New opportunities revealed by biotechnological explorations of extremophiles
Mircea Podar¹ and Anna-Louise Reysenbach²

Over the past few decades the extremes at which life thrives has continued to challenge our understanding of biochemistry, biology and evolution. As more new extremophiles are brought into laboratory culture, they have provided a multitude of potential applications for biotechnology. More recently, innovative culturing approaches, environmental genome sequencing and whole genome sequencing have provided new opportunities for the biotechnological exploration of extremophiles.

Addresses
¹ Metagenomica, San Diego, CA 92128, USA
² Department of Biology, Portland State University, Portland, OR 97201, USA

Corresponding author: Reysenbach, Anna-Louise (reysenbach@pdx.edu)

Introduction
Organisms that live at the extremes of pH (>pH 8.5, <pH 5.0), temperature (>45 °C, <15 °C), pressure (>500 atmospheres), salinity (>1.0 M NaCl) and in high concentrations of recalcitrant substances or heavy metals (extremophiles) represent one of the last frontiers for biotechnological and industrial discovery. As we learn more about the extremes at which life can survive and thrive, more of these extremophiles are brought into culture and their genomes sequenced. In many cases, biotechnological applications of extremophiles and their biomolecules (e.g. enzymes) have been the driving force in both academic and industrial research of these organisms. Extremophiles and extremozymes occupy an important place in the multibillion dollar environmental biotechnology industry, with applications spanning agricultural, biomedical and industrial sectors. Owing to the highly competitive nature of industrial R&D, in most cases the path from an extremophile to a successful commercial application is not documented in peer reviewed scientific publications and can be partially followed through patents, biotechnology meetings and company websites. In this review, we will consider a few new developments as several recent reviews have focused on the biotechnological applications of extremophiles [1–4,5***].

Extremophiles and biomolecules
The most direct application of extremophiles in biotechnological processes involves the organisms themselves. Among the most established is biomining (bioleaching), in which microbial consortia are used to extract metals such as copper, cobalt, gold and uranium from ores (reviewed in [6,7]). This area has received considerable interest lately, propelled by developments in microbial isolation and the application of genomic approaches for studying the individual organisms and their community [8]. These processes involve iron- or sulfur-oxidizing acidophilic microorganisms adapted to different temperature ranges, from mesophiles (bacteria such as Acidithiobacillus, Leptospirillum and the archaea Ferroplasma) to thermophiles (archaea from the genera Sulfolobus, Metallosphaera and Acidianus). When the activity of such extreme acidophiles is not controlled in biomining operations, this can lead to acid mine drainage (AMD), which causes considerable environmental damage. Although the microbes responsible for bioleaching and AMD are inherently adapted to extremely low pH and high concentrations of metals, they are part of an open and dynamic consortia and continuously adapt to the particular conditions they are facing. In commercial bioleaching operations this can result in the selection of more robust and efficient strains, with reduced sensitivity to metal toxicity. For example, arsenic resistance genes have been horizontally transferred via a transposon from an unidentified bacterium to Acidithiobacillus caldus and Leptospirillum ferriphilum, resulting in substantially increased resistance to arsenic in gold-bearing arsenopyrite biooxidators [9*]. This demonstrates that phylogeny alone is not necessarily predictive of the physiological fitness of an individual organism and that the community gene pool can impact the adaptability of the constituent members across phylogenetic barriers. Such strain engineering and selection for improved biotechnological characteristics can be applied to a wide range of bioremediation projects, as has been done for contaminating petroleum hydrocarbons accidently released in arctic environments [10,11].

Most applications involving extremophiles are based on their biomolecules, primarily enzymes but also other proteins (e.g. cryoprotectant antifreeze proteins), lipids, and various small molecules. The most well-known example of a successful application of an extremophile-derived product is Taq DNA polymerase, which was isolated from Thermus aquaticus — first isolated from a
Extremophiles in biotechnology

Podar and Reysenbach

251

geothermal spring in Yellowstone National Park. This enzyme approaches sales of about half a billion dollars per year. Molecular biology research tool enzymes not only include a wide range of other thermostable polymerases and ligases, but also include enzymes isolated from psychrophilic (cold-loving) organisms, such as the ‘Antarctic phosphatase’ from New England Biolabs. Genencor commercialized one of the first industrial extremozymes for use in textile detergents, a cellulase isolated from an alkalophilic bacterium from an east African soda lake. Numerous other examples are reviewed in the article by Antranikian [5**].

