ARTÍCULO ORIGINAL

ISSN 0325-7541
Revista Argentina de Microbiología (2011) 43: 94-103

Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements

RENAUD BERLEMONT[1], DELPHINE PIPERS[1], MAUD DELSAUTE[1], FEDERICO ANGIONO[3], GEORGES FELLER[2], MORENO GALLENI[1], PABLO POWER[1, 3*]

[1]*Laboratoire de Macromolécules Biologiques and* [2]*Laboratoire de Biochimie, Centre d'Ingénierie des Protéines, Université de Liège, Belgium;* [3]*Laboratorio de Resistencia Bacteriana, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Argentina*
Correspondence. E-mail: ppower@ffyb.uba.ar

ABSTRACT

Metagenomic library PP1 was obtained from Antarctic soil samples. Both functional and genotypic metagenomic screening were used for the isolation of novel cold-adapted enzymes with potential applications, and for the detection of genetic elements associated with gene mobilization, respectively. Fourteen lipase/esterase-, 14 amylase-, 3 protease-, and 11 cellulase-producing clones were detected by activity-driven screening, with apparent maximum activities around 35 °C for both amylolytic and lipolytic enzymes, and 35-55 °C for cellulases, as observed for other cold-adapted enzymes. However, the behavior of at least one of the studied cellulases is more compatible to that observed for mesophilic enzymes. These enzymes are usually still active at temperatures above 60 °C, probably resulting in a psychrotolerant behavior in Antarctic soils. Metagenomics allows to access novel genes encoding for enzymatic and biophysic properties from almost every environment with potential benefits for biotechnological and industrial applications. Only *intI*- and *tnp*-like genes were detected by PCR, encoding for proteins with 58-86 %, and 58-73 % amino acid identity with known entries, respectively. Two clones, BAC 27A-9 and BAC 14A-5, seem to present unique syntenic organizations, suggesting the occurrence of gene rearrangements that were probably due to evolutionary divergences within the genus or facilitated by the association with transposable elements. The evidence for genetic elements related to recruitment and mobilization of genes (transposons/integrons) in an extreme environment like Antarctica reinforces the hypothesis of the origin of some of the genes disseminated by mobile elements among "human-associated" microorganisms.

Key words: metagenomics, cellulases, psychrophilic enzymes, integrases, activity-driven metagenomics

RESUMEN

Explorando el metagenoma del suelo antártico como fuente de nuevas enzimas adaptadas al frío y elementos génicos móviles. A partir de muestras de suelo antártico se obtuvo la metagenoteca PP1. Esta fue sometida a análisis funcionales y genotípicos para el aislamiento de nuevas enzimas adaptadas al frío con potenciales aplicaciones, y para la detección de elementos génicos asociados a la movilización de genes, respectivamente. Por tamizaje fenotípico se detectaron 14, 14, 3 y 11 clones productores de lipasas/esterasas, proteasas, amilasas y celulasas, respectivamente, con actividades máximas aparentes de 35 °C para las amilasas y lipasas, y de 35-55 °C para las celulasas, tal como se observó para otras enzimas adaptadas al frío. Sin embargo, una celulasa parece ser compatible con enzimas mesófilas, las que usualmente se mantienen activas hasta por sobre 60 °C. Este hecho probablemente esté asociado a un comportamiento psicrotolerante en los suelos antárticos. La metagenómica permite acceder a una nueva miríada de productos metabólicos con potenciales beneficios para aplicaciones biotecnológicas e industriales. Se detectaron los genes tipo *intI* y *tnp* por PCR, y sus productos génicos deducidos tuvieron identidades del 58 al 86 % y del 58 al 73 % con secuencias conocidas, respectivamente. Dos clones, BAC 27A-9 y BAC 14A-5, parecen presentar organizaciones sintéticas únicas, lo cual sugiere la existencia de rearreglos génicos probablemente debidos a divergencias evolutivas dentro del género o facilitados por la asociación de elementos de transposición. La evidencia de elementos génicos relacionados con el reclutamiento y la movilización de genes en ambientes extremos como la Antártida refuerza la hipótesis sobre el origen de algunos genes diseminados por elementos móviles entre los microorganismos asociados al ser humano.

Palabras clave: metagenómica, celulasas, enzimas psicrofílicas, integrasas, metagenómica funcional

INTRODUCTION

Metagenomics includes a series of methodological manipulations developed for accessing a vast and unexplored genetic pool coming from essentially uncultured bacteria living in different samples, including environmental (water, soil, sediments), clinical (tissues, fluids, cadaveric samples) and other samples containing hypothetically uncultured microorganisms.

The power of metagenomics relies in the access, without prior sequence information, to the so far uncultured majority of microorganisms on Earth, which is estimated

to be more than 95 % of the microorganisms in some habitats (14, 26).

The metagenomic approach is generally based on the direct cloning of environmental DNA (extracted by using either direct or indirect techniques and cleaved in "easy-to-use" fragments) for the construction of metagenomic libraries obtained in suitable vectors and introduced in cultivable hosts like *Escherichia coli*, bypassing the necessity of dealing with culturing techniques (20, 46). In this way, metagenomics allows the isolation of novel sequences, genes, complete pathways and consequently their products (when expression is possible) by multiple screening approaches, paving the way for elucidating the functions of microbial communities, genomic analyses of uncultured soil microorganisms, and to search for new genes coding for various proteins from unbiased gene pools (8, 42).

Screening of metagenomic libraries includes a functional approach (also known as phenotype-based, activity-driven or functional metagenomics) in which a specific activity is directly or indirectly detected from the metagenomic library, if the expression of the encoding gene is feasible (2, 5, 8, 9, 19, 34, 43, 48); the other alternative is the identification of interesting genes based on sequence homology through PCR, random sequencing and large-scale shotgun sequencing, hybridization, etc. (genotypic or sequence-driven metagenomics), by which new genes can be accessed without the need of any activity-based detection system but where a minimal information of the sequences of interest is mandatory (19, 28, 31, 35, 37, 45).

Since evolution and natural selection have occurred in the environment for billions of years, the metagenomic approach allows the isolation of enzymes harboring tailor-made properties, fitting the physicochemical conditions of the studied habitats.

In this perspective, new enzymes having special biophysic features were isolated from metagenomic libraries built from different environments, many of them having good potential for being used in industrial and biotechnological applications (6, 13, 21, 29, 32, 36, 40, 41, 47).

