Advances in Experimental Medicine and Biology 883

Nevan J. Krogan
Mohan Babu *Editors*

Prokaryotic Systems Biology

Springer

Advances in Experimental Medicine and Biology

Volume 883

More information about this series at http://www.springer.com/series/5584

Nevan J. Krogan • Mohan Babu
Editors

Prokaryotic Systems Biology

Springer

Editors
Nevan J. Krogan
Cellular and Molecular Pharmacology
University of California, San Francisco
San Francisco, California, USA

Mohan Babu
Department of Biochemistry
University of Regina
Regina, Saskatchewan, Canada

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-319-23602-5 ISBN 978-3-319-23603-2 (eBook)
DOI 10.1007/978-3-319-23603-2

Library of Congress Control Number: 2015955839

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Ruedi Aebersold Institute of Molecular Systems Biology, Zurich, Switzerland

Faculty of Science, University of Zurich, Zurich, Switzerland

Mohan Babu Department of Biochemistry, University of Regina, Regina, SK, Canada

Vaibhav Bhandari Department of Biochemistry, University of Toronto, Toronto, ON, Canada

Steven Bowden Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama, Ikoma, Nara, Japan

Cedoljub Bundalovic-Torma Department of Molecular Structure and Function, Hospital for Sick Children, Toronto, ON, Canada

Charles Calmettes Department of Biochemistry, University of Toronto, Toronto, ON, Canada

Catherine S. Chan Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

Tatyana N. Chernikova School of Biological Sciences, Bangor University, Gwynedd, UK

Colin A. Cooper Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Viktor Deineko Department of Biochemistry, University of Regina, Regina, SK, Canada

Karine Dufresne Department of Microbiology, Infectiology and Immunology, Université de Montréal, Montreal, QC, Canada

Andrew Emili Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

Peter N. Golyshin School of Biological Sciences, Bangor University, Gwynedd, UK

Alla Gagarinova Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

Olga V. Golyshina School of Biological Sciences, Bangor University, Gwynedd, UK

Walid A. Houry Department of Biochemistry, University of Toronto, Toronto, ON, Canada

Andrew Judd Department of Biochemistry, University of Toronto, Toronto, ON, Canada

Ashwani Kumar Department of Biochemistry, University of Regina, Regina, SK, Canada

M. Joanne Lemieux Department of Biochemistry, University of Alberta, Edmonton, AB, Canada

Igor Libourel The Biotechnology Institute, University of Minnesota, St. Paul, MN, USA

Iain L. Mainprize Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Trevor F. Moraes Department of Biochemistry, University of Toronto, Toronto, ON, Canada

G. Moreno-Hagelsieb Department of Biology, Wilfrid Laurier University, Waterloo, ON, Canada

Hirotada Mori Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama, Ikoma, Nara, Japan

Ai Muto Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama, Ikoma, Nara, Japan

Nicholas N. Nickerson Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Nicholas N. Nickerson Department of Infectious Diseases, Genentech Inc., South San Francisco, CA, USA

Yuta Otsuka Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama, Ikoma, Nara, Japan

Rashmi Panigrahi Department of Biochemistry, University of Alberta, Edmonton, AB, Canada

Catherine Paradis-Bleau Department of Microbiology, Infectiology and Immunology, Université de Montréal, Montreal, QC, Canada

John Parkinson Department of Molecular Structure and Function, Hospital for Sick Children, Toronto, ON, Canada

Leopold Parts EMBL-EBI (South Building), Wellcome Trust Genome Campus, Saffron Walden, United Kingdom

Ana Popovic Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, Canada

Seesandra Venkatappa Rajagopala J. Craig Venter Institute, Rockville, MD, USA

G. Santoya Instituo de Investigaciones Químico Biológicas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, Mexico

Olga T. Schubert Institute of Molecular Systems Biology, Zurich, Switzerland

Systems Biology Graduate School, Zurich, Switzerland

Rikiya Takeuchi Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama, Ikoma, Nara, Japan

Anatoly Tchigvintsev Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, Canada

Hai Tran School of Biological Sciences, Bangor University, Gwynedd, UK

Raymond J. Turner Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

James Vlasblom Department of Biochemistry, University of Regina, Regina, SK, Canada

Joseph T. Wade New York State Department of Health, Wadsworth Center, Albany, NY, USA

Department of Biomedical Sciences, University at Albany, Albany, NY, USA

Omar Wagih EMBL-EBI (South Building), Wellcome Trust Genome Campus, Saffron Walden, United Kingdom

Barry L. Wanner Department of Biological Sciences, Purdue University, West Lafayette, IN, USA

Michail M. Yakimov Institute for Coastal Marine Environment, Messina, Italy

Alexander F. Yakunin Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, Canada

Katsushi Yokoyama Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama, Ikoma, Nara, Japan

Chapter 1
Metagenomics as a Tool for Enzyme Discovery: Hydrolytic Enzymes from Marine-Related Metagenomes

Ana Popovic, Anatoly Tchigvintsev, Hai Tran, Tatyana N. Chernikova, Olga V. Golyshina, Michail M. Yakimov, Peter N. Golyshin, and Alexander F. Yakunin

Abstract This chapter discusses metagenomics and its application for enzyme discovery, with a focus on hydrolytic enzymes from marine metagenomic libraries. With less than one percent of culturable microorganisms in the environment, metagenomics, or the collective study of community genetics, has opened up a rich pool of uncharacterized metabolic pathways, enzymes, and adaptations. This great untapped pool of genes provides the particularly exciting potential to mine for new biochemical activities or novel enzymes with activities tailored to peculiar sets of environmental conditions. Metagenomes also represent a huge reservoir of novel enzymes for applications in biocatalysis, biofuels, and bioremediation. Here we present the results of enzyme discovery for four enzyme activities, of particular industrial or environmental interest, including esterase/lipase, glycosyl hydrolase, protease and dehalogenase.

Keywords Metagenome • Gene library • Gene discovery • Enzyme screening • Hydrolase

1.1 Introduction to Metagenomics and Its Applications

Prokaryotes constitute the largest fraction of individual organisms on Earth, accounting for up to 10^8 separate genotypes, with conservative estimates of up

A. Popovic • A. Tchigvintsev • A.F. Yakunin (✉)
Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON M5G 1L6, Canada
e-mail: a.iakounine@utoronto.ca

H. Tran • T.N. Chernikova • O.V. Golyshina • P.N. Golyshin
School of Biological Sciences, Bangor University, Gwynedd, LL57 2UW, UK

M.M. Yakimov
Institute for Coastal Marine Environment, CNR, 98122 Messina, Italy

N.J. Krogan, M. Babu (eds.), *Prokaryotic Systems Biology*, Advances in Experimental Medicine and Biology 883, DOI 10.1007/978-3-319-23603-2_1

to 52,000 microbial species residing in just one gram of soil, and several hundred to several thousand species in just one millilitre of sea water (Simon and Daniel 2011; Roesch et al. 2007; Kemp and Aller 2004; Ravenschlag et al. 1999; Schloss and Handelsman 2005). Less than one percent of these microorganisms, however, are culturable in the laboratory, and amenable to traditional experimental studies (Giovannoni et al. 1990). Through the advent of metagenomics, we are just now starting to gain insight into the rich microbial worlds thriving within distinct habitats. One of the first heralded successes of metagenomics was the discovery of bacteriorhodopsin in marine bacterioplankton (Beja et al. 2000). This chapter gives an overview of the importance and applications of function-based metagenomic studies, and describes enzyme screening of metagenome libraries and findings to demonstrate what metagenomes have to offer.

