Similarity scores computed with ARB using corresponding filters for each bacterial group.
[b] GNSB, green non-sulphur bacterium.
[c] GPB, gram positive bacterium.

Specific amplification of 16S rRNA genes and phylogenetic analysis of the enrichments

All enrichments contained bacterial 16S rRNA genes that were used to construct six *Bacteria* 16S *rRNA* libraries. No 16S rRNA genes from *Archaea* could be amplified. A total of 208 clones were successfully sequenced with the R805 primer so that sequences could be aligned to each other and to reference sequences from the Ribosomal Database Project. Only sequences with reliable nucleotide sequence were edited and aligned with reference strains. No chimeric artifacts were found among all sequences. Operational taxonomic units (OTU, phylotype) were established by pairwise similarity calculations, using 98% DNA-DNA similarity as cutoff. Subsequently, one representative of each phylotype was selected and its partial or whole sequence (about 700–1500 bp, respectively) was determined. Figure 1 shows a maximum-likelihood phylogenetic tree containing 17 OTUs obtained by low nutrient enrichment (LNE clones) and the 14 OTUs that were found by direct phylogenetic analysis of the hot spring biomass (SRI clones) (Skirnisdottir *et al*. 2000). Table 2 presents the closest database matches for each 16S rRNA library with corresponding similarity scores computed by ARB.

Six OTUs (LNE-1, LNE-2, LNE-8, LNE-9, LNE-10 and LNE-17) showed closest database match to *Thermus* and *Meiothermus* species. In particular, clones LNE-1, LNE-8, and LNE-10 were very closely related to SRI clones previously identified in the hot spring microbial mat (Skirnisdottir *et al*. 2000). Phylotypes LNE-1 and LNE-8 were closely related to *Thermus* sp. SRI-96 but sequence comparison showed that LNE-8 was phylogenetically distinct from LNE-1. However, as LNE-8 clusters within the *Thermus scotoductus* cluster, the results suggest that LNE-8 may represent a new subspecies of *T. scotoductus* (Kristjansson *et al*. 1994; Skirnisdottir *et al*. 2001). LNE-2 was showing very high homology to *Thermus oshimai*, which has been isolated from the same geographical hot spring location in Iceland but also in the Azores (Williams *et al*. 1996). LNE-9 and LNE-10 were showing closest homology to *Thermus kawarayuensis* and *Thermus* sp. SRI248 respectively, two new uncultivated *Thermus* species which branched deeply among the *Thermales*. Despite clustering with *Meiothermus* (Logi-