Enzymes that have optimal activity at extreme temperatures and pH are widely used in household detergents and in the food, textile, pulp and paper, leather processing, and chemical industries. For each application, the enzymes have to fulfill numerous requirements related to features such as activity and stability, substrate specificity and enantioselectivity. As a result, natural enzymes are often not optimal for the desired biotechnological application. Consequently, a variety of approaches have been used to modify enzyme properties. These include error-prone PCR, saturation mutagenesis, structure-based protein engineering and in vitro evolution approaches [12,13]. Such approaches are best combined with genetic selection or high-throughput screening, to identify the rare mutants that approach the target characteristics, followed by an iterative process of building fitness into the resulting variants. For example, thermostability was built into a suite of engineered pectinases used in cotton fabric processing [14] and increased alkaline stability was obtained for a previously engineered xylanase from the fungus Trichoderma reesei [15]. Furthermore, based on genomics, structural data and computational modeling, certain protein architectures and trends in amino acid usage provide clues as to mechanisms of protein thermostability [16] and promise to lead to predictive protein thermostabilization [17,18**].

Extremophile genomics expose biotechnological potential

Extremophiles, including eukarya (e.g. Alcinella pompejana, Tetrahymena thermophila and Dunaliella salina), have been prime subjects for genomic sequencing projects in an effort to understand the fundamental mechanisms of adaptation to specific environments and to develop practical applications (e.g. reviewed in [19–21]). Recently, the complete genome sequences of four psychrophilic bacteria have been published: two from arctic sediments, Desulfotalea psychrophila [22] and Colwellia psychrerythraea [23], one from coastal Antarctic waters, Psychrothermus haloplankis [24], and one from Photobacterium profundum, a deep sea bacterium that is adapted to both low temperatures and high pressure [25*]. Although there does not appear to be a distinct genomic trait unifying these cold-adapted organisms, they share several characteristics, most notably a membrane with an increased proportion of polyunsaturated and branched fatty acids to increase fluidity at low temperatures (and high pressure) and the presence of cold shock proteins, which are believed to increase translation efficiency by destabilizing secondary structures in mRNA. Additionally, cryoprotectants increase the capacity for nutrient uptake. Owing to the higher solubility of oxygen at low temperatures, and hence potential for oxidative damage within the cell, metabolic reactions that generate reactive oxygen are reduced and molybdopterin-dependent metabolism is eliminated (in P. haplokantis) or the number of catalase and superoxide dismutase genes increased (in C. psychrerythraea and D. psychrophila). Although there are specific trends in amino acid usage that appear to correlate with the psychrophilic proteomes, the most important adaptation for cold-adapted enzymes appears to be a high specific activity at low temperatures. This is achieved through a highly flexible catalytic center at the expense of overall reduced protein stability and susceptibility for thermal denaturation [26]. Such enzymatic characteristics are currently exploited in food biotechnological applications that require low temperatures (e.g. milk and fruit juice processing). Beside cold adaptation, P. profundum is also able to withstand high pressure and has become a model organism for understanding piezophily. For example, transport processes and energy generation by cytochrome respiration are impacted by high pressure [27].