1.1.1 Metagenomics and Its Approaches

Metagenomics is the study of community genetics through the extraction and direct analysis of environmental DNA, most often via creating large or small insert DNA libraries transformed into *E. coli* as a surrogate host. It allows us to circumvent the problems associated with culturing environmental bacteria and to study the biodiversity and biogeochemical roles of the communities through sequence analysis and function-based enzyme screens (Fig. 1.1). The increasingly more accessible and economical Next-Generation Sequencing platforms and continuous advances in computational biology allow us to analyse ever larger sets of sequence data, but prediction and annotation of new genes still relies on sequence similarity to already characterized genes and pathways in the public databases (GenBank, UniProt, KEGG, etc.). As a result, 40–50 % of genes in genomes are routinely labelled as "hypothetical" or proteins of unknown function (Koonin and Galperin 2003; Ferrer et al. 2007; Pelletier et al. 2008). In this scenario, function-based metagenomics is invaluable. The magnitude of microbial and protein diversity of marine metagenomes was demonstrated by two landmark papers by Venter et al., which together revealed over 500 new species, over 6 million protein encoding genes, and almost 2000 new protein families with unknown function (Venter et al. 2004; Yooseph et al. 2007). A more recent high-throughput metagenomics project identified over 27,000 putative carbohydrate-active genes in the cow rumen metagenome and demonstrated the presence of glycosyl hydrolase activity in 51 of 90 tested proteins.

The experimental approaches of functional metagenomics include developing new cultivation methods, meta-transcriptomics, meta-proteomics, meta-metabolomics, and enzyme screening (Rondon et al. 2000; Ferrer et al. 2007; Simon and Daniel 2011; Uchiyama and Miyazaki 2009). The enzyme screening approach involves gene expression and directly assaying metagenomic gene libraries for the ability to modify or hydrolyze a specific chemical substrate. Most often, this means expressing metagenomic enzymes from native or inducible promoters in *E. coli* and detecting enzymatic activity using chromogenic or insoluble substrates in

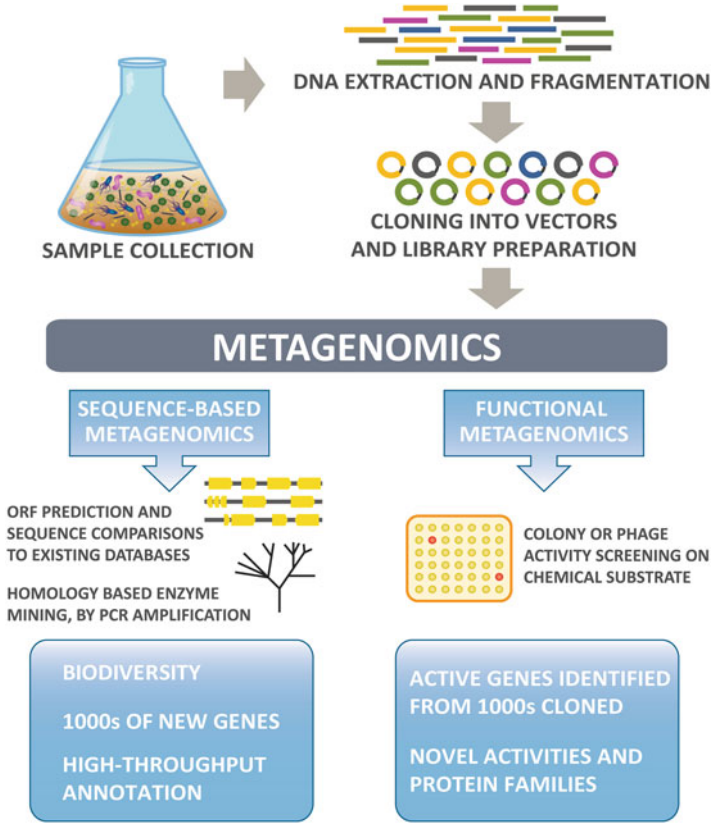

Fig. 1.1 Overview of sequence-based and functional metagenomics

agar (Rondon et al. 2000). An alternate approach is to clone environmental DNA fragments into a lambda phage-based expression vector and to screen for particular enzymatic activities directly on phage plaques (Ferrer et al. 2005). Enzymatic screening of metagenome libraries allows mining for new enzyme activities, and offers the possibility to discover novel families of enzymes with no sequence similarity to previously characterized enzymes found in BRENDA or Uniprot. It also offers an immense repository of new enzymes with an incredible variety of characteristics evolved to accommodate the unique environments that the microbes reside in.

Screening of metagenome gene libraries has greatly expanded the number of novel enzymes, including over 130 new nitrilases and many cellulases, carboxyl esterases, and laccases (Robertson et al. 2004; Lorenz and Eck 2005; Beloqui et al. 2006). Recently, metagenomes of several extreme environments have also been explored and revealed a rich biochemical diversity of enzymes adapted to function under extreme conditions, such as low or high temperatures, low or high pH, and high salt concentrations (Ferrer et al. 2007). Biochemical and structural

characterization of these enzymes revealed different molecular mechanisms of adaptation to extreme environmental conditions (Feller and Gerday 2003; Olufsen et al. 2005; Siddiqui and Cavicchioli 2006). The potential for enzyme discovery in metagenomes has not gone unnoticed in the industrial sector, and several companies such as Diversa, Genencor International (now part of DuPont), Henkel, Degussa (now Evonik Industries), among others, have already made efforts in this area (reviewed by Lorenz and Eck 2005). The efforts will only increase as comparative studies have shown that replacing conventional industrial processes with enzymatic processes in a multitude of industries have indeed lead to savings and cleaner production (reviewed by Jegannathan and Nielsen 2013, of Novozymes).

1.1.2 Metagenomics and Enzyme Discovery

Examples of industries which employ enzymatic processes include pulp and paper production, household detergent production, the textile industry, food and beverage industries, animal feed production and biodiesel production, among others (Jimenez et al. 1999; Nguyen et al. 2008; Hemachander and Puvanakrishnan 2000; Saeki et al. 2007; Aly et al. 2004; Osma et al. 2010; Okamura-Matsui et al. 2003; Gado et al. 2009; Monsan and Donohue 2010; Hernández-Martín and Otero 2008). Novozymes, which holds a 48 % share of the global market for industrial enzymes, has reported enzyme business sales up by 5 % in 2013, at 1574 million euros, with the strongest sales to household care (including most importantly detergents) and bioenergy industries. They also calculated that customers saved approximately 52 million tons of CO_2 (editor Bedingfield 2013). In the trend toward alternative, cleaner, cheaper and more efficient processes, the potential for enzyme application is great.