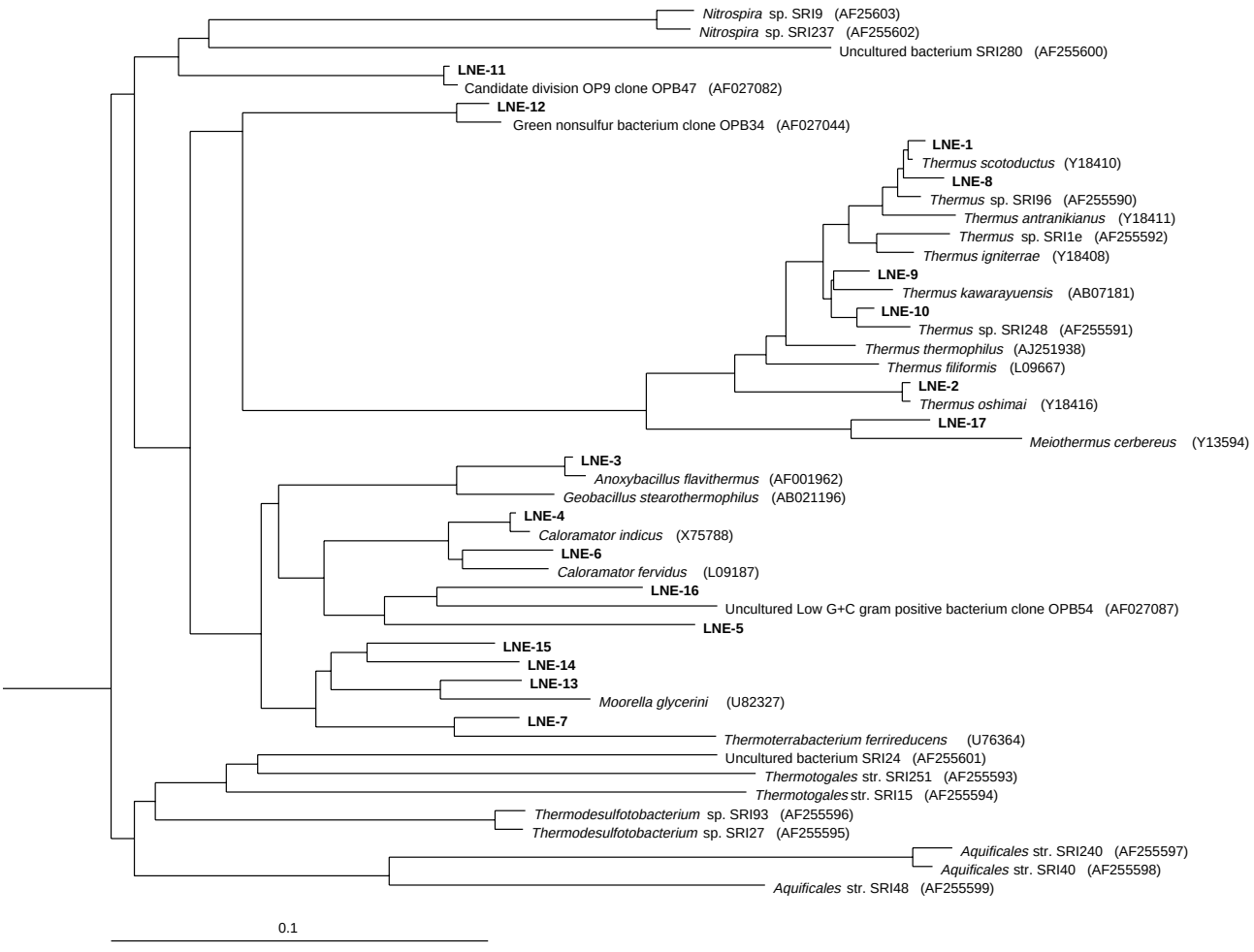

Figure 1. Evolutionary maximum-likelihood phylogenetic dendrogram of the 16S rRNA sequences detected in the sulphide-rich hot spring biomass (designated SRI) and by low nutrient enrichments (designated LNE). *Sulfolobus acidocaldarius* was used as an outgroup. The scale bar represents the expected number of substitution per nucleotide position.

nova *et al*. 1984), LNE-17 did not seem to be related to any *Meiothermus* strains identified so far, and may well be representative of a new species.

Phylotypes LNE-11, LNE-12 respectively showed closest database match to division candidates OP9 clone OPB47 and green non-sulphur bacterium (GNSB) clone OPB34 (Hugenholtz *et al*. 1998). Both LNE-11 and LNE-12 clustered ideally with their respective closest database match. The small sequence dissimilarity between LNE-12 and GNSB clone OPB34, as displayed on the phylogenetic tree, could yet be due to clonal variation and the sensitivity of the phylogenetic computation.

Five phylotypes were found to be related to gram-positive bacteria. LNE-3 was closely related to *Anoxybacillus flavithermus* (Rainey *et al*. 1994; Pikuta *et al*. 2000), LNE-4 and LNE-6 clustered with *Caloramator indicus* (Chrisostomos *et al*. 1996) and *Caloramator fervidus* (Collins *et al*. 1994) respectively. Clone LNE-13 was related to *Moorella glycerini* (Slobodkin *et al*. 1997a) and LNE-7 was affiliated to the iron-reducing *Thermoterrabacterium ferrireducens* (Slobodkin *et al*. 1997b). Low similarity scores were found for LNE-14 and LNE-15 which showed closest homology to *Thermoterrabacterium ferrireducens*. Similarly, LNE-5 was distantly related to *Caloramator* sp. clone IrT-R5M2-31 (unpublished). The three phylotypes LNE-14, LNE-15 and LNE-5 branched within the gram positives although their respective position in this group remained uncertain. LNE-16 showed homology to uncultured low G + C gram-positive bacterium (GPB) clone OPB54 from Obsidian Pool in Yellowstone Park (Hugneholtz *et al*. 1998). With the exception of LNE-3 and LNE-4, all gram positive LNE clones are probable representatives of new species (≤96–97%), or even new candidate divisions where very low similarity scores were found (≤92%) (Stackebrandt & Goebel 1994).