Cold-adapted enzymes isolated from psychrophilic microorganisms represent an interesting field of study because these proteins are generally characterized by having high specific activities at relatively low temperatures associated with a fairly high thermal sensitivity, making them useful for biotechnology (17). Previously studied cold-adapted enzymes include xylanases (EC 3.2.1.8) (7), cellulases (EC 3.2.1.4) (5, 15), amylases (EC 3.2.1.1) (7), and esterases (EC 3.1.1.1) (4), among others.

In addition, Antarctica represents a very attractive location for the application of metagenomic approaches aimed at the search for novel cold-adapted enzymes, due to the fact that its human-associated activity is still kept to a minimum.

The aim of this work was therefore to use both functional and genotypic metagenomics for the isolation of novel cold-adapted enzymes with potential applications in biotechnology and industrial processes, and the detection of genetic elements associated with the recruitment and mobilization of genes.

Materials and methods

Soil samples

The samples were collected from Pointe Geologie archipelago (Ile des Petrels), Terre Adélie, Antarctica (66°40'S-140°01'E) during the austral summer 1999-2000 expedition organized by IPEV (www.institut-polaire.fr/). Three different samples were used: a humid, crude oil-contaminated soil, and two diesel-oil-contaminated, fertilizer-amended (~5.5 g/kg Inipol Eap-22; Elf Atochem) soils (one of them also containing a 20 % bird soil). The temperature of soils during the sampling process was 12 °C. The samples were kept at -70 °C until the construction of the metagenomic library.

Environmental DNA (eDNA) extraction

The eDNA extraction was performed using the direct method developed by Zhou *et al.* (50) with modifications described by Henne *et al.* (23). Briefly, 50 g of each soil were mixed with 135 ml DNA extraction buffer (100 mM Na_2.EDTA, 100 mM Tris-HCl, 100 mM Na_2HPO_4, 1.5M NaCl, 1 % CTAB) containing 10 mg proteinase K, and were incubated for 30 min at 37 °C. Subsequently, 15 ml SDS (20 %) were added and the suspension incubated for two hours at 65 °C with gentle inversion every 20 minutes. After centrifugation for 10 min at 6,000 × *g* the aqueous phase was recovered in one volume chloroform/isoamylic alcohol (24:1). The solution was then centrifuged as described above and the upper aqueous phase was recovered in 0.6 volumes of isopropanol, and incubated at room temperature for 24 h to precipitate the DNA. An eDNA pellet was obtained after centrifugation at 16,000 × *g* for 20 min. The pellet was washed twice with ethanol (70 %) and centrifuged as before. Finally, the precipitated eDNA was air-dried and resuspended in 4 ml of sterile MilliQ water.

Direct PCR-based screening of integron elements

A direct PCR-based screening was attempted from the environmental DNA 2G, independently from the library construction. For this purpose, degenerated primers targeting at conserved sequences in the structure of integrons (*intI*, 59-be) were used in PCR reactions (Table 1). Resulting amplicons were cloned in a pPCR-Script vector (Fermentas) and used for transformation in competent *E. coli* DH10B (Invitrogen). Finally, plasmid DNA from selected positive clones was extracted and the insert was sequenced.

Metagenomic library construction

The purified eDNA was partially digested using *Eco*RI (Fermentas, St. Leon-Rot, Germany) and the resulting fragments were separated by overnight ultracentrifugation at 4 °C in a sucrose gradient (10-40 %) at 27,000 × *g* DNA fractions were resolved in 1 % agarose gels and those having molecular sizes higher than 5 kb were pooled and precipitated with a solution of 20 % polyethylene glycol (PEG) 6,000 and 2.5 M NaCl, centrifuged at 16,000 × *g* for 30 min, washed with 70 % ethanol and resuspended in a suitable volume of sterile MilliQ water. Selected eDNA fragments were then cloned in the CopyControl pCC1BAC-*Eco*RI (Epicentre, Madison, WI, USA). After dialysis in 1 % agarose containing 1.8 % glucose (3), the ligation mixture was used to transform electrocompetent *E. coli* TransforMax EPI300 cells (Epicentre) using the following conditions: 2.5 V, 200 Ω, and 25 μF. Resulting recombinant clones were obtained on LB-agar plates supplemented with 12.5 μg/ml chloramphenicol, the "LBA-Chl basic medium", and different substrates (see below).

Characterization of the PP1 metagenomic library

A fraction of the recombinant clones was obtained on LB-agar plates containing chloramphenicol (12.5 µg/ml), 30 µM isopropyl-β-thio-galactoside (IPTG), and 50 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (Fermentas, USA) in order to estimate the efficiency of transformation by "blue-white" evaluation (clones harboring insert-containing vectors versus self-ligated BAC). Moreover, for assessing the average insert size, BAC constructions from 40 randomly selected recombinant clones were isolated, EcoRI (Fermentas, USA) digested for releasing the insert, and resolved by electrophoresis in 1 % agarose gels.

Functional screening for hydrolytic activities

Functional screening of the PP1 metagenomic library was performed by direct selection of *E. coli* recombinant clones on LBA-Chl basic medium supplemented with different substrates depending on the activities evaluated. For detection of cellulose-hydrolyzing clones, 0.5 % carboxymethyl-cellulose (CMC; Sigma, USA) and 0.01 % Trypan Blue (Sigma-Aldrich, USA) were added, and positive clones were visualized as blue colonies surrounded by a pale hydrolytic halo. Lipolytic enzyme-producing clones were detected by addition of 1 % emulsified Tributyrin (Sigma; USA), giving the medium an opaque aspect whereas hydrolytic colonies produced a clear hydrolysis halo. For amylolytic enzyme-producing clones, 0.5 % starch (Sigma, USA) was added. Positive clones were detected by presenting a clear hydrolytic halo which was highlighted by exposing plates to sublimated iodine vapors. Xylanase-producing clones were evaluated by addition of 0.5 % insoluble AZCL-Xylan (Megazyme, Ireland) and positive clones gave a diffuse blue precipitate due to the AZCL substrate. Protease-producing clones were detected with 1 % casein (Sigma, USA) and the presence of a protease was evidenced by a clear hydrolysis zone. All incubations were carried out overnight at 37 °C, and followed for up to five additional days at 18-20 °C.