Cold-active enzymes offer particular advantages in the household cleaner and food industries, including primarily energy savings compared to traditional high-temperature processes, both for consumers, in the case of detergents marketed for cold-water cleaning, and manufacturing processes (Cavicchioli et al. 2002). These enzymes maintain high levels of activity and specificity, and many have the convenient property of being thermolabile, allowing for easy inactivation prior to subsequent processing steps. As a result of their inherent flexibility, in order to remain active and mobile at low temperatures, they have been proposed as candidates for organic synthesis in non-aqueous or mixed aqueous-organic solvents, where the absence of water stabilizes and inhibits activity of many mesophilic enzymes. Similarly halophilic enzyme instability in aqueous low salt environments has made them attractive for non-aqueous synthesis (van den Burg 2003; Sellek and Chaudhuri 1999; Cavicchioli et al. 2002). Marine environments offer an ideal opportunity to sample microbial communities which have evolved to thrive at both cold temperatures and hypersaline conditions.

The Earth's biosphere is predominantly aqueous (70 % water) and cold (around 5 °C). Marine microorganisms that are able to grow at low temperatures have evolved different adaptation mechanisms to survive under these conditions.

Temperature is one of the most important factors for enzyme activity as the reaction rate can be reduced 30–80 times when the temperature drops from 37 to 0 °C (Lonhienne et al. 2000). To explore the biochemical diversity of marine metagenomes, we have screened twelve metagenomic libraries from a diverse set of marine-related environments for one or more of the following hydrolytic functions: esterase/lipase, glycosyl hydrolase and protease, which have applications in one or more of the industries listed above, and lastly dehalogenase, which together with the above mentioned enzyme activities, plays an important role in bioremediation and detoxification of halogenated organic pollutants (Brisson et al. 2012; Jegannathan and Nielsen 2013).

1.2 Metagenome Gene Library Preparation

1.2.1 Environments and Library Preparation

Immense sequence diversity has so far been documented in environmental metagenomes, suggesting significant metabolic and biochemical variety as well (Dinsdale et al. 2008; Yooseph et al. 2007). To exploit this diversity, we selected a set of marine-related metagenomes to search for cold-active and salt-tolerant esterases, lipases, proteases, glycosyl hydrolases and dehalogenases.

We sampled 12 marine-related environments, from communities thriving under extreme anoxic deep sea conditions (Urania, Kryolo, Medee, Rimicaris Gill and Gut) or extremely low pHs (Vulcano), to various regions associated with heavy industrialization and oil-contamination of the Mediterranean or Barents Sea (Messina, Milazzo, Priolo, Haven, Kolguev, Murmansk). As examples of the environmental diversity, the Rimicaris Gut and Gill libraries were prepared from dense bacterial communities inhabiting *Rimicaris exoculata* shrimp, which lives on chimney walls of hydrothermal vents in the Mid-Atlantic ridge, over 3 km beneath the surface (Williams and Rona 1986). The Haven library, though, is derived from samples of tar collected on the coast of Genoa, Italy, where the Amoco Milford Haven tanker exploded in 1991 releasing thousands of tonnes of crude oil. The temperatures of all of selected environments range from 3 to 15 °C and water salinity ranges from 3.1 to 3.8 %. Selected communities were treated for 1 month with phenanthrene and pyrene (Milazzo, Messina) or crude oil (Kolguev, Murmansk), prior to extraction of large molecular weight DNA.

Two types of metagenomic libraries were prepared from the marine samples—large DNA insert (40,000 bases) fosmid libraries and small DNA insert (4000–7000 bases) lambda phage libraries, using commercially available kits (Epicentre's CopyControl Fosmid and Stratagene's, now Agilent's, Lambda-ZAP). The names and descriptions of the 12 metagenomic libraries are provided in Table 1.1.

Table 1.1 Metagenomic libraries prepared and screened for enzyme activity

Metagenomic library	FosmidTotal clones	Lambda-ZAPTotal clones	Notes
Contaminated source			
Kolguev Island	N/A	100,000	Crude oil-degrading psychrophilic community; Barents Sea
Port of Murmansk	N/A	100,000	Crude oil-degrading psychrophilic community; Barents Sea
Milazzo enrichment	50,000	2400	Marine-based enrichment cultures with phenanthrene and pyrene; Mediterranean Sea
Messina enrichment	21,000	1000	Marine-based enrichment cultures with phenanthrene and pyrene; Mediterranean Sea
Priolo sediment	3000	40,000	Anoxic community, heavy industrialization and crude/refined oil contamination; harbor of Priolo Gargallo, Italy
Haven sediment	9200	25,000	Petroleum contamination; harbor of Arenzano, Italy
Specialized environment			
Rimicaris exoculata gut	11,100	150,000	Deep sea shrimp metagenome; Mid-Atlantic Ridge
Rimicaris exoculata gill	20,000	350,000	Deep sea shrimp metagenome; Mid-Atlantic Ridge
Urania basin interface	N/A	100,000	Deep hypersaline anoxic lake; Mediterranean Sea
Kryos brine interface	9300	620,000	Deep hypersaline anoxic lake; Mediterranean Sea
Medee basin interface	18,432	N/A	Deep hypersaline anoxic lake (salinity 170–190 g/l); Mediterranean Sea, Cycloclasticus naphthalene enrichment
Vulcano acidic pool	3456	N/A	Enrichment made from acidic pool sand/gravel; Mediterranean Sea

1.2.2 Advantages and Disadvantages of Large and Small DNA Insert Libraries

Each of the two types of metagenome gene libraries (large DNA insert fosmid libraries and small DNA insert phage libraries) offers particular advantages and disadvantages. Although the lambda phage libraries may be converted to phagemid clones and screened as colonies, one of the biggest advantages to the phage is lysis of *E. coli* cells at the end of the infection cycle and release of translated metagenomic proteins to the extracellular matrix, and consequently the substrate. Lambda phage are stable for long periods, both at −80 °C and at 4 °C, and library contamination with other laboratory strains is a smaller concern due to the specificity of the host–virus interaction. Perhaps most importantly, toxic effects of metagenomic enzymes generally do not pose problems as they would in cell-based libraries, since gene expression in cells is only driven during the short interval of viral infection. In addition, the lambda libraries, although containing small inserts of on average 4–8 kilobases (4–8 genes), have IPTG-driven expression which allows for higher concentrations of metagenomic proteins. This is important for phage screening, since there is only a short window of infection (approximately 50 min for a wildtype lambda virus) during which the metagenomic genes can be expressed prior to cell lysis. In screening phage, it is possible to screen large numbers of clones (600–2000 per 10 cm plate, depending on the substrate) very quickly and easily. Because the enzyme is released to the extracellular matrix upon cell lysis, the activity is often seen earlier than with fosmid libraries.

Alternatively, the large insert libraries have the advantage of just that—the presence of a longer metagenomic DNA insert. In some cases, more than one gene is required for activity or correct expression and folding of the enzyme of interest, and fosmid libraries offer the advantage of screening larger gene clusters (up to 50 genes), in many cases entire operons. In the particular cases of two esterases we have identified during functional screening, we have also found predicted lipase chaperones immediately downstream of the active genes. Sequencing the genetic neighbourhood of an active fosmid clone can also offer hints about the native metabolic role of the enzyme in question, and more accurately predict the taxonomy of the source organism. Finally, vectors used for cloning large-insert DNA fragments are typically low copy which, unless induced by engineered copy-control, minimize effects of toxic genes (Taupp et al. 2011) and allow basal expression levels from native promoters, avoiding inclusion body formation often associated with overexpression in *E. coli*.