Primer design for screening of amylases from glycoside-hydrolase family 13

Multiple alignments of amylase sequences obtained with ClustalX v1.8 were manually analysed. Homology groups were established based on amino acid conservation at designated sites from β strands of the catalytic domain (Table 3) (Horvatova *et al*. 2001). One forward priming site was chosen in strand β3, where a [FLIV-YM]NH motif was found for a majority of amylolytic enzymes. Three motifs G[FW]R, [AYV][NK]H and [NKSV]HD, from strands β4, β4 and β7 respectively, were targeted as reverse priming sites. Alpha-glucan branching enzymes (EC 2.4.1.18) retain the specific motif WVP[GS]H within the β3 strand and for which both forward and reverse primers were designed. Ideally, a maximum degeneracy score of 32 was chosen in order to limit unspecific annealing during gene amplification. Manual editing of the primers was then performed where higher degeneracy scores (≥ 64) were found. Higher screening potential was obtained by

Table 3. Screening primers used for detecting amylolytic genes in low nutrient enrichment samples and the hot spring biomass.

Primer [a]	Sequence (5′ → 3′) [b]	Codon usage [c]	Target sequence [d]
GH13-F1	GCATGTTATGCTGGATGCAgtnttyaayca	Bac	β3
GH13-F3	AAATGTGCAAGTGTATATGGATTTTgtnytnaayca	Bac	β3
GH13-F4	CATTAAAGTTTATGTGGATGCGgtnathaayca	Bac	β3
GH13-F15	GGTGATGCTGGACGCGrtnttyaayca	Equ	β3
GH13-F18	GTCTTCGTGGACGCCgtnathaayca-3′	Equ	β3
CH13-F19	TGAAGATCATCATGGACCTGgtngtnaayca	Equ	β3
GH13-F20	CGGGTCTTCATCGACTTCgtnatgaayca	Equ	β3
GH13-F27	GCCCCTACGACCTGTACgayytnggnga	Equ	β3
GH13-F28	CGCCGACGCCGTGntnaaycayatg	Equ	β3
GH13-F33	GCCGATGTCGtnttygayca	Equ	β3
GH13-F7	TTTTGATTATGGATCCtggggntayca	Bac	β3/BE
GH13-F8	TGAATTTATGCCGGTTACGgarcaycaycc	Bac	β3/BE
GH13-F21	TCATGCCCGTCACCgarcaycaycc	Equ	β3/BE
GH13-F22	CGACTACGGGTCCtggggntayca	Equ	β3/BE
GH13-R12	AATATGTTTCACCGCATCAAATckraanccrtc	Bac	β4
GH-13-R13	GCCAGCCAAACGGCatrtgyttnac	Bac	β4
GH13-R14	GATCAACTTAATTAGCAACATCCATTckccanccrtc	Bac	β4
GH13-R26	GTTGGCCACGTCCATCckccanccrtc	Equ	β4
GH13-R31	GGGCAGCTCCATGTGCttngcngcrtc	Equ	β4
GH13-R9	TCCCAATACAGCATATTTGCAacngcrtcnac	Bac	β3/BE
GH13-R23	CAGTACAGCATGTTGGCCacngcrtcnac	Equ	β3/BE
GH13-R30	GCCCCGCTGGGTGtcrtgrttntc	Equ	β7

[a] F, forward primer; R, reverse primer.
[b] Letters in lower case correspond to the core region and upper case letters to the consensus clamp region. Degenerate alphabet: Y, C or T; R, A or G; K, G or T; H, A, C or T; N, A, C, G or T.
[c] Equ, equal codon usage; Bac, *Bacillus* codon usage.
[d] Sites found in the á-amylase-type $(\beta/á)_8$ –barrel, nomenclature by Horvathova et al. (14). β3 /BE, target sequences found only in 1,4-á-glucan branching enzymes.

selecting two codon usage biases, Equal and *Bacillus subtilis*, as it was found that many *Bacilli* were positive for amylolytic activity. The predicted size of the amylase fragments varied between 200 and 600 bp, on the basis of comparison with reference sequences.

