Metagenomics of marine and mangrove sediments: Phylogenetic diversity and characterization of amylase obtained by functional screening

Thesis submitted to

Cochin University of Science and Technology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Under the Faculty of Science

By

HARISREE P. NAIR
Reg. No. 4832

Under the Guidance of
Dr. Sarita G. Bhat

MICROBIAL GENETICS LABORATORY

DEPARTMENT OF BIOTECHNOLOGY

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COCHIN - 682022, KERALA, INDIA.

SEPTEMBER 2015

Declaration

I hereby declare that the thesis entitled **"Metagenomics of marine and mangrove sediments: Phylogenetic diversity and characterization of amylase obtained by functional screening"** is the authentic record of research work carried out by me for my doctoral degree, under the supervision and guidance of Dr. Sarita G. Bhat, Professor, Department of Biotechnology, Cochin University of Science and Technology and that no part thereof has previously formed the basis for the award of any degree, diploma, associateship or other similar titles or recognition.

Cochin - 682022 **Harisree P. Nair**
23/09/ 2015

DEPARTMENT OF BIOTECHNOLOGY
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN - 682 022, KERALA, INDIA.
Ph: 0484 – 257667 |Email: saritagbhat@gmail..com | Fax: 91-484-2576267, 2577595

Dr. Sarita G. Bhat
Professor

Date: 23/09/2015

Certificate

This is to certify that the thesis entitled "**Metagenomics of marine and mangrove sediments: Phylogenetic diversity and characterization of amylase obtained by functional screening**" is a record of bonafide research work done by Ms. Harisree P. Nair under my supervision and guidance, in partial fulfilment of the requirement for the degree of Doctor of Philosophy, under the Faculty of Sciences of Cochin University of Science and Technology.

I certify that all the suggestions made by the doctoral committe during her presynopsis are included in the thesis, and that no part thereof has been presented for the award of any degree.

Dr. Sarita G. Bhat
Supervising Guide

Acknowledgement

First and foremost I thank God for all the blessings you showered on me throughout the journey of life making my dreams come true.

I express my heartfelt and profound sense of gratitude to my research guide and mentor, Dr. Sarita G. Bhat, for her encouragement, inspiring guidance, valuable suggestions and wonderful backing throughout the course of my doctoral research. I am taking this opportunity to thank you for accepting me as a Ph.D student. I really thank you for the wonderful support and friendship throughout my research, especially during our eight day cruise through Arabian Sea for sample collection and also during our Indian science congress seminars in Jammu and Mumbai. I am thanking you for giving me all freedom for designing my experiments and for allowing me to follow my own working hours in our lab! All your suggestions and ideas especially our Friday reports helped me a lot to finish my work on time I am truly indebted to you for all the emotional and intellectual support at times when needed. You are the best teacher I have ever seen and I am lucky to have you as my guide. Thanking you once again for making my dream come true.

I express my profound gratitude to Dr. C.S. Paulose for his encouragement and motivation throughout the research period. I thank you for all the valuable suggestions for the review presentations during initial stages of research. I would also like to thank you for admitting me as a Guest lecturer after my Master's degree which served as a stepping stone to my research career. I would also like to thank Dr. M. Chandrasekharan, for the encouragement and valuable suggestions during his visits to India. I specially thank Dr. Hari Krishnan K of Rajiv Gandhi Centre of Biotechnology for introducing me to the world of metagenomics for my MSc dissertation.

I express my gratitude to my doctoral committee members, Dr. Padma Nambisan and Dr. Ammini Joseph for their valuable suggestions and support throughout my research work. The valuable suggestions offered by Dr. Elyas K.K. and Dr. E. Vijayan are gratefully acknowledged. I thank all the teachers of my school days, graduation and post-graduation for laying my foundations especially in biological science.

I am thankful to Cochin University for providing the financial support in the form of Junior Research Fellowship. I also thank University Grants Commission (UGC) for the major research project sanctioned to my guide and proving financial assistance in the form of Junior and Senior Research Fellowship for three years. I thank all the present and past office staffs of the Department of Biotechnology for their prompt support and help. My special thanks to the higher authorities and administrative staffs of Cochin University of Science & Technology for their help and co-operation.

Words fail to express my sincere gratitude to my colleagues at the Microbial Genetics Lab. I am deeply indebted to all my seniors in our lab for providing friendly and motivating environment. First, I would like to thank Dr. Alphonsa Vijaya Joseph, our Vijaya Miss for all the support, care and friendship you offered me. You are one among my strongest pillar of support and the greatest gift I got from CUSAT! You are such a wonderful person I have ever met and all the lovely and fun filled moments spent with you are really cherished. Thank you for all the concerns and enquiries about my work progress even during your busy schedule. I gratefully acknowledge Dr. Siju M Varghese for all the advice during my research career. Thank you for teaching me Sigma Plot and Endnote. I also extend my gratitude to Dr. Jeena Augustine for the encouragement and support throughout my work. My special appreciation to Dr. Raghul Subin for his support and care, and for introducing me to various bioinformatics tools. Indeed he is my evening companion in lab during my initial stages of work; I really thank you for the friendship and brotherly affection. All the silly fights and joyful moments with you are really treasured.

I express my sincere gratitude to Dr. Helvin Vincent, my trainer and the greatest friend of mine, she is moreover like an elder sister to me. Thank you Helvin chechi for making me who I am. I really appreciate your patience in listening all my work plans and helping and supporting me in making all my works a success. You always found time to spend with me even after submitting your thesis, and constantly motivated me throughout my research career by frequent enquires about my work progress. I am grateful for all your support and care and thank you for helping me with all calculations which made my work easy. Thank you for finding time to proof read most of my thesis chapters with such dedication. My heartfelt thanks to Linda Louis, my MSc teacher turned lab mate and my best buddy. I would not have stepped into this department if you have not informed me about the Guest Lecturer vacancy. I am really thankful for your sisterly affection, care and moral support throughout my research. I appreciate you for all the valuable practical suggestions and for the mental support you gave me and I know well that you are my greatest well-wisher and all your prayers helped me a lot. My special thanks to Dr. Smitha S for all the support and care you offered me and your guidance in my protein work need a special mention.

I specially thank Mridula V.G, my lecturer colleague turned lab mate for all the loving support and affection throughout my research. All the fun filled moments with you are really cherished. I sincerely acknowledge Cikesh P.C. for his support and care. Thank you for being a great friend and for all the joyful atmosphere created by you. I express my gratitude to Noble K. Kurian for his endless support during the lab hours. I am greatly indebted to you for all the suggestions you made on my work plans and helping me in clarifying all my silly doubts. My special thanks to Laxmi M for all her support and care.

My sincere gratitude to Rinu madhu Puthusseri or Rinu mon! My trainee turned lab mate and moreover my younger brother. I really thank you for being my right hand in all the experiments right from the time you joined the lab as summer trainee. Your helping hands made my library screening easy and your assistance

during amylase assays requires special mention. I really enjoyed your companionship and thanks for all the mental support offered by you when needed. Thanking you for all the wonderful and joyful lunch time and tea breaks. I am deeply indebted to you for spending almost one week in my thesis work for proof reading and helping me in checking references. My special thanks for designing the thesis cover with such dedication. I express my truthful appreciation to Bindiya E.S for all the support and care throughout my research. Thank you Bindyechi for all the valuable suggestion and for the patience in finding TLC plates and sprayer! My special thanks to Sritha K.S for all your support from the time you joined our lab with a special mention for proof reading my thesis. I thank Nanditha M, Tina K. J and Anu M.A for their friendship and support.

I am thankful to Dr. Manzur Ali P.P and Dr. Sapna K for their valuable practical suggestions in the protein work. I really thank Rekha Mol K.R for her constant support throughout my research. Thanks for sharing all my worries during the last stages. Special thanks to Dr. Abraham Mathew and Ramesh Sir for helpful advices.

My sincere thanks to seniors of the Microbial Technology lab, Dr. Soorej M. Basheer, Dr. Sreeja and Dr. Beena P. S for their suggestions, encouragement and unconditional help during various stages of my research work. Special thanks to Soorejettan for all the practical suggestion and support for MALDI analysis of my protein. I am grateful to Beena chechi for your support for clarifying all my doubts regarding DNA sequencing and helping me with NGS for my diversity studies. I thank Dr. Roseline Alex, Dr. Karthikeyan, Dr. Manjula P, Mr. Ajith M. Thomas, Mr. Sajan and Mr. Doles for all the friendship and support.

I am grateful to all the research scholars of Plant Biotechnology lab for the co-operation and friendship. My special thanks to Soumya S for all the support especially during metal ion and pH studies. My sincere thanks to Arinnina Anto for the friendship and care with special thanks for all the help in clarifying my doubts regarding protein modeling. I thank Kiran Lakshmi, Anala, Aiswarya, Nayana, Anuja, Dr. Jasmine Koshy, Dr. Jikku Jose, Sudha, Aneeta Habeeba and Regi chechi for their encouragement and support throughout the work. My thanks to all the research scholars of the Neuroscience laboratory especially Dr. Anju T. R for the friendship, help and co-operation.

