Fall 2011

Identification and characterization of a novel thermohalophilic esterase from the Red Sea: Atlantis II brine pool

Yasmine Mustafa Mohamed
The American University in Cairo AUC

The American University in Cairo
School of Sciences and Engineering

Identification and Characterization of a Novel Thermohalophilic Esterase from the Red Sea; Atlantis II Brine Pool

A Thesis Submitted to
The Biology Department

in partial fulfillment of the requirements for
the degree of Master of Science

by Yasmine Mustafa Mohamed

under the supervision of Dr. Rania Siam
January 2012

The American University in Cairo

Identification and Characterization of a Novel Thermohalophilic Esterase from the Red Sea; Atlantis II Brine Pool

A Thesis Submitted by

Yasmine Mustafa Mohamed

To the Biotechnology Graduate Program

Month/ Year

In partial fulfillment of the requirements for
The degree of Master of Science

Has been approved by

Thesis Committee Supervisor/Chair _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/External Examiner _____

Affiliation _____

_____	_____	_____	_____
Dept. Chair/Director	Date	Dean	Date

DEDICATION

To My Beloved Parents For Their Constant Support And For Being My Biggest And Truest Fans.

To My Brother, Who Manages To Put A Smile On My Face No Matter What.

Last, But Definitely Not Least, To My Partner, Mohamed Abouelsoud For Being The Amazing Person He Is.

Acknowledgements

I am particularly grateful to my advisor Dr. Rania Siam, for her continuous support, guidance and advice. Without her constant encouragement and enthusiasm, this work would have not been possible.

Much gratitude to Dr. Hamza El Dorry, the principal investigator of the Red Sea Metagenomics Project.

I am grateful to KAUST (King Abdullah University for Science and Technology) for providing financial support for the project.

I am heartily thankful to Al-Alfi foundation for funding my studies at AUC.

Many thanks to Dr. Mohamed Ghazy, Dr. Ahmed Sayed and mr. Amged Ouf for their help and support in the lab.

I would also like to express my gratefulness to my family and to my friends in the biotechnology program for making this journey enjoyable.

Abstract

The American University in Cairo

Identification and Characterization of a Novel Thermohalophilic Esterase from the Red Sea; the Atlantis II Brine Pool

by Yasmine Mustafa Mohamed
under the supervision of Dr. Rania Siam

Industrial (white) biotechnology poses an increasing demand for novel biocatalysts that are robust under a wide range of conditions. Traditionally, biocatalysts were isolated from cultured isolates, however, less than 1% of microorganisms are culturable. Therefore, it became evident that the unculturable majority holds a great potential for the discovery of novel biocatalysts. Metagenomics is an invaluable tool for accessing the genomes of the uncultured majority and has led to the isolation of a large number of biocatalysts from various environments.

Extreme environments such as hydrothermal vents, brine pools and glaciers, are an attractive source for biocatalysts. Biocatalysts from these environments almost always reflect in their characteristics the environment from which they originated, and therefore may exhibit high stability and activity in the aggressive conditions required by some processes. The Atlantis II deep is a brine pool in the Red Sea and is characterized by high temperature (almost 70^{o}C), high salinity (7.5 times that of normal sea-water), high metal concentration and anoxia. Such extreme conditions make the Atlantis II deep an attractive site for mining for biocatalysts.

Using lipolytic enzymes as biocatalysts in industrial and biotechnological processes is estimated to be a billion dollar business. Their applications in industry include, and are not limited to, biodiesel formation, pulp and paper industry, detergent industry and flavor development and therefore, the demand for novel lipolytic enzymes is increasing continuously.

Several studies attempted to, and successfully isolated novel lipolytic enzymes with unique characteristics using metagenomic approaches. However, the Atlantis II deep was not previously mined for lipolytic enzymes. In this study, samples were collected from the Atlantis II deep and were used for the construction of a large-insert fosmid clone library. The library was screened for lipolytic activity using a function-based approach. Sequencing of positive clones identified a novel lipolytic enzyme (EstATII), which was then subcloned from the original fosmid into a high copy number plasmid to allow simple overexpression and characterization of the enzyme. Characterization of EstATII revealed that it's a novel thermophilic (optimum temperature = 60^{o}C) and halophilic esterase, with potential applications in processes requiring extreme conditions such as biodiesel production and resolving of racemates.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

LCL	Lower Convective Layer
Psu	Practical Salinity Unit
PCR	Polymerase Chain Reaction
rRNA	Ribosomal Ribonucleic Acid
SSU	Small Subunit
SDS	Sodium Dodecyl Sulfate
BAC	Bacterial Artificial Chromosome
LB	Luria-Bertani
iTOL	Interactive Tree Of Life
ORF	Open Reading Frame
BLAST	Basic Local Alignment Search Tool
IPTG	Isopropy-beta-D-thiogalactoside
X-Gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside
pNP	*p*-nitrophenol
pNPB	*p*-nitrophenyl Butyrate
pNPP	*p*-nitrophenyl Palmitate
RFLP	Restriction Fragment Length Polymorphism
COG	Clusters of Orthologous Group
HSL	Hormone-Sensitive Lipase

Chapter 1: Literature Review

1. Atlantis II Deep: The Largest Brine Pool in the Red Sea

Until lately, and despite its uniqueness, the Red Sea has received little attention among marine environments. The Red Sea formed 3-5 million years ago when the Arabian and African plates started to split [1]. It is characterized by high temperature and salinity owing to the high rate of evaporation, lack of major river inflows and a low rate of rainfall [1].

The Red Sea is characterized by the presence of deep-sea hypersaline anoxic basins; called brine pools. To date, twenty five brine pools have been found in the Red Sea [1, 2]. Deep-sea brine pools are large bodies of water at the bottom of the ocean and are characterized by high temperature and salinity. Owing to its high salinity, the brine pool is denser than the surrounding waters and blends poorly with it. This creates a distinct surface for the pool and hence, brine pools are sometimes referred to as "submarine lakes" [3]. The high temperature and salinity of the Red Sea brine pools is believed to be a function of tectonic activity. Localized tectonic activity creates fissures in the seabed sediment which is penetrated by deep-sea water. This water is heated by submarine magma and absorbs minerals from the sediment. The heated water is driven back to the seafloor surface by convective currents, and due to its high density forms a distinctive layer of brine [3] (Figure 1).

Atlantis II Deep (Figure 2) is the largest brine pool in the Red Sea, has the highest temperature and is the most dynamic [2, 4]. It has a maximum depth of 2,194 m and is stratified into several layers that increase in temperature and salinity with increasing depth; the brine-seawater interface, upper convective, middle convective and lower convective layers (LCL) [2, 5]. The lowest layer; LCL is characterized by a temperature of 68.2°C, pH value of 5.3 and salinity of 270 psu, which is 7.5 times that of normal seawater [2, 5]. Atlantis II Deep is nearly anoxic and has high concentrations of iron, zinc,

copper and other heavy metals [2, 5]. Together, these extreme conditions make the Atlantis II brine pool an attractive site for mining for biocatalysts, such as lipolytic enzymes, which are predicted to possess desirable traits, including and not limited to, thermo-tolerance, halo-tolerance, pH plasticity and resistance to inhibition by heavy metals.

2. Metagenomics: Exploring the, So-Far, Unculturable

Microorganisms are continuously gaining a growing attention and appreciation. They are ubiquitous major components of the earth's biota; inhabiting various environments, performing indispensible functions and contributing largely to the genetic diversity of life[6]. They also have a huge impact on our health and well-being. It has been estimated that there are 100 trillion microbial cells in a human body (10 times the number of human cells) encoding for 100 times more genes than our genome [7]. Classically, the main approach by which microorganisms were studied was culturing in the lab. The realization that as much as 99% of the microorganisms inhabiting almost every environment cannot be cultured under standard laboratory conditions revolutionized the concepts and practices in the field of microbiology [8, 9]. In 1990, a phylogenetic study on the Sargasso Sea employed a PCR-based approach to amplify almost entire 16 rRNA genes directly from the environmental sample without the need for culturing. A new bacterial group; the SAR11 cluster, was described in this study as a seemingly significant component of the habitat [10]. This technical breakthrough was then employed in several other habitats in an accelerated manner, leading to the discovery of new diverse taxa [11, 12]. Excluding the culturing bias by this technique, demonstrated the high microbial diversity of the uncultured majority. Cloning DNA directly (without PCR amplification) from environmental samples for the purpose of preservation and study, was first proposed as an idea by Pace in 1985. However, it was applied for the first time in 1991, where environmental DNA was extracted, fragmented, subjected to size fractionation and cloned into bacteriophage lambda [12]. In 1998, a term was coined by Handelsman and colleagues to describe this technique: Metagenomics. The term refers to analyzing a collection of genomes derived from a sample [13].

An advancement to studying the microbial diversity of a community by amplifying small subunit (SSU) rRNA hypervariable region(s) is using the pyrosequencing technology. Instead of cloning the amplified hypervariable region(s) of SSU rRNA gene and sequencing by Sanger chain termination approach, PCR products are subjected to massively parallel pyrosequencing producing what is called pyrotags [14]. This method provides higher in-depth coverage and a lower cost than conventional Sanger sequencing. It also avoids skewing of the results towards the more dominant species in the community, caused by PCR [14]. Another application to pyrosequencing is shotgun sequencing, where environmental DNA is isolated and randomly sequenced. A vast amount of sequences (almost a million per run) with an average read-length of about 400 bp is achieved. This huge amount of data is subjected to extensive computational analysis for numerous purposes such as conducting diversity analysis through protein-based phylogeny, mining for genes of interest and gaining insight about the genetic make-up of the microbial community. In cases where the environment is inhabited by a small number of microbial species, it could be possible to obtain enough coverage for assembling whole genomes. In a study, DNA from acidophilic biofilm was isolated and subjected to shotgun sequencing. Near-complete genomes of *Leptospirillum* group II and *Ferroplasma* type II were reconstructed [15].

In addition to exploring the diversity of microbial communities inhabiting various environments, metagenomics is an important tool for the discovery of novel genes, bioactive compounds and biocatalysts. Extreme environments such as hydrothermal vents, brine pools, deep-sea sediments, glaciers and many others are being explored in an accelerated manner by metagenomic approaches, due to their potential as sources of novel genes and biocatalysts that can withstand extreme conditions [16-18]. DNA from such environments is isolated and used for the construction of metagenomic clone libraries which could be screened by different approaches. A sequence based approach involves employing primers or probes targeting a gene of interest, while a function based approach involves testing metagenomic library clones on a medium supplemented with a substrate

that only the clone of interest can uptake [8, 19, 20]. Construction of metagenomic libraries and screening approaches will be discussed later on in a detailed manner.

2.1. Construction and Screening of Metagenomic Libraries

In order to facilitate access to the metagenome, metagenomic libraries are constructed to be further screened for biocatalysts or any other gene(s) of interest. To construct and screen a metagenomic library, DNA is isolated from the environment while taking into consideration several points. First, it should be representative of the microbial community inhabiting the environment. In addition, shearing of DNA should be avoided to prevent the formation of chimeric products between smaller DNA fragments. It should also be free from inhibitors such as humic substances which interfere with downstream processes including PCR, restriction digestion and transformation [8, 19].

Various methods are used for DNA isolation. Physical methods for cell lysis include freezing and thawing, ultrasonication and homogenization; however, these methods result in shearing of the DNA into smaller fragment ranging from 5-10 kb which as mentioned above could result in chimeric products. Chemical methods involve the use of SDS for cell lysis, which is gentler than physical methods. It aids in the recovery of a higher yield of DNA than that obtained by physical methods, without excessive shearing. A mixture of physical and chemical methods could be used to tailor the environmental sample from which DNA is isolated [8].

Constructing a clone library depends mainly on the purpose of the study. For example, a small insert library could be suitable for identifying single genes, while large insert libraries are required if the aim of the study is to identify gene clusters or entire operons [19]. Plasmid vectors such as pBluescript SK+, pUC19 and pZero-2 are used to maintain small insert (less than 10 kb) libraries. For large insert libraries, cosmids (25-35 kb), fosmids (25-40 kb) or BACs (100-200 kb) are used [18]. Several hosts can be used in metagenomic libraries such as *Escherichia coli, Pseudomonas putida*, *Streptomyces lividans , Sinorhizobium meliloti* and *Rhizobium leguminosarum*, however, *E. coli* is used

preferably due to its high transformation efficiency [18]. A small insert library has 10 times more clones than a large insert library on average, to cover the same amount of isolated DNA which makes large insert libraries more convenient for screening owing to the lower number of clones that has to be screened. However, genes that exhibit low expression levels in a surrogate host could be missed by screening large insert libraries due to the large size of the insert and the low copy number, in which case, small insert libraries become a better alternative for screening [18]. There are two main approaches by which a metagenomic library can be screened: the function or activity-based approaches which are discussed further below (Figure 3).

2.1.1. Sequence-based approach

This approach depends on designing primers or probes based on conserved domains of known gene or protein sequences. PCR or probe hybridization is then employed to screen the metagenomic library for prospective novel genes followed by primer walking or sequencing the whole clone to obtain the full length of the gene [8]. Accordingly, it is expected that genes identified through this approach will be genetically similar to already characterized genes which is considered a limitation given that the main aim of a metagenomic study is to identify novel genes that differ substantially from what we already know [8, 18]. Nonetheless, sequence-based approaches are independent of expression of the cloned gene in the surrogate host, and therefore, overcome the limitations of heterologous gene expression experienced in function-based approaches [8].

2.1.2. Function-based approach

This approach involves testing metagenomic clones on indicator agar plates, through which the activity can be detected by change of color or appearance of a halo zone around the colony producing the gene sought. The indicator assay used is most often derived from classical microbiology methods for detection of an activity from pure isolates [18]. The advantage of this approach over the sequence-based one is its dependence on the function of the gene regardless of its sequence; therefore it allows for

identification of entirely novel genes that might not share a significant sequence homology with known genes that have the same function [16, 18]. Nonetheless, function-based approaches require the expression of the gene in the chosen host (*E. coli* in most cases) and therefore, are limited by the drawbacks of heterologous gene expression [8, 17, 18]. In a study involving 32 prokaryotic genomes, it was found that only 40% of the genes could be expressed in *E. coli* [18]. It's conceivable that this number could be less in case of metagenomic samples, especially samples that are collected from extreme environments.