1.3 Agar-Based Enzyme Screens for Hydrolytic Enzymes

Agar plate-based screening of metagenomic libraries provides a direct and simple approach to mine for industrially useful enzymes that function under diverse conditions—low temperature, extreme pH, nonaqueous, anoxic, hypersaline, among

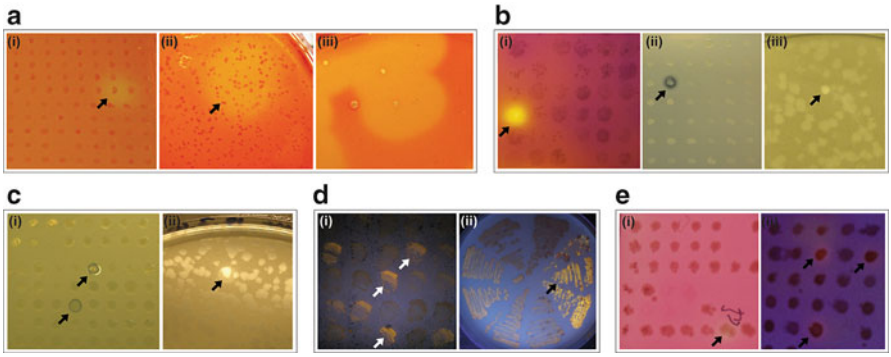

Fig. 1.2 Agar-based enzymatic screening of metagenomic libraries. Panels show positive hits on fosmid or phage library screens as follows: (**a**) Glycosyl hydrolase positive fosmid (i) and phage (ii and iii) clones screened on carboxymethyl cellulose. (**b**) Skim milk screens for protease and glycosyl hydrolase activities, on fosmid libraries with and without a pH indicator (i and ii), and on phage libraries (iii). (**c**) Esterase or lipase positive fosmid (i) and phage (ii) hits screened on tributyrin. (**d**) Lipase positive fosmid (i) and excised phagemid (ii) clones screened on olive oil. (**e**) Fosmid clones positive for dehalogenase activity

others. This method has successfully produced a large number of novel esterases, lipases, proteases, glycosyl hydrolases, laccases, and other enzymes (Lorenz and Eck 2005; Ferrer et al. 2009; Steele et al. 2009). Below we describe the five screens we have used to screen up to 100,000 clones per experiment using Lambda-ZAP or over 6000 clones for all activities using fosmid libraries. Examples of enzyme screens and positive clones are shown in Fig. 1.2.

1.3.1 Esterases and Lipases

We use two substrates to screen for esterase and lipase activity. A commonly used substrate for detection of esterase or lipase activity is the simple ester tributyrin. To specifically detect lipase activity, we use olive oil, which is comprised primarily of the long carbon chain (C >16) triglyceride esters oleic acid, linoleic acid and palmitic acid. The screens are adapted from previously published protocols (Kok et al. 1993; Kouker and Jaeger 1987).

In fosmid screens, the substrates, 1 % tributyrin and 3 % olive oil, are emulsified with 0.5 % gum arabic in standard Luria-Bertani (LB) broth and agar plate media containing appropriate antibiotics. The clones, containing metagenomic DNA, are grown in LB in microtiter plates for several hours, plated on substrate plates using 96-pin or 384-pin replicators and incubated overnight at 37 °C. Emulsified tributyrin gives a turbid appearance to the plates, and hydrolysis by esterases or lipases is seen as a clearing or halo around the colony, or plaque in phage screening (Fig. 1.2c). In the case of olive oil plates, 0.001 % w/v Rhodamine B dye is also added to the

plates, and activity is detected as orange fluorescence under UV light (Fig. 1.2d), presumably caused by formation of complexes of Rhodamine B molecules and the hydrolyzed fatty acids (Kouker and Jaeger 1987). During phage screening, phage clones are preincubated with host cells, added to several millilitres of soft LB-agar containing emulsified tributyrin and gum arabic, and plated on LB-agar plates, containing 1 mM IPTG, overnight at 37 °C.

All plates are kept for an additional 2–4 days at room temperature to 37 °C, and monitored for positive clones.

1.3.2 Glycosyl Hydrolases

One rather simple color-based method to screen for glycosyl hydrolase activity uses carboxymethyl cellulose as a substrate and Congo red dye as an indicator (adapted from Teather and Wood 1982). In this assay, 0.3 % carboxymethyl cellulose is added to conventional plate media (for fosmid screening) or to soft LB-agar (for phage screening), and cells or phage are plated as described in the screen above. After 2–4 days, cellulose is stained with a 0.1 % Congo red solution, and unstained haloes are observed around positive colonies (Fig. 1.2a).

1.3.3 Proteases

The skim milk-based agar screen has been proposed for detecting proteases in soil metagenome libraries (Rondon et al. 2000). In these screens, 1 % skim milk is added to LB-agar plates for colony screening, and 3–4 % skim milk is added to soft LB-agar for phage library screening, as described for screens above (adapted from Rondon et al. 2000). Activity is detected as a clearing in the turbidity (Fig. 1.2b), as the protease degrades milk proteins, mostly caseins. Later, Jones et al. showed that skim milk screens can also detect the metagenomic clones expressing glycosyl hydrolases or releasing acid (Jones et al. 2007). Therefore, to some screens we have added pH indicator dyes (0.5 mM phenol red and/or bromothymol blue) to increase sensitivity and detect the acidic shift during hydrolysis of casein by proteases or lactose by glycosyl hydrolases (Jones et al. 2007). In order to distinguish proteases from glycosyl hydrolases, we rescreened positive clones from the skim milk assay on plates containing X-gal. In pH based assays, including also the assay described below for dehalogenases, however, we have found a higher tendency for false positives, and positive clone genes must be further tested to ensure the enzymatic activity is present.

1.3.4 Dehalogenases

Detection of dehalogenase activity has been described using a variety of haloalkane and haloacid substrates (Holloway et al. 1998). The basic concept involves freeing the halogen (e.g. chloride or bromine) from the substrate and detecting the ensuing pH change with an indicator dye, which results from haloacid formation (HCl, HBr, etc.). The described agar based screen is set up in a similar way to those of the other activities above with the following changes. Colonies or phage are plated on LB-agar plates containing 0.5 mM phenol red and/or bromothymol blue pH indicator dyes, but no substrates. After overnight incubation, or in the case of fosmid libraries, possibly after 2–3 days of incubation, a mixture of melted 0.4 % agarose, 20 mM EPPS buffer (pH 8.0) and a 2.5 mM combination of haloalkanes and haloacids, such as bromoacetic acid, 3-bromo-2-methylpropanoic acid, iodoacetic acid, ethyl-3-(bromomethyl)propanoate, 3-dibromopropanol, 1-iodopropane, is layered on top of the colonies or phage. The plates are incubated at 30 °C, and checked every few minutes for colour change (Fig. 1.2e). The pH change on solid media is short lived as a result of diffusion, therefore requires close monitoring.

In the above described screens, we have found phage particularly useful for screens involving turbid substrates (tributyrin or skim milk), as well as color-based end-point assays (such as the glycosyl hydrolase screen described above). Positive plaques are difficult to spot in screens involving transient colour changes, such as the dehalogenase screen. For these screens, liquid-based microplate assays with fosmid or excised phagemid libraries are perhaps best.