In addition, cytoplasmic fractions from each clone expressing a specific phenotype were prepared from 5-10 ml of 72 h cultures at 18 °C. Cells were harvested by centrifugation at 13,000 × *g* for 5 min, resuspended in 1 ml of 20 mM sodium phosphate buffer (pH 7.0) and sonicated in an ice bath (3 cycles of 30 sec each at amplitude 10-12 µm). After centrifugation, as above, supernatants were recovered and used for subsequent activity screening at various temperatures. When available, hydrolytic activities were determined from crude extract preparations using chromogenic substrates. Assays were performed by incubation of crude extracts containing the tested cytoplasmic fraction with the substrate and incubated at room temperature for 30 min. Cellulolytic activity was detected using CMC (as substrate) and 3,5-dinitro salicylic acid (DNS) quantitation assay; a positive result was the appearance of a dark yellow/orange color (negative: light yellow) (38). For lipase enzymes, *p*-nitrophenol butyrate (pNPB) was used and a positive result was visualized as a yellow color (negative: no color) (18). Finally, the presence of amylolytic activity was detected using ethylidene-*p*-nitrophenyl-α,D-maltoheptaoside, obtaining a yellow/pale green as positive (negative: no color) (33).

BAC extraction from cell pools

Pools of transformed cells containing the metagenomic library were inoculated in 20 ml LB supplemented with 12.5 µg/ml chloramphenicol and 6 µl/ml culture of an induction solution (Autoinduction Solution, Epicentre), and incubated overnight at 37 °C, under agitation. After incubation, BACs were extracted using the GeneJet Plasmid Miniprep kit (Fermentas).

Determination of thermal dependence of hydrolytic activities

The effects of temperature on lipolytic, amylolytic and cellulolytic activities were evaluated by incubating cell lysates for 30 min at 15-70 °C, and remaining activity was determined as above at 20 °C.

PCR-screening of integrase-producing metagenomic clones

In order to screen for integrase-encoding genes and to determine the associated genetic background, a PCR-screening using different degenerate primers was performed. Briefly, a first PCR was attempted using total BAC extracted from each pool, 0.5 U GoTaq polymerase (Promega, USA), and 0.8 µM primers intI-255F and intI-948R (Table 1). Positive BAC discriminated the pools to be tested in the second step, where positive metagenomic pools were spread on LB agar plates supplemented with 12.5 µg/ml chloramphenicol. After overnight incubation at 37 °C,

Table 1. Primers used for direct screening of integron elements

Name	Sequence (5´-3´)	Target DNA	Reference
intI-255F	ACSCAGAACCAGGCGYTSKCSGCN	Conserved TQNQALSA in environmental genes	(11)
intI-480F	GGGTCAAGGAYSTSGAYTTCGG	Position 461-482 in *intI1*	(11)
intI-528F	CGNGAYGGYAARGGSRNVAAGGAYCGS	GKGG(N)KDR at patch II from environmental genes	(11)
intI-864R	YAGCAGATGNGTGGCRAAVSWRTGSCG	RHSFATHLL at box II in environmental genes	(11)
intI-948R	NARTACRTGNGTRTADATCATNGT	925-948 (compl.) in *intI1*	(11)
attCdir	GCSGCTKANCTCVRRCGTTAGSC	Gene cassettes – set 1	This study
attCrev	TCSGCTKGARCGAMTTGTTAGVC		This study
CGP1	GCSGCTKANCTCVRRCGTTRRRY	Gene cassettes – set 2	(11)
CGP2	TCSGCTKGARCGAMTTGTTRRRY		(11)

a representative number of clones was sub-cultured and used in colony-PCR assays in the same conditions as above but using primers intI-480F or intI-528F (forward primers) and intI-864R (Table 1). Positive clones were recovered and the BAC was extracted and sequenced by genome walking. Alternatively, the inserts from positive clones were digested with *Eco*RI/*Hin*dIII/*Bam*HI, the gel was purified and sub-cloned (after polishing of the resulting sticky ends) using the CloneJet PCR Cloning Kit (Fermentas).

DNA sequencing and sequence analyses

DNA sequencing was performed at the GIGA sequencing platform of the University of Liège (Belgium) using universal M13 reverse and forward primers. Additional primers were used to complete the sequencing of the inserts by a genome walking strategy as necessary. Comparative sequence analysis was carried out using the BLAST2.0 tool (http://www.ncbi.nlm.nih.gov/BLAST/). Analysis of the encoded proteins was done using the Expasy Molecular Biology Server (http://www.expasy.org/).

Nucleotide sequences accession numbers

Sequence data of BAC 27A-9 and BAC 14A-5 clones were deposited in the Genbank/EMBL nucleotide databases under the accession numbers FR716471 and FR716472, respectively.

Results and discussion

The metagenomic library PP1 from Antarctic soils was obtained in a BAC system (pCC1BAC) after transformation into *E. coli* EPI300 cells. The library contains 113,742 clones harboring inserts ranging between 5-10 kb (average size of 5.4 kb) (Figure 1), conserved at -70 °C in 41 pools, each of them including between 2,000-4,000 clones. The fact that relatively short insert sizes were selectively cloned in the library probably indicates either an over digestion or a previous shearing of the eDNA.

The total amount of genetic information covered in the PP1 library is approximately 650 Mb, which is equivalent to 130X, 155X, and 72X the complete genomes of *E. coli* (4.9 Mb), *Bacillus licheniformis* (4.2 Mb), and *Streptomyces coelicolor* (9.02 Mb), respectively.

In order to evaluate the diversity of genes cloned in the metagenomic library PP1, a random selection of ten different clones was performed and a BAC-end sequencing approach was carried out (39). All the selected clones harbored different types of eDNA inserts matching a variety of prokaryotic classes, such as Alphaproteobacteria (*Sphingomonas* sp., *Erythrobacter* sp., *Silicibacter* sp.), Gammaproteobacteria (*Photobacterium profundum*, *Azotobacter vinelandii* and *Pseudomonas mendocina*), Actinobacteria (*Mycobacterium gilvum*, *Nocardia farcinica*, *Rhodococcus* sp.), and Aquificae (*Sulfurihydrogenibium* sp.), among others.

From these results, it is expected that the metagenomic library PP1 will possess a high diversity of genes, and therefore the possibilities of screening for an interesting activity and/or gene of interest is also feasible.

Screening for different “non-selective” biochemical properties was directly performed on the transformed cells, using differential media containing substrates for detection of enzymatic activities with potential biotechnological or industrial applications.