The complex microbial community in a metagenome is expected to have novel unexplored characteristics including transcriptional and translational signals that might not by recognized by the host. For example, in *E. coli*, AUG is the most common codon for initiation of translation (91 %), versus GUG and UUG (or other novel initiation codons that we do not know about) which are favored by some organisms [17]. Other limitations that could lead to weak or no expression of a gene are misfolding due to lack of specific chaperones that are required for protein folding or poor secretion of the protein [8]. All these factors account for the low number of positive clones detected per total screened clones**.**

2.2. Industrial Biocatalysts from Metagenomes

Industrialized societies are moving towards white (industrial) biotechnology, which has proven to be environmentally sound and commercially efficient [20]. This poses a continuous demand for novel biocatalysts, preferably biocatalysts that demonstrate high activity over a wide range of conditions such as temperature, salinity, pH and metal concentration. In other words, it poses a demand for the "ideal biocatalyst". Biocatalysts of microbial origin represent the majority of biocatalysts used in industrial and biotechnological processes [17]. This owes to the capability of prokaryotes to populate and adapt to different environments, from hydrothermal vents to Antarctic desert soil, from which we can derive a wide array of biocatalysts that are robust within a flexible range of conditions; making them desirable for industry [19]. Traditionally, biocatalysts were obtained from pure microbial isolates; however, less than 1% of microorganisms can be

accessed through culture-dependant approaches [17]. It became evident that a huge potential for discovering new biocatalysts lies within the majority of the prokaryotic community that is unculturable [18].

Metagenomics serves as a powerful tool to access the genomes of the unculturable majority of prokaryotes, and to investigate their potential as sources of novel biocatalysts. It has led to the identification and characterization of a vast number of biocatalysts that are active under a wide range of conditions which might reflect the environment from which they originate, making them desirable for industrial use. The choice of the explored environment often falls into one of three main categories. The first is an environment that is enriched for the biocatalyst sought; such as searching insect gut for xylanases. The second is selecting environments that demonstrate high microbial diversity, such as soil samples where 1 gram of soil corresponds to a range of 3,000 - 11,000 genomes [18, 19]. The third is obtaining samples from extreme conditions, such as the Atlantis II brine pool in this study. The biocatalysts obtained from these environments are thought to be stable under the extreme conditions experienced in them [18]. Metagenomics is considered a tool by which the "ideal biocatalyst" can be reached. In every industrial application that involves a biocatalyst, enzymatic restrictions (such as optimum pH, temperature, etc) control the process**.** The diversity of biocatalysts derived from different environments offer a platform to either obtain a natural ideal biocatalyst for a specific process or a natural biocatalyst that, by further *in vitro* evolution techniques, can serve as a backbone from which an ideal "designer" biocatalyst can be derived [20]. In addition, this diversity offered by the prokaryotic community inhabiting different environments, could be helpful in avoiding issues regarding intellectual property rights, as novel enzymes can be continuously accessed and characterized through the metagenome [20].

3. Bacterial Lipolytic Enzymes

Lipolytic enzymes, which include carboxylesterases (EC 3.1.1.1) and true lipases (EC 3.1.1.3), belong to the family of hydrolases which catalyze bond cleavage by addition of

water [21]. Carboxylesterases mainly act on small water soluble esters and the reaction kinetics follow a pattern that can be explained by the classical Michaelis-Menten equation. On the other hand, lipases preferentially act on long-chain triacylglycerols, which are mostly insoluble in water. However, the reaction kinetics does not follow the Michaelis-Menten equation, but rather form a sigmoid curve where the activity of the enzyme rapidly increases when the substrate concentration reaches a certain point [21]. This property of lipases is referred to as "interfacial activation": when the substrate reaches its critical micellar concentration and starts to form aggregates, an interface is formed between the hydrophobic core of the micelle and water i.e., an emulsion is formed. At this point, the activity of the lipase increases rapidly. Accordingly, lipases are sometimes referred to as "interfacial enzymes" [21, 22].

This distinctive property of lipases over esterases has been described as early as 1958 by Sarda and Desnuelle [23]. However, a probable underlying structural explanation of the phenomenon was not described until 1990 with the elucidation of three-dimensional structures of lipases from the fungus *R. miehei* and the human pancreatic lipase by Brady *et al.*, and Winkler *et al.*, respectively [24, 25]. These structures showed that a flexible α-helix covered a tunnel leading to the active site. Upon contact with the interface, this α-helix (referred to as the "lid") undergoes a conformational change that exposes the substrate to the active site leading to the rapid increase in enzymatic activity [21, 22] These two properties of lipases; interfacial activation and possession of a lid structure that undergoes a conformational change upon contact with a lipid/water interface, were classically used to define a lipolytic enzyme as a "true" lipase [26]. However, with the identification and characterization of other lipases, exceptions were found. For example, LipA from *Bacillus subtilis*, although a true lipase, does not possess a lid nor does it exhibit interfacial activation. Also, LipA from *Pseudomonas auruginosa* is a true lipase as well that has a lid structure but does not show interfacial activation [21, 26].

Therefore, lipases are defined loosely as carboxylesterases that catalyze the hydrolysis (and synthesis) of long-chain glycerolesters with trioleoylglycerol being the

standard substrate. Even the term “long-chain” does not have a strict definition, however, an acyl chain that has more than 10 carbon atoms is considered to be “long-chain”. On the other hand, hydrolysis of glycerolesters with a chain length that is less than 10 carbon atoms indicates esterase activity with tributyrylglycerol (tributyrin) being the standard substrate. It should be noted that many true lipases possess esterolytic activity and are capable of hydrolyzing short-chain substrates. In addition, several thiol esters, amides, low- and high-molecular weight esters, etc are also accepted as substrates by lipases [26].

3.1. Primary Sequence Diversity and Three-Dimensional Structural Conservation within Lipolytic Enzymes

Although there is a continuous increase in the number of identified lipase and esterase sequences, a specific sequence similarity was not found to be shared by all known lipolytic enzymes, conversely, they appear to have a high degree of variability. It was initially thought that a consensus pentapeptide; GXSXG (Gly-X-Ser-X-Gly, where X is any amino acid) which contains the nucleophilic serine is shared by all lipolytic enzymes [22]. However, in 1999, an amino acid sequence comparison of 53 known lipases and esterases classified them into eight families , some of which displayed other motifs [27]. Since then, the number of enzymes belonging to each family has increased at least six times [21]. In addition, a number of metagenomic studies screening for lipolytic enzymes, report lipases and esterases that cannot be classified under any of the eight families and cluster independently with other homologous sequences, proposing evidence for the presence of other unexplored families [28]. For example a number of lipases and esterases which constitute family II, were found to share a GDSL (Gly-Asp-Ser-Leu) consensus sequence. Other lipases and esterases were found to have a significant sequence homology to the mammalian hormone-sensitive lipase. These enzymes are grouped in family IV or the hormone-sensitive lipase family. They display a consensus pentapeptide GDSAG (Gly-Asp-Ser-Ala-Gly) that contains the nucleophilic serine residue. In addition, another motif; HGGG was found in this family. It contributes to the formation of the

oxyanion hole (discussed later) which participates in the catalytic activity of the enzyme [27, 29].

Despite the high degree of variability observed in the amino acid sequences of lipolytic enzymes, they display a conserved three-dimensional structure, which is conserved among all hydrolases and referred to as the α/β hydrolase fold (Figure 4). The structure of the α/β hydrolase fold is formed of a mostly parallel β-sheet in the centre, comprising eight β strands which are all parallel except for β strand number 2 that is anti-parallel. β strands from 3 to 8 are connected by α helices on both sides of the sheet [22]. In α/β hydrolases, the active site comprises a nucleophilic serine (mostly found in the GXSXG consensus), an acidic residue and a histidine always in this order and together they form a catalytic triad **[22, 26].**

3.2. Folding and Secretion

Lipases and esterases are extracellular enzymes which require secretion through bacterial membranes to reach their site of activity. To date, lipolytic enzymes appear to employ at least two of the known secretion pathways [26, 30]. In addition, proteins such as Lif chaperones and accessory proteins involved in disulfide bond formation could also be needed for proper folding of the enzyme to allow for its subsequent secretion. Secretion depends on whether the enzyme has a signal peptide or not. Lipolytic enzymes lacking a signal peptide are secreted via the type I secretion system which involves an ABC transporter consisting of three proteins [26, 30]. Lipolytic enzymes possessing a signal peptide are secreted differently. In gram-positive bacteria, extracellular enzymes need to be secreted through a single cytoplasmic membrane, however, in gram-negative bacteria; an additional outer membrane is present. Secretion of lipolytic enzymes possessing a signal peptide through single cytoplasmic membrane in gram-positive bacteria (or the inner membrane of gram-negative bacteria) involves the Sec machinery or the Tat pathway. Secretion through the outer membrane in gram-negative bacteria involves a multimeric (12-14 proteins) complex machinery called the secreton which forms type II secretion pathway [26, 30].

3.3. Mechanism of Hydrolysis

Hydrolysis of the acyglycerol substrate starts by a nucleophilic attack from the oxygen of the serine residue on the carbonyl carbon in the ester bond of the substrate, forming a tetrahedral intermediate. The nucleophilicity of the serine residue is increased by donating a proton to the catalytic histidine and the oxyanion hole which involves the formation of two hydrogen bonds between the now negatively charged carbonyl oxygen of the substrate and two main-chain NH groups. The oxyanion hole also stabilizes the formed intermediate. The proton of the histidine is then transferred to the carbonyl oxygen in the ester bond of the substrate which is cleaved releasing the alcohol component of the substrate which comprises the first step of the hydrolysis reaction. Upon the release of the alcoholic component, a covalent intermediate is formed between the acyl component of the substrate and the enzyme, which leads to the next step: deacetlyation of the enzyme. A proton is transferred from water in the reaction medium to the catalytic serine producing a hydroxide ion that attacks the carbonyl carbon atom of the substrate forming again a tetrahedral intermediate. Finally, histidine donates a proton to the oxygen atom of the catalytic serine, releasing the acyl component of the substrate [22, 26, 29].

3.4. Physiological Functions of Lipolytic Enzymes in a Bacterial Cell

The physiological roles of many lipases and esterases are still not fully understood. Intracellular lipases are involved in central metabolic pathways such as lipid degradation, modification and synthesis. Many lipolytic enzymes are secreted into the environment to degrade lipid substrates and supply the bacterial cell with carbon source enabling it to colonize harsh environmental niches such as *P. auruginosa* YS-7 which can colonize oil emulsions with water content as low as 1% [21]. In *Streptomyces* species, esterases provide a carbon source for antibiotic biosynthesis. However, expression of lipolytic enzymes often displays a complex regulation that implies that their main role in the cell is not merely supplying the cell with a carbon source.

Other roles of lipolytic enzymes include acting as virulence factors in several pathogenic bacteria [21, 22]. For example, lipases from *P. auruginosa*, a pathogen which infects the respiratory tract, degrades dipalmitoyl-phosphatidylcholin; the main component of the lung surfactant which functions to reduce the surface tension of alveoli membranes to facilitate breathing [21]. Esterases are also involved in the virulence of pathogenic bacteria. For example, the development of peptic ulcers and stomach cancer is linked to EstV from *Helicobacter pylori* [21]. Esterases are also involved in regulating interspecies communication and interaction, such as esterases from *P.fluorescens* which degrade lactones that act as quorum sensing signals in many bacterial species.

3.5. Attractiveness of Lipolytic Enzymes for Industry

Microbial lipolytic enzymes possess a huge potential as industrial biocatalysts. In general, enzymes from microbial origin are diverse (reflecting the diversity of the microbial community), robust, easy to manipulate by protein engineering and to produce by fermentation [22]. Lipases and esterases introduce another level of diversity in their primary structures as described above, which is reflected in a high degree of biochemical variability. Lipases and esterases from different organisms and environments, and even isoenzymes from the same organism vary greatly in their molecular mass, optimum temperature and pH, substrate specificity as well as other characteristics. This extensive diversity of lipolytic enzymes accounts for their attractiveness for industry as biocatalysts, as they offer a wide range of specificities and optima and therefore can adapt to different process conditions. In addition to their extensive diversity, lipases in particular are characterized by substrate specificity, regio- and enantioselectivity that surpasses that of any other enzyme, making their application potential boundless [31].

3.5.1. Substrate Specificity

Mostly, lipases maintain a conserved three-dimensional structure despite their primary sequence diversity. The active site formed by the catalytic triad is embedded within the protein and substrate access to it involves binding of the substrate to binding site(s) located in pocket(s) around the catalytic site [22]. Despite the conservation in the

three-dimensional structure, these substrate binding pockets differ among lipases in terms of length, shape and degree of hydrophobicity of the residues lining the pockets. All these factors have been related to substrate chain length preference i.e. substrate specificity [22]. In addition, the lid structure was proven to be related to substrate specificity in several studies. For example, in a study by *Carrière et al.,* a comparison between the human pancreatic lipase (which acts only on triglycerides) and the guinea pig pancreatic lipase (which has additional phospholipase and galactolipase activities) showed that the main structural difference between the two enzymes lies in the lid structure which has a significantly reduced size (5 amino acids) in the guinea pig lipase. Exchanging lid structures between both enzymes, confirmed that the lid is a major structural determinant of substrate specificity [32].

3.5.2. Enantio- and Regioselectivity

Enantioselective catalysis refers to a reaction in which one enantiomer of a chiral product is preferentially produced. On the other hand, regioselectivity refers to the preference towards formation or cleavage of a specific bond over the other. The structural basis underlying these desirable properties of lipases is not completely elucidated. A structural study on the lipase from *Burkholderia cepacia* which is widely used for enantiomeric resolution of esters of secondry alcohols was one of the studies conducted for this purpose. It was found that this lipase has four binding sites for its lipid substrate; the oxyanion hole and three binding pockets lined with hydrophobic residues to accommodate the three acyl groups of the substrate (*sn*-1, *sn*-1 and *sn*-3). A large hydrophobic pocket accommodates *sn*-3, a mixed hydrophilic/ hydrophobic pocket accommodates *sn*-2 and *sn*-1 fits into a smaller hydrophobic pocket [22, 26]. The differences in size and hydrophobicity/ hydrophilicity of the pockets is a major determinant of the enzyme's enantio- and regioselectivity and therefore are important targets for protein engineering and directed evolution [22, 26, 31]. An example to this would be increasing the enantioselectivity of *Candida antarctica* B lipase from E=14 to 28 by a single amino acid substitution [22]

3.6. Screening Metagenomic Libraries for Lipolytic Activity

As discussed earlier, metagenomic libraries can be screened both by sequence-based and function-based approaches. Since lipolytic enzymes do not share significant homology with their identified counterparts, the function-based approach is considered a more convenient method for screening [16]. Currently, the most commonly used substrate to screen for lipolytic activity is tributyrin (Glyceryl tributyrate) (Table 1). Being a short-chain glycerol ester (4-carbon side chains), it is suitable for detecting the activity of both lipases and esterases [33].