Screening of amylase genes

As a control, primer stringency experiments were performed against our *Bacillus* strain collection. No inconsistencies were observed and new thermophilic amylase genes were amplified (data not shown). Based on the species composition found in the enrichments, four LNE samples (HS2, HS3, HS6, LS7) were screened for α-amylase genes. Enrichments HS5 and LS10 were excluded since they had lower overall diversity and overlapped the bacterial species found in the other enrichments.

Clone libraries were constructed by using every 112 possible pairs of forward and reverse primers. About 30–50% of the PCR reactions were successful and yielded on average 2 bands in agarose gel electrophoresis. All bands corresponding to the predicted size of PCR amplicons (70% in average) were cloned, and 12 of the resulting clones were picked. Roughly 800 clones were thus obtained and sequenced for each of the 4 enrichments screened. Depending on the sample, about 20–100 protein sequence groups were generated by pairwise similarity calculation, using a 85% amino acid identity cutoff. Only 24 primer pairs, out of all 112 gave amylase genes from the enrichments. Notably, same primer pairs resulted in the amplification of different amylase sequence types, and similarly, certain sequence types were amplified with more than one primer pair (Table 4). A total of 18 amylase sequence types (from a total of 16 OTUs) were detected in all 4 enrichments,

Table 4. Frequency of detected α-amylase genes in low nutrient enrichment samples (HS, LS) and hot spring biomass sample (SRI).

Sample	Sequence type	Enzyme type [a]	Closest Database Match (identity %)	Database Acc. Number	GH13 Primers [c]
Enrichments					
HS2	AmC3	AA	*Lactobacillus manihotivorans* (62)	AAC24760	F8-R30
	AmB6	AG	*Geobacillus thermoglucosidasius* (56)	A41707	F1-R12
	AmB7	BE	*Bacillus caldolyticus* (74)	CAA78440	F7-R9
	AmC2	MA	Geobacillus *stearothermophilus* (72)	AAC46346	F21-R30
	AmC8	NP	*Bacillus flavocaldarius* (85)	BAB18516	F21-R26
HS3	AmD5	BE	*Clostridium thermocellum* (64) [b]	ZP_00061166	F3-R31
	AmB8	GH	*Clostridium perfringens* str. 13 (56) [b]	BAB80030	F8-R9
	AmB9	GH	*Thermococcus sp*. Rt3 (41)	AAB87860	F8-R30
	AmE3	MA	*Thermus* sp. IM6501 (86)	AAC15072	F28-R26
	AmC0	NP	*Thermoactinomyces vulgaris* (72)	1911217A	F4-R14
HS6	AmD3	GH	*Chloroflexus aurantiacus* (98) [b]	ZP_00018024	F28-R26
	AmC8	NP	*Bacillus flavocaldarius* (85)	BAB18516	F21-R26
LS7	AmD2	BE	*Clostridium thermocellum* (72) [b]	ZP_00061166	F22-R23
	AmD7	BE	*Pseudomonas aeruginosa* (73)	AAG05541	F22-R23
	AmE2	CMDH	*Alicyclobacillus acidocaldarius* (85)	CAB40078	F20-R26
	AmC6	GH	*Streptomyces coelicolor* A3(2) (36) [b]	CAA15803	F33-R12
	AmC5	MA	*Thermus* sp. IM6501 (52)	AAC15072	F33-R14
	AmC8	NP	*Bacillus flavocaldarius* (85)	BAB18516	F21-R26
	AmD0	NP	*Synechocystis* sp. PCC 6803 (50)	BAA17809	F15-R26
	AmD1	NP	*Bacillus flavocaldarius* (54)	BAB18516	F15-R26
Biomass					
SRI	AmC8	NP	*Bacillus flavocaldarius* (85)	BAB18516	F21-R26