Distribution of hydrolytic clones is shown in Figure 1. In summary, 14 tributyrin-, 14 starch-, 3 casein-, and 11 cellulose-hydrolyzing clones were detected, representing 0.044, 0.051, 0.019 and 0.12% of the total clones screened for each activity, respectively.

Active clones were replicated in selective media for phenotype confirmation. Crude cell extracts were obtained from amylolytic, cellulolytic and lipolytic clones, and hydrolytic activity was observed for all of them when tested with specific chromogenic substrates. In addition, the same phenotype was observed (data not shown) after purification of BAC-constructions from each clone and re-transformation in a different host (*E. coli* DH10B).

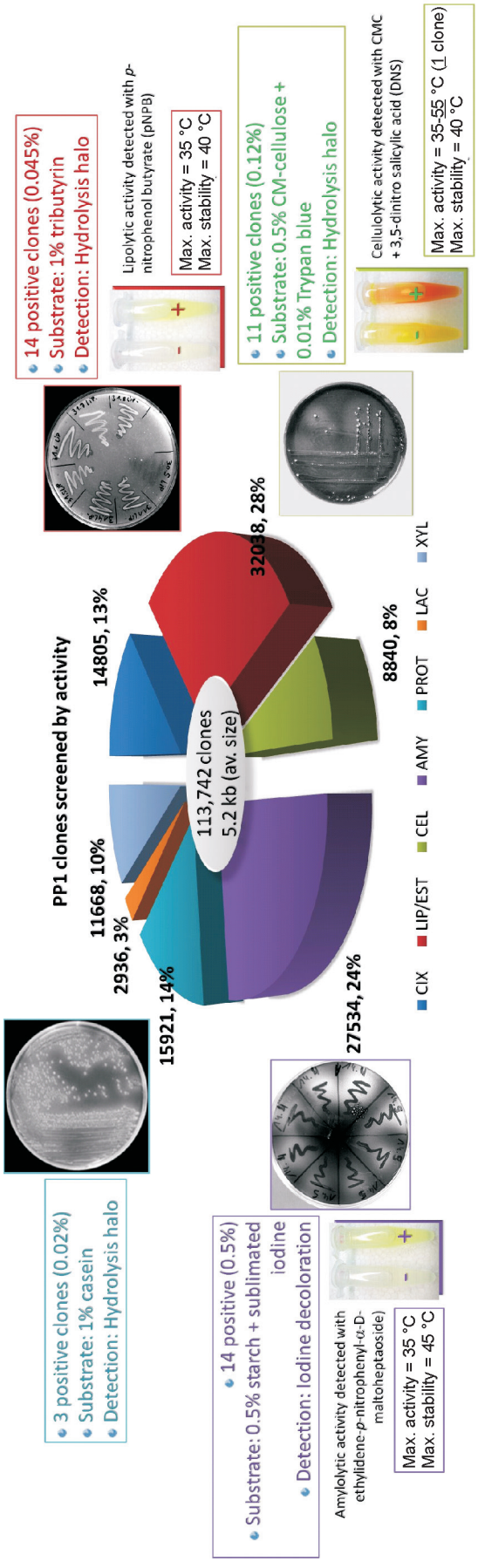
Selected clones were analyzed by DNA sequencing. For example, *E. coli* 39.5 clone harbors a 4.5 kb-insert, including ORFs with a relevant identity with putative proteins from *Pseudomonas stutzeri* A1501. The first ORF of the insert encodes a putative 351-amino acids endo-1,4-β-glucanase named RBcel1, which was extensively studied at a biochemical level and published elsewhere (5).

Another selected clone showing lipolytic activity was analyzed. The *E. coli* 31.8 clone harbored a 7 kb-DNA insert. An ORF of 1,908 bp (deduced protein of 635 amino acids) was sequenced and showed 63 % identity with an auto transporting lipolytic enzyme from *Pseudomonas* spp. Specifically, the putative protein (named DPestI) is strongly related to proteins from different species of *Pseudomonas* such as *P. syringae* (YP_272949, YP_237674, NP_790416), *P. aeruginosa* (CAC14200, EAZ61448, EAZ55624, NP_253799, ZP_01368126, ZP_00966082, YP_001351164, AAB61674) and *P. putida* (YP_001265806). Among these enzymes, the vast majority are lipases/esterases although some others have attributed functions of hemolysins/phospholipases.

These lipases/esterases are serine triacylglicerol lipases (EC 3.1.1.3) containing a consensus amino acid motif composed of Gly-Asp-Ser-Leu around their active-site-serine, and therefore named as GDSL-lipases (1). In addition, DPestI includes the conserved residues Ser, Gly, Asn, and His in boxes I, II, III and V, respectively (Figure 2), which are typically present in the SGNH-subfamily of GDSL hydrolases and play important roles in their catalytic activity (1, 27).

Metagenomics allows to access novel genes encoding for particular enzymatic and biophysic properties from almost every environment, including those in which human-related activities are kept to a minimum rate such as Antarctica.

In addition, metagenomics expands the availability of a novel myriad of metabolic products with potential benefits for biotechnological and industrial applications. Among these enzymes, several esterases (EC 3.1.1.1) and lipases (EC 3.1.1.3) (10, 24, 30), cellulases (EC 3.2.1.4) (22), and amylases (EC 3.2.1.1) (49), xylanases (EC 3.2.1.8) (25) have been isolated for this purpose.



As an example, the isolation of new putative cellulases, including those derived from extreme environments, offers an attractive feature for use in the production of biofuels (lignocellulosic ethanol) from cellulose-derived substrates. Also, lipase/esterase enzymes represent an important part of the industrial enzymes, being currently used for the production of various food products and in fine chemistry.

Due to the fact that the soil sample comes from a low-temperature environment, it is expected that the isolated clones will express new psychrophilic enzymes. Thus, we examined the effects of temperature on the hydrolytic activities of both recombinant clones and their derivative cell lysates to verify the thermal dependence of their activity. Assays were performed by incubating either pure cultures or cell lysate preparations at various temperatures (15 °C to 75 °C) and monitoring their ability to hydrolyze specific substrates.