My heartfelt thanks to contract lecturers Dr. Sreekanth, Dr Anoop and Dr. Manjusha for their constant support and valuable suggestions during my presynopsis. I am greatly indebed to Sreekanth Sir for all the suggestions, support and advices offered by you. My sincere thanks to Anoopettan for all the help and support offered by you. Even though you joined here during the final stages of my work your contributions meant a lot. Thank you for being a great friend, and for the joyful tea breaks.

I am grateful to the Captain and all the crew members of FORV Sagar Sampada (Cruise #305) for all the help and support during my sample collection.

Special thanks cruise members and scholars of CMLRE, Rajeesh, Thommy, Vinu and Jijith for the support, care and friendship offered during the cruise which helped me a lot to overcome the sea-sickness. Special thanks to Serene Sara John, Ramya R. Nair and Sanyo Sabu, scholars of NCAAH, CUSAT for the unconditional love and affection toward me. Thank you all for the great companionship during our Jammu and Mumbai trips. All the wonderful and fun filled moments spent with you are really priceless.

Once again special thanks to Sreekanth Sir, Anoopettan, Helvin chechi, Linda Miss, Bindiyechi, Sritha, Rinu, Akhil and Rajesh for proof reading my thesis and helping in cross checking the references. I thankfully remember all the M.Sc. students of the Department especially Amritha, Anjali and Akhil. I am grateful to Dr. Divya, St.Xavier's College for helping me with bioinformatic tool for phylogenetic diversity analysis.

My genuine thanks to Rajesh Nair, my MSc class mate and my best friend. Thank you for being a constant supporter since the time we met. You are good motivator and my greatest strength. You are the one who taught me to face life with a smiling face. Thanks for listening to all my stories especially my endless talks about uncultured microbes! Thank you for all the encouragement, support and care and for all the help with my thesis work from reference formatting to proof reading and you never hesitated to make an ISD call daily to help and support me. I also thank Aswathy N for all her support, encouragement and care in achieving my goals. Thanks for sharing all my tensions during early stages of work and as a research student she can understand it better. I am grateful to you for being such a loving friend throughout my journey.

Words cannot express my gratitude to my dear parents for their love, motivation and prayers. I am deeply indebted to them for their persistent support for accomplishing my goals. Thank you Achan and Amma for having faith in me and providing me complete freedom for following my destiny. Special thanks to my elder sisters Haritha and Hani for the constant support and love since my childhood. My heartfelt love and thanks for the three little angels of our family, Devika, Aardra and Avanthika- my nieces! my Devutty, Appu and Avanthu. Thank you for your little prayers and smiling face which provides me greatest relief from all the tensions and worries. Specially mentioning Appu for her enthusiasm in selecting colors for my phylogeny pie charts and constant enquiries about my thesis writing progress even though she doesn't know the spelling or meaning of it! Last things are always the best, Vipin my better half. Thank you for being such an understanding person; you know better who you are to me. Special thanks to my in-laws for accompanying him in my absence.

Harisree P. Nair

Contents

List of Tables

List of Figures

Abbreviations

%	-	Percentage
~	-	Approximately
<	-	less than
>	-	greater than
°C	-	Degree Celsius
A_{230}	-	Absorbance at 230 nm
A_{260}	-	Absorbance at 260 nm
A_{280}	-	Absorbance at 280 nm
APS	-	Ammonium per sulfate
BAC	-	Bacterial artificial chromosome
BLAST	-	Basic Local Alignment Search Tool
bp	-	Base pair
BSA	-	Bovine serum albumin
cm	-	Centimetre
CTAB	-	Cetyl trimethyl ammonium bromide
Da	-	Dalton
DMSO	-	Dimethyl sulphoxide
DNA	-	Deoxyribonucleic acid
DTT	-	Dithiothreitol
DW	-	Distilled water
e.g.	-	for example
EDTA	-	Ethylene diamine tetra acetic acid
et al.	-	and others
EtBr	-	Ethidium bromide
Fig	-	Figure
g	-	Grams
h	-	Hours
i.e.	-	that is
IEF	-	Isoelectric focusing
IPG	-	Immobilised pH Gradient
IPTG	-	Isopropyl β-D-1-thiogalactopyranoside
kb	-	Kilobase
kDa	-	Kilo Dalton
L	-	Litre
LB	-	Luria Bertani
M	-	Molar
m	-	Metre
Mb	-	Megabases
mg	-	Milligram
min	-	Minutes
mL	-	Millilitre

mm	-	Millimetre
mM	-	Millimolar
N	-	Normality
NaCl	-	Sodium chloride
NaOH	-	Sodium hydroxide
NBS	-	N-bromosuccinimide
NCBI	-	National Center for Biotechnology Information
ng	-	Nanogram
nm	-	Nanometer
No.	-	Number
OD	-	Optical density
ORF	-	Open reading frame
PAGE	-	Polyacrylamide gel electrophoresis
PCR	-	Polymerase Chain Reaction
pH	-	Power of Hydrogen
pI	-	Isoelectric point
rpm	-	Revolutions per minute
rRNA	-	Ribosomal RNA
sec	-	Seconds
S	-	Svedberg
SDS	-	Sodium dodecyl sulphate
sp.	-	Species
SSU	-	Small sub unit
TAE	-	Tris-acetate-EDTA
TE	-	Tris-EDTA
TEMED	-	N-N-N’-N’-Tetramethyl ethylene diamine
UV-VIS	-	Ultraviolet-Visible
V	-	Volts
v/v	-	Volume/volume
w/v	-	Weight/volume
X-gal	-	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
µg	-	Microgram
µL	-	Microlitre
µM	-	Micromolar
µM	-	Micromole
µm	-	Micrometer

A-	Ala-	Alanine
R-	Arg-	Arginine
N-	Asn-	Asparagine
D-	Asp-	Aspartic acid
C-	Cys-	Cysteine
E-	Glu-	Glutamic acid
Q-	Gln-	Glutamine
G-	Gly-	Glycine

H-	His-	Histidine
I-	Ile-	Isoleucine
L-	Leu-	Leucine
K-	Lys-	Lysine
M-	Met-	Methionine
F-	Phe-	Phenyl alanine
P-	Pro-	Proline
S-	Ser-	Serine
T-	Thr-	Threonine
W-	Trp-	Tryptophan
Y-	Tyr-	Tyrosine
V-	Val-	Valine

Chapter 1
INTRODUCTION

Microorganisms are integral to the function of life on Earth, playing key roles in various geological, climatic, geochemical and biological processes. Fossil records have identified that microbial life was present on the planet more than 3,85 billion years ago (Mojzsis *et al.,* 1996). The environmental microbial flora include protozoa, fungi, algae, bacteria, archaea and also viruses. They are essential for the biogeochemical cycles of key elements like carbon, nitrogen, phosphorous, sulphur and oxygen thereby contributing to the global primary production (Azam, 1998). In the history of biological sciences, the significance of prokaryotes has only been recognized in recent times. During 17^{th} century, Antonie van Leeuwenhoek, the founder of microbiology, discovered 'miniscules'. But it took about 200 more years to realize the massive distribution and impact of these microorganisms. It is estimated that the total number of prokaryotic cells on Earth is approximately 4-6 x 10^{30} (Whitman *et al.*, 1998). However, it is estimated that about 95% to 99.9% of microorganisms have not been cultured by standard laboratory techniques (Amann *et al.,* 1995). This diversity represents a huge genetic bounty that may be exploited for the discovery of novel genes, entire metabolic pathways and potentially important end-products. In order to harness these huge genetic diversity in toto, metagenomic methods can be employed.

Metagenomics refers to culture independent methods based on the genomic analysis of microbial DNA that is extracted directly from communities in environmental samples, a process that circumvents traditional culturing methods. The term "metagenomics" was first used by Jo Handelsman and coworkers of the Department of Plant Pathology, University of Wisconsin, Madison, in their study

of natural products from soil microbes (Handelsman *et al.,* 1998).The approach has been also termed as zoolibrary construction (Healy *et al.,* 1995), environmental DNA cloning (Stein *et al.,* 1996), environmental genomics (Beja *et al.,* 2000), recombinant environmental cloning (Courtois *et al.,* 2003) and community genome analyses (Tyson *et al.,* 2004). Direct cloning of DNA from environmental samples was proposed in the mid 80s (Pace *et al.,* 1985). With the advent of modern molecular biology tools it was understood that the uncultured world is much more diverse and also a reservoir of many potent biomolecules. Metagenomics involves DNA isolation from an environmental sample, cloning of the DNA into a vector, and transforming the clones into a host bacterium. Depending on the scientific needs and questions, there are two approaches to metagenomic analysis, viz., sequence based approach and function driven approach.