[a] AA, α-amylase (EC 3.2.1.1); AG: α-glucosidase (EC 3.2.1.20); BE: 1,4-α-glucan branching enzyme (EC 2.4.1.18); CMDH: cyclomaltodextrin hydrolase (EC 3.2.1.54); GH: unspecified glycoside-hydrolase family 13; MA: maltogenic amylase (EC 3.2.1.133); NP: neopullulanase (EC 3.2.1.135).

[b] Described as probable or hypothetical protein.

[c] Cf. Table 3.

representing seven different activity types including one α-amylase, one α-glucosidase, four 1,4-α-glucan branching enzymes, one cyclomaltodextrin hydrolase, three maltogenic amylases, four neopullulanases and four were unspecified family 13 glycoside-hydrolases (enzyme classes, see Table 4). Classification of the sequence types within the amylase family was sometimes uncertain because of low identity scores (≤50–60%), yet it suggested a high degree of novelty for the respective enzymes.

For comparison, we used all 24 primers pairs that gave amylase genes from the enrichments to screen the DNA isolated directly from the hot spring biomass. Analysis of the biomass DNA was shown to contain a total of 14 bacterial OTUs based on a similar 16S rRNA gene diversity study (Skirnisdottir *et al*. 2000). A total of 22 bands were obtained from 13 of the 24 primer pairs. All were cloned, and 230 resulting clones were sequenced. A total of 46 protein sequence groups were generated but only one unique group composed of two clones was found corresponding to an amylase sequence. The amylase sequence was then identified as neopullulanase AmC8 (Table 4).

Discussion

In this study, we have confirmed that microbial diversity studies based on direct DNA extraction, also called culture-independent analysis, have intrinsic methodological limitations as only dominant species in any particular microbial habitat may be detected (Ward *et al*. 1998). Comparably, screening strategies for detecting novel genes from the environment using sequence-based molecular methods are highly dependent on the sensitivity of the amplification method, the quality and composition of the DNA and the ability to amplify genes with extremely low copy numbers in complex DNA mixtures. Such methods alone will therefore underestimate species and gene richness.

In order to obtain slow growing heterotrophs carrying genes encoding for novel amylolytic enzymes, we used low concentrations of nutrients and different non-standard growth conditions such as very long incubation time. Untreated hot spring water was used in all enrichments to mimic *in situ* conditions, and starch was chosen as a recalcitrant carbon source that will selectively favour the growth of starch-degrading organisms. However, it was found that the cultivation conditions were too stringent to obtain sufficient cell density when using only starch. The addition of a very low concentration of yeast extract was thus performed to meet vitamin and mineral requirements (Cote & Gherha 1994). No significant growth could be observed in test enrichments with yeast extract concentrations lower than 0.01% as the only nutrient source (data not shown).

The inferred metabolism of the LNE OTUs found in the enrichments reflected the cultivation conditions. Aerated enrichments HS2, HS5, HS6 and LS10 were typically dominated by aerobic, organotrophic *Thermus*/*Meiothermus* and *Anoxybacillus* species. Anaerobic enrichments HS3 and LS7 displayed a diversity of gram-positive, putative fermentative microorganisms or related, with 5 and 7 phylotypes respectively. Interestingly, neither aerobic nor anaerobic microorganisms belonging to the *Archaea* were detected. Cultivation conditions, in particular temperature, may have been less favorable for the *Archaea,* or they may have been overgrown by the *Bacteria* so their 16S rRNA genes could not be amplified from the enrichment DNA.