The apparent maximum activities were determined around 35 °C for both amylolytic and lipolytic enzymes, whereas for cellulose-hydrolyzing enzymes the optimal activity seems to be observed between 35-55 °C. An increase of the incubation temperature induces a complete loss (after 30 min incubation) of the hydrolytic activity beyond 40 °C, 45 °C and 40 °C for lipolytic, amylolytic and cellulolytic enzymes, respectively, except for a single cellulolytic clone that retained activity even after incubation at 55 °C; this clone (*E. coli* 39.5) was further characterized and the cellulose was studied in detail due to its enlarged thermal tolerance (5).

The importance of isolating novel enzymes from cold environments relies in their unique and not-yet-completely-understood behavior. Cold-adapted enzymes have generally been observed to possess high activities at low temperatures (equivalent to that for mesophilic enzymes at 37 °C in some cases), which is associated with thermal instability due to a localized increase in the active site flexibility (12).

The major part of cold-adapted enzymes is characterized by showing a shift in their apparent T_{opt} (optimum

Figure 1. Description of metagenomic library PP1, containing 113,742 clones that harbor inserts of average size of 5.2 kb, along with phenotypic screening and analysis of active clones. The chart shows the number (and percentages) of clones screened for each activity (for example: "11,668, 10 %", etc). Activities screened include: lipase/esterase (LIP/EST), cellulose (CEL), amylase (AMY), protease (PROT), laccase (LAC), and xylanase (XYL); clones without specific activity-based screening are shown as CIX (chloramphenicol, IPTG/XGal system). Details about the screening system, number of positive clones, screening on crude lysates and temperature influence on enzymatic activity is shown next to each activity slice, matching the corresponding color (light blue: proteases; violet: amylases; red: lipases/esterases; green: cellulases). All plates contained LB-agar as the basic media, supplemented with 12.5 μg/ml chloramphenicol (CopyControl pCC1BAC resistance).

temperature) to lower temperatures resulting in stability decrease and a compensatory high-reaction rate by decreasing the activation free-energy barrier between the ground state and the transition state (16, 44). The effect of temperature on hydrolytic activities from tested clones seems to be well correlated with what has been reported for other cold-adapted enzymes, presenting an apparent optimal activity around 35-45 °C and gradual loss of activity or instability over 50 °C. However, the behavior of at least one of the studied cellulases is more compatible to that observed for mesophilic enzymes that are usually still active at temperatures above 60 °C (5). Such a mesophilic property could result from psychrotolerant bacteria inhabiting Antarctic soils, and deserves to be further studied.

The possibility of isolating novel enzymes relies on the "taxonomic" diversity present in the samples studied, which is in turn dependent on the reliability of the DNA purification methodologies (direct versus indirect, relative abundance of bacterial, eukaryotic and archaeal DNA), efficiency of the cloning strategy (diversity and size of

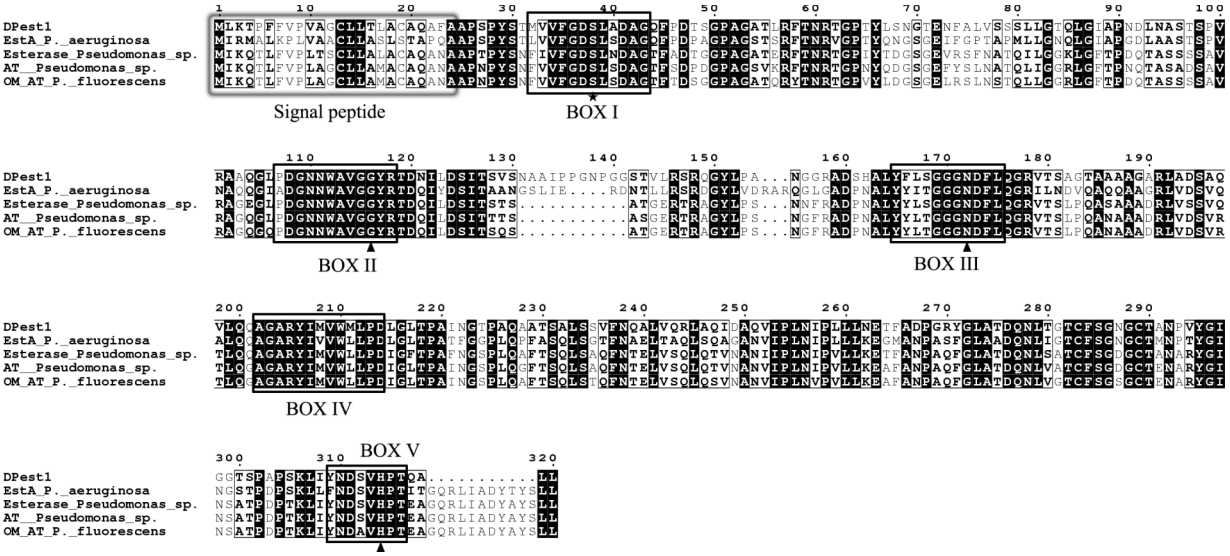


Figure 2: Amino acid sequence alignment of N-terminal domains of membrane esterase DPest1 and other members of auto transporting lipolytic enzymes. Signal peptides are boxed in light grey, and five conserved domains in dark squares. Active-site serine is highlighted with a star, and other conserved residues playing important roles in catalytic activity for SGNH-subfamily of GDSL-hydrolases are shown with black arrows.

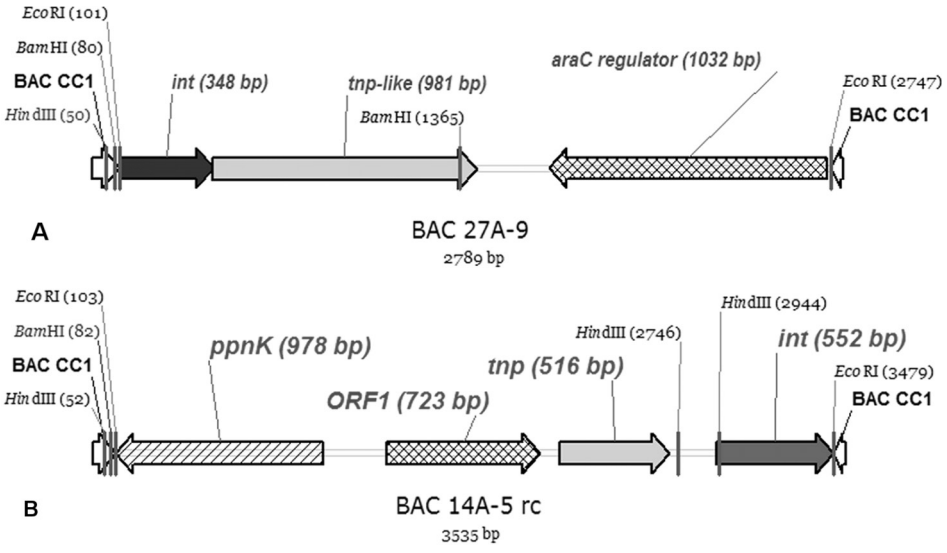


Figure 3. Schematic representation of gene organization and restriction map of BAC 27A-9 (A; GenBank FR716471) and BAC 14A-5 (B; GenBank FR716472) clones. Elements compatible with genetic mobile elements and selected restriction sites are shown, and their respective position in the insert are in parenthesis. BAC CC1 indicates both ends of the pCC1BAC CopyControl cloning system flanking the insert.