Similar to the situation in cold adaptation, the intrinsic properties of nucleic acids, lipids and enzymes/proteins allow thermophiles to flourish in high temperature environments, and specific composition biases and structural adaptations have been identified [16,28–31]. Recent publications describing the genomes of two hyperthermophilic archaea (Nanoarchaeum equitans and Thermococcus kodakarenensis [32*,33]), one archaeal thermoacidophile (Sulfolobus acidocaldarius) [34] and one thermophilic bacterium (Carboxydothermus hydrogenoformans) [35], although not significantly changing the perception of high temperature adaptation, bring together several exciting novelties on various aspects of the biology, genome evolution and metabolic versatility in specific thermophilic environments. The genome of N. equitans became available shortly after the discovery of this unusual organism, which represents the first case of an archaeon that is an obligate symbiont on another archaeon, a species of the marine crenarchaeote Ignicoccus [36**]. The genome abounds in oddities, from being at the smallest end of the spectrum of cellular genome sizes and encoding virtually no metabolic pathways, to containing a large number of split genes including uniquely split tRNA genes [37**]. One of the split genes is represented by two separate open reading frames encoding DNA polymerase B (also fused with a split intein), which trans splice at high temperature to restore a functional polymerase [38].
The \textit{S. acidocaldarius} genome sequence represents an important landmark in archaeal genomics, because this organism also represents an Archaeal laboratory genetic system [34]. The \textit{S. acidocaldarius} genome has several similarities and differences with two other \textit{Sulfolobus} species, and will facilitate future experimental studies and overall biotechnological applications of these \textit{Crenarchaeota}. The bacterial thermophile, \textit{C. hydrogenoformans} [35], has a remarkable efficiency to carry out carbon monoxide oxidation owing to the presence of five anaerobic carbon monoxide dehydrogenase complexes. The genomic blueprint of this microorganism should allow detailed studies of hydrogenogenesis, which could become an important industrial process for generating hydrogen. The most thermoacidophilic organism known, \textit{Picrophilus torridus}, inhabits solfataric environments with a pH below 1 and a temperature of about 60 °C [39]. The low pH-adapted enzymes of this organism will most likely be found useful for biotechnological applications requiring acidic conditions. To cope with such conditions, \textit{P. torridus} has evolved a membrane with low proton permeability and special lipid composition together with efficient transport mechanisms to maintain the internal pH at values compatible with biochemical functions. Unlike other organisms, which use sodium ion and ATP-driven primary transporters, \textit{P. torridus} predominantly uses the internal high proton concentration (pH 4) to power a large number of solute secondary transporters.

At the other extreme, the genome of a haloalkaliphilic archaean from highly alkaline soda lakes, \textit{Natronomonas pharaonis}, has revealed several adaptations to this environment [40]. These include an overall modification of the proteome to increase the fraction of acidic amino acids and reduce protein hydrophobicity, a coating of the cell membrane with glycoproteins and secreted enzymes attached by lipid anchors, and an efficient transport system for heavy metals and nitrogen compounds which are scarce in hypersaline environments. The halophilic bacterium \textit{Salinibacter ruber} [41] displays similar adaptation mechanisms to hypersaline environments and some of these could have been acquired via lateral gene transfer. No doubt many more surprises are hidden in these genomes that can be exploited for biotechnological purposes.

**Tapping into the hidden biotechnological potential through metagenomics**

Most microbial communities are highly complex and consist of hundreds of species, few of which have been cultured. Even if their cultivation were feasible, the sheer numbers makes individual studies of every species impractical for the foreseeable future. Nevertheless, to unlock the vast amount of genetic information from the uncultured microbial majority for fundamental scientific and potential practical application, a wide range of approaches have been used over the past two decades or so. These approaches have been collectively described as environmental, community genomics or metagenomics. Few areas of biology have witnessed such a surge in interest as metagenomics (e.g. [42**,43]) and this field has already had a significant impact on biotechnology [44**,45].

The use of environmental genomic expression libraries to identify specific enzymatic activities using specific selection schemes or high-throughput screening has been the method of choice. Strategies can often be tailored to directly identify enzymes that will have the desired specificity and physico-chemical optima. Such approaches also enable the rapid screening of a large number of libraries that can represent highly complex communities. For example, over a hundred novel nitrate hydrolases, encompassing several previously unknown subfamilies, have been selected from numerous environments around the world, including the deep sea, arctic regions and Antarctica [46]. Combining enzymatic specificity analysis with biogeographical and genomic data not only made possible phylogenetic-based predictions of enzymatic specificities that are important for biotechnological applications, but also resulted in new evolutionary hypotheses regarding this gene family [47]. In a different study, several esterases with high salt tolerance were identified using screening of a metagenomic library from a hypersaline environment [48]. These types of approaches are, however, limited to enzymatic activities for which very focused screening/selection schemes can be developed, which generate little information about the biology of the community.