Although the sampling of the sulphide-rich microbial mat for this experiment was performed after the phylogenetic study of the hot spring biomass (Skirnisdottir *et al*. 2000), repeated observation and sampling in the Grensdal area showed good temporal stability of both physical and chemical parameters (temperature, pH) of this particular hot spring. The sampling conditions used in this study were the same as performed previously and the appearance of the biomass was identical. The direct phylogenetic analysis of the microbial mat revealed low diversity (Skirnisdottir *et al*. 2000). The resulting 16S rRNA gene library (SRI clones) was dominated by three *Aquificales* species that were 68% of a total of 171 clones analysed. Some clones (14%) belonged to a putative heterotrophic sulphate-reducing bacterium of the genus *Thermodesulfobacter*, 4.5% of the clones to apparent fermentative heterotrophs of the order *Thermotogales*, and the remaining sequences to *Thermus* (4.1%) and *Nitrospira* species (approx. 1%). Additionally, the two *Thermus* clones LNE-1 and LNE-10 were very closely related to *Thermus* SRI-96 and *Thermus* SRI-248 clones reported in the previous study. This provides strong evidence that, apart from *Thermus* species, all other LNE phylotypes were rare and in minority in the undisturbed microbial mat and were therefore not detectable by direct 16S rRNA-based analysis. Moreover, our results demonstrated that the enrichment conditions have induced the selective growth of microorganisms able to degrade starch. Eighteen sequence types of amylase genes were detected in the four enrichments screened, whereas only one type was found in the biomass sample.

Amylases from glycoside-hydrolase family 13 have little amino acid conservation outside well-known catalytic and structural domains (Henrissat & Davies 1997; Horvatova *et al*. 2001). Also, the intrinsic sequence diversity of the conserved sites among different amylases was the basis for designing multiple sequence variants of screening primers (Table 3). A large part of the enrichment clone libraries contained unspecific sequences (on average 70% of total amplified sequences; data not shown) due to mispriming and degeneracy of the screening primers. Nevertheless, the parameters chosen for primer design and PCR conditions were proven necessary to contribute to the specific amplification of the targeted genes.

The number of the amylase sequence types found was roughly proportional to the 16S rRNA diversity in the corresponding enrichments (Tables 2 and 4).

Enrichment HS2 yielded 5 amylase sequence types from 3 OTUs, enrichment HS3 gave 5 sequence types from 5 OTUs, 2 sequence types came from 3 OTUs in enrichment HS6, and the anaerobic enrichment LS7 yielded 8 sequence types from 7 OTUs. Neopullulanase AmC8 was the only enzyme sequence found endemic to the non-cultivated biomass and also to all of the selected enrichments except HS3. It is difficult to establish any direct correlation between 16S rRNA genes and amylase sequences in the different samples. Moreover, little phylogenetic information can be deduced on the basis of homology searches for proteins, since few amylase sequences from thermophilic microorganisms are available in protein sequence databases. No definite conclusion can therefore be drawn from the comparison of the pattern of amylase diversity and the pattern of 16S rRNA gene diversity in the enrichments. However, it is clear that the hot spring biomass was apparently dominated by autotrophic *Aquificales* types (Skirnisdottir *et al*. 2000) whereas the enrichments were dominated by heterotrophic bacteria. This dominant autotrophy in the biomass was furthermore reflected by the significantly lower amylase diversity detected.

Gene diversity studies, i.e. ribosomal genes or metabolic genes, are limited by three independent parameters: gene copy number, primer degeneracy and priming competition between the targeted sequences. Sequences with good primer match and high copy numbers will be selectively and preferably amplified. Increasing the primer degeneracy, and/or lowering the annealing temperature will allow mismatches, hence increase the diversity of the PCR products but also increase the amount of non-specific by-products (Dahllof 2002). The uniqueness of the neopullulanase sequence found in the hot spring biomass sample emphasizes the difficulty in detecting specific metabolic genes in complex DNA mixtures, such as found in environmental samples. De facto, less than 1% of the biomass clone library (2 out of 230 clones) contained the amylase sequence, despite that comparable numbers of clones were sequenced for both enrichments and biomass screening. Nevertheless, we have shown that rational enrichment process can be an efficient and targeted way to increase gene copy number, and therefore allow better access to the otherwise silent microbial and metabolic gene diversities. Thus, a more comprehensive assessment of the overall diversity can be obtained by a combination of novel culture-related methods and molecular techniques, rather than by using either method alone.