Table 2. ORFs and closest matches in the inserts from intI-positive clones

BAC/Clone	ORF	Encoded protein	Closest match / Description	Amino acid identity (%)	Accession number
	tnpB	Hypothetical transposase	I*S6110 transposase TnpB from Mycobacterium tuberculosis*	61	GenBank EFD78722.1
7A-182	*tnpA*	Hypothetical transposase	IS*Myma03*_aa1-like transposase from *Mycobacterium* sp.	58	NCBI YP_001849722.1
7A-284 10A-42	*greA*	Hypothetical transcription elongation protein	Transcription elongation factor GreA from *Mycobacterium tuberculosis*	35	NP_215596.1
	ORF1	Hypothetical alkylmercury lyase	Alkylmercury lyase from *Mycobacterium gilvum*	34	NCBI YP_001134100.1
	ORF2	Transcriptional regulator	Fis family regulator from *Mycobacterium* sp	39	YP_639762.1
10A-148	*lysC*	Putative Asp-kinase	Aspartate kinase from *Clavibacter michiganensis* subsp *michiganensis*	79	NCBI YP_001221642.1
	asdA	Putative Asp-dehydrogenase	Aspartate-semialdehyde dehydrogenase from *Clavibacter michiganensis* subsp *michiganensis*	76	NCBI YP_001221643
	pbp-like	Putative PBP	Penicillin binding protein 1a (carboxipeptidase) from C*lavibacter michiganensis* subsp *michiganensis*	37	NCBI YP_001221659.1
	ORF1	Putative phosphohydrolase	Predicted phosphohydrolase from *Clavibacter michiganensis* subsp *michiganensis*	54	NCBI YP_001221658.1
14A-5	*int*	Integrase	Integron catalytic subunit from *Psychrobacter cryohalolentis*	86	NCBI YP_579984
	tnpA	Hypothetical transposase	IS*3*/IS*911* transposase from *Psychrobacter cryohalolentis*	73	NCBI YP_579985
	*ORF*1	Hypothetical protein	Predicted membrane protein from *Psychrobacter cryohalolentis*	97	NCBI YP_580905
	ppnK	NAD(+) kinase	Inorganic polyphosphate/ ATP-NAD kinase from *Psychrobacter cryohalolentis*	97	NCBI YP_580904
27A-9	*int*	Putative integrase	Phage integrase from *Rhodococcus opacus*	79	NCBI YP_002784538
	ORF1	Hypothetical recombinase/ topoisomerase	Hypothetical DNA breaking-rejoining enzyme from *Rhodococcus opacus*	42	NCBI YP_002784537
	ORF2	Transcriptional regulatory protein	Hypothetical HTH AraC transcriptional regulator from *Rhodococcus erythropolis*	75	NCBI ZP_04384921

the inserts, efficiency of ligation and transformation), and availability of screening approaches.

By direct PCR-amplification from eDNA it was possible to screen for *intI*-type genes without the need of cloning strategies or metagenomic approaches. The amplified product is part of an *intI*-type gene encoding a putative integrase having 58 % amino acid identity (E-value $6e^{-49}$) with an integrase from *Marinobacter* sp.

The main limitation of this approach is that only known genes are screened by this methodology, more than one PCR-product could be obtained having the same size (and therefore only those sequenced are detected), and genes having a low degree of similarity with the known genes may escape from the discrimination power of the primers used.

On the other hand, the absence of positive results for β-lactamase-encoding genes (*bla*) indicates that either eDNA does not contain this type of gene, or the previously mentioned factors are involved (presence of unknown *bla* genes, failures in the PCR, etc.).

In addition, no gene cassettes were detected by direct PCR, using different primer combinations. As previously reported, the detection of gene cassettes by this approach in different samples is feasible (37), suggesting that the

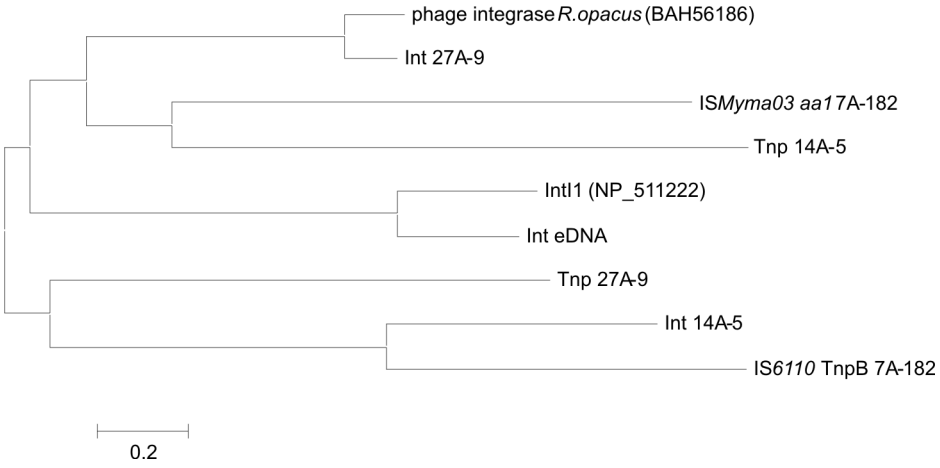


Figure 4. Aminoacid sequence relationship between metagenomic-derived hypothetical integrases (Int 27A-9, Int 14A-5, and Int eDNA) and transposases (ISMyma03 aa1 7A-182, Tnp 27A-9 and IS6110 TnpB 7A-182).

intI-like genes detected in the eDNA studied in this work could be associated with other types of genetic elements non homologous to the already known gene cassettes.