The diversity of biological species in the metagenome is measured usually through sequence-driven analysis. Sequence based analysis involves sequencing of clones carrying phylogenetic anchors that indicate the taxonomic group to which they belong. Also, random sequencing can be conducted to find a gene of interest which is followed by sequencing of the flanking DNA for any phylogenetic anchors so as to provide a link of phylogeny with the functional gene (Riesenfeld *et al.,* 2004; Hoff *et al.,* 2008). Application of PCR technique reduced the culturing bottleneck for microbial diversity studies. The uncultured majorities were shown to be highly diverse with few representatives distantly related to the culturable ones (Rappé and Giovannoni, 2003). The culture-independent methods for accessing microbial diversity offer many new insights into the uncultured world, but at the same time it cannot invalidate culturing efforts. Discoveries based on phylogenic anchors especially 16S rRNA signatures, prompted even more culturing efforts, in order to more properly describe these key organisms prevailing in a locale (Handelsman, 2004). Genome reconstruction studies from

metagenomic samples can focus the analysis to single organisms or specific clades in a culture-independent way (Venter *et al.,* 2004, Meyerdierks *et al.,* 2010). Such analysis might reveal key parameters necessary for the successful isolation of yet-to-be-cultured organisms. Culture-independent method reported the discovery of an archaeal 16S rRNA gene in a metagenomic library constructed from seawater (Stein *et al.,* 1996). Similarly, bacterial rhodopsins were found in an uncultured γ-Proteobacterium (Beja *et al.,* 2000). These studies proved that marine autotrophs possess a light-driven proton pump based on other pigments than chlorophyll. Further metagenomic studies discovered the high diversity of bacterial proteorhodopsins (Venter *et al.,* 2004).

The use of molecular methods to investigate uncultivated microbes from natural environments has revolutionized our views of microbial biodiversity and ecology in recent years with the development of next generation sequencing platforms. Next-generation sequencing technologies promise cheaper, faster and more accurate sequences with longer read length (Metzker, 2010) allowing better assembly and the application of established *in silico* methods which are designed for longer sequences (Martin *et al.,* 2008, Weber *et al.,* 2011). Unprecedented data obtained with next generation sequencing methods allows high taxonomic resolution providing insights to understand even the rare microbes in different environmental habitats (Gobet *et al.,* 2012; Mandal *et al.,* 2015)

Functional metagenomics includes screening of metagenomic libraries for novel bioactive compounds through heterologous expression (Schloss and Handelsman, 2003). Heterologous gene expression and secretion of the gene product determines the success of the approach. These methods promise to provide new molecules and novel enzymes with diverse functions and enhanced features compared to the enzymes obtained from the culturable microorganisms. The novel biocatalysts obtained by metagenomic approach includes esterase (Henne *et al.,* 2000), nitrilase (DeSantis *et al.,* 2002), protease (Gupta *et al.,*

2002), β-glucosidase (Gabor *et al.,* 2003), amidase (Gabor *et al.,* 2004), β-lactamases (Song *et al.,* 2005), lipase (Ranjan *et al.,* 2005) amylase (Sharma *et al.,* 2010) and monooxygenase (Singh *et al.,* 2010) to name a few. Apart from these enzymes antimicrobial compounds, antibiotics and various pathway genes were identified by the successful applications of metagenomics (Wang *et al.,* 2000; MacNeil *et al.,* 2001; Gillespie *et al.,* 2002).

Study of bacterial diversity of an environment is important for understanding their distribution, community structure and thereby the functioning of the ecosystem. Subsequently it will provide hints about the potential of novel biocatalysts. Soil represents one of the most diverse habitats and is considered to be the largest prokaryotic diversity rich niche. Among the major habitats of the biosphere, the marine realm covers 70% of the Earth's surface providing the largest habitable space for living organisms, particularly microorganisms. Marine microbes survive from the surface waters of the sea to the lower and abyssal depths, and from coastal to the offshore regions. Study of marine microbial biodiversity is of vital importance for understanding the different processes of the ocean, which may present potent microorganisms for screening of novel bioactive compounds (Zhao, 2011; Felczykowska *et al.,* 2012). Oceanic microorganisms are adapted to survive in extreme environmental conditions such as high salinity, low temperature and extreme pressures which in turn will allow them to produce special metabolites to survive. Thus, the general life conditions are reflected into the metabolites they produce, that potentially may be endowed of unique properties.

Mangroves are boundary land-form ecosystems present in tropical and subtropical regions, located in the intersection between the land and the sea. They are highly productive ecosystems with immense ecological values. The majority (60-70%) of the world's tropical and subtropical coastlines are covered with mangrove ecosystems. Mangrove sediments form a unique environment, with

varying salinity and nutrient availability and are predominantly anaerobic. They harbor diverse groups of organisms, including microorganisms with important roles in nutrient cycling and mineralization (Alongi, 2002; Lyimo *et al*., 2009). Exploration of these sediment samples provides an efficient way for the rapid generation of metagenomic libraries, to screen for hitherto unknown enzyme candidates of industrial importance. With the advent of these new techniques the enormous potential of microbial communities are explored well, whereby the conventional chemical processes can be switched to biotechnological routes, requiring the discovery and development of novel enzymes and biocatalysts for application (Johannes and Zhao 2006).

Amylases (EC 3.2.1.) are among the most important hydrolytic enzymes belonging to family 13 (GH-13) of the glycoside hydrolase group of enzymes (Bordbar *et al.,* 2005). It is a widespread group of enzymes which hydrolyses starch molecules to give diverse products, including dextrins and progressively smaller polymers composed of glucose units. Among industrial enzymes, amylases represent one of the largest group and account for 30% of the overall worldwide sale of enzymes (Van Der Maarel *et al.,* 2002). Although amylases can be derived from several sources, including animals, plants and microorganisms, microbial enzymes generally meet industrial demands. These enzymes are of great importance in biotechnology with applications in all starch-based industries ranging from food, fermentation, and paper and textile industries to alcohol production in breweries. With the advent of new tools in biotechnology like PCR, improved cloning strategies etc, the spectrum of amylase applications has expanded into many other fields, such as clinical, medical as well as analytical chemistry (Pandey *et al.,* 2000).

Alpha-amylases (EC 3.2.1.1) are hydrolases acting on α-1,4 glycosidic bonds of starch and other related compounds with the retention of α-anomeric configuration in the products. They are mostly metallo-enzymes, with calcium

ions in the active site, necessary for amylolytic activity and stability. The rate of starch hydrolysis by α-amylase depends on many process conditions, such as temperature, pH, nature of substrate, substrate concentration, enzyme concentration, presence of Ca^{2+} ions and other stabilizing agents. Hence, the diversity of the application creates the need to search for novel α-amylases with novel and improved properties.

A great deal of work has been done on the cloning of α-amylase genes obtained from various cultivable microbial representatives and the expression of the recombinant enzyme in separate host systems. Metagenomic methods also contributed to the discovery of α-amylase with improved properties from diverse habitats. A novel α-amylase has been obtained from a soil metagenome that retained 90% of activity even at low temperature (Sharma *et al.*, 2010). Similarly a thermostable and calcium-dependant amylase was isolated from a soil metagenome, with suggested applications in destarching and baking (Vidya *et al.*, 2011). There is need for thermostable amylases for starch liquefaction, which employs high temperature and metagenomics has proved to be inevitable in unlocking the same from nature. In addition, a cold-adapted alpha-amylase has been reported from a metagenomic library suggesting its usefulness as a detergent enzyme in environmentally friendly, low-temperature laundry processes (Vester *et al.*, 2015)

OBJECTIVES

Marine and mangrove ecosystem are distinctive ecological niche, with a variety of microbes playing important roles in nutrient recycling and other ecological processes, thereby requiring a thorough exploration of their microflora. Similarly Arabian Sea sediments are hotspots for microbial diversity and therefore requires a detailed investigation. Mangalavanam, a conserved mangrove ecosystem located in Kochi, Kerala, India, known as the "Green lungs of Kochi" hitherto not studied for its microbial diversity may be a treasure trove of unexplored microflora. Therefore, by employing metagenomic methods the phylogenetic diversity and the potential for novel biomolecules of both the sediments were explored with the following specific objectives:

1. **Extraction of metagenomic DNA by different methods from various environmental samples including marine and mangrove sediments and their evaluation.**
2. **Phyologenetic diversity analysis of marine and mangrove sediments based on 16S rRNA gene sequence analysis by Sanger and Next generation sequencing.**
3. **Construction of metagenomic libraries to screen for amylase enzyme production.**
4. **Characterization of amylase gene obtained from metagenomic library utilizing bioinformatics approaches.**
5. **Purification and characterization of amylase enzyme obtained from metagenomic clone and its application studies.**

Chapter 2

REVIEW OF LITERATURE

Microbial life dominates all aspects of life on Earth, as they are the most ubiquitous organisms on earth, present in all habitats, including soil, sediment, water and in extreme environments, playing a key role in the biogeochemical cycles of the biosphere and representing a huge reservoir of novel biomolecules Bacteria and archaea are capable of living in all environments and in many cases are the solitary inhabitants of extreme conditions. Microbes are essential to all life forms as they are the primary source for nutrients, and the primary recyclers converting dead matter into available organic form. Estimates reveal that less than 1% of the total microbial communities from the environment are readily cultivable by standard microbiological methods. The unculturable microbes remain uncharacterised, the deficiency of information about their culturing parameters, allowing their continuation as unexplored reservoir of metabolic and genetic diversity. Metagenomic methods help to bypass and overcome the limitations of traditional culturing method, allowing the exploration of the unexplored.