Tributyrin is emulsified mechanically into the solid growth medium (such as LB agar) and poured into petri dishes. Clones that are capable of hydrolyzing tributyrin, display a clear halo zone indicative of lipolytic activity [34]. Despite the widespread use of tributyrin agar in screening metagenomic libraries for lipolytic activity, it has been reported in the literature that this method of screening can result in false positives. In a metagenomic study comprising 10,000 clones, tributyrin agar was employed for screening for lipolytic activity. 18 clones displayed clear halo zones, however, sequencing revealed that only 3 were esterases, the rest were unrelated to lipolytic enzymes [33]. The authors of this study recommend the use of other short-chain glycerol esters such as tricaproin (C6),tricaprylin (C8) and tricaprin (C10) [33]

Other substrates that are used less frequently include triolein, which is supplemented to the solid medium together with the fluorescent dye rhodamine B. Lipolytic activity results in the formation of an orange fluorescent halo upon UV irradiation [35]. Another substrate is Bacto Lipid, which upon hydrolysis by lipolytic clones, results in a clear halo zone [36]. Although the function-based approach remains the method of choice for screening for lipolytic enzymes, it still has the drawbacks of heterologous gene expression. This is evident in the very small number of positive clones detected per number of clones screened in most studies.

The sequence-based approach is less often employed in the search for lipolytic enzymes. In most cases, enrichment cultures are carried out first, followed by testing the obtained pure isolates on tributyrin agar (or any other medium) to select for lipolytic isolates. 16S rRNA analysis is then carried out to identify the isolate. Once identified, lipases and esterases from closely related organisms are retrieved and aligned to identify putative conserved regions that could be used for primer design [37, 38]. A few attempts to isolate lipolytic enzymes by amplifying directly from the environmental DNA are also reported [39, 40].

3.7. Industrial and Biotechnological Applications

Lipolytic enzymes, as mentioned above, are endowed with substrate specificity, regio- and enantioselectivity which results in minimization of unwanted side reactions and byproducts. Based on total sales volume, lipases are considered the third largest group after proteases and carbohydrates. Using lipases in industrial and biotechnological applications is estimated to be a billion dollar business [41]. The diversity of these enzymes and the environments from which they originate offers a variety of biochemical characteristics and optima which can be suited to a wide array of processes. In addition, a wide range of substrates is accepted by these enzymes. Lipases in particular, are capable of catalyzing the reverse synthesis reaction as efficiently as hydrolysis and in some cases are more suited for synthesis applications than hydrolysis [42]. We can therefore classify reactions catalyzed by lipases into:

a. Hydrolysis
b. Synthesis, which can be sub-classified into: Esterficiation and Transesterification. The latter refers to Interesterification, Alcoholysis and Acidolysis.

Industrial and biotechnological applications that involve lipolytic enzymes can be classified accordingly to synthesis and hydrolysis processes. Below, examples of applications that belong to each category are mentioned.

3.7.1. Hydrolysis

This process involves the hydrolysis of a lipid/ester into its constituent acid and glycerol/alcohol as shown in the reaction: RCOOR' + H_2O → RCOOH + R'OH. In such processes, the ultimate goal could be (1) the breakdown of lipid/ester or (2) obtaining the produced alcohol or acid [42].

Leather Manufacture

This process makes use of the ability of lipases to break down fats. An important step in the processing of skins and hides is the removal of fats and protein debris (or degreasing) that is associated with the hides and hair. Traditionally, chemical methods such as liming were used but did not result in an efficient removal. The use of a mixture of proteases and lipases is now common practice in leather manufacture [41, 42].

Waste Treatment

In aerobic waste treatment processes such as activated sludge, fat layers form at the surface that should be continuously removed in order to maintain aeration of the biomass. The fat layer is skimmed and digested by lipases. In sewage disposal plants in USA, a lipase, Lip-MY from *C. rugosa* is employed for such purpose [41, 42].

Detergents

One of the main industries that employ lipases for their fat-hydrolyzing potential is textile washing and cleaning as well as dishwashing. Lipases, together with other enzymes are important components of detergent mixtures. Lipases from *Candida* and *Chromobacterium viscosum* are among others used for this purpose [41, 42].

Flavor Development in the Dairy Industry

The use of lipolytic enzymes in flavor development is well established in the dairy industry. The aroma and texture of dairy products is a product of the lactose, protein and fat metabolism in milk. Enzymes such as lipases, esterases and proteases are therefore used to accelerate these metabolic processes for the maturation of cheese and formation of certain flavors. In addition, free fatty acids and peptides are formed during this process and contribute to the flavor [42, 43].

Medical Applications

Lipases can be used as digestive aids, and along with other components, can be used for the treatment of gastrointestinal disturbances and dyspepsia. Like other enzymes, lipases could be immobilized on pH or oxygen electrodes to determine blood cholesterol and triglycerides [42]. Lipases can also be used for the enzymatic determination of serum triglycerides. The activity of the enzyme will release glycerol which can then be determined by colorimetric reactions [41].

Pulp and Paper Industry

Historically, the contribution of microbial enzymes for the pulp and paper manufacture was merely limited to areas such as raw starch modifications. However, the enzymatic pitch method which employs lipase has been fully established and implemented in the industry since the 1990s [41]. In this method, lipase is added to hydrolyze triglycerides into glycerol and free fatty acids, preventing the coagulation of resinous matter, which was a major problem in the industry. Resinous matter used triglycerides as cores for coagulation [44]. Another aspect of the industry in which lipases are employed, is wastepaper deinking for the purpose of recycling. Using lipases (such as lipase from *Pseudomonas* species (KWI-56)) in deinking increases whiteness, decreases the use of chemicals, saves energy and time, and prolongs equipment life [41].

Oil Biodegradation

Microbial species producing lipolytic enzymes can be used for ecorestoration by degrading oil spills in coastal environments [41].

3.7.2. Ester Synthesis

As mentioned above, lipases can catalyze the reverse synthesis reaction as efficiently as hydrolysis. Synthesis reactions that employ lipolytic enzymes are superior to their chemical counterparts. In an enzymatically-driven ester synthesis reaction, only the substrates and the enzyme are present in the reaction mixture and water is the only byproduct. Such reaction can proceed without the presence of a solvent and can be driven to complete conversion of the substrate to the product by continuous removal of water.

On the other hand, hazardous solvents and acid biocatalysts are needed for a chemical synthesis reaction [42]. In esterification reactions, only ester and water are formed. While, interesterification, acidolysis or alcoholysis (collectivey referred to as transesterification), an ester, acid or alcohol is produced instead of water which could be the desired products of the reaction. A few examples are discussed below.

Pharmaceuticals and cosmetics

Esters are important compounds in the industry of cosmetics, fragrances and pharmaceuticals. Mono-,di-, and triglyceride esters of octanoic and decanoic acid are used as dyes and perfume bases. Immobilized lipase from *Rhizomucor meihei* was used as a biocatalyst for the production of isopropyl myristate, isopropyl palmitate and 2-ethylhexylpalmitate which are used as emollients in skin care products such as sun-tan creams and skin creams [41].

Biopolymers

An increased attention is directed towards biopolymeric materials such as polysaccharides, polyphenols and polyesters since they are biodegradable and produced from sustainable natural resources. Lipases and esterases are used as catalysts in such polymeric reactions [30]

Biodiesel

The rapidly diminishing resources of fossil fuels, in addition to the environmental hazards that they pose, are among the main reasons for experimenting with vegetable oils as alternative fuels. Biodiesel reduces sulfur oxide production and is produced from renewable natural resources [30]. It has been produced chemically from various vegetable oils. Lipases can covert vegetable oil to methyl or other short chain alcohol esters in a signle transeterification reaction [30, 45]. Tranesterification of soyabean oil with ethanol and methanol was conducted using immobilized lipase from *P. cepacia* [41].

Chapter 2: Materials and Methods

1. Sample collection

Samples were collected during the KAUST Red Sea spring 2010 expedition, from the lower convective layer of the Atlantis II deep (latitudes 21° 13'N and 21° 30'N and longitudes 37° 58'E and 38° 9'E) [5]. Large water volumes (approximately 100 liters) were collected and serially filtered to fractionate the prokaryotic community using Millipore Mixed Cellulose Esters filter (Nitrocellulose/Cellulose acetate) of sizes 3 μm, 0.8 μm and 0.1 μm.

2. DNA isolation and Fosmid Library Construction

DNA extraction was carried out as previously described in Rusch et al. [46] (with minor modifications) from 0.1 μm filters. Metagenomic fosmid library was constructed using the CopyControl™ HTP Fosmid Library Production Kit (Epicentre, Cat.# CCFOS059) according to the manufacturer's instructions. Microbial DNA extracted from the LCL was resolved on 1% low-melting point agarose by gel electrophoresis and fragments with an approximate size of 40 kb were gel purified. Purified DNA was end-repaired to have 5'-phosphorylated blunt ends then ligated to the pCC2FOS vector provided with the kit. The ligates were then packaged into lambda phage using the lambda packaging extract supplied by the kit which was used to transfect EPI300-T1R phage T1-resistant *E.coli* host with the ligates. The transformants were plated on Luria-Bertani (LB) agar plates supplemented with 12.5 μg chlormaphenicol/ml and were manually placed in 96- well plates for ease of handling. Sampling was done by Red Sea Metagenomics team from both AUC and KAUST, while DNA extraction was done by Mr. Amged Ouf in the Genomic laboratory at AUC.

3. Functional Screening for Lipolytic Clones

Functional screening for lipolytic activity was performed by growing the transformants on LB agar plates supplemented with 12.5µg chloramphenicol/ml and 1% Tributyrin (Sigma-Aldrich, Cat.# W222305). Plates were incubated at 37^{o}C for 3 days. Appearance of a clear halo zone around a transformant was indicative of a candidate lipolytic activity. Candidate transformants were selected for fosmid isolation using the Wizard® Plus SV Minipreps DNA Purification System (Promega, Cat.# A1330). Fosmids were digested using *Bam*HI to assess their diversity.

4. Identification of Putative Lipolytic Gene

4.1. Pyrosequencing

Fosmids extracted from the five putative lipolytic clones were subjected to shotgun sequencing. DNA libraries were constructed according to the instructions provided in the GS FLX Titanium library guide. DNA fragment size selection was done using the Double SPRI Method and sequencing was done using the GS FLX Titanium pyrosequencer (454 Life Sciences). Sequencing was done by Dr. Ahmed Sayed, Dr. Mohamed Ghazy and Mr. Amged Ouf in the Genomics Laboratory at AUC.

4.2. Assembly and ORF detection

Assembly was done using GS FLX de novo assembler. Open reading frames (ORFs) were identified using the ORF Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) provided by the National Center for Biotechnology Information (NCBI). The putative function of each ORF was annotated by comparing the amino acid sequences to the non-redundant protein database using the Basic Local Alignment Search Tool (BLASTP) [47].

4.3. Sequence Analysis and Phylogenetic Tree Construction

Domain search was conducted by the Conserved Domain (CD)-search tool provided by NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Prediction of signal peptide sequence was performed using SignalP 3.0 servers [48]. For phylogenetic analysis, sequences of 43 bacterial lipolytic enzymes were retrieved from the GenBank sequence database. Multiple sequence alignment of retrieved sequences and EstATII was performed using ClustalW version 1.83 [49].

Phylogenetic tree was constructed using the neighbor-joining method using the software MEGA version 5.05 [50]. Bootstrapping (10,000 replicates) was used to estimate the confidence of the tree. The tree was visualized using the Interactive Tree Of Life (iTOL) v2 server [51].

5. Identification and Subcloning of Fosmid Harboring EstATII

5.1. PCR amplification of EstATII

To amplify EstATII, the forward primer (EstF) 5'-ATG TCC AGG TAC GTT GAT GAG C-3' and the reverse primer (EstR) 5'-TCA GCT TAC CGA GTC GGT CT-3' were designed. PCR conditions were as follows: initial denaturation at 95^{o}C for 5 minutes, 35 cycles (95^{o}C for 30 seconds, 49^{o}C for 45 seconds and 72^{o}C for 1 minute) and a final extension stage at 72^{o}C for 10 minutes.

5.2. Subcloning of Lipolytic Clone Harboring EstATII

The recombinant fosmid was randomly sheared by sonication (SONIFIER® 150, Branson). Fragments of size 1.5-5 kb were recovered from the gel and end-repaired using T4 DNA polymerase. Addition of 3' dA overhangs was done by incubating end-repaired fragments with dATP and *Taq* polymerase for 1 hour at 72^{o}C. The A-tailed fragments were then ligated to pGEM®-T Easy (Promega, Cat.#A1360) according to manufacturer's instructions. 2 ul of the ligation reaction was used to transform electrocompetent *E. coli* TOP10 cells using MicroPulser™

Electroporation Apparatus (Bio-Rad) according to manufacturer instructions. Transformants were plated on LB/Ampicillin/IPTG/X-Gal agar plates, and 96 white colonies were picked and tested for lipolytic activity on tributyrin agar.

5.3. Sequencing

Chain termination sequencing was carried out in the Applied Biosystems 3730xl DNA Analyzer using the BigDye® Terminator v3.1 Cycle Sequencing Kit. Two reactions were prepared each containing: Big dye terminator, 5X Sequencing Buffer, primer (forward or reverse), DNA template and the volume was completed using water. The primers used EstF and EstR. Thermocycler conditions were initial denaturation step at 96^{o}C for 1 minute, 25 cycles (denaturation at 96^{o}C for 10 seconds, annealing at 49^{o}C for 5 seconds and extension at 60^{o}C for 4 minutes). The DNA sample to be sequenced was purified using the Edge-Biosystem Columns, dried and loaded into the 96 well optical plate.