By PCR-screening of DNA extracted from the clone pools containing the metagenomic library PP1, six pools: 7, 10, 14, 21, 22 and 27, yielded a positive result when integrase-specific (degenerate) primers were used.

The second step was to isolate clones included in each positive pool, and colony-PCR was performed using different primer combinations (Table 1) on nearly 3,500 clones from the PP1 library. Two positive clones were detected in pools 7 (7A-182 and 7A-284) and 10 (10A-42 and 10A-148), and one positive clone in pools 14 (14A-5) and 27 (27A-9). None of the tested clones from the remaining pools yielded a positive result, probably indicating that false positives had been obtained in the first screening.

Positive clones were isolated, BACs extracted and *Eco*RI-digested to assess the molecular size of the inserts; BAC 7A-182, 7A-284 and 10A-42 possess an insert of around 6 kb, and BACs 10A-148, 14A-5 and 27A-9 contain inserts of 8 kb, 3.5 kb and 2.5 kb, respectively. The inserts were partially sequenced by genome walking and the results are shown in Figure 3 and Table 2.

The BAC 27A-9 insert (Figure 3A) includes three ORFs encoding for a hypothetical integrase having 78 % amino acid identity with a phage integrase family from *Rhodococcus opacus* (GenBank YP_002784538), a putative recombinase having 41 % identity with a hypothetical recombinase/topoisomerase also from *R. opacus* (GenBank YP_002784537), and a putative regulatory protein with 74 % identity with a helix-turn-helix (HTH) domain-containing regulator from the AraC family of transcriptional regulators from *Rhodococcus erythropolis* (GenBank EEN87802). The GC content of the whole insert and the individual genes was calculated as 64 % (insert), 62 % (ORF1), 70 % (ORF2), and 60 % (ORF3). From these results, it seems highly probable that the ORFs contained in the insert are strongly related to *Rhodococcus* (GC% complete genome: 68 %). In addition, the first two ORFs are part of a 111-kb plasmid (pKNR) from *R. opacus* B4 strain (GenBank NC_012523), keeping the same arrangement observed in the BAC clone, whereas the third one is absent from the pKNR plasmid, being the closest match an ORF present in the genome from a *Rhodococcus erythropolis* SK121 strain (GenBank ACNO01000037). These results suggest that the gene organization in BAC 27A-9 appears to be unique, and a gene rearrangement within the genus could have occurred, probably facilitated by the presence of plasmids harbored by some strains like *R. opacus* B4.

On the other hand, BAC 14A-5 harbors an insert containing four genes closely related to genes from *Psychrobacter cryohalolentis* (Figure 3B and Table 2), encoding for a putative integrase, a hypothetical IS*3*/IS*911* transposase, a predicted membrane protein and an inorganic polyphosphate/ATP-NAD kinase, having 86 %, 73 %, 97 % and 97 % amino acid identity (GC contents of 41 %, 42 %, 45 % and 49 %) with equivalent genes in the chromosome of *Psychrobacter cryohalolentis* K5 (GenBank NC_007969; GC content 43 %), respectively. As for the gene arrangement observed in BAC 27A-9, BAC 14A-5 presents an apparently unique genetic synteny different from that from the *P. cryohalolentis* reference sequence, suggesting that the observed lack of syntenic linkage or arrangement could be due to evolutionary divergence within the species.

The most remarkable features found in the *intI*-positive BACs are the presence of several transposase-encoding genes which could be related to the mobilization of transposon-associated genes between different microorganisms. In addition, the *intI*-specific primers seem to be

also suitable for the detection of transposon-associated elements. This could be related to a high variability within both families of genes (*tnp* and *int*) and/or to a low-discrimination power of the used primers. In fact, *in silico* analysis including sequence multi-alignment on both reference sequences and recombinant clones harboring either *int* or *tnp* genes shows that primer hybridization could occur indistinctively (data not shown) and are somewhat related (Figure 4). A more complete analysis on the sequences and their features is needed in order to look for similarities between transposases and integrases, and to know if a common evolution process has taken place.

The evidence for the presence of genetic elements related to recruitment and mobilization of genes such as transposons and integrons in an extreme environment like Antarctica (as in many other settings and environments) reinforces the hypothesis for the origin of some of the most "familiar" genes, which could be already present in different "gene pools", waiting to be captured by different recombinases and further disseminated by mobile elements among "human-associated" microorganisms (especially pathogens) in various settings.

Acknowledgements: PP is a member of the Carrera del Investigador Científico (CIC) from CONICET, Argentina, and was a recipient of a Postdoctoral fellowship from the Belgian Science Policy Office (BELSPO). This work was mainly supported by a grant from FNRS (FRFC, 2.4561.07) to MG, and partially by a grant from CONICET (PIP 11220090100533) to PP.