2.1 Metagenomic anlaysis to study the unexplored

The word meta in Greek is “beyond” and genomics refers to the analysis of genomic DNA from an individual organism or cell. So the term literally means “beyond the single genome study”. The principal definition of metagenomics is the investigation of genomic DNA of a total community. The term was first published in the late nineties in a study of soil microbes, using random cloning of environmental DNA (Handelsman *et al.*, 1998). Later on, definitions were modified to include any study whereby a total community is analyzed, e.g., diversity studies targeting 16S rRNA gene of total DNA from environmental

samples without prior cultivation (Chen and Pachter, 2005). So metagenomics is the culture independent analysis of communities which involves the analyisis of total genomes present in an environment that cannot be, or have not been cultured, by isolating the total DNA, cloning it into culturable organisms, followed by bioprospecting and diversity analysis.

In 1898, an Austrian microbiologist Heinrich Winterberg made a inquisitive observation that the number of microbial cells in his samples did not match the number of colonies formed on nutrient media (Winterberg, 1898). A decade later, studies quantified this mismatch suggesting that non-growing cells outnumbered the cultivable ones almost 150 times (Amann, 1911). These findings hinted at some of the earliest steps towards the discovery of an important phenomenon known today as the Great Plate Count Anomaly (Staley and Konopka, 1985). Later on, the phenomenon was repeatedly confirmed by all microbiologists who had compared the initial cell counts to colonies formed in Petri dishes. Celebrating its 117^{th} anniversary, the Great Plate Count Anomaly today is debatably the oldest unresolved microbiological phenomenon. This vast majority of microbes remains uncultured mainly because of the lack of ability to provide appropriate culture conditions, which may be impossible as its hard to mimick all natural milieus. Also, some organisms require interdependence with other organisms in nature and hence they cannot be grown as pure culture in laboratory conditions. Thus, we can only speculate about the economical value, and environmental importance of the majority of organisms that have remained unexplored so far. To a certain extent, metagenomic approach can be helpful for mining the unexplored diversity of different microbial polulations, particularly by removing the previous restriction of cultivation and by allowing equal access to the large population of the community comprising yet-to-be-cultivated members.

2.2 History of metagenomics

The great plate count anomaly pointed that cultured microorganisms did not represent much of the microbial world and for a long time microbiologists ignored the challenge to identify and characterize uncultured organisms. Instead the focus was on the rich source of discovery in the readily cultured organisms, which contributed to the outburst of knowledge in microbial physiology and genetics in the 1960s to mid-1980s. Meanwhile, the research on uncultured microbes remained in the hands of a few scientists who began to accumulate hints that flitted at the boundaries of microbiological realization, suggesting that culturing did not represent the full spectrum of microbial diversity.

During late 1960's attempts were made to culture microorganisms in Yellowstone hotsprings, but they were unable to culture many of the organisms because their temperature requirements exceeded the melting point and their behavior in culture did not reflect their activities *in situ* (Brock, 1967; Bott and Brock, 1969). The *in situ* population size was estimated by immersion of microscope slides in the spring for 1 to 7 days, followed by microscopic examination. Further evidence about the uncultured world was gathered during the 1970s and 1980s. Studies by Whang and Hattori (1988) pointed that incubation times longer than 25 days enhanced the recovery of certain oligotrophic organisms. The concept of viable but not culturable organisms emerged during 1990s from the work of Colwell and colleagues, who showed that strains of *Vibrio cholerae* were virulent and alive when isolated from aquatic environments, but did not grow in culture after passage through a mouse or human intestine (Colwell *et al.,* 1990). Soil bacterial diversity analysis using DNA-DNA reassociation techniques suggested that the complexity of the bacterial DNA in the soil was 100-fold greater than could be accounted by culturing which exceeded all previous estimates (Torsvik *et al.,* 1990). These early discoveries provided persuasive evidence that drew microbiologists to wrestle with the intimidating challenge of devising strategies to access these uncultured organisms.

Later, direct analysis of 5S and 16S rRNA gene sequences in the environment was used to describe microbial diversity in an environmental sample without culturing (Stahl *et al.,* 1985; Pace *et al.,* 1986). The next technical breakthrough was the use of PCR technology and the design of primers to amplify almost the entire gene (Giovannoni *et al.,* 1990). The new technique accelerated the discovery of diverse taxa across the earth (Eden *et al.,* 1991; Schmidt *et al.,* 1991; Barns, *et al.,*1994) revealing that the uncultured majority is highly diverse containing members greatly divergent from the readily cultivable minority. Pace and colleagues (1986) proposed the idea of cloning DNA directly from environmental samples and in 1991, the first such cloning in a phage vector was reported (Schmidt *et al.,* 1991). Later, metagenomic library was constructed with DNA derived from a mixture of organisms enriched on dried grasses in the laboratory. These libraries were referred to as zoolibraries, and clones expressing cellulolytic activity were found in these libraries (Healy *et al.,* 1995). Metagenomic libraries constructed from prokaryotes in seawater laid the groundwork for environmental phylogenies based on 16S rDNA sequences. A 40 kb insert clone carrying 16S rRNA gene was identified in a mteagenomic library indicating that the clone was derived from an archaeon that had never been cultured before (Stein *et al.,* 1996).

Construction of libraries with DNA extracted from soil was delayed due to difficulties associated with maintaining DNA integrity during its extraction and purification from the soil matrix (Handelsman *et al.,* 1998; Krsek and Wellington, 1999; Berry *et al.,* 2003). Rondon *et al.,* (1999) reported cloning of soil metagenome in BAC vector and identified clones with amylase, lipase, nuclease, antibacterial and hemolytic activities. Important information about marine *Archaea* were obtained from marine metagenomic libraries constructed in BAC vector (Beja *et al.,* 2000). In 2001, Beja and colleagues created fosmid libraries to characterize the marine archaeal phylum *Crenarchaeota* from the Antarctic Ocean and from deep waters of the temperate Pacific Ocean. In 2002, environmental

shotgun sequencing was used to show that 200 liters of seawater contains over 5000 different viruses (Breitbart *et al.,* 2002).

2.3 Metagenomics to study microbial diversity in different environments

Metagenomic studies have already been conducted on a broad range of environments, from oceans to humans, to assess the genetic potential as well as the phylogenetic diversity of the microbial communities present using a variety of techniques.

2.3.1 Soil environment

Soil represents the most demanding environmental niche for microorganism. Soil environment harbors an extraordinary high prokaryotic diversity that are largely undescribed beyond the level of ribotyping and serves as a huge resource for natural product discovery. By employing metagenomics approach several landmark studies have been performed in the soil environment. Soil is one of the most challenging environmental sources to evaluate microbial diversity, as several parameters of soil, such as particle size, porosity, permeability, water content, mineral composition, and plant litter, can influence microbial composition. Microbial populations in soil are altered due to changes in the water content and other environmental factors such as pH, temperature and availability of oxygen.

Metagenomic diversity analysis based on 16S rRNA gene in soil revealed the abundance of bacterial members belonging to phylum *Proteobacteria, Actinobacterium, Acidobacteria, Firmicutes* and *Verrucomicrobia* and also large number of the 16S rDNA clones that did not fit into any taxonomic hierarchy, suggesting the possibility of uncultured bacterial species (Janssen, 2006; Delmont *et al.,* 2011a). Microbial composition may vary greatly based on the soil types. For example, the forest soils are dominated by phylum *Acidobacteria* but with fewer β-proteobacterial members. On the other hand, agricultural or farm

soils harbor higher numbers of β-proteobacteria but less *Acidobacteria* (Lim *et al.,* 2005). A recent comparative metgenomic study in different soil types such as desert soil, forest soil, grassland soil, Artic soil and mangrove sediment revealed a total of 63 clades with 11 phyla and 53 genera. Proteobacteria was the most dominated phylum in the microbial community of soil, ≥70% abundance were identified in all soil sites except for desert sample (Xu *et al.,* 2014a).

It was estimated that microbial content in sea is less compared to soil environments, ranging up to 4×10^6/ton of soil, whereas bacteria counts are unlikely to exceed 2×10^6 in the sea (Curtis *et al.,* 2002).