1.4 Screening and Sequencing Results for Marine Metagenome Libraries

1.4.1 General Functional Screening Statistics

Over 1.3 million clones from the 12 described marine-related metagenomic libraries were screened for esterase/lipase, glycosyl hydrolase, protease and dehalogenase activities, yielding 545 positive hits. Over half of these were putative esterases or lipases identified with tributyrin screens, the quickest and most successful assay, followed by glycosyl hydrolases, putative proteases or glycosyl hydrolases from skim milk screens, and finally dehalogenases (Table 1.2). The tributyrin plate screen for esterases and lipases is the most common metagenomic screen, which depending on the metagenomic library has been reported to have a hit rate in the range of 1 positive per 5 to 4000 Mb of DNA screened (Lorenz and Eck 2005; Steele et al. 2009; Uchiyama and Miyazaki 2009). In our work, this screen produced 1 hit per 9 Mb of screened DNA (Table 1.3), which is within the reported range. The glycosyl hydrolase and dehalogenase screens produced comparable frequencies of positive hits (1 hit per 28.4 and 23.9 Mb DNA, respectively). The highest frequency

Table 1.2 Enzyme screening statistics for marine metagenomic libraries

	Esterase/lipase		Glycosyl hydrolase		Protease		Dehalogenase	
Metagenomic library	Screened	Positive	Screened	Positive	Screened	Positive	Screened	Positive
Contaminated source								
Kolguev (phage)	154,000	30	111,500	5	8200	6	–	–
Murmansk (phage)	108,000	41	103,000	9	–	–	–	–
Milazzo (phage)	20,000	8	15,000	0	24,000	0	–	–
Messina (phage)	24,000	18	–	–	25,000	3	–	–
Priolo (phage)	118,500	5	54,000	0	–	–	–	–
Haven (phage)	36,800	11	94,500	4	16,000	2	–	–
Specialized environment								
Rimicaris Gill (fosmid)	–	–	8400	27	8400	38	8400	17
Rimicaris Gut (phage)	137,500	7	–	–	–	–	–	–
Rimicaris Gill (phage)	21,100	8	–	–	–	–	–	–
Urania (phage)	90,800	36	130,000	5	–	–	–	–
Kryos (fosmid)	–	–	3456	49	3456	16	3456	17
Medee (fosmid)	4992	147	4992	0	4992	1	4992	0
Vulcano (fosmid)	1920	32	3456	1	3456	2	3456	0
Total	717,612	343	528,304	100	93,504	68	20,304	34

Table 1.3 Frequency of positive hits (hit per Mb of DNA screened)[a]

Metagenomic library	Esterase/lipase	Glycosyl hydrolase	Protease	Dehalogenase
Contaminated environment				
Kolguev (phage)	1/20.5	1/89.2	1/5.5	–
Murmansk (phage)	1/10.5	1/45.7	–	–
Milazzo (phage)	1/10	0	0	–
Messina (phage)	1/5.3	–	1/33.3	–
Priolo (phage)	1/94.8	0	–	–
Haven (phage)	1/13.3	1/94.5	1/32	–
Specialized environment				
Rimicaris Gill (fosmid)	–	1/12.4	1/8.8	1/19.8
Rimicaris Gut (phage)	1/78.6	–	–	–
Rimicaris Gill (phage)	1/10.5	–	–	–
Urania (phage)	1/10.1	1/104	–	–
Kryos (fosmid)	–	1/2.8	1/8.6	1/8.1
Medee (fosmid)	1/1.4	0	1/200	0
Vulcano (fosmid)	1/2.4	1/138	1/69	0
Total	1/9.1	1/28.4	1/16.2	1/23.9

[a]For calculation of hit frequency, an average of 4000 bp/phage clone and 40,000 bp/fosmid clone were used

of positive esterase hits (1 hit per 1.4 Mb DNA) was obtained from the Medee fosmid library, whereas the Kryos fosmid library produced the highest frequencies of positive glycosyl hydrolase and dehalogenase clones (1 hit per 2.8 and 8.1 Mb DNA, respectively) and the Rimicaris Gill and Kryos fosmid libraries revealed the highest frequencies of positive hits on skim milk screens (1 hit per 8.8 and 8.6 Mb DNA, respectively) (Table 1.3). In some libraries, the high frequency of positive hits can be explained by the presence of multiple copies of a limited number of positive genes (redundancy) as was revealed by DNA sequencing (see below), or a high propensity for false positives in enzyme activity screens using transient pH changes.

1.4.2 Clone Sequencing

Due to the size differences between phagemids (on average 4–7 kilobases) and fosmids (on average 40 kilobases), we used different strategies for sequencing the positive clones. Simple gene walking was used to sequence phagemids. This approach, however, would have been tedious and impractical for fosmids. Fosmid DNA samples were pooled into mixtures of 40–100 clones and submitted for Next Generation sequencing by Illumina. The massive amount of sequence data obtained, the substantial decrease in cost of this sequencing platform, and the increasing number of companies offering the service have made this approach practical. In theory, it is possible to obtain 1000× sequence coverage for samples of 500 pooled fosmids, assuming 200 million reads per lane (Genome Quebec, personal communication), however one must keep in mind inevitable downstream problems of contig assembly, including overlapping clones, underrepresented sequences, short-sequence DNA repeats, or stretches of DNA inherently difficult to sequence. For these reasons, we were relatively conservative in our DNA pool sizes. There are a variety of assemblers for Illumina data (Velvet, ABySS, DNASTAR). We used Geneious, a powerful sequence analysis program, which uses a Velvet based algorithm for sequence assembly. In order to trace the assembled contigs to particular clones, we also sequenced the ends of each fosmid. In lieu of end sequencing, it is also possible to barcode the fosmids, in house or at added expense by the sequencing companies.

We obtained sequences for most isolated phage and phagemid positives and 50 % of our fosmids. Open reading frames were predicted and annotated using a combination of Geneious' gene prediction, Glimmer, BLASTx and MG-RAST. Gene annotation is laborious and requires significant manual curation to accurately resolve and annotate predicted overlapping open reading frames. Genes with predicted enzyme activities of interest were cloned and rescreened in standard *E. coli* expression vectors. Where the putative enzyme could not be identified through sequence-based searches, multiple, if not all, genes were cloned and retested.

Sequencing revealed a high proportion of marine hydrocarbon degrading bacteria in the tested metagenomic libraries. Murmansk and Kolguev libraries both had

a large proportion of the bacterium *Alcanivorax borkumensis*, an oil-degrading bacterium. The deep sea Urania library had predominantly *Marinobacter aquaeolei* and *Marinobacter hydrocarbonoclasticus*, the most abundant sea dwelling bacteria capable of degrading oil or other hydrocarbons, or species with high nucleotide similarity to these. Finally, a majority of positive clones isolated from the Medee deep sea basin were predicted to contain DNA from one or more of the marine polycyclic aromatic hydrocarbon degrading *Cycloclasticus* species. Positive hits isolated from libraries with geographical proximity showed some overlap, as expected, with seven enzymes recovered from both Murmansk and Kolguev screening, two enzymes from Milazzo and Messina libraries, and the same *Cycloclasticus* enzyme isolated from Messina, Vulcano as well as the Medee libraries.