Acknowledgements

The authors would like to thank Jóna B. Jónsdóttir, and Steinunn Á. Magnúsdóttir for their valuable contribution in this project. This work was supported by a grant from the European Union, Contract number QLK3-CT-2000-01068, a grant from the European Program *Leonardo da Vinci* and a grant from the University of Iceland.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. & Lipman, D.J. 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Amann, R.I., Ludwig, W. & Schleifer, K.-H. 1995 Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**, 143–169.

Chrisostomos, S., Patel, B.K.C., Dwivedi, P.P. & Denman, S.E. 1996 *Caloramator indicus* gen. nov., sp. nov., a new thermophilic anaerobic bacterium isolated from a non-volcanically heated waters of an Indian artesian aquifer. *International Journal of Systematic Bacteriology* **46**, 497–501.

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. & 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 Research* **31**, 442–443.

Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. & Farrow, J.E.A. 1994 The phylogeny of the genus *Clostridium*: Proposal of five new genera and eleven species combinations. *International Journal of Systematic Bacteriology* **44**, 812–826.

Cote, R.J. & Gherna, R.L. 1994 Nutrition and Media. In *Methods for General and Molecular Bacteriology*, eds. Gerhardt, P., Murray, R.G.E., Wood, W.A. & Krieg, N.R. pp. 155–178. Washington DC, USA: American Society of Microbiology. ISBN: 1-55581-048-9.

Coutinho, P.M. & Henrissat, B. 1999a Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering*, eds. Gilbert, H.J., Davies, G., Henrissat, B. & Svensson, B. pp. 3–12. Cambridge, UK: The Royal Society of Chemistry. ISBN 0-85404774-3.

Coutinho, P.M. & Henrissat, B. 1999b The modular structure of cellulases and other carbohydrate-active enzymes: an integrated database approach. In *Genetics, Biochemistry and Ecology of Cellulose Degradation*, eds. Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S. & Kimura, T. pp. 15–23. Tokyo, Japan: Uni Publishers Co. ISBN 0-85404774-3.

Dahllof, I. 2002 Molecular community analysis of microbial diversity. *Current Opinion in Biotechnology* **13**, 213–217.

Hall, T.A. 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Henikoff, S., Henikoff, J.G., Alford, W.J. & Pietrokovski, S. 1995 Automated construction and graphical presentation of protein blocks from unaligned sequences. *Gene* **163**, GC17–GC26.

Henrissat, B. & Davies, G. 1997 Structural and sequence-based classification of glycoside-hydrolases. *Current Opinion in Structural Biology* **7**, 637–644.

Hjorleifsdottir, S., Skirnisdottir, S., Hreggvidsson, G.O., Holst, O. & Kristjansson, J.K. 2001 Species composition of cultivated and noncultivated bacteria from short filaments in an Icelandic hot spring at 88°C. *Microbial Ecology* **42**, 117–125.

Horvathova, V., Janecek, S. & Sturdik, E. 2001 Amylolytic enzymes: molecular aspects of their properties. *General Physiology and Biophysics* **20**, 7–32.

Hugenholtz, P., Pitulle, C., Hershberger, K.L. & Pace, N.R. 1998 Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of Bacteriology* **180**, 366–376.

Jackson, C.R., Roden, E.E. & Churchill, P.F. 1998 Changes in bacterial species composition in enrichment cultures with various dilutions of inoculum as monitored by denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology* **64**, 5046–5048.

Kristjansson, J.K., Hjorleifsdottir, S., Marteinsson, V.T. & Alfredsson, G.A. 1994 *Thermus scotoductus*, sp. nov., a pigment-producing thermophilic bacterium from hot tap water in Iceland and including *Thermus* sp. X-1. *Systematic and Applied Microbiology* **17**, 44–50.

Loginova, L.G., Egorova, L.A., Golovacheva, R.S. & Seregina, L.M. 1984 *Thermus ruber* sp. nov. nom rev. *International Journal of Systematic Bacteriology* **34**, 498–499.