2.3.1.1 Marine sediments

Marine environments are extremely diverse and marine microorganisms are exposed to extreme temperature, pressure, salinity and nutrient availability. Community genomics studies of marine sediments based on 16S rRNA gene has proved the extreme diversity with distinct prokaryotic communities in marine sediments (Rappe *et al.,* 1997; Campbell *et al.,* 2001; Ronquist and Huelsenbeck, 2003). *Proteobacteria* represent the dominant bacterial phylum in marine sediment, at 50% of the total microbial biomass (Bowman *et al.,* 2000; Ravenschlag *et al.,* 2001).

Phylogenteic analysis of various coastal and deep-sea oceanic marine sediments identified the presence of several large and so far uncultivated clades within class *Gammaproteobacteria,* designated as Gammaproteobacterial marine sediment (GMS) clades (Li *et al.,* 1999; Cifuentes *et al.,* 2000). Dot blot hybridization with group-specific oligonucleotide probes on sediments of Arctic Ocean identified the predominance of sequences similar to bacteria involved in sulfur cycle (Ravenschlag *et al.,* 1999). Community profiling in the Pacific Northwest marine sediments revealed the dominance of Betaproteobacterial ammonia oxidizers (Nold *et al.,* 2000). Members of the GMS clades are assumed

to be involved in sulfur cycling in marine sediments, probably as syntrophs with sulfate-reducing bacteria (Bowman *et al.,* 2003). Deep sea sediments of northeastern Pacific Ocean indicated the dominance of bacteria belonging to phylum *Proteobacteria,* with a predominance of class *Gammaproteobacteria* followed by *Alphaproteobacteria* (Hongxiang *et al.,* 2008). Bacterial diversity and community structure analysis in the Oxygen Minimum Zone (OMZ) sediments of the eastern Arabian Sea through 16S rDNA clone library analysis demonstrated of the presence of phylum *Proteobacteria, Planctomycetes, Chloroflexi* and an unidentified bacterial group (Divya *et al.,* 2011)

2.3.1.2 Mangrove sediments

Mangrove ecosystems are unique ecological niches situated at the inter-phase between the marine and terrestrial environment which maintain genetically diverse groups of aquatic and terrestrial organisms. Mangrove sediments are generally nutrient rich with a variety of microbes playing important roles in nutrient recycling and various other ecological processes. Culture independent analysis of bacterial communities in Sundarban mangrove sediment, the world's biggest coastal mangrove forest which extends one million hectares in the south-eastern parts of India and the southern parts of Bangladesh, identified representatives from phylum *Proteobacteria, Flexibacteria, Actinobacteria, Acidobacteria, Firmicutes, Chloroflexi, Gammatimonadates* and *Planctomycetes* (Ghosh *et al.,* 2010). Bacterial diversity analysis in Amazonian mangrove ecosystem identified the abundance of bacteria belonging to phylum *Proteobacteria* (Pureza *et al.,* 2012).

2.3.2 Hot springs

Many hyperthermophilic archaea and bacteria have been isolated from hot spring environments. Research on hot springs in the Yellowstone National Park, USA, revealed the large archaeal diversity with a predominance of members similar to *Crenarchaeota* phylotype. On the basis of environmental studies from

other hydrothermal environments, it is believed that these archaeal members were dominating hot springs (Barns *et al.,* 1994; Barns *et al.,* 1996). An analysis of the bacterial community in the same hot spring sample revealed the remarkable bacterial diversity with twelve new division-level lineages (Hugenholtz *et al.*, 1998a) and members of the bacterial domain outnumbered the Archaea in this hydrothermal environment. It was reported that all the hot springs in close geographical proximity in the Yellowstone National Park had comparable temperatures (between 85°C and 95°C) and pH value (7.8–8.9), but they differed noticeably with regard to overall microbial diversity (Blank *et al.,* 2002). Metagenomic analysis revealed the presence of *Cyanobacteria, Synechococcus*, an inhabitant of microbial mats in hot springs, where ecologically diverse subpopulations have been found (Bhaya *et al.,* 2007). It was also found that *Thermocrinis ruber*, a member of the order *Aquificales* is ubiquitous, in all the springs suggesting that the primary production in these springs was by chemoauxotrophic hydrogen oxidation. Microbial community analysis of the Gedongsongo hot spring, GS-2, revealed the presence of unique strains belonging to *Betaproteobacteria and Firmicutes* (Aminin *et al.,* 2008). Metagenomics based phylogenetic analyses of Kamchatkan hot spring revealed a new distinct phylum within the Archaea, the *Thaumarchaeota* (Eme *et al.,* 2013).

2.3.3 Polar ice caps

Around 13% of the Earth's surface area is covered by sea ice of the polar oceans. These environments are good sites for preservation of biomarkers as they preserve organic matter deposited onto surfaces (Phillips and Parnell, 2006). Polar ice caps are fertile habitat for bacteria, microscopic plants and animals (Thomas and Dieckmann, 2002). Studies on diversity and associated bacteria in Antarctic sea ice revealed several psychrotrophic as well as psychrophilic isolates belonging to the alpha and gamma subdivisions of the phylum *Proteobacteria*, members of *Flexibacter-Bacteroides-Cytophaga* phylum and the Gram-positive branch were obtained. Psychrotrophic strains were noted to be members of the genera

Pseudoalteromonas, Psychrobacter, Halomonas, Pseudomonas, Hyphomonas, Sphingomonas, Arthrobacter, Planococcus, and *Halobacillus*. The psychrophilic isolates identified include *Colwellia, Shewanella, Marinobacter, Planococcus,* and novel phylogenetic lineages adjacent to *Colwellia* and *Alteromonas* and within the *Flexibacter-Bacteroides-Cytophaga* phylum (Bowman *et al.,* 1997). Further studies on bacterial biodiversity of seven Antarctic sea-ice samples and one Arctic sea-ice sample identified several members belonging to proteobacterial phylum and the *Cytophaga-Flavobacterium-Bacteroides* division (Brown and Bowman, 2001). Research on Antarctic cryoconite hole, which is formed when particles on the surface of a glacier are warmed by solar irradiation and melt into the ice revealed several heterotrophic bacterial taxa (Christner *et al.,* 2003). Several phylotypes were obtained from studies on subglacial ice above Lake Vostok, Antarctica (Priscu *et al.,* 1999; Christner *et al.,* 2001) and from the permanent Antarctic lake ice (Gordonrid *et al.,* 2000)

2.3.4 Hypersaline habitats

Hypersaline environments contain elevated concentrations of sodium chloride, in which only archaeal microorganisms can survive. Studies on hypersaline environments suggested the predominance of archaeal halophilic microorganisms (Benlloch *et al.,* 2002: Ochsenreiter *et al.,* 2002). Nevertheless, the overall diversity in these hypersaline environments is generally low (Benlloch *et al.,* 2002). Studies on halophilic bacteria in crystallizer ponds from solar salterns identified a new bacterium, *Salinibacter ruber* to be an important component of the crystallizer microbial community (Antón *et al.,* 2000), that comprise up to 25% of the total prokaryotic diversity of hypersaline environments. Metagenome sequencing of prokaryotic microbiota from two hypersaline ponds in Santa Pola, Spain revealed the abundance of phylum *Euryarchaeota* in both (Fernandez *et al.,* 2013).

2.3.5 Acid mine drainage

Investigations on microbial diversity of non-thermal environments with extreme acidic conditions like acid mine drainage have identified few archaeal members playing key roles in the ecosystem. Nearly complete metagenome sequencing of one most extreme environments on Earth, the Richmond mine revealed the predominance of genera like *Leptospirillum*, *Sulfobacillus*, and sometimes *Acidomicrobium*, and one archaeal species, *Ferroplasma acidarmanus*, and other members of its group, the *Thermoplasmatales* (Bond *et al.*, 2000; Baker and Banfield, 2003). In 2004, Tyson and co-workers employed whole genome shotgun sequencing approach for DNA extracted from a natural acidophilic biofilm of an acid mine drainage system and reported reconstruction of near-complete genomes of *Leptospirillum* group II and *Ferroplasma* type II. Comparative genomics in acid mine drainage biofilm revealed shared characteristics of organisms from the *Thermoplasmatales* lineage providing new insights into the functioning of acid mine drainage communities (Yelton *et al.*, 2013)

2.3.6 Hydrothermal vents

Deep-sea hydrothermal vents are regarded as one of the main habitats for thermophiles since the discovery of hydrothermal vents 1970s (Weiss *et al.*, 1977; Corliss *et al.*, 1979). Research on deep-sea vent sites have identified several thermophilic bacterial and archaeal species (Huber *et al.*, 2000). A molecular analysis on the hydrothermal vent population revealed uncultured archaeal phylotypes (Takai and Horikoshi, 1999). Archaeal diversity studies from different vents on the East Pacific Rise indicated that the microbial composition changed between different sites and over time (Huber *et al.*, 2002; Nercessian *et al.*, 2003). In 2003, Schrenk and co-workers investigated the microbial diversity in the wall of an active deep-sea sulphide chimney and identified observed that the microbial composition varied greatly from the exterior to the interior of the chimney. To recognize the microbial diversity in these deep-sea environments, sampling

procedures and measuring geochemical parameters are crucial (Schrenk *et al.,* 2003). Metagenomic pyrosequencing approach in the Mothra hydrothermal vent field at the Juan de Fuca Ridge revealed the phylogenetic diversity in this location (Xie *et al.,* 2011)