In our experience, one quarter of all initially cloned metagenomic genes could be expressed in *E. coli* and purified sufficiently for biochemical analysis. Quite often, it became necessary to clone protein fragments and remove predicted N-terminal signal peptides or transmembrane domains, identified using prediction programs such as TMHMM or SignalP. Based on sequence, nearly one half (45 %) of the experimentally confirmed metagenomic esterases were predicted to have N-terminal signal or transmembrane sequences. A combination of removing these sequences and using chemical chaperones, such as sorbitol or glycerol, for recombinant protein expression (Prasad et al. 2011), greatly aided in obtaining soluble proteins.

1.4.3 Sequence Analysis of Esterase Positive Genes

Thirty-nine so far cloned and confirmed esterases from marine-related metagenomes belong to a diverse set of protein families (Fig. 1.3). As expected, a majority (70 %) is predicted as α/β-hydrolase family proteins, the largest proportion belonging

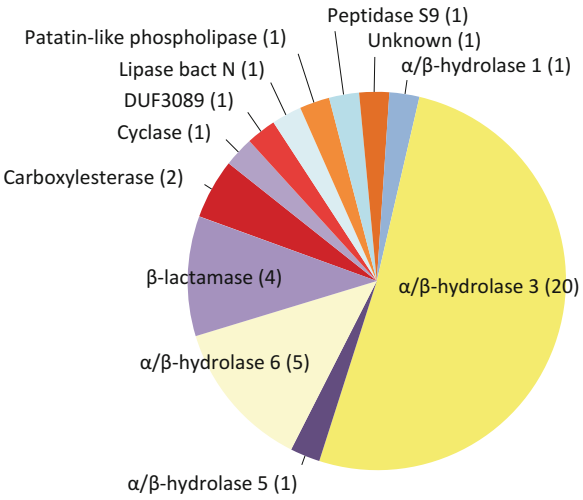

Fig. 1.3 Protein family prediction for 39 confirmed esterases

to the α/β-hydrolase 3 family. Four enzymes are predicted β-lactamases, two carboxylesterases, and the remaining are a predicted lipase, patatin, and interestingly a predicted cyclase, prolyl oligopeptidase, DUF3089 family protein and one completely unknown.

Most of the biochemically characterized α/β-hydrolases contain a Ser-His-Asp catalytic triad in the active site. Sequence analysis of the identified α/β-hydrolases and carboxylesterases reveals conserved GxSxG motifs, suggesting the positions of the catalytic Ser residues, as well as several conserved candidates for the His and Asp residues of the catalytic triad. The β-lactamases each have the typical N-terminal SxxK motif, as well as a conserved Tyr and an additional GxSxG or GxSxx near the C-terminus described for Family VIII esterases (reviewed by Hausmann and Jaeger 2010). The predicted lipase, patatin and DUF3089 proteins also show conserved GxSxG motifs and several conserved His and Asp residues suggesting that these proteins also belong to serine hydrolases. The enzyme of unknown protein family (internally identified as MGS0084) shares low sequence similarity (20–23 % of sequence identity) with only four predicted proteins in Genbank and Uniprot. These proteins have several conserved Ser residues suggesting that these sequences might represent another family of Ser-dependent hydrolases. Finally, both esterase and peptidase activities have been previously detected in select serine proteases, specifically prolyl oligopeptidases predicted to contain an α/β hydrolase fold (Wang et al. 2006). It was found that a single conserved Arg residue can discriminate between the two enzymatic activities. The active residues of the latter enzymes will need to be confirmed through mutagenesis.

Thus, while a majority of the lipolytic enzymes so far mined belong to one of the well-known esterase families, a significant 4 (10 %) are either unknown or predicted to have alternate activities. For esterases from characterized families, we have the opportunity to compare and study particular amino acid substitutions or structural changes that confer psychro or halophilicity.

1.4.3.1 Case Example of Venturing into Novel Sequence Space

MGS0012, one of several enzymes isolated in our functional assays that were predicted as domains of unknown function, shares 99 % sequence identity with a predicted hypothetical protein from *Kordiimonas gwangyangensis*, an organism isolated in the Gwangyang Bay of Korea capable of degrading high-molecular-mass polycyclic aromatic hydrocarbons (Kwon et al. 2005). It also shares 40 % sequence identity with another recently and independently isolated soil metagenomic enzyme EstWSD (Uniprot K9RYB0), which exhibits salt tolerant and solvent tolerant activities (Wang et al. 2013). It is particularly useful that as we isolate and annotate new enzymes in functional metagenomics, we can extend these findings and predictions through protein sequence space to other interesting and environmentally important organisms. A protein sequence alignment (Fig. 1.4) shows that MGS0012 and EstWSD, predicted members of the DUF3089 family, share sequence similarity with uncharacterized proteins from *Dehalococcoides*

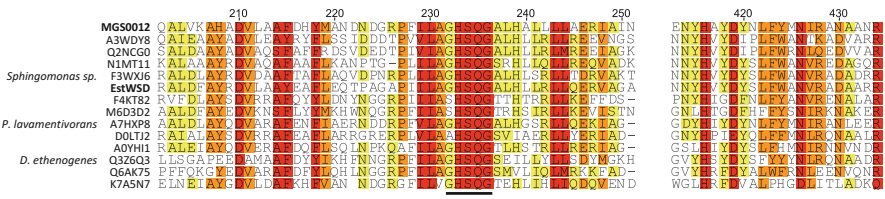

Fig. 1.4 Sequence alignment of the conserved Ser motif in identified metagenomic carboxyl esterase MGS0012 and DUF3089 family proteins

ethenogenes (Q3Z6Q3), an organism known to reductively dechlorinate the groundwater pollutants tetrachloroethene and trichloroethene (Maymo-Gatell et al. 1997), *Parvibaculum lavamentivorans* (A7HXP8), a microbe which degrades the commercial laundry surfactant linear alkylbenzenesulfonate (Schleheck et al. 2000), and proteins from various species of *Sphingomonas* (F3WXJ6), some of which are known to degrade aromatic compounds (Fredrickson et al. 1991, 1995; Baraniecki et al. 2002). In fact, MGS0012 has >30 % sequence identity to 197 proteins in Uniprot. Thus slowly, but surely, we are venturing into novel sequence space, and making progress into those 40 % hypothetical or uncharacterized proteins currently in the database.

1.5 Biochemical Properties of Carboxyl Esterases from Marine Metagenomes

Eight genes encoding metagenomic carboxyl esterases were cloned for overexpression in *E. coli* and purified using affinity chromatography. Detailed biochemical characterization of these enzymes revealed a high prevalence of cold-adapted activity and salt tolerance, in accordance with the microbial environment. The enzyme activities of purified proteins were assayed using a set of *p*-nitrophenyl ester substrates with different acyl chain lengths (C2–C16). They were also tested for temperature optimum, pH preference, as well as salt and organic solvent tolerance, in addition to the substrate specificity.