References

1. Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF. GDSL family of serine esterases/lipases. Prog Lipid Res 2004; 43**:** 534-52.
2. Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. ISME J 2009; 3**:** 243-51.
3. Atrazhev AM, Elliott JF. Simplified desalting of ligation reactions immediately prior to electroporation into *E. coli*. Biotechniques 1996; 21**:** 1024.
4. Aurilia V, Parracino A, Saviano M, Rossi M, D'Auria S. The psychrophilic bacterium *Pseudoalteromonas halosplanktis* TAC125 possesses a gene coding for a cold-adapted feruloyl esterase activity that shares homology with esterase enzymes from gamma-proteobacteria and yeast. Gene 2007; 397**:** 51-7.
5. 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: 1070-81.
6. Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM, Heidelberg JF. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J 2007; 1**:** 703-13.
7. Collins T, De Vos D, Hoyoux A, Savvides SN, Gerday C, Van Beeumen J, Feller G. Study of the active site residues of a glycoside hydrolase family 8 xylanase. J Mol Biol 2005; 354**:** 425-35.
8. Daniel R. The metagenomics of soil. Nat Rev Microbiol 2005; 3**:** 470-8.
9. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F. Functional metagenomic profiling of nine biomes. Nature 2008; 452**:** 629-32.
10. Elend C, Schmeisser C, Leggewie C, Babiak P, Carballeira JD, Steele HL, Reymond JL, Jaeger KE, Streit WR. Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl Environ Microbiol 2006; 72**:** 3637-45.
11. Elsaied H, Stokes HW, Nakamura T, Kitamura K, Fuse H, Maruyama A. Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. Environ Microbiol 2007; 9**:** 2298-312.
12. Feller G, Gerday C. Psychrophilic enzymes: hot topics in cold adaptation. Nat Rev Microbiol 2003; 1**:** 200-8.
13. Feng Y, Duan CJ, Pang H, Mo XC, Wu CF, Yu Y, Hu YL, Wei J, Tang JL, Feng JX. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. Appl Microbiol Biotechnol 2007; 75**:** 319-28.
14. Frostegard A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P. Quantification of bias related to the extraction of DNA directly from soils. Appl Environ Microbiol 1999; 65**:** 5409-20.
15. Garsoux G, Lamotte J, Gerday C, Feller G. Kinetic and structural optimization to catalysis at low temperatures in a psychrophilic cellulase from the Antarctic bacterium *Pseudoalteromonas haloplanktis*. Biochem J 2004; 384**:** 247-53.
16. Georlette D, Blaise V, Collins T, D'Amico S, Gratia E, Hoyoux A, Marx JC, Sonan G, Feller G, Gerday C. Some like it cold: biocatalysis at low temperatures. FEMS Microbiol Rev 2004; 28**:** 25-42.
17. Gerday C, Aittaleb M, Bentahir M, Chessa JP, Claverie P, Collins T, D'Amico S, Dumont J, Garsoux G, Georlette D, Hoyoux A, Lonhienne T, Meuwis MA, Feller G. Cold-adapted enzymes: from fundamentals to biotechnology. Trends Biotechnol 2000; 18**:** 103-7.
18. Gupta R, Rathi P, Gupta N, Bradoo S. Lipase assays for conventional and molecular screening: an overview. Biotechnol Appl Biochem 2003; 37**:** 63-71.
19. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 2004; 68**:** 669-85.
20. Handelsman J, Liles M, Mann D, Riesenfeld C, Goodman RM. Cloning the metagenome: culture-independent access to the diversity and functions of the uncultivated microbial world. In: Wren B, Dorrell N, editors. Methods in Microbiology - Functional Microbial Genomics. London. Academic Press, 2002, p. 241-55.
21. Hardeman F, Sjoling S. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. FEMS Microbiol Ecol 2007; 59**:** 524-34.
22. Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. Appl Microbiol Biotechnol 1995; 43**:** 667-74.
23. Henne A, Daniel R, Schmitz RA, Gottschalk G. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl Environ Microbiol 1999; 65**:** 3901-7.
24. Henne A, Schmitz RA, Bomeke M, Gottschalk G, Daniel R. Screening of environmental DNA libraries for the presence

of genes conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbiol 2000; 66: 3113-6.
25. Hu Y, Zhang G, Li A, Chen J, Ma L. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. Appl Microbiol Biotechnol 2008; 80: 823-30.
26. Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 1998; 180: 4765-74.
27. Knizewski L, Steczkiewicz K, Kuchta K, Wyrwicz L, Plewczynski D, Kolinski A, Rychlewski L, Ginalski K. Uncharacterized DUF1574 leptospira proteins are SGNH hydrolases. Cell Cycle 2008; 7: 542-4.
28. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P. A bioinformatician's guide to metagenomics. Microbiol Mol Biol Rev 2008; 72: 557-78.
29. Lee MH, Lee CH, Oh TK, Song JK, Yoon JH. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. Appl Environ Microbiol 2006; 72: 7406-9.
30. Lee SW, Won K, Lim HK, Kim JC, Choi GJ, Cho KY. Screening for novel lipolytic enzymes from uncultured soil microorganisms. Appl Microbiol Biotechnol 2004; 65: 720-6.
31. Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl Environ Microbiol 2003; 69: 2684-91.
32. Lim SP, Gan SN, Tan IK. Degradation of medium-chain-length polyhydroxyalkanoates in tropical forest and mangrove soils. Appl Biochem Biotechnol 2005; 126: 23-33.
33. Lorentz K. Routine alpha-amylase assay using protected 4-nitrophenyl-1, 4-α-D-maltoheptaoside and a novel α-glucosidase. Clin Chem 2000; 46: 644-9.
34. MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D, Osburne MS. Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 2001; 3: 301-8.
35. Manichanh C, Chapple CE, Frangeul L, Gloux K, Guigo R, Dore J. A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. Nucleic Acids Res 2008; 36: 5180-8.
36. Moreira D, Rodriguez-Valera F, Lopez-Garcia P. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. Environ Microbiol 2004; 6: 959-69.
37. Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, Nevalainen KM, Stokes HW. Recovery of new integron classes from environmental DNA. FEMS Microbiol Lett 2001; 195: 59-65.
38. Percival Zhang YH, Himmel ME, Mielenz JR. Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv 2006; 24: 452-81.
39. Poulsen TS, Johnsen HE. BAC end sequencing. In: Zhao S, Stodolsky M, editors. Bacterial Artificial Chromosomes. Totowa, NJ, Humana Press, 2004, p. 157-62.
40. Ranjan R, Grover A, Kapardar RK, Sharma R. Isolation of novel lipolytic genes from uncultured bacteria of pond water. Biochem Biophys Res Commun 2005; 335: 57-65.
41. Rhee JK, Ahn DG, Kim YG, Oh JW. New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. Appl Environ Microbiol 2005; 71: 817-25.
42. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet 2004; 38: 525-52.
43. Rondon MR, Raffel SJ, Goodman RM, Handelsman J. Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. Proc Natl Acad Sci USA 1999; 96: 6451-5.
44. Siddiqui KS, Cavicchioli R. Cold-adapted enzymes. Annu Rev Biochem 2006; 75: 403-33.
45. Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, Bali V, Batra N. Metagenomics: Concept, methodology, ecological inference and recent advances. Biotechnol J 2009; 4: 480-94.
46. Streit WR, Schmitz RA. Metagenomics-the key to the uncultured microbes. Curr Opin Microbiol 2004; 7: 492-8.
47. Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, Pootanakit K, Eurwilaichitr L. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. J Biotechnol 2008; 133: 42-9.
48. Uchiyama T, Miyazaki K. Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr Opin Biotechnol 2009; 20: 616-22.
49. Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, Ryu S. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Appl Environ Microbiol 2004; 70: 7229-35.
50. Zhou A, Jiang X, Xu X. Improved alkaline lysis method for rapid isolation of plasmid DNA. Biotechniques 1997; 23: 592-4.

Recibido: 26/10/10 – Aceptado: 22/02/11