2.3.7 Extreme pH environments

Microbial diversity in extreme pH environments is usually low compared to other environments. Non-thermal, extreme acidic environments are often dominated by different members of the bacterial and archaeal genera (Gonzalez *et al.,* 2003). Molecular analysis in Iron Mountain, California an extremely low pH environment, revealed that about 85% of the microbial population constituted *Ferroplasma* species (Edwards *et al.,* 2000), whereas studies in Tinto River, Spain revealed these Archaea comprised only a small fraction of the identified (Gonzalez *et al.,* 2003). Research in alkaline environments, such as soda lakes, identified the dominant haloalkaliphilic Archaea such as *Thioalcalovibrio, Roseonatronobacter* etc. (Rees *et al.,* 2004; Humayoun *et al.,* 2003)

2.3.8 Gut and skin microbiome

Metagenomics proved to be a powerful tool for studying gut flora from humans to fishes. Gut microbial community are extremely complex and dynamic including archaea, bacteria, viruses and eukaryote. Identifying human intestinal microbial diversity can infer relationship to health and disease. The gut microbial community protects the host against pathogenic microbes and helps regulate metabolic processes. A metagenomic study of the gut microbiome of African and European children revealed that the microbiome of the former was enriched with *Actinobacteria* and *Bacteroidetes* whereas the latter microbiome was enriched with *Firmicutes* and *Proteobacteria,* suggesting that the gut microbial content was influenced by host dietary habits (De Filippo *et al.,* 2010). Metagenomic studies also revealed the differences in microbial composition in Crohn's disease patients

in comparison with healthy individuals (Kang *et al.,* 2010; Pflughoeft and Versalovic, 2012)

Skin acts as a good source of microbes and include both commensals and pathogenic bacteria. Various studies revealed the diversity profile of human skin microbiota (Grice *et al.,* 2008; Foulongne *et al.,* 2012). Bacteria belonging to *Proteobacteria*, such as *Janthinobacterium* species and *Pseudomonas* species were abundant in both human and mice skin biopsies. Other dominant ones are *Actinobacteria* species, such as *Pripionibacteria* species and *Kocuria* species, *Firmicutes*, and *Bacteroidetes* (Courtois *et al.,* 2003; Kong *et al.,* 2012)

2.4 Pioneering projects in metagenomics

Over the past decade metagenomics has benefited the scientific world in various ways by rapidly analyzing changes in microbial communities and identification of new microbial genes from different environments. Metagenomics is rapidly advancing with new techniques and has become a most effective tool in this area of investigation. Few pioneering metagenomics projects are discussed.

2.4.1 Acid Mine Drainage Project

Metagenomic analyses explored the distribution and diversity of metabolic pathways involved in Acid Mine Drainage (AMD). Due to extensive mining activity the sulfide minerals exposed to air undergo oxidation to form acids. These are referred to as AMD. Metagenomic studies investigated the metabolic pathways involved in nitrogen fixation, sulfur oxidation, and iron oxidation thereby understanding the mechanisms by which the microorganisms tolerated the extremely acidic environment (Tyson *et al.,* 2004; Allen and Banfield, 2005; Tyson and Banfield, 2005; Ram *et al*., 2005). Molecular analysis revealed the complexity of the AMD microbiota, with five major microbes including three bacterial and two archaeal species which produced dense biofilm at these sites. Shotgun metagenomic DNA sequencing allowed complete assembly

of two genomes and partial recovery of three others. *In-silico* analysis of the sequences revealed the collective interaction among individual community members and provided evidences of more long-term genetic interaction at the level of recombination and lateral gene transfer. These sequencing and metabolic analysis provided new opening for 'proteogenomic' analysis. Proteins of acid mine drainage biofilm communities were extracted digested with trypsin followed by shotgun mass spectrometry of the fragmented protein. By combining the metagenome and the proteome, a more powerful outcome was obtained, the identification from the *Leptospirillum* group II sequences of a novel acid-stable iron-oxidizing c-type cytochrome with an adsorption maximum wavelength at 579 nm (Cyt_{579}). It was identified that Cyt_{579} acts as the primary iron-oxidizing enzyme in the microbial community and mediates the rate-limiting step in acid production. This study suggested that proteogenomic approach enables quantification of protein production from each ORF and to validate the DNA-derived metabolic model, and identify key processes involved in ecosystem maintenance. The AMD project moved quickly and relatively easily, because of the very low complexity of the microbial assemblage studied (Ram *et al*., 2005).

2.4.2 The Sargasso Sea Metagenomic Survey

The Sargasso Sea is a complex ecosystem compared with the contained acid mine drainage system. Craig Venter and co-workers embarked on one of the largest metagenomic sequencing endeavours conducted to date, in which they sequenced over 1 billion bp and discovered 1.2 million new genes (Venter *et al.,* 2004). They grouped 794,061 genes in a conserved hypothetical protein group. The next group contained 69,718 genes, all of which were involved in energy transduction. Among these 782 genes were rhodopsin-like photoreceptors, resulting in 10-fold increasing the number of sequenced proteorhodopsin genes. Linkage of the rhodopsin genes to genes encoding subunits of RNA polymerase provided phylogenetic affiliations, specifying that the proteorhodopsins were distributed among taxa that were not previously known to contain light-harvesting

functions, including the phylum *Bacteroides*. Initial observation obtained from Sargasso Sea dataset suggested that many of the genomes in the Sargasso Sea contain genes with resemblance to those involved in utilization of polyphosphates and pyrophosphates, which are usually present in phosphate limited ecosystem (Venter *et al.*, 2004). The phosphorus cycle is not clearly understood, and this genomic data collection provides new avenues for discovery of the mechanism underlying phosphorus acquisition and transformation. The sequence dataset provides means to reassemble a number of genomes and to find similarities with previously sequenced genomes. The structures of these genomes will individually and collectively provide further information about nutrient cycling and thereby serving as a gold mine for further analysis.

2.4.3 The Soil-Resistome Project

Most of the clinically used antibiotics were discovered from soil bacteria, so it is possible that the resistance genes also developed from soil. This invites the use of metagenomics to assess the "resistome" or collection of antibiotic-resistance genes in the soil (D'Costa *et al.*, 2006). The soil-resistome project employed a functional metagenomics approach, in which fragments of DNA are cloned from soil, and the clones are screened for expression of antibiotic resistance. This differs from the metagenomic analysis discussed so far in which the genes are recognized by their activity rather than by their sequence. The soil-resistome project involved culturing the clones with the antibiotic of interest; those able to grow in the presence of the antibiotic were retested and selected for further analysis. The soil-resistome project led to isolation of new groups of antibiotic-resistance genes. The strategy utilized cloning of metagenomic DNA from soil in temperate sites with natural vegetation, mixed grassland in Wisconsin, and a boreal forest in central Alaska. The study revealed genes encoding an enzyme called acetyltransferases which provides aminoglycoside resistance. The study also gave clues regarding the genes that encode resistance to β-lactam antibiotics that were phylogenetically distinct from previously identified enzymes.

A gene that encodes an acetyltransferase was also discovered in the Alaskan forest soil, and its closest homologues in the sequence database were the genes discovered in the Wisconsin soil (D'Costa *et al.,* 2006).

2.4.4 The Human-Microbiome Project

Microbial communities play important role in health and disease in various anatomical locations. The Human-Microbiome Project (HMP) was a National Institutes of Health (NIH) funded initiative with the goal of identifying and characterizing microorganisms that are found in association with both diseased and healthy humans. As human microbiota has not fully explored, little knowledge was obtained from the comparisons of germ-free animals suggesting that the gut microbes regulates numerous processes like energy balance, biotransformations and modulates the maturation and activity of the innate and adaptive immune systems (Turnbaugh *et al.,* 2006 ; Samuel and Gordon 2006). In 2006, the first true metagenomic survey of a component of the human microbiota was performed (Gill *et al.,* 2006), which involved sequencing the microbial communities from the colons of two healthy adults. A total of 78 million base pairs of unique DNA sequence were identified. Comparative genomic analysis of human genome and previously sequenced microbial genomes revealed that the gut metagenome is enriched with genes involved in the fermentation and breakdown of otherwise indigestible plant-derived polysaccharides that are components of modern diets and genes involved in detoxification of xenobiotics consumed intentionally and the synthesis of essential amino acids and vitamins suggesting that human metabolome is in fact a composite of human and microbial attributes.