6. Characterization of EstATII

Subclone 14G2F5 was cultured overnight in LB Broth supplemented with ampicillin. Cells were harvested by centrifugation at 5000 rpm for 15 minutes. The culture supernatant showed positive lipolytic activity on tributyrin agar and was used directly for characterization of EstATII. Enzyme activity was determined by measuring the formation of *p*-nitrophenol (pNP) from the enzymatic hydrolysis of the *p*-nitrophenyl ester; *p*-nitrophenyl butyrate (Sigma-Aldrich, Cat.# N9876). Measurements were done at 410 nm using (Ultrospec 3100 pro, Amersham Biosciences). Amano lipases from *Aspergillus niger* (Sigma-Aldrich, Cat.# 534781) and *Pseudomonas fluorescens* (Sigma-Aldrich, Cat.#534730) were used as a positive control. Culture supernatant of *E. coli* TOP10 harboring pGEM®-T Easy vector was used as negative control. All experiments were performed in triplicates.

6.1. Effect of Temperature on EstATII activity

Enzyme activity was measured after incubation of the reaction mixture at different temperatures (25-80^{o}C). The reaction mixture [890 ul of 100 mM Tris-HCL (pH 8.0), 10 ul of *p*-nitrophenyl butyrate (pNPB) to a final concentration 0.1 mM and 100 ul of culture supernatant] was incubated for 10 minutes at each temperature, followed by measuring pNP formation at 410 nm.

6.2. Effect of pH on EstATII activity

Enzyme activity was assayed at a pH range (3-9) using the following buffers: 50 mM sodium acetate (pH 3-5), 50 mM sodium phosphate (pH 6, 7), 50 mM Tris-HCL (pH 8) and 50 mM glycing-NaOH (pH 9). The reaction mixture [890 ul of assigned buffer, 10 ul of *p*-nitrophenyl butyrate (pNPB) to a final concentration 0.1 mM and 100 ul of culture supernatant] was incubated at 60^{o}C for 10 minutes. Formation of pNP was measured at 348 nm (the pH-independent isosbestic wavelength of pNP).

6.3. Effect of NaCl concentration on EstATII activity

Enzyme activity was assayed at different NaCl concentrations (0-4 M) in 100 mM Tris-HCL (pH 8.0) using 100 ul culture supernatant and pNPB at final concentration 0.1 mM. The reaction mixture was incubated for 10 minutes at 60^{o}C and pNP formation was measured at 410 nm.

6.4. Substrate Specificity of EstATII

Substrate specificity of EstATII was determined using two *p*-nitrophenol esters; pNPB (C4) and *p*-nitrophenyl palmitate (pNPP) (C16). The reaction mixture [890 ul of Tris.Hcl pH= 8, 10 ul of pNPB or pNPP to a final concentration of 0.1 mM and 100 ul of culture supernatant] was incubated at 60^{o}C for 10 minutes. pNP formation was measured at 410 nm.

Chapter 3: Results

1. Screening Metagenomic Library for Lipolytic Activity

The constructed fosmid library comprised 10,656 clones that were manually placed into 111 96-well plates for ease of handling. Restriction fragment length polymorphism (RFLP) analysis using *Bam*HI of randomly selected fosmids confirmed the diversity of the library (data not shown). Functional screening of the fosmid library on tributyrin agar detected a total of five recombinant clones forming a clear halo zone which is indicative of putative lipolytic activity (Figure 5). The clones were designated 10E2, 12C11, 14G2, 16G7 and 35H9.

Fosmids of the five positive recombinant clones were pyrosequenced to identify the genes responsible for putative lipolytic activity. Following assembly, four large contigs (> 10 kb) and seven smaller contigs (ranging from 1.4 – 4.9 kb) were obtained; the largest contig (Contig 1) obtained being approximately 32 kb. ORFs were detected in all contigs by the ORF finder tool and annotated by BLASTP. Analysis revealed that Contig 1 (Figure 6) contains a 945 bp ORF encoding a putative esterase/lipase (designated EstATII) that shows significant similarities to lipolytic enzymes from cultured species such as *Pseudomonas sp.* and *Marinobacter sp.* It also shows significant similarities to lipolytic enzymes isolated through similar metagenomic approaches from uncultured soil and marine sediment bacteria. The maximum identity to sequences in the database was 65% with an alpha/beta hydrolase domain-containing protein from *Pseudomonas mendocina*. The highest identity to a lipolytic enzyme in the database was 56% to an esterase from *Pseudomonas aeruginosa* (Figure 7). A total of 29 ORFs were identified in Contig 1. The source organism for 20 ORFs as deduced from BLASTP analysis was a *Pseudomonas sp.* (Table 2).

2. Identification of EstATII

2.1. Sequence Analysis of EstATII

EstATII consists of 945 bp corresponding to 314 amino acids. A domain search conducted using CD-search tool detected an alpha/beta hydrolase fold domain [Pfam ID: pfam07859] between residues 85 and 286 which is a catalytic domain found in members of the alpha/beta hydrolases family. An esterase/lipase domain (cd00312) was also detected. This domain is characteristic of esterases and lipases which have a catalytic triad consisting of a serine, a glutamate or an aspartate and a histidine residue. In addition, two prokaryotic Clusters of Orthologous Groups (COGs) were identified; COG0657 and COG2272 which are involved in lipid metabolism (Figure 8). The catalytic triad residues, which are conserved in lipolytic enzymes, were identified in EstATII as Ser160, Asp204 and His282. The catalytic nucleophilic residue Ser160 was found in the consensus pentapeptide GDSAG which is characteristic of the hormone sensitive lipase (HSL) family. Another motif characteristic of the HSL family (HGG), which contributes to the formation of the oxyanion hole, was also identified in the sequence (Figure 9). A signal peptide was not detected in EstATII, and the protein was predicted to be soluble.

2.2. Phylogenetic Analysis and Classification of EstATII

Bacterial lipolytic enzymes were classified into eight families by Arpigny & Jaeger in 1999 [27]. In order to determine whether EstATII classifies as a member of one of these families, a multiple sequence alignment of EstATII together with 43 sequences of bacterial lipolytic enzymes, representing the eight families, was performed. A phylogenetic tree was constructed and EstATII grouped with members of family IV which is also known as the HSL family (Figure 10).

3. Subcloning of the Fosmid Harboring EstATII

PCR using the EstF-EstR primer pair showed amplification at expected size (~ 1 kb) with fosmid 14G2 (Figure 11). 14G2 was subjected to further subcloning in pGEM®-T Easy. 96 subclones were tested for lipolytic activity on tributyrin agar, out of which 7 subclones showed clear halo zones (Figure 12). Subclone 14G2F5, showing highest lipolytic activity, was selected for activity assays. PCR using the EstF-EstR primer pair on plasmid DNA extracted from subclone 14G2F5 showed amplification at expected size (data not shown) and sequencing confirmed that the subclone harbors EstATII.

4. Characterization of EstATII

Culture supernatant of subclone 14G2F5 was found positive for lipolytic activity and was used directly for the characterization of EstATII. The effects of temperature, pH and NaCl concentration, as well as, substrate specificity were assayed.

4.1. Effect of Temperature on EstATII Activity

The effect of temperature on the activity of EstATII was assayed at temperatures ranging from 25^{o}C to 80^{o}C. The activity of the enzyme increased reproducibly with the increase in temperature until 60^{o}C, after which the activity started to drop. High activity of the enzyme (>70%) was observed at temperatures ranging from 45^{o}C to 65^{o}C. The apparent optimum temperature of EstATII is 60^{o}C. The enzyme remained active even after reaching 80^{o}C (Figure 13). Results are displayed as percentage relative activity, where the highest activity exhibited by the enzyme is defined as 100%.

4.2. Effect of pH on EstATII Activity

The effect of pH on the activity of EstATII was assayed at pH range 3-9. The enzyme exhibited significant activity at pH= 6-8, with the highest activity at pH=7. No activity was observed at pH=3 and pH=4, while very low activity (<10%) was observed at

pH=5 and pH=9 (Figure 14). Results are displayed as percentage relative activity, where the highest activity exhibited by the enzyme is defined as 100%.

4.3. Effect of NaCl concentration on EstATII Activity

The activity of EstATII was assayed at different molar concentrations of NaCl ranging from 0M – 4M. The enzyme showed highest activity in the presence of 2M NaCl (Figure 15). Enzyme activity was maintained up to 4M NaCl and remained higher than 100%. Results are displayed as percentage relative activity, where the activity of the enzyme in the absence of NaCl is defined as 100%.

4.4. Substrate Specificity of EstATII

Substrate specificity was determined using pNPB (short-chain *p*-nitrophenyl ester = C4) and pNPP (long-chain *p*-nitrophenyl ester = C16). EstATII was active towards pNPB (Mean of UV spectrophotometric measurements= 0.737); however, it did not show any activity towards pNPP (Mean of UV spectrophotometric measurements= 0.012) under the conditions tested.

Chapter 4: Discussion

In this thesis work, we describe the identification and characterization of a novel thermohalophilic lipolytic enzyme (EstATII) from the Atlantis II Deep. Atlantis II Deep is one of the brine pools found in the Red Sea, approximately 2200m below the surface of the water. It is characterized by high temperature (68^{o}C), high salinity (270 psu), in addition to high concentrations of heavy metals and anoxia [2]. Despite the uniqueness of the Atlantis II brine pool, it represents an untapped environment that holds great potential for discovering novel biocatalysts with unique characteristics that make them suited for several industrial and biotechnological applications. To the best of our knowledge, this is first attempt to mine for lipolytic enzymes from the Atlantis II Deep.

A fosmid library constructed using DNA from the LCL of Atlantis II Deep was screened for lipolytic activity. Lipolytic fosmid clone was subcloned into a high copy number plasmid (pGEM®-T Easy) which is a simple approach to achieve sufficient quantities of the enzyme for further characterization. Subclones were screened for lipolytic activity for the selection of the subclone with the highest activity to be used for characterization of the enzyme. The feasibility and high success rate of this strategy makes it ideal for characterizing various candidates in a high-throughput manner, to select the candidate(s) that might have potential applications [30].

BLASTP analyses of EstATII revealed its significant similarity to several lipolytic enzymes isolated from cultured species such as *Pseudomonas sp*. and *Marinobacter sp.,* as well as from uncultured bacteria from soil and marine sediment samples identified through metagenomic approaches that resemble the work described here. EstATII showed little match with any of the sequences in the database, with the highest maximum identity being 65%; reflecting the novelty of the enzyme. Domain search by CD-search tool detected the presence of an esterase/lipase domain (cd00312) characteristic of lipolytic enzymes, in addition to an alpha/beta hydrolase domain (pfam07859) that is found in members of the alpha/beta hydrolases family that includes lipolytic enzymes [26].

Phylogenetic analysis showed that EstATII groups with members of family IV (based on the classification by Arpigny and Jaeger [27]) which is also referred to as the Hormone-Sensitive Lipase (HSL) family. Members of this family strikingly resemble the mammalian HSL (hence the name) and have several conserved motifs which were also found in EstATII. Multiple sequence alignment of EstATII with other members of the HSL family, revealed the presence of the nucleophilic catalytic serine residue in the consensus GDSAG, as well as the motif HGG, which is involved in the formation of the oxyanion hole. Both motifs are highly conserved within members of the HSL family [27]. These findings suggest that EstATII is a novel member of the HSL family which includes psychrophilic, mesophilic and thermophilic members [27].

EstATII was among 29 other ORFs detected in Contig 1 (approximately 32 kb). Interestingly, the source organism (as deduced by BLASTP analysis) for 20 of the 29 ORFs was a *Pseudomonas* species, suggesting that this DNA insert might have originated from a microorganism that is closely related to *Pseudomonas*.

The thermophilic and halophilic properties of EstATII were assessed. The enzyme was assayed at a wide temperature range (25^{o}C - 80^{o}C) and maximal activity was achieved at 60^{o}C. EstATII activity was also assayed at different salt (NaCl) concentrations (0M – 4M) and showed enhanced activity in the presence of NaCl (0.5M – 4M) and peaked at 2M. EstATII is an esterase and not a lipase, since it preferentially hydrolyzes pNPB over pNPP. The amino acid identity of EstATII with esterases and the preference for hydrolyzing short-chain esters agrees with its esterase properties. The improved activity at both high temperature and salt concentration reflects the environment from which EstATII was isolated. Further characterization of EstATII is required to fully understand its potential applications. Nonetheless, based on its thermophilic and halophilic properties, it has a potential application in processes which may require extreme conditions including high temperature and/or high salt concentration such as resolving racemic mixtures and biodiesel production [52]. To the best of our knowledge, this is the first study to identify and characterize a novel enzyme from the Atlantis II brine pool in the Red Sea.

Although we were successful in isolating a novel lipolytic enzyme, the coverage of the fosmid library was insufficient as reflected by the low hit rate (1:10,656). This low hit rate could

be attributed to the limitations of heterologous gene expression such as codon usage, improper folding and secretion. In addition, using *E. coli* as a host limits the screening process to the optimum conditions for *E. coli* growth, which may not be favorable for the activity of the enzyme sought. Further studies can be performed to improve the hit rate, this could be achieved through expanding the library to comprise more clones, or construction of the library following enrichment processes [53]. In addition, development of host/vector systems that are adapted to high-temperature could be helpful in screening for biocatalysts from environments characterized by high temperatures such as the focus of this work: the Atlantis II Deep.

It is worth noting that the use of tributyrin as a substrate for the screening process is debatable. Despite its wide use in similar studies, tributyrin was reported to detect false positives [33]. In a study, 18 out of 10,000 clones displayed lipolytic activity on tributyrin agar, however sequencing revealed that only 3 out of the 18 positive clones (1:6) were esterases [33]. Almost the same rate of false positives was encountered in our study (1:5). However, Tributyrin (C4) is preferred as a substrate because it's a short-chain acyl ester and therefore would detect both esterases and lipases. Other short-chain esters should be considered as substrates, such as tricaproin (C6) and tricaprylin (C8).

This work supports the potential usage of metagenomic approaches to identify novel biocatalysts with unique characteristics and potential applications from extreme environments.

Chapter 5: Conclusion and Future Prospects

In conclusion, we have identified a novel esterase (EstATII) from the Atlantis II brine pool in the Red Sea, using a function-based approach. Sequence analysis of this enzyme revealed that is a new member of the Hormone-Sensitive Lipase family. Characterization of EstATII revealed its thermophilic and halophilic characteristics which could make EstATII a potentially useful biocatalyst in several industrial processes.

This work demonstrates that the Atlantis II Deep and other similar extreme environments hold great potential for the discovery of novel biocatalysts. The characteristics of these biocatalysts almost always reflect the conditions of the environment from which they were isolated. Therefore, these biocatalysts can be potentially used for industrial and biotechnological applications that employ extreme conditions in terms of temperature, pH, salt concentration and heavy metal concentration among others.