The metagenomic carboxyl esterases were active over a broad pH range at 30 °C (pH 7–9). These enzymes display a wide range of chain length preferences, ranging from acyl chain length C2 (*p*NP-acetate for MGS0109) to C8 (*p*NP-octanoate for MGS0010). The preference for short acyl chain substrates is typical of carboxyl esterases (Hausmann and Jaeger 2010). However, two enzymes, MGS0012 from the Messina library and ABO1197 from the oil-degrading *Alcanivorax borkumensis* of the Murmansk library, exhibit high activity (up to 30 μmoles/min per mg protein) against the C16 *p*NP-palmitate, a model substrate used for lipase activity. However, the preferred substrates for these two enzymes are α-naphthyl acetate (61 μmoles/min per mg protein) and *p*NP-valerate (nearly 300 μmoles/min per mg

Table 1.4 Biochemical characteristics of purified metagenomic carboxyl esterases

Clone	Metagenomic library	Predicted Pfam	Optimal temp	Activity at 4 °C	Optimal salt conc	Activity	
						0 M NaCl	2 M NaCl
MGS0006	Messina	α/β-hydrolase 3	30 °C	55 %	0 M	100 %	3 %
MGS0010	Messina	β-lactamase	30 °C	32 %	>3.5 M	100 %	144 %
MGS0012	Messina	DUF3089	40 °C	9 %	2 M	100 %	187 %
MGS0018	Rimicaris Gill	α/β-hydrolase 3	30 °C	37 %	0 M	100 %	9 %
MGS0105	Kolguev	β-lactamase	15 °C	41 %	0 M	100 %	0 %
MGS0109	Kolguev	α/β-hydrolase 6	30 °C	50 %	0 M	100 %	0 %
ABO1197	Murmansk	α/β-hydrolase 6	30 °C	66 %	0 M	100 %	40 %
ABO1251	Murmansk	Carboxylesterase	35 °C	46 %	0 M	100 %	23 %

protein), respectively. A weak lipase activity was also detected for MGS0012 in an agar-based olive oil assay, which contains 50–80 % oleic acid (C18).

Temperature optima for seven out of eight enzymes were found to range from 15 to 35 °C, with 32–66 % activity at 4 °C (Table 1.4) suggesting that these proteins are cold-adapted (psychrophilic) enzymes. MGS00012 alone showed an optimal temperature of 40 °C, and only 9 % activity at 4 °C, reminiscent of a mesophilic enzyme. As intracellular salt concentrations vary species by species, and can differ from external salt concentrations (Oren 2002; Christian and Waltho 1961, 1962), we would expect enzymes recovered from different species to reflect these differences. For the eight characterized metagenomic enzymes (Table 1.4), we found a range of salt effects on activity including stimulation and various levels of inhibition. As shown in Table 1.4, three enzymes were completely inhibited by 2 M salt (MGS0006, MGS0105 and MGS0109), while the activity of two other enzymes was stimulated by addition of NaCl, and in fact have optimal activity at 2 M or over 3.5 M concentrations (MGS0012 and MGS0010, respectively). The remaining enzymes show intermediate salt tolerance and lie on various points along this spectrum. Thus, many carboxyl esterases from marine metagenomes are cold-adapted enzymes showing different levels of salt resistance.

1.6 Conclusion

In summary, enzymatic screening of metagenomic marine libraries identifies genes from diverse organisms and protein families with enzymatic properties that reflect the environmental conditions of the microbial community. Most of the esterases we have biochemically characterized are halotolerant or halophilic, cold adapted enzymes, as one would expect for proteins from a marine environment. We have identified several esterases which belong to uncharacterized families or proteins annotated to have alternate functions, which could not have been identified through sequence analysis alone.

Since only 40 % of enzymes from environmental DNA have been suggested to express in *E. coli*, (Gabor et al. 2004), this leaves a large proportion of the environmental gene pool unsampled during a standard enzyme activity screen, missing some potentially very exciting enzymes. In order to increase physiologic and metabolic diversity and therefore close this expression gap, multi-host shuttle vectors have already been designed for expression in *Bacillus subtilis*, *Pseudomonas putida*, *Streptomyces lividans*, and *Rhizobium leguminosarum* (Martinez et al. 2004; Staskawicz et al. 1987; Troeschel et al. 2012; Li et al. 2005; Wexler et al. 2005). Enzymatic screening of metagenomic libraries expressed in these hosts revealed different gene expression profiles indicating that additional metagenomic enzymes can be identified using this approach. For example, these hosts have successfully identified a novel alcohol/aldehyde dehydrogenase in *R. leguminosarum*, and multiple hemolytic clones that were active in *S. lividans* but not *E. coli*.

Presently, DNA sequencing and sequence-based studies of metagenomes far outpace functional metagenomic studies. As the computational load has now fallen on improving high-throughput annotation of thousands of metagenomes, functional studies offer us a targeted approach to find the needles in the haystack. Activity based studies remain crucial in identifying novel enzymes and enzyme families, and allow us to isolate and clone those genes which we know can be expressed in industrially-relevant hosts from gigabases of environmental DNA. Selecting appropriate environmental communities enriched in the activities we wish to mine for, and being creative in designing screens to identify enzymes with desired characteristics are key in the field.

Acknowledgments This work was supported in part by the Government of Canada through Genome Canada and the Ontario Genomics Institute (2009-OGI-ABC-1405 and the research agreement OGI-055), Ontario Research Fund (ORF-GL2-01-004), the EU FP7 project MAMBA (FP7-KBBE-2008-226977) and MicroB3 (OCEAN-2011-287589). We thank all members of the Structural Proteomics in Toronto (SPiT) Centre for help in conducting the experiments and Dr. Manuel Ferrer (Institute of Catalysis, Madrid) for helpful discussions.

References

Aly AS, Moustafa AB, Hebeish A (2004) Bio-technological treatment of cellulosic textiles. J Clean Prod 12(7):697–705

Baraniecki CA, Aislabie J, Foght JM (2002) Characterization of Sphingomonas sp. Ant 17, an aromatic hydrocarbon-degrading bacterium isolated from Antarctic soil. Microb Ecol 43(1):44–54

Bedingfield J (ed) (2013) The Novozymes Report. Novozymes. http://report2013.novozymes.com Accessed 24 Feb 2014

Beja O et al (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289(5486):1902–1906

Beloqui A et al (2006) Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. J Biol Chem 281(32):22933–22942