2.4.5 Viral Metagenomics

Metagenomic analyses of marine virus populations have previously provided considerable insight into the viral metagenome. Shotgun viral DNA library construction was employed in studies of naturally occurring phages in aquatic systems, followed by random sequencing. Studies using seawater

confirmed the huge diversity of viral assemblages, suggesting that 65% of the sequences examined from the first seawater viral libraries were novel with no significant similarity to previously known genes in the databases. Greater novelty were obtained from studies on marine sediments with more than 75% of the viral sequences recovered resembling nothing in the databases. Studies on double-stranded viral DNA sequences identified in sediments suggested a significant role for temperate phages, for example, viruses that can integrate into their host's genome and recovered major families, including those with bacterial or algal hosts. Comparative analysis of seawater viral assemblages collected from diverse locales indicated that marine viral species have a global distribution (Angly *et al.*, 2006). Studies targeting RNA-based virus assemblages from seawater have identified new groups of RNA-viruses infecting marine planktonic protists and animals. The early viral metagenomic analyses provided a solid platform for exploring and interpreting the genomic diversity in naturally occurring viruses.

2.5 Soil metagenomic DNA isolation

The key step in metagenomic study is to choose an appropriate method for community DNA extraction from the environmental samples. The type of environment, sampling methods employed, transport and storage of samples until analysis, are the main factors affecting the efficiency of the research methods used in metagenomics. Extracting pure DNA from environmental samples is practically as important as yield; however it is also one of the most complex problems associated with the application of molecular techniques on environmental samples.

Heterogeneous nature of the environmental samples requires each extraction procedure to be precise and optimized for every sample. Most DNA extraction procedures co-extract humic acids, pigments, heavy metals, and other contaminants.

Soil DNA extraction methods are classified as direct (*in situ*) and indirect (*ex situ*) methods. In direct methods, cells are lysed within the soil sample, followed by consequent separation of DNA from cell debris and soil matrix (Ogarm *et al.*, 1987), and indirect method employs cell separation followed by cell lysis, DNA recovery and purification (Holben *et al.,* 1988; Courtois *et al.,* 2001; Robe *et al.,* 2003). These approaches have advantages as well as disadvantages concerning DNA yields, purity for molecular analysis and biased representation of the entire genome (Tsai and Olson, 1991; Courtois *et al.,* 2001). Four key parameters that define the suitability of the metagenomic DNA extracted by each method for subsequent analysis have been identified as yield of DNA, purity, fragment size, and representativeness (Ekkers *et al.,* 2012). On the other hand, these factors often stand in negative relation to one another, increase in one will often have a negative impact on other factors followed by extractions often resulting in either low yield extracts containing large DNA fragment sizes or high yield with small fragmented DNA. A low average fragment size mostly encumbers the subsequent analysis of larger operons, for which larger insert libraries are needed (Williamson *et al.*, 2011).

2.5.1 Direct DNA extraction methods

The direct extraction methods involving *in situ* lysis have been extensively used during the last decade. These methods generally provide the highest DNA yields within acceptable processing time and involve complete *in situ* lysis of all microorganisms within the environmental sample. Microbial cell wall disruption is the initial step, which releases all nucleic acids within the sample to the extraction buffer, followed by separation of the extraction buffer from soil particles and nucleic acids are precipitated from the extraction buffer. This is the most challenging step because of the co-extraction of contaminants such as humic acids, heavy metal ions, and proteins. Components of the extraction buffer greatly determine the expected quantity and purity of the DNA (Robe *et al.,* 2003).

2.5.1.1. Methods of cell lysis

Microbial cell disruption can be accomplished by different methods including thermal, physical, chemical and enzymatic lysis which is used either alone or in combination. Physical treatments such as homogenization, bead-beating, vortexing, sonication (Steffan *et al.,* 1988; Miller *et al.,* 1999; Niemi *et al.,* 2001; Miller, 2001), and thermal treatments such as freezing-thawing, freezing-boiling (Tsai and Olson, 1991; More *et al.,* 1994; Porteous *et al.,* 1997, Kauffmann *et al.,* 2004) aids in destruction of soil structure, giving access to the whole bacterial community which is buried deep within soil microaggregates. These methods are efficient in disruption of vegetative forms, small cells and spores, but they often result in considerable DNA shearing (More *et al.,* 1994). Other physical methods for cell lysis are grinding with liquid nitrogen (Volossiouk *et al.,* 1995), mortar mill grinding (Tebbe and Vahjen, 1993), and microwave thermal shock (Orsini and Romano-Spica, 2001; Lakay *et al.,* 2007).

Most commonly employed physical method is bead beating, which has the advantage of increased DNA yields with longer beating times, but can increase the chances of DNA shearing (Bürgmann *et al.,* 2001). This method can yield DNA of very different sizes, *e.g.* 0.1‑0.5 kb (Picard *et al.,* 1992), 2-5 kb (Gillespie *et al.,* 2005), and even 20 kb (Yeates *et al.,* 1997). Many commercial kits are currently available based on bead beating like Power Soil™ DNA Isolation Kit (MO BIO, USA), FastDNA® SPIN Kit for Soil (MP Biomedicals, USA) and ISOIL Soil DNA Extraction Kit (Nippon Gene, Japan) and ZR Soil Microbe DNA Kit™ (Zymo Research, USA).

Chemical lyses either alone or in combination with physical methods are widely used. It requires preliminary physical processing of the material which allows the extraction lysis buffer to reach the cells imbedded within soil particles. The most common chemical used is sodium dodecyl sulfate (SDS), an anionic detergent that dissolves the hydrophobic part of cell membranes. Detergents have

often been used in combination with heat-treatment and with chelating agents such as EDTA, Chelex 100 (Robe *et al.,* 2003) and various buffers (Krsek and Wellington, 1999). Increasing EDTA concentration can result in higher DNA yields, but it lowers purity. Other chemical agents used are cetyltrimethylammonium bromide (CTAB), that form insoluble complexes with denatured proteins, polysaccharides and cell debris (Saano *et al.,* 1995) and can moderately remove humic acids (Zhou *et al.,* 1996). Polyvinylpolypyrrolidone (PVPP) can also help to remove humic acid contaminants, but it lowers DNA yield, thus it is commonly used during nucleic acids purification step (Krsek and Wellington, 1999).

Enzymatic methods employ the property of various enzymes like lysozyme, proteinase K, pronase and achromopeptidase. They affect DNA in the mildest way and are particularly useful in the case of Gram positive bacteria, which are resistant to physical and chemical methods, and when the size of isolated DNA is of importance. Enzymes can be also used for destroying DNA nucleases, and for removal of RNA. Lysozyme treatment is commonly used and its hydrolytic action on the glycosidic bonds enhances DNA purity (Bruce *et al.,* 1992; Niemi *et al.,* 2001; Rochelle *et al.,* 1992; Tebbe and Vahjen, 1993; Tsai and Olson, 1991). Proteinase K digests contaminating proteins (Niemi *et al.,* 2001; Porteous and Armstrong, 1991; Zhou *et al.,* 1996), whereas another enzyme achromopeptidase improve the lysis of lysozyme resistant bacteria (Simonet *et al.,* 1984).

2.5.2 Indirect DNA extraction methods

The first and the most important step in indirect lysis methods, is to disperse the soil particles as much as possible in order to isolate intact bacterial cells, representing maximum microbial diversity as possible. The next step is cell lysis, followed by extraction and purification of DNA. Dispersion of soil particles can be achieved by both physical and chemical methods. The most common

physical method employed were homogenization, sonication, shaking or a rotating pestle procedure (Robe *et al.,* 2003). This can done using blenders (Bakken and Lindahl, 1995; Faegri *et al.,* 1977; Hardeman and Sjoling, 2007), mild dispersal by shaking (Turpin *et al.,* 1993), sonication (Ramsay, 1984), and rotating pestle (Lindahl and Bakken, 1995). Care has to be taken not to exceed the time of mechanical impact on cells over 18 minutes, as some microorganisms are disintegrated (Lindahl and Bakken, 1995). Another method employed to separate the bacterial cells from the soil matrix is centrifugation based on differences in sedimentation between the individual components of the sample (Robe *et al.,* 2003). The method consists of two successive centrifugations. The first centrifugation performed at low acceleration, serves to remove large pieces of soil followed by a second one, performed at high speed, employing supernatant obtained from the first centrifugation to collect the bacterial sediment. After one cycle it is possible to separate about 10% bacteria present in the soil sample, and according to the authors, this represents the whole biological diversity of the sample (Holben *et al.,* 1988). Subsequent cycles of centrifugation will increase yield accordingly. An alternative method is density gradient centrifugation (Bakken, 1985) to separate bacteria according to their buoyant density. Percoll, metrizamide, Nycodenz (Robe *et al.,* 2003), or sucrose (Pillai *et al.,* 1991) can be used as density gradient medium. Nycodenz gradients are useful as bacteria settle on top of the Nycodenz gradient, and organic and mineral particles of greater density sediment to the bottom of the tube. The efficiency of this method of separation varies from 6 to 50% and depends mainly on the composition of the soil.