Further characterization of EstATII is required to understand its full potential as a biocatalyst. For example, its activity in the presence of detergents, heavy metals and solvents should be assayed. In addition, in this study the activity of EstATII towards pNPB (C4) and pNPP (C16) only was assessed. Substrate activity towards a wider array of *p*-nitrophenol esters of different carbon chain lengths should be explored.

Another important aspect is the enantioselectivity and regioselectivity of EstATII. Enantioselectivity ensures the formation of one enantiomer over the other if the product is chiral. On the other hand, regioselectivity refers to the preference of the enzyme to break a specific bond over the other. These properties are therefore of high importance in industrial and biotechnological processes as they ensure the formation of a specific desired product and avoid the formation of unwanted side products.

Finally, mining for more lipolytic enzymes, in addition to other biocatalysts, from the Atlantis II Deep should be sought. Other brine pools in the Red Sea such as Kebrit Deep, Shaban Deep and Discovery Deep should be explored as well.

References

1. Qian, P.Y. et al. Vertical stratification of microbial communities in the Red Sea revealed by 16S rDNA pyrosequencing. *ISME J* **5**, 507-18 (2010).
2. Antunes, A., Ngugi, D.K. & Stingl, U. Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environmental Microbiology Reports* **3**, 416–433 (2011).
3. http://krse.kaust.edu.sa/spring-2010/research.html.
4. Danielsson, L.G., Dyrssen, D. & Granéli, A. Chemical investigations of Atlantis II and discovery brines in the Red Sea. *Geochimica et Cosmochimica Acta* **44**, 2051-2065 (1980).
5. Bower, A.S. R/V Oceanus Voyage 449-6 Red Sea Atlantis II Deep Complex Area. *Woods Hole Oceanog. Instit. Technical Report, WHOI-KAUST-CTR-2009-01* (2009).
6. Whitman, W.B., Coleman, D.C. & Wiebe, W.J. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**, 6578-83 (1998).
7. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65 (2010).
8. Singh J. et al. Metagenomics: Concept, methodology, ecological inference and recent advances. *Biotechnology Journal* **4**, 480–494 (2009).
9. R. D. Sleator, C. Shortall & Hill, C. Metagenomics. *Letters in applied microbiology* **47**, 361-366 (2008).
10. Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. & Field, K.G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60-3 (1990).
11. Eden, P.A., Schmidt, T.M., Blakemore, R.P. & Pace, N.R. Phylogenetic analysis of Aquaspirillum magnetotacticum using polymerase chain reaction-amplified 16S rRNA-specific DNA. *Int J Syst Bacteriol* **41**, 324-5 (1991).
12. Schmidt, T.M., DeLong, E.F. & Pace, N.R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* **173**, 4371-8 (1991).
13. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**, R245-9 (1998).
14. Huse, S.M. et al. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* **4**, e1000255 (2008).
15. Tyson, G.W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43 (2004).
16. Guazzaroni M. E. et al. in Handbook of Hydrocarbon and Lipid Microbiology 2911-2927 (Springer-Verlag Berlin Heidelberg, 2010).
17. Uchiyama T. & Miyazaki K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology* **20**, 616–622 (2009).
18. Steele H. L., Jaeger K. E., Daniel R & Streit W. R. Advances in Recovery of Novel Biocatalysts from Metagenomes. *Journal of Molecular Microbiology and Biotechnology* **16**, 25-37 (2009).
19. Schmeisser C., Steele H. & Streit W. R. Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology* **75**, 955–962 (2007).
20. Lorenz P. & Eck J. Metagenomics and industrial applications. *Nature Reviews Microbiology* **3**, 510-516 (2005).

21. Hausmann S. & Jaeger K. E. in Handbook of Hydrocarbon and Lipid Microbiology 1100-1126 (Springer-Verlag Berlin Heidelberg, 2010).
22. Lotti M. & Alberghina L. in Industrial Enzymes 263–281 (Springer, 2007).
23. Sarda L. & Desnuelle P. Actions of pancreatic lipase on esters in emulsions. *Biochimica et Biophysica Acta* **3**, 513-521 (1958).
24. Brady L. et al. A serine protease triad forms the catalytic center of a triacylglycerol lipase. *Nature* **343**, 767–770 (1990).
25. Winkler F. K., D'Arcy A. & Hunziker W. Structure of human pancreatic lipase. *Nature* **343**, 771–774 (1990).
26. Jaeger K .E., Dijkstra B. W. & Reetz M. T. Bacterial Biocatalysts**:** Molecular Biology, Three-Dimensional Structures,and Biotechnological Applications of Lipases. *Annual Review of Microbiology* **53**, 315-351 (1999).
27. Arpigny J. L. & Jaeger K. E. Bacterial lipolytic enzymes: classification and properties. *Journal of Biochemistry* **343**, 177-183 (1999).
28. Lee M. H. et al. Isolation and Characterization of a Novel Lipase from a Metagenomic Library of Tidal Flat Sediments: Evidence for a New Family of Bacterial Lipases. *Applied and Environmental Microbiology* **72**, 7406–7409 (2006).
29. Bornscheuer U. T. Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiology Reviews* **26**, 73-81 (2002).
30. Jaeger K .E. & Eggert T. Lipases for biotechnology. *Current Opinion in Biotechnology* **13** (2002).
31. Salameh M. & Wiegel J. Lipases from extremophiles and potential for industrial applications. *Applied and Environmental Microbiology* **61**, 253-283 (2007).
32. Carrière F. et al. Structural basis for the substrate selectivity of pancreatic lipases and some related proteins. *Biochimica et Biophysica Acta* **1376**, 417-432 (1998).
33. Litthauer D., Abbai N. S., Piater L. A. & van Heerden E. Pitfalls using tributyrin agar screening to detect lipolytic activity in metagenomic studies. *African Journal of Biotechnology* **9**, 4282-4285 (2010).
34. Hasan F., Shah A. A. & Hameed A. Methods for detection and characterization of lipases: A comprehensive review. *Biotechnology Advances* **27**, 782–798 (2009).
35. Kouker, G. & Jaeger, K.E. Specific and sensitive plate assay for bacterial lipases. *Appl Environ Microbiol* **53**, 211-3 (1987).
36. Rondon, M.R. et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**, 2541-7 (2000).
37. Sifour, M., Saeed, H.M., Zaghloul, T.I., Berekaa, M.M. & Abdel-Fattah, Y.R. Isolation of Lipase Gene of the Thermophilic Geobacillus stearothermophilus Strain-5. *Biotechnology* **9**, 55-60 (2010).
38. Leow, T.C., Rahman, R.N., Basri, M. & Salleh, A.B. High level expression of thermostable lipase from Geobacillus sp. strain T1. *Biosci Biotechnol Biochem* **68**, 96-103 (2004).
39. Jiang, Z., Wang, H., Ma, Y. & Wei, D. Characterization of two novel lipase genes isolated directly from environmental sample. *Appl Microbiol Biotechnol* **70**, 327-32 (2006).
40. Bell, P.J. et al. Prospecting for novel lipase genes using PCR. *Microbiology* **148**, 2283-91 (2002).
41. Hasan F., , Shah A. A. & Hameed A. Industrial applications of microbial lipases. *Enzyme and Microbial Technology* **39** 235-251 (2006).
42. Gandhi N. N. Applications of Lipase. *Journal of the American Oil Chemists' Society* **74**, 621-634 (1997).

43. Aravindan R., Anbumathi P. & Viruthagiri T. Lipase applications in food industry. *Indian Journal of Biotechnology* **6**, 141-158 (2007).
44. Hata K., Matsukura M., Taneda H. & Fujita Y. in Enzymes for Pulp and Paper Processing 280–296 (American Chemical Society, 1996).
45. Dias R. B., de Castro A. M., Coelho M. A. Z. & Freire D. M. G. Production and Use of Lipases in Bioenergy: A Review from the Feedstocks to Biodiesel Production. *Enzyme Research* **2011**, 1-16 (2011).
46. Rusch, D.B. et al. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**, e77 (2007).
47. Ye, J., McGinnis, S. & Madden, T.L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res* **34**, W6-9 (2006).
48. Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-95 (2004).
49. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80 (1994).
50. Tamura, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9 (2011).
51. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**, W475-8 (2011).
52. Rocha, S.N., Abrahao-Neto, J., Cerdan, M.E., Gombert, A.K. & Gonzalez-Siso, M.I. Heterologous expression of a thermophilic esterase in Kluyveromyces yeasts. *Appl Microbiol Biotechnol* **89**, 375-85 (2011).
53. Lee, S.W. et al. Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biotechnol* **65**, 720-6 (2004).
54. Henne, A., Schmitz, R.A., 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* **66**, 3113-6 (2000).
55. Rhee, J.K., Ahn, D.G., Kim, Y.G. & Oh, J.W. New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl Environ Microbiol* **71**, 817-25 (2005).
56. Lee, M.H., Lee, C.H., Oh, T.K., Song, J.K. & Yoon, J.H. 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* **72**, 7406-9 (2006).
57. 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* **59**, 524-34 (2007).
58. Chu, X., He, H., Guo, C. & Sun, B. Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl Microbiol Biotechnol* **80**, 615-25 (2008).
59. Tirawongsaroj, P. et al. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J Biotechnol* **133**, 42-9 (2008).
60. Zhang, T. & Han, W.J. Gene cloning and characterization of a novel esterase from activated sludge metagenome. *Microb Cell Fact* **8**, 67 (2009).
61. Glogauer, A. et al. Identification and characterization of a new true lipase isolated through metagenomic approach. *Microb Cell Fact* **10**, 54 (2011).

62. Morohoshi, T. et al. Isolation and characterization of novel lipases from a metagenomic library of the microbial community in the pitcher fluid of the carnivorous plant Nepenthes hybrida. *J Biosci Bioeng* **112**, 315-20 (2011).
63. Jeon, J.H. et al. Novel lipolytic enzymes identified from metagenomic library of deep-sea sediment. *Evid Based Complement Alternat Med* **2011**, 271419 (2011).

TABLES

Table 1: Screening Metagenomic Libraries for Lipolytic Activity Using Tributyrin Agar

Sample	Number of Screened Clones	Number of Positive Clones	Vector/Host used for Library Construction	Reference
Samples were collected from 3 different locations, a library was constructed from each. Sample I: from a meadow near Northeim in Germany Sample II: from a sugar beet field near Göttingen Sample III: from the valley of the Nieme River	I: 73,000 II:115,000 III: 98,000	I: 1 II: 0 III: 2	pBluescript SK(+) plamid vector/ *E. coli* DH5a	54
Mixtures of mud and sediment-rich water collected from hot springs and mud holes in solfataric fields in Indonesia.	2000	4	pCC1FOS fosmid vector/ EPI300-T1R phage T1-resistant *E.coli*	55
Soil samples from forest topsoil from Yuseong at Daejeon, Korea	33,700	8	pEpiFOS-5 fosmid vector/ *E. coli* EPI-100	53
Korean tidal flat	386,400	4	pCC1FOS fosmid vector/ *E. coli* EPI300-T1	56
Baltic Sea sediment	7000	70	pCC1FOS fosmid vector/ *E. coli* EPI300TM-T1R	57
Surface seawater collected from South China Sea	20,000	4	pIndigoBAC-5 BAC vector/ *E. coli* EPI300	58
Soil from the Jae Sawn hot spring	36,000	2	pZErO-2 plasmid vector/ *E. coli* TOP10	59
Activated sludge from a low temperature sequencing batch bioreactor	100, 000	1	pCC2FOS Fosmid vector/ *E. coli* EPI300 -T1R	60
Anaerobic lagoon of wastewater treatment plant	500,000 clones	2,661	pCC2FOS fosmid vector/ *E. coli* EPI300™-T1R	61
Symbiotic bacteria from the fluid of two pitchers of a single *N. hybrida* pitcher plant	55,500	2	pUC118 plasmid vector/ *E. coli* DH5α	62
Deep-sea sediment sample from the southern clam beds area around the summit of Edison Seamount in the New Ireland Fore-arc	81,100	6	pCC1FOS fosmid vector/ *E. coli* EPI300-T1R	63

Table 2: BlastP results for ORFs identified in contig 1

ORF no.	Length (amino acids)	Putative Function	Source Organism	Accession number of homologue	%identity/ similarity	E Value
1	94 (truncat-ed)	ABC transporter permease	*Pseudomonas stutzeri* ATCC 17588	YP_004715921	86/95	2e-49
2	309	ABC transporter permease	*Pseudomonas stutzeri A1501*	YP_001174049	80/87	4e-163
3	56	Lipoprotein	*Pseudomonas fulva 12-X*	YP_004473638	55/76	1e-08
4	259	short-chain dehydrogenase	*Pseudomonas stutzeri LMG 11199T*	YP_004716473	66/77	1e-115
5	314	*Putative lipolytic enzyme. BlastP results are shown in Figure 7.*				
6	491	flavin-binding monooxygenase	Pseudomonas aeruginosa PAO1	NP_250787	61/77	0.0
7	344	AraC family transcriptional regulator	Pseudomonas mendocina ymp	YP_001187384	50/66	9e-97
8	394	possible acyl-CoA dehydrogenase	marine gamma proteobacterium HTCC2207	ZP_01225007	76/87	0.0
9	380	putative acyl-CoA dehydrogenase	marine gamma proteobacterium HTCC2207	ZP_01225008	57/72	3e-149
10	396	hypothetical transposase	Vibrio harveyi ATCC BAA-1116	YP_001436060	51/71	3e-119
11	287	hypothetical protein Pfl01_1698	Pseudomonas fluorescens Pf0-1	YP_347430	62/74	5e-108
12	168	MutT-like protein	Pseudomonas sp. ADP	NP_862501	48/63	2e-29
13	841	cyclic di-GMP signal transduction protein	Azotobacter vinelandii DJ	YP_002800011	41/61	0.0
14	154	LrgA family protein	Pseudomonas mendocina NK-01	YP_004381897	69/80	6e-37
15	235	LrgB-like protein	Pseudomonas syringae pv. oryzae str. 1_6	ZP_04586011	65/78	2e-93
16	518	cytochrome bd-type quinol oxidase, subunit 1	Rheinheimera sp. A13L	ZP_08569239	75/89	0.0