Brisson VL et al (2012) Metagenomic analysis of a stable trichloroethene-degrading microbial community. ISME J 6(9):1702–1714
Cavicchioli R et al (2002) Low-temperature extremophiles and their applications. Curr Opin Biotechnol 13(3):253–261
Christian JH, Waltho JA (1961) The sodium and potassium content of non-halophilic bacteria in relation to salt tolerance. J Gen Microbiol 25:97–102
Christian JH, Waltho JA (1962) Solute concentrations within cells of halophilic and non-halophilic bacteria. Biochim Biophys Acta 65:506–508
Dinsdale EA et al (2008) Functional metagenomic profiling of nine biomes. Nature 452(7187):629–632
Feller G, Gerday C (2003) Psychrophilic enzymes: hot topics in cold adaptation. Nat Rev Microbiol 1(3):200–208
Ferrer M, Martínez-Abarca F, Golyshin PN (2005) Mining genomes and "metagenomes" for novel catalysts. Curr Opin Biotechnol 16(6):588–593
Ferrer M et al (2007) Mining enzymes from extreme environments. Curr Opin Microbiol 10(3):207–214
Ferrer M et al (2009) Metagenomics for mining new genetic resources of microbial communities. J Mol Microbiol Biotechnol 16(1–2):109–123
Fredrickson JK et al (1991) Isolation and characterization of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds. Appl Environ Microbiol 57(3):796–803
Fredrickson JK et al (1995) Aromatic-degrading Sphingomonas isolates from the deep subsurface. Appl Environ Microbiol 61(5):1917–1922
Gabor EM, Alkema WBL, Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ Microbiol 6(9):879–886
Gado HM et al (2009) Influence of exogenous enzymes on nutrient digestibility, extent of ruminal fermentation as well as milk production and composition in dairy cows. Anim Feed Sci Technol 154(1–2):36–46
Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63
Hausmann S, Jaeger EK (2010) Lipolytic enzymes from bacteria. In: Timmis KN (ed) Handbook of hydrocarbon and lipid microbiology. Springer, Berlin/Heidelberg, pp 1100–1126
Hemachander C, Puvanakrishnan R (2000) Lipase from Ralstonia pickettii as an additive in laundry detergent formulations. Process Biochem 35(8):809–814
Hernández-Martín E, Otero C (2008) Different enzyme requirements for the synthesis of biodiesel: Novozym 435 and Lipozyme TL IM. Bioresour Technol 99(2):277–286
Holloway P, Trevors JT, Lee H (1998) A colorimetric assay for detecting haloalkane dehalogenase activity. J Microbiol Methods 32(1):31–36
Jegannathan KR, Nielsen PH (2013) Environmental assessment of enzyme use in industrial production – a literature review. J Clean Prod 42:228–240
Jimenez L et al (1999) Biobleaching of cellulose pulp from wheat straw with enzymes and hydrogen peroxide. Process Biochem 35:149–157
Jones BV, Sun F, Marchesi JR (2007) Using skimmed milk agar to functionally screen a gut metagenomic library for proteases may lead to false positives. Lett Appl Microbiol 45(4):418–420
Kemp PF, Aller JY (2004) Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. FEMS Microbiol Ecol 47(2):161–177
Kok RG et al (1993) Growth-phase-dependent expression of the lipolytic system of Acinetobacter calcoaceticus BD413: cloning of a gene encoding one of the esterases. J Gen Microbiol 139:2329–2342
Koonin EV, Galperin MY (2003) Sequence - evolution - function: computational approaches in comparative genomics. Kluwer Academic, Boston
Kouker G, Jaeger KE (1987) Specific and sensitive plate assay for bacterial lipases. Appl Environ Microbiol 53(1):211–213

Kwon KK et al (2005) Kordiimonas gwangyangensis gen. nov., sp. nov., a marine bacterium isolated from marine sediments that forms a distinct phyletic lineage (Kordiimonadales ord. nov.) in the "Alphaproteobacteria". Int J Syst Evol Microbiol 55(Pt 5):2033–2037

Li Y et al (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of Rhizobium leguminosarum and of Escherichia coli reveals different classes of cloned trp genes. Environ Microbiol 7(12):1927–1936

Lonhienne T, Gerday C, Feller G (2000) Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. Biochim Biophys Acta 1543(1):1–10

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

Martinez A et al (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl Environ Microbiol 70(4):2452–2463

Maymo-Gatell X et al (1997) Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. Science 276(5318):1568–1571

Monsan P, Donohue MJO (2010) Industrial biotechnology in the food and feed sector. In: Soetaert W, Vandamme EJ (eds) Industrial biotechnology: sustainable growth and economic success. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, p 350

Nguyen D et al (2008) Bleaching of kraft pulp by a commercial lipase: accessory enzymes degrade hexenuronic acids. Enzyme Microb Technol 43(2):130–136

Okamura-Matsui T et al (2003) Discovery of alcohol dehydrogenase from mushrooms and application to alcoholic beverages. J Mol Catal B: Enzym 23(2–6):133–144

Olufsen M et al (2005) Increased flexibility as a strategy for cold adaptation: a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase. J Biol Chem 280(18):18042–18048

Oren A (2002) Halophilic microorganisms and their environments. Kluwer Academic, Dordrecht

Osma JF, Toca-Herrera JL, Rodríguez-Couto S (2010) Uses of Laccases in the Food Industry. Enzyme Research 2010:918761. doi: 10.4061/2010/918761

Pelletier E et al (2008) "Candidatus Cloacamonas Acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. J Bacteriol 190(7):2572–2579

Prasad S, Khadatare PB, Roy I (2011) Effect of chemical chaperones in improving the solubility of recombinant proteins in Escherichia coli. Appl Environ Microbiol 77(13):4603–4609

Ravenschlag K et al (1999) High bacterial diversity in permanently cold marine sediments. Appl Environ Microbiol 65(9):3982–3989

Robertson DE et al (2004) Exploring nitrilase sequence space for enantioselective catalysis. Appl Environ Microbiol 70(4):2429–2436

Roesch LFW et al (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1(4):283–290

Rondon MR et al (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66(6):2541–2547

Saeki K et al (2007) Detergent alkaline proteases: enzymatic properties, genes, and crystal structures. J Biosci Bioeng 103(6):501–508

Schleheck D et al (2000) An α-proteobacterium converts linear alkylbenzenesulfonate surfactants into sulfophenylcarboxylates and linear alkyldiphenyletherdisulfonate surfactants into sulfodiphenylethercarboxylates. Appl Environ Microbiol 66(5):1911–1916

Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71(3):1501–1506

Sellek GA, Chaudhuri JB (1999) Biocatalysis in organic media using enzymes from extremophiles. Enzyme Microb Technol 25(6):471–482

Siddiqui KS, Cavicchioli R (2006) Cold-adapted enzymes. Annu Rev Biochem 75:403–433

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

Staskawicz B et al (1987) Molecular characterization of cloned avirulence genes from race 0 and race 1 of Pseudomonas syringae pv. glycinea. J Bacteriol 169(12):5789–5794

Steele HL et al (2009) Advances in recovery of novel biocatalysts from metagenomes. J Mol Microbiol Biotechnol 16(1–2):25–37

Taupp M, Mewis K, Hallam SJ (2011) The art and design of functional metagenomic screens. Curr Opin Biotechnol 22(3):465–472

Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43(4):777–780

Troeschel SC et al (2012) Novel broad host range shuttle vectors for expression in Escherichia coli, Bacillus subtilis and Pseudomonas putida. J Biotechnol 161(2):71–79

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

Van den Burg B (2003) Extremophiles as a source for novel enzymes. Curr Opin Microbiol 6(3):213–218

Venter JC et al (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304(5667):66–74

Wang Q et al (2006) Discrimination of esterase and peptidase activities of acylaminoacyl peptidase from hyperthermophilic Aeropyrum pernix K1 by a single mutation. J Biol Chem 281(27):18618–18625

Wang S et al (2013) Isolation and characterization of a novel organic solvent-tolerant and halotolerant esterase from a soil metagenomic library. J Mol Catal B: Enzym 95:1–8

Wexler M et al (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. Environ Microbiol 7(12):1917–1926

Williams AB, Rona PA (1986) Two new caridean shrimps (Bresiliidae) from a hydrothermal field on the Mid-Atlantic Ridge. J Crustac Biol 6(3):446–462

Yooseph S et al (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5(3):e16