Marteinsson, V.T., Kristjansson, J.K., Kristmannsdottir, H., Dahlkvist, M., Saemundsson, K., Hannington, M., Petursdottir, S.K., Geptner, A. & Stoffers P. 2001 Discovery and description of giant submarine smectite cones on the seafloor in Eyjafjordur, northern Iceland, and a novel thermal microbial habitat. *Applied and Environmental Microbiology* **67**, 827–833.

Niehaus, F., Bertoldo, C., Kahler, M. & Antranikian, G. 1999 Extremophiles as a source of novel enzymes for industrial applications. *Applied Microbiology and Biotechnology* **51**, 711–729.

Pikuta, E., Lysenko, A., Chuvilskaya, N., Mendrock, U., Hippe, H., Suzina, N., Nikitin, D., Osipov, G. & Laurinavichius, K. 2000 *Anoxybacillus pushchinensis* gen. nov., sp. nov., a novel anaerobic, alkaliphilic, moderately thermophilic bacterium from manure, and description of *Anoxybacillus flavithermus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* **50**, 2109–2117.

Rainey, F.A., Fritze, D. & Stackebrandt, E. 1994 The phylogenetic diversity of thermophilic members of the genus *Bacillus* as revealed by 16S rDNA analysis. *FEMS Microbiology Letters* **115**, 205–211.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M. & Henikoff, S. 1998 Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. *Nucleic Acids Research* **26**, 1628–1635.

Santegoeds, C.M., Nold, S.C. & Ward, D.M. 1996 Denaturing gradient gel electrophoresis used to monitor the enrichment culture of aerobic chemoorganotrophic bacteria from a hot spring cyanobacterial mat. *Applied and Environmental Microbiology* **62**, 3922–3928.

Skirnisdottir, S., Hreggvidsson, G.O., Hjorleifsdottir, S., Marteinsson, V.T., Petursdottir, S.K., Holst, O. & Kristjansson, J.K. 2000 Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Applied and Environmental Microbiology* **66**, 2835–2841.

Skirnisdottir, S., Hreggvidsson, G.O., Holst, O. & Kristjansson, J.K. 2001 Isolation and characterization of a mixotrophic sulphur-oxidizing *Thermus scotoductus*. *Extremophiles* **5**, 45–51.

Slobodkin, A., Reysenbach, A.-L., Mayer, F. & Wiegel, J. 1997a Isolation and characterization of the homoacetogenic thermophilic bacterium *Moorella glycerini* sp. nov. *International Journal of Systematic Bacteriology* **47**, 969–974.

Slobodkin, A., Reysenbach, A.-L., Strutz, N., Dreier, M. & Wiegel, J. 1997b *Thermoterrabacterium ferrireducens* gen. nov., sp. nov., a thermophilic anaerobic dissimilatory Fe(III)-reducing bacterium from a continental hot spring. *International Journal of Systematic Bacteriology* **47**, 541–547.

Sonnhammer, E.L.L., Eddy, S.R. & Durbin, R. 1997 Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**, 405–420.

Stackebrandt, E. & Goebel, B.M. 1994 Taxonomic note; A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology* **44**, 846–849.

Suzuki, M.T. & Giovannoni, S.J. 1996 Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology* **62**, 625–630.

Thompson, J.D., Gibson, T.G., Plewniak, F., Jeanmougin, F. & Higgins, D.G. 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**, 4876–4882.

Von Wintzingerode, F., Gobel, U.B. & Stackebrandt, E. 1997 Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21**, 213–229.

Ward, D.M., Ferris, M.J., Nold, S.C. & Bateson, M.M. 1998 A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiology and Molecular Biology Reviews* **62**, 1353–1370.

Williams, R.A.D., Smith, K.E., Welch, S.G. & Micallef, J. 1996 *Thermus oshimai* sp. nov., isolated from hot springs in Portugal, Iceland, and the Azores, and comment on the concept of a limited geographical distribution of *Thermus* species. *International Journal of Systematic Bacteriology* **46**, 403–408.