17	378	cytochrome d oxidase cyd, subunit II	Rheinheimera sp. A13L	ZP_08569238	70/81	0.0
18	37	cyd operon protein YbgT	Rheinheimera sp. A13L	ZP_08569237	86/94	2e-15
19	585	ABC transporter, CydDC cysteine exporter (CydDC-E) family	Psychrobacter sp. PRwf-1	YP_001279231	48/63	2e-162
20	544	ABC-type transporter involved in cytochrome bd biosynthesis	Dichelobacter nodosus VCS1703A	YP_001209031	50/66	5e-144
21	1094	pyruvate carboxylase., propionyl-CoA carboxylase	Pseudomonas mendocina ymp	YP_001189321	71/81	0.0
22	566	acyl-CoA synthetase	Pseudomonas fluorescens Pf-5	YP_260134	68/81	0.0
23	316	LysR family transcriptional regulator	Pseudomonas aeruginosa PA7	YP_001349295	60/78	5e-123
24	799	penicillin-binding protein, 1A family	Pseudomonas fulva 12-X	YP_004472498	62/76	0.0
25	261	type IV pilus assembly protein PilM	Pseudomonas fulva 12-X	YP_004472499	82/92	2e-147
26	192	type 4 fimbrial biogenesis protein PilN	Pseudomonas aeruginosa PAO1	NP_253730	66/79	3e-87
27	207	type 4 fimbrial biogenesis protein PilO	Pseudomonas aeruginosa UCBPP-PA14	YP_793511	67/79	4e-96
28	176	type 4 fimbrial biogenesis protein PilPz	Pseudomonas aeruginosa PA7	YP_001351097	72/87	8e-90
29	74 (truncat-ed)	type IV pilus secretin PilQ	Pseudomonas mendocina ymp	YP_001186048	62/84	3e-13

FIGURES

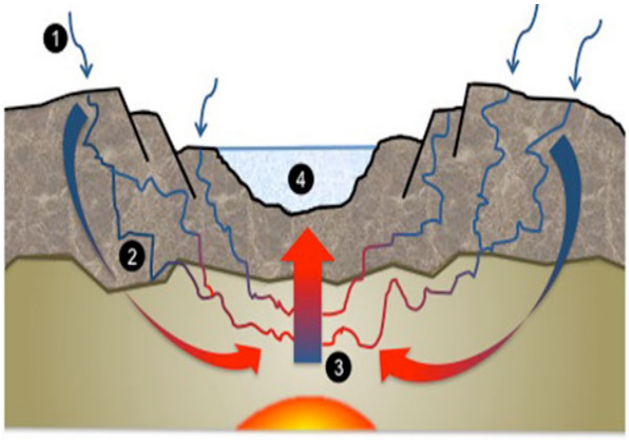

Figure 1: Formation of Brine Pools by Tectonic Activity

Localized tectonic activity (1) leads to the formation of brine pools by the formation of fissures that are penetrated by deep-sea water (2). Water is heated by submarine magma (3) and absorbs minerals as it is driven back to the seafloor by convective currents (4) [3].

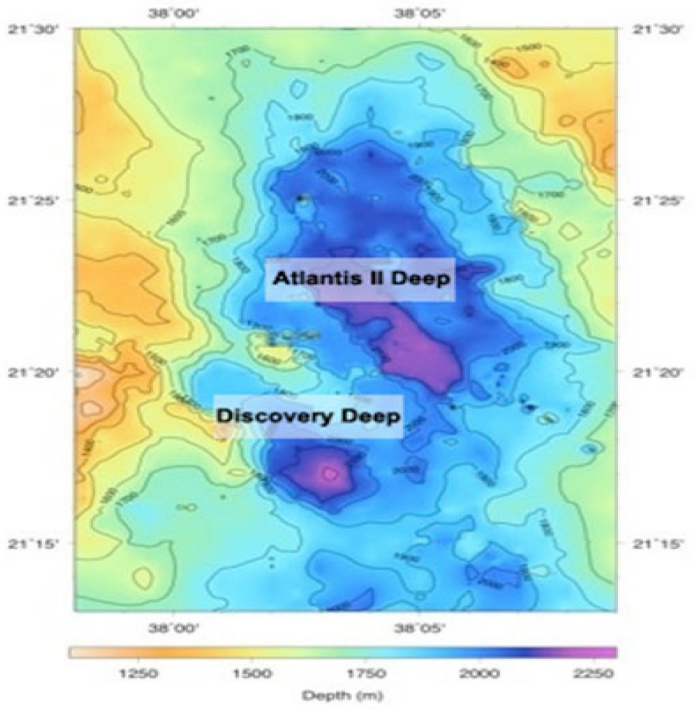

Figure 2: Locations and Depths of Atlantis II Deep and Discovery Deep in the Red Sea

The locations of two of the brine pools in the Red Sea; the Atlantis II deep (the largest brine pool discovered to date in the Red Sea) and Discovery deep. The color code at the bottom refers to the depth [3].

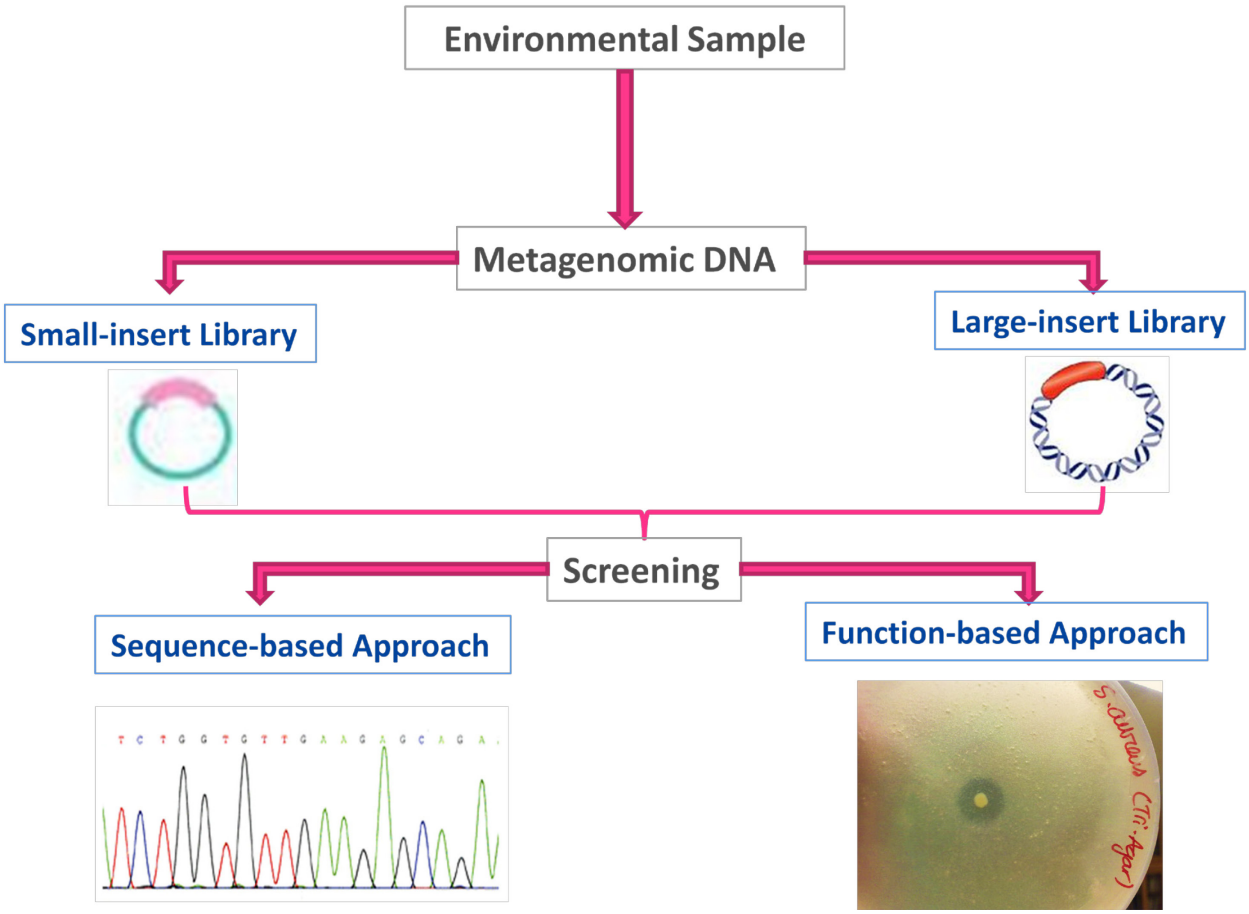

Figure 3: Construction and Screening of Metagenomic Clone Library

DNA is isolated from environmental samples and used to construct metagenomic clone libraries. A library could be a small-insert library (<10 kb) if the purpose of the study is to isolate single genes or a large-insert library (>10 kb) if the purpose of the study is to isolate full operons or gene clusters. The library can then be screened by the sequence-based approach using primers or probes, or by the function-based approach using a substrate supplemented to the medium, that can only be utilized by the gene sought.

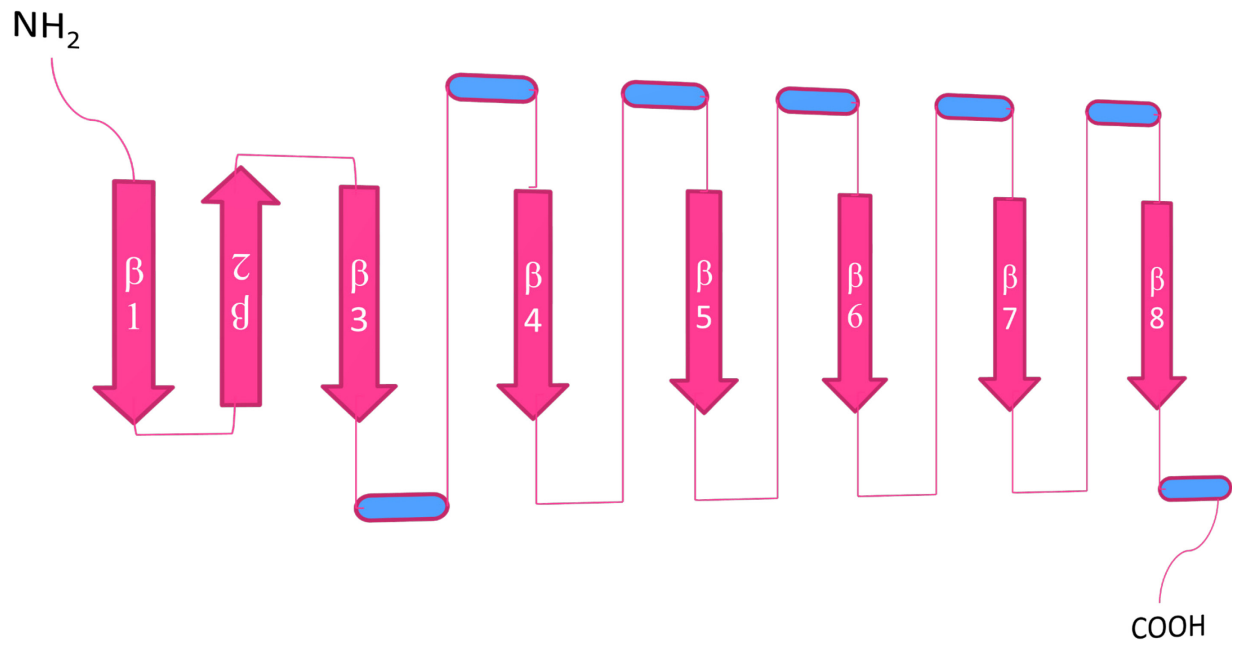

Figure 4: Conserved Three-Dimensional Structure of Lipolytic Enzymes

The alpha/beta hydrolase fold consists of an almost parallel beta sheet, formed by 8 beta strands. Strands 3-8 are connected by alpha helices. This three-dimensional conformation is conserved among lipolytic enzymes despite their primary structure diversity.

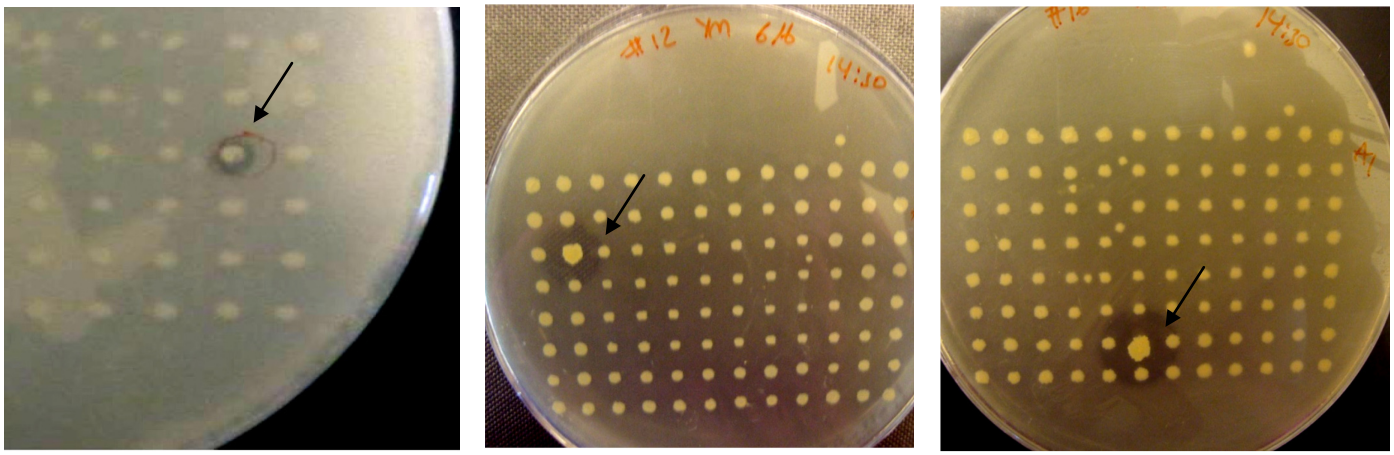

Figure 5: Screening Metagenomic Fosmid Library for Lipolytic Activity

A total of 111 96-well plates were screened for lipolytic activity. A replica of each plate was made on LB agar supplemented with tributyrin. The plates were incubated at 37°C for 3 days. Five clones (designated 10E2, 12C11, 14G2, 16G7 and 35H9) showed clear halo zones which indicate putative lipolytic activity. Three positive clones are shown in this figure; 10E2, 12C11 and 16G7.