Sequence-based approaches to study the metabolism of microbial communities have also been in use for over a decade and involve hybridization as well as random or targeted sequencing of cloned genomic inserts. These approaches can yield a vast amount of information about the genomic potential of the community, which can translate into potential biotechnological applications. For example, an acid mine drainage low-diversity biofilm from the Iron Mountain [49**] was used to generate over 76 Mbp of sequence data. This community was dominated by \textit{Leptospirillum} species and also contained the archaean \textit{Ferroplasma acidarmanus}. The study represented a landmark in genomic research, as a genome for an uncultivated microbe was almost completely assembled and reconstructed. Important insights were gained into the community metabolism and population genetics and this set the stage for subsequent genomics-driven studies of this model acidophilic community. One such follow-up study was the successful cultivation of a novel bacterium, \textit{Leptospirillum ferroidiazotrophum}, a minor representative of the community responsible for nitrogen fixation in the biofilms and predicted to have that function solely on the basis of sequence analysis [50**]. This example of ‘reverse metagenomics’ demonstrates how microbial cultivation and physiology can directly benefit from community shotgun sequencing. The sequence data...
also allowed the first metaproteomic analysis of a microbial community, going beyond the genomic blueprint of the community [51].

Several other large-scale metagenomic sequencing efforts have been directed at communities from extreme environments (e.g. [52,53]). Tringe et al. [53] targeted three deep sea communities associated with decomposing whale carcasses. These remains, owing to the amount of organic carbon and lipids they contain, represent important ‘oases’ for life on the seafloor. The authors propose a gene-centric approach to analyze microbial communities, focusing not on species, genomes or even full-length genes, but instead on gene signatures (‘environmental genomic tags’ or EGTs). Analyzed in a comparative metabolic framework, EGTs can signal specific traits of a community defining adaptations of microbes to specific conditions. For example, the whale carcass communities had an increased frequency of genes associated with lipid degradation as compared with other communities (e.g. soil), indicating the presence of microbes specialized in using the high lipid content in whale bones. Such genomic markers (EGTs), used in the design of microarrays for example, could provide quick screening of samples that might have certain metabolic properties of biotechnological interest. Furthermore, a combination of environmental genomic sequencing with direct phenotypic detection (transcriptomics, proteomics and metabolomics) could soon become standard approaches to analyze and monitor communities, and hence provide further insights into potential novel applications.

Unexplored frontiers and future prospects
Microbial communities continue to be discovered in environments once thought to be too hostile to support any form of life (e.g. [54,55]). Often overlooked are the viruses of some of these extremophiles, which also have significant interest for controlling microbial community structure or as sources of extremozymes [56]. With the development of innovative culturing technologies and new tools for exploring remote areas such as the deep-sea and ice-covered oceans, tapping into the biodiversity of extreme environments will continue to provide a rich resource for biotechnology.

Acknowledgements
This work was partially supported by an US-NSF grant to ALR.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


A predictive protein thermostabilization study combined with structural and metabolic analyses.


37. Randau L, Munch R, Hohn MJ, Jahn D, Soll D: Nanoarchaeum equitans creates functional tRNA sequences from separate genes for their 5'- and 3'-halves. Nature 2005, 433:537-541. Reports the computational prediction and experimental confirmation of the presence of split tRNA genes in N. equitans, the only organism known so far in which such tRNA trans-splicing occurs.


The authors used selective enrichment techniques and metagenomic sequencing and analysis to identify genes and infer metabolic characteristics of deep-sea archael methanotrophs.


A metagenomic study of three whale fall microbial communities and a soil sample. Comparative analysis of the data with other public metagenomic sequences revealed specific gene signatures and metabolic features that can be used to identify and characterize different consortia.