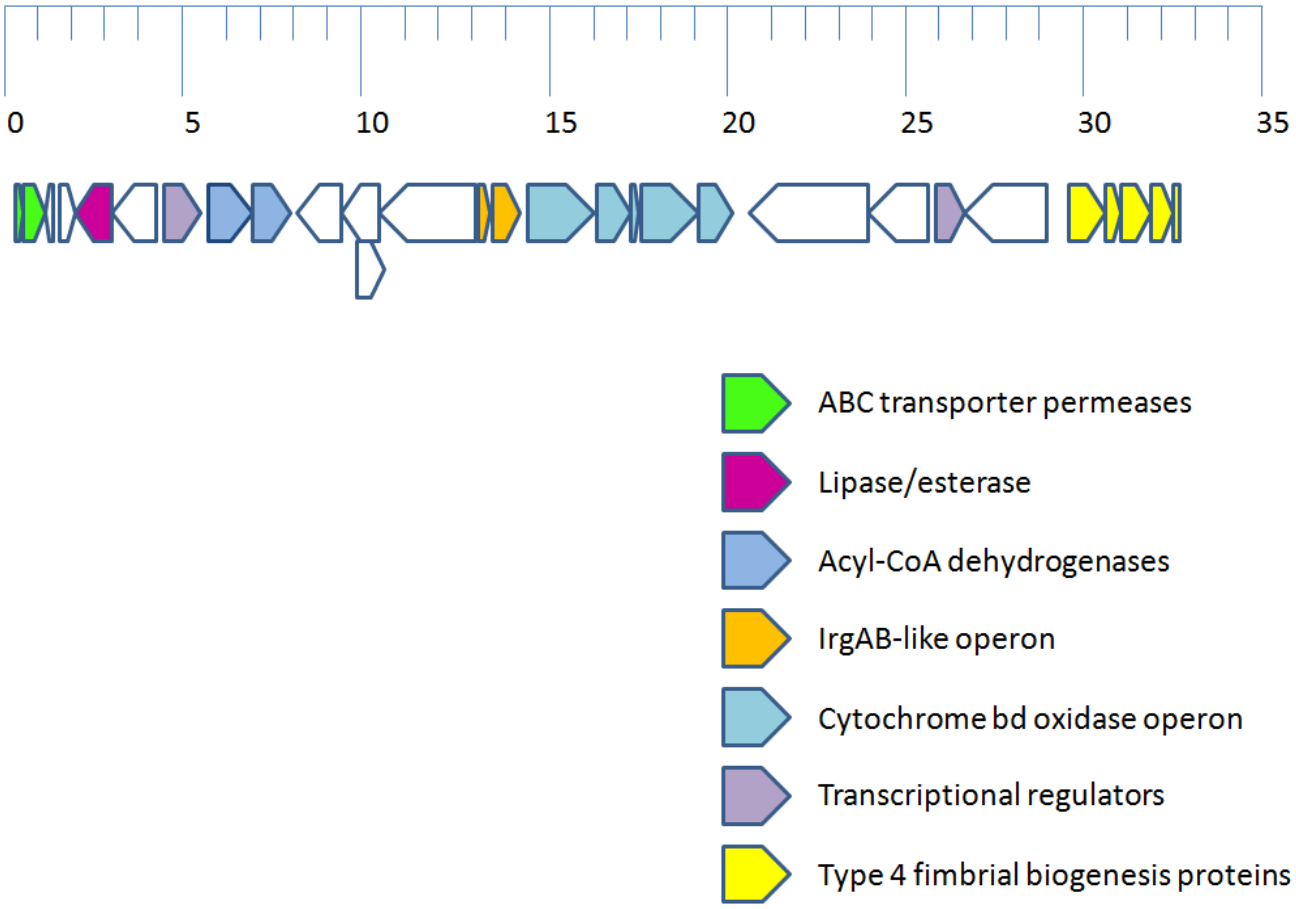

Figure 6: Graphical Representation of Contig 1

The figure shows open reading frames detected in Contig 1 (32374 bp) by the software tool ORF finder, and some of their putative functions as deduced from BLASTP results. EstATII, a 945 bp ORF with putative lipolytic function is demonstrated (deep purple). The contig also harbors ORFs encoding for ABC transporter, and three putative operons; the IrgAB-like operon, the Cytochrome bd oxidase operon and the type 4 fimbrial operon. In total, 29 ORFs were detected in Contig 1. The source organism for 20 ORFs based on the BLASTP analysis is *Pseudomonas* sp.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
YP_001187386.1	alpha/beta hydrolase domain-containing protein [Pseudomonas mendoci	374	374	92%	4e-128	65%
YP_004380790.1	alpha/beta hydrolase domain-containing protein [Pseudomonas mendoci	362	362	92%	2e-123	62%
ZP_06878960.1	putative esterase [Pseudomonas aeruginosa PAb1]	269	269	88%	1e-86	56%
YP_791141.1	unnamed protein product [Pseudomonas aeruginosa UCBPP-PA14] >gb|	268	268	90%	2e-86	55%
ZP_04928456.1	hypothetical protein PACG_01017 [Pseudomonas aeruginosa C3719] >g	267	267	90%	5e-86	55%
NP_250788.1	unnamed protein product [Pseudomonas aeruginosa PAO1] >ref|YP_002	267	267	90%	7e-86	55%
ZP_09055640.1	hypothetical protein HMPREF1030_04726 [Pseudomonas sp. 2_1_26] >g	266	266	90%	1e-85	55%
ZP_01365444.1	hypothetical protein PaerPA_01002569 [Pseudomonas aeruginosa PACS	263	263	90%	2e-84	55%
BAK89926.1	putative esterase [Pseudomonas aeruginosa NCGM2.S1]	224	224	71%	4e-70	58%
YP_693828.1	unnamed protein product [Alcanivorax borkumensis SK2] >emb|CAL1755	224	224	97%	2e-69	44%
ZP_05042810.1	alpha/beta hydrolase fold domain protein [Alcanivorax sp. DG881] >gb|E	224	224	97%	3e-69	44%
ZP_07796906.1	putative esterase [Pseudomonas aeruginosa 39016] >gb|EFQ42002.1|	220	220	70%	1e-68	57%
ZP_09160728.1	alpha/beta hydrolase domain-containing protein [Marinobacter sp. MnI7-	222	222	93%	1e-68	45%
ADP98321.1	alpha/beta hydrolase fold-3 domain protein [Marinobacter adhaerens HP	220	220	93%	9e-68	45%
ZP_01894884.1	putative esterase [Marinobacter algicola DG893] >gb|EDM47026.1| puta	220	220	92%	1e-67	43%
CBW44451.1	putative Acetyl-hydrolase [Marinobacter hydrocarbonoclasticus]	218	218	90%	8e-67	45%
YP_960085.1	alpha/beta hydrolase domain-containing protein [Marinobacter aquaeole	215	215	90%	7e-66	45%
YP_001412624.1	alpha/beta hydrolase domain-containing protein [Parvibaculum lavament	202	202	70%	7e-61	50%
ABY61092.1	esterase [uncultured bacterium]	201	201	80%	4e-60	42%
YP_002874800.1	putative flavin-binding monooxygenase-like protein [Pseudomonas fluore	204	204	91%	5e-57	42%
YP_003111476.1	alpha/beta hydrolase fold protein-3 domain-containing protein [Catenuli	190	190	78%	5e-56	46%
AAS77236.1	lipase/esterase [uncultured bacterium]	188	188	81%	3e-55	41%
AAS77238.1	lipase/esterase [uncultured bacterium]	187	187	81%	4e-55	41%
AAX37295.1	lipase/esterase [uncultured bacterium]	187	187	76%	5e-55	43%
AAX37296.1	lipase/esterase [uncultured bacterium]	187	187	82%	5e-55	41%
YP_003184482.1	unnamed protein product [Alicyclobacillus acidocaldarius subsp. acidocalc	187	187	94%	5e-55	40%
AAS77233.1	lipase/esterase [uncultured bacterium]	187	187	81%	8e-55	41%
AEM45113.1	hypothetical protein [uncultured organism]	186	186	82%	1e-54	40%

Figure 7: BLASTP Results of EstATII

BlastP results for EstATII show significant similarities to lipolytic enzymes from *Pseudomonas* sp., *Marinobacter* sp., and several uncultured bacteria from soil and marine sediment samples. The maximum identity to any sequence in the database is 65%. The highest identity to a lipolytic enzyme in the database is 56% which is an esterase from *Pseudomonas auruginosa.*

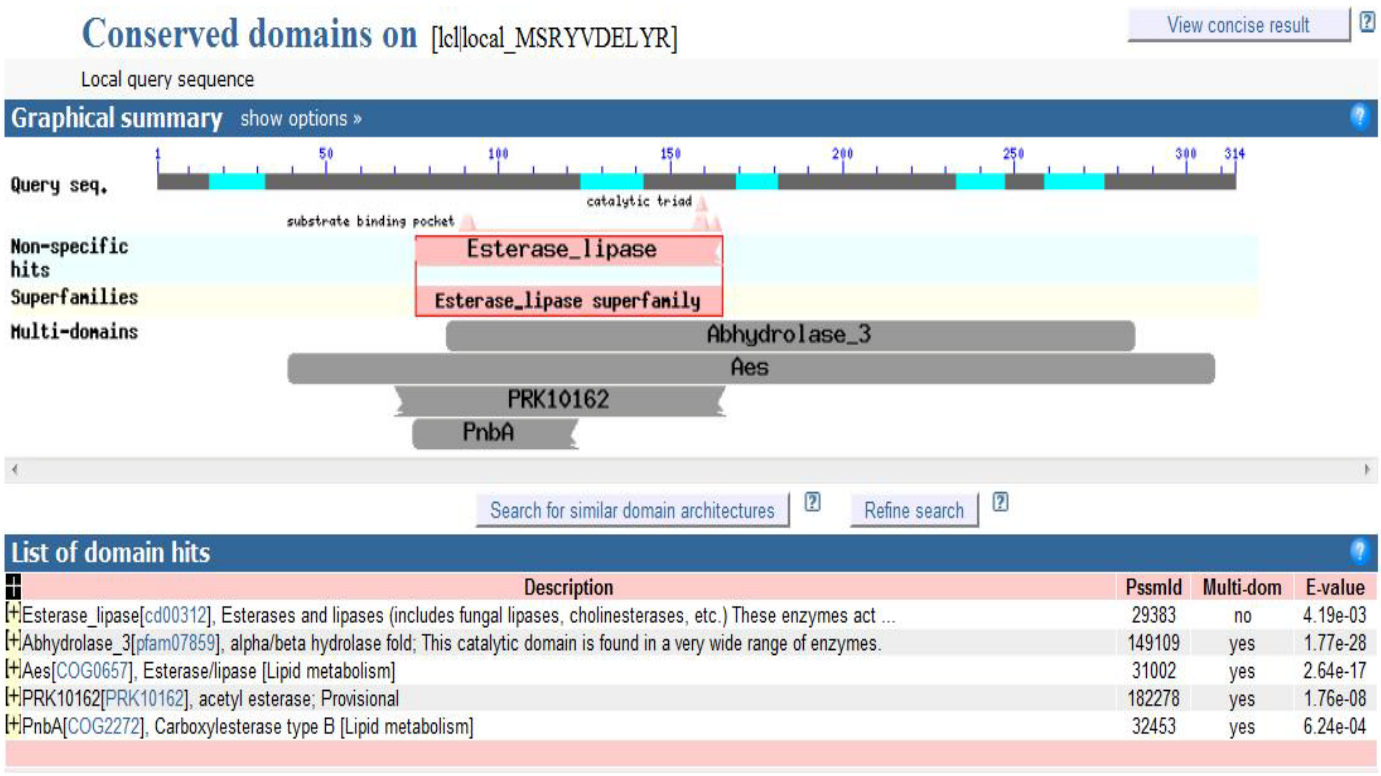

Figure 8: Domain search in EstATII

The domain search by CD-search software tool detected an alpha/beta hydrolase fold domain (pfam07859) which is characteristic of members of the alpha/beta hydrolases family to which lipolytic enzymes belong. In addition, the search detected an esterase/lipase domain (cd00312) which is also found in lipolytic enzymes.

Figure 9: Conserved Motifs in Family IV: The Hormone Sensitive Lipase (HSL) Family

Multiple sequence alignment of EstATII with other members of the HSL family: *Pseudomonas* sp. B11-1 (AF034088), *Archaeoglobus fulgidus* (AE000985), *Alcaligenes eutrophus* (L36817), *Escherichia coli* (AE000153) and *Moraxella* sp. (X53868) was performed using ClustalW and visualized by MEGA v. 5.05. The alignment shows the conserved motif HGG which is involved in the formation of the oxyanion hole. It also shows the nucleophilic catalytic serine residue in the pentapeptide GDSAG which is also conserved in HSL family.

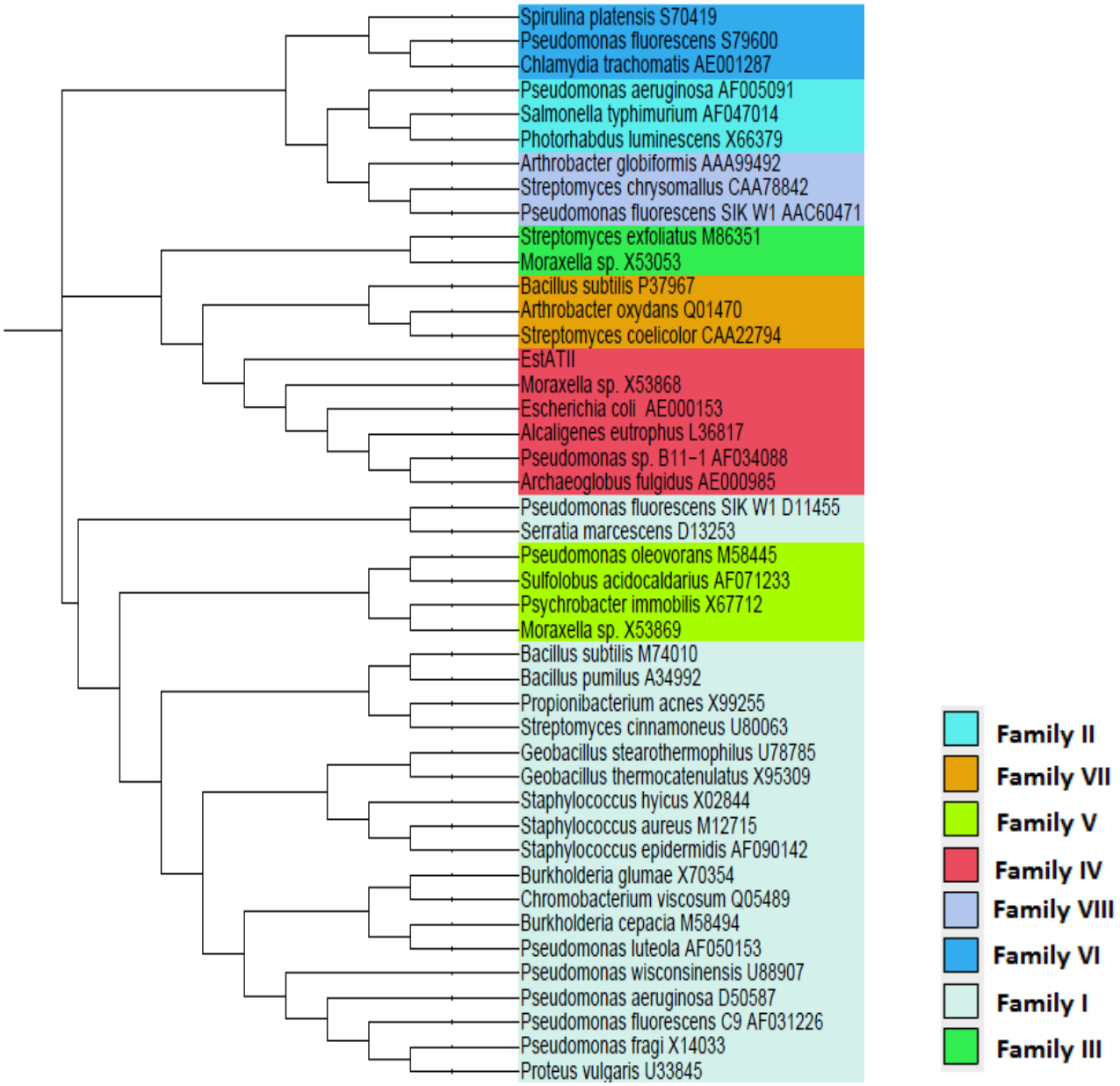

Figure 10: Phylogenetic Analysis and Classification of EstATII

A multiple sequence alignment of EstATII together with 43 lipolytic enzymes(representing the eight families of bacterial lipolytic enzymes as classified by Arpigny and Jaeger[27]) was used to construct a phylogenetic tree to classify EstATII. EstATII groups with members of family IV also known as the HSL family. The confidence level of the tree was estimated by bootstrapping (10,000 replicates).

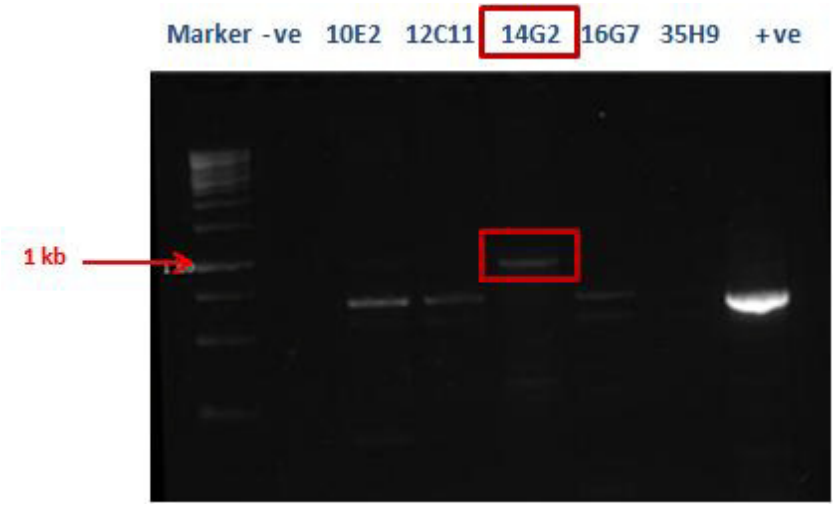

Figure 11: Identification of the Fosmid Harboring EstATII

Based on sequencing results, primers were designed to amplify EstATII (945 bp). PCR showed amplification at expected size (~1 kb) with fosmid 14G2 but not other clones. Lane 1 shows 1 kb molecular weight marker (Fermentas), lanes 3-7 show PCR results for fosmids of the five putative lipolytic clones and lane 8 is a positive control for the PCR mix.

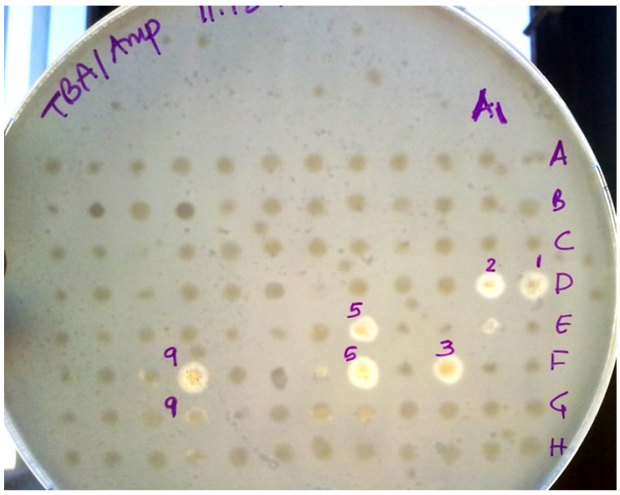

Figure 12: Subcloning of Fosmid 14G2

Fosmid 14G2 (which harbors EstATII) was subcloned into pGEM®-T Easy. The secondary library was tested for lipolytic activity on tributyrin agar. Seven subclones showed positive lipolytic activity and the subclone with the highest activity (14G2F5) was used for characterization of EstATII.

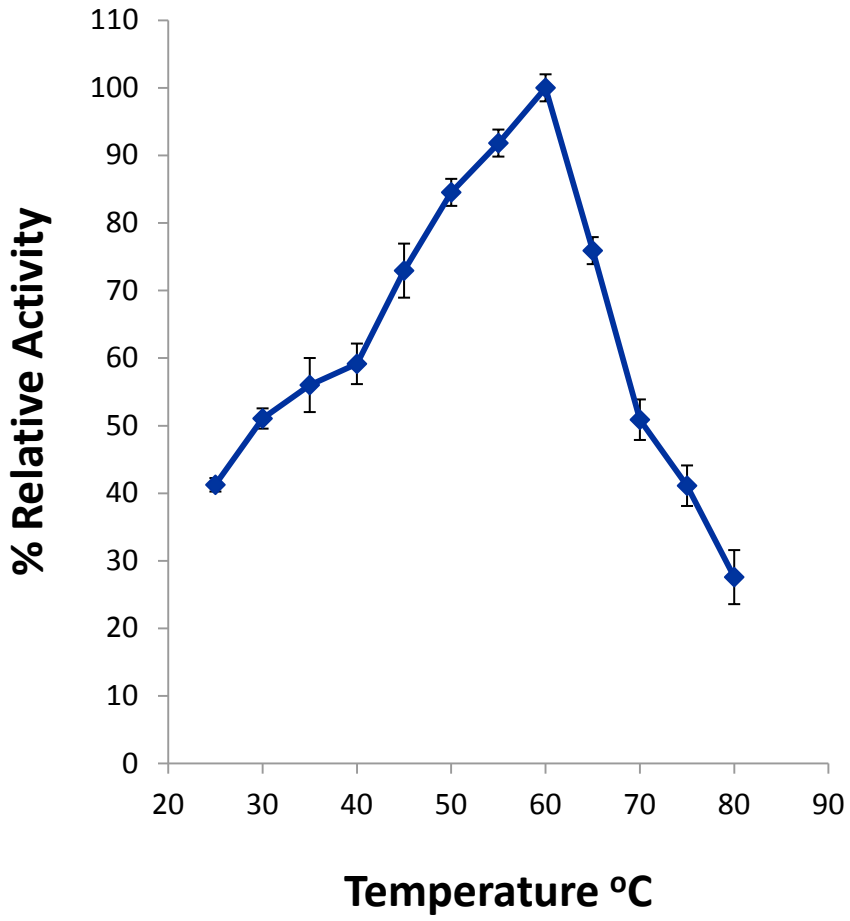

Figure 13: Effect of Temperature on the Activity of EstATII

The effect of temperature on pNPB hydrolysis by EstATII was assayed at temperatures 25°C to 80°C. Activity increases with the increase in temperature till 60°C (apparent optimum temperature). The enzyme exhibits highest activity (>70%) between 45°C and 65°C.

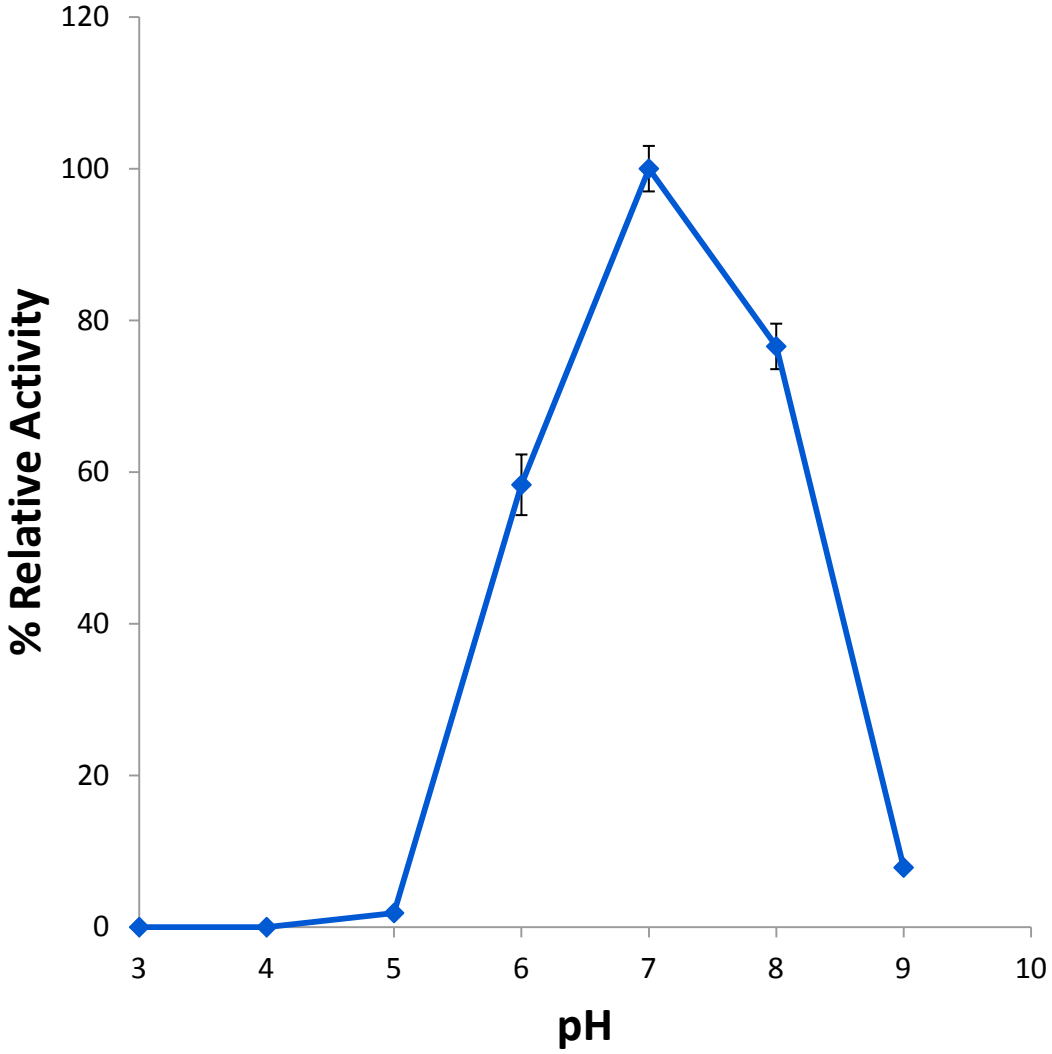

Figure 14: Effect of pH on the Activity of EstATII

The effect of pH on the hydrolysis of pNPB by EstATII was assayed at pH range = 3-9. The enzyme showed significant activity at pH=6-8, and displayed the highest activity at pH=7.

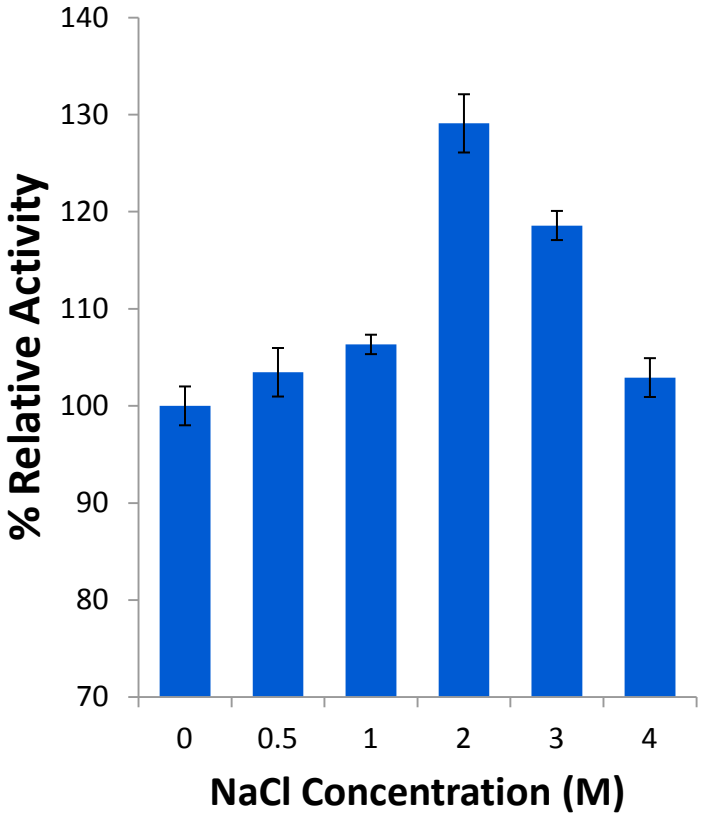

Figure 15: Effect of NaCl Concentration on the Activity of EstATII

The effect of NaCl concentration on the hydrolysis of pNPB by EstATII was assayed at concentrations 0M – 4M. The activity in the absence of NaCl is defined as 100%. Activity in the presence of NaCl (0.5M – 4M) was higher than activity in the absence of NaCl. Highest enzyme activity was seen at 2M NaCl.