After separation of microbial cells, cell lysis results in DNA release, which can be purified. Cesium chloride–Ethidium bromide equilibrium density centrifugation has been successfully used to recover pure DNA of large size, at least 48 kb (Jacobsen and Rasmussen, 1992; Tien *et al.,* 1999). Embedding bacteria in agarose plugs followed by a gentle bacterial lysis helps to recover

DNA fragments of a few hundred kilobases with limited mechanical shearing. This approach can be used to establish Bacterial Artificial Chromosome (BAC) libraries (Diaz-Perez *et al.,* 1997; Brosch *et al.,* 1998; Rondon *et al.,* 1999). An integrated approach combining centrifugation based cell separation from soil particles followed by plug lysis and pulsed field gel electrophoresis (PFGE) has been successfully applied to non-culturable bacteria from environmental samples (Berry *et al.,* 2003; Gillespie *et al.,* 2005). Pure DNA fragments of more than 300 kb size can be recovered using this method. Another method for separating DNA from highly contaminated sample is called synchronous coefficient of drag alteration which applies a rotating dipole and quadruple electric field in an aqueous gel by which DNA is concentrated at a focal point, while contaminants are pushed outwards (Pel *et al.,* 2009).

Metagenomic analysis requires high concentration of DNA for restriction digestion and ligation reactions. Samples that yield less DNA can be processed by pre-amplification methods like multiple displacement amplification using random hexamers and phage phi29 polymerase, which will successfully amplify femtograms of DNA to produce micrograms of products (Lasken, 2009; Eisen, 2007).

Despite choosing direct or indirect methods, nucleic acids obtained are contaminated to different degrees with humic acids, polysaccharides, proteins, minerals, lipids, as well as eukaryotic DNA (Kozdrój, 2010). The choice of lysis method, which in turn depends directly on the soil type, will result in various degree of fragmentation of DNA and affect its quality. Majority of the factors mentioned above tend to hinder molecular techniques like PCR and hybridization, restriction enzymes and ligases (Tebbe and Vahjen, 1993). In order to remove unwanted contamination, additional protocols have to be developed and applied at different steps of extraction and purification of DNA. There is no agreement as to

which method is the most effective one. Many of the protocols appear to be very specific and only effective for the type of soil for which they were developed.

2.5.3 Extraction and purification of nucleic acids

The most common contaminant of DNA isolated from soil is humic acids. Humic contaminants due to their three dimensional structure and functional reactive groups bind with organic compounds (Stevenson, 1976) and are one of the major problems associated with any soil community DNA isolation. The phenolic groups in humic acids denature biological molecules by bonding to amides or are oxidized to form a quinone which covalently bonds to DNA (Young *et al.,* 1993). In addition, due to similar physicochemical properties they easily co-precipitate with nucleic acid (Sharma *et al.,* 2007; Hu *et al.,* 2010). These contaminants may not only hinder PCR, restriction digestion and ligation reactions, but also can degrade the DNA during storage. Humic content also interferes with DNA quantification because they exhibit absorbance at both 260 nm (used to quantitate DNA) and 230 nm (Sharma *et al.,* 2007). An absorbance ratio of 260/230 nm is commonly used to evaluate the purity of metagenomic DNA. Different soil types are characterized by different composition and humic content, so it is necessary to optimize a specific protocol for each soil sample, which is a time-consuming and difficult task (Peršoh *et al.,* 2008).

DNA extraction and purification steps are more or less complex depending on soil structure, quantity of organic matter and other potential enzyme inhibitors employed in molecular reactions (Milling *et al.,* 2005). The primary step following cell lysis is extraction of nucleic acid which is achieved by using organic solvent, either phenol alone or combinations like phenol:chloroform (50:50, v:v) (Ranjan *et al.,* 2005). Isoamyl alcohol is sometimes added to prevent foaming as a combination of phenol: chloroform: isoamyl alcohol (25:24:1, v:v:v) (Zhou *et al.,*1996). Prolonged treatment with phenol is avoided due to its toxicity and its ability to degrade DNA together with other compounds (Porteous and

Armstrong, 1991). Chloroform mixed with phenol is more efficient in denaturing proteins. Hydroxyapatite columns were successfully used for extracting DNA (Ogram *et al.,* 1987; Steffan *et al.,* 1988) and rRNA (Purdy *et al.,* 1996) from soil and sediment samples.

Isolated nucleic acids can be purified by precipitation with potassium/sodium acetate, polyethylene glycol (PEG) (Porteous and Armstrong, 1991; Roose-Amsaleg *et al.,* 2001). The addition of sodium acetate during precipitation improves the purity of the DNA. PEG often replaces isopropanol, since alcohol has the ability to precipitate DNA along with humus (Porteous and Armstrong, 1991). Cesium chloride (CsCl) density gradient centrifugation was also used for purification of nucleic acids (Ogram *et al.,* 1987; Porteous and Armstrong, 1991). Agarose gel electrophoresis assisted by polyvinylpyrrolidone (PVP) facilitates separation of DNA from humic acids (Hugenholtz *et al.,* 1998b; LaMontagne *et al.,* 2002), as PVP retards the phenolic compounds that usually co-migrate with nucleic acids (Lee *et al.,* 1996; Young *et al.,* 1993).

Various gel filtration resins such as sephadex G150 (Kuske *et al.,* 1998) and G200, Sepharose 2B, 4B and 6B, Biogel P100 and P200 (Jackson *et al.,* 1997; Niemi *et al.,* 2001; Miller, 2001) Microspin Sephacryl S-300 and S-400 columns were used to purify crude DNA extracts contaminated with humic acids (Edgcomb *et al.,* 1999; Frostegard *et al.,* 1999). Other commercial purification products employed include Wizard DNA clean-up system (Promega) (Henne *et al.,*1999) and Centricon™ 50 and Microcon™ 100 concentrators (Amicon) (Zhou *et al.,* 1996), Elutip™ D column from Schleicher and Schuell (Degrange and Bardin,1995; Frostegard *et al.,* 1999), silica-based DNA binding SpinBind Columns from FMC BioProducts (Miller *et al.,* 1999) and Tip-100 and Tip-500 columns from Qiagen (Hurt *et al.,* 2001; Tebbe and Vahjen, 1993) and MoBio UltraClean soil DNA isolation kit (Delmont *et al.,* 2011b). Magnetic capture hybridization (MCH) approach successfully removes the PCR-inhibitory effect of

humic acids (Jacobsen, 1995). MCH separates a specific DNA target from all other DNA and contaminants, including humic acids. Although recovery of large DNA fragments (40–90 kb) using gentle lysozyme–SDS-based methods was reported (Krsek and Wellington, 1999).

By employing "ready to use" DNA extraction and purification kits, different types of soil samples can be processed to get relatively pure DNA in short time. Currently we are witness to progressive efforts for the improvement of DNA purification methods from environmental samples. Still, there is around 50% DNA loss employing these methods (Carrigg *et al.*, 2007). Therefore it is very important not only to choose an appropriate lysis method, but also a suitable extraction and purification method to obtain high quality DNA.

2.6 Strategies for metagenomic analysis

Metagenomic analysis involves two main strategies, function-based and sequence-based approach. The choice of screening method depends on factors like type of library constructed, functional activity of interest, availability of resources and time to characterize the library. Once metagenomic libraries are constructed it can be screened for phylogenetic markers such as 16S rRNA gene, *recA* or for other conserved genes by multiplex PCR or hybridization or for expression of specific traits, such as enzyme production, antibiotic production etc, or they can be sequenced randomly. Altogether these approaches have enriched our understanding about the uncultured world, providing insight into prokaryotic diversity that is otherwise entirely unknown.

2.6.1 Function-based screening (Metagenome expression libraries)

Metagenome expression libraries are constructed by ligating fragmented metagenomic DNA into expression vectors such as plasmids, cosmids, fosmids and Bacterial Artificial Chromosome (BAC) and screened for gene expression in a

suitable host system**.** Function-based metagenome library screening has uncovered a range of novel biomolecules.

For small insert libraries, DNA fragment of sizes ranging from 1 and 10 kilobase (kb) were constructed in plasmids or lambda vectors, and then screened for phenotypic expression. Larger gene clusters, preferentially require expression libraries with inserts between 20 and 40 kb in fosmid and cosmid vectors, and up to 100 to 200 kb in BAC vectors. Although common *E. coli* host strains used for transformation have relaxed requirements for promoter recognition and translation initiation, many genes from environmental samples may not be expressed efficiently in heterologous hosts due to differences in transcription and translation initiation signals, codon usage, protein-folding elements, post-translational modifications, such as glycosylation, or toxicity of the active enzyme. To circumvent this, selecting suitable vector systems containing appropriate transcription and translation initiation sequences, and suitable expression hosts systems, such as the *E. coli* Rosetta strains is pertinent. Alternate host systems such as insect cells, the yeast *Pichia pastoris*, and bacterial hosts such as, *Streptomyces lividans*, *Bacillus subtilis* or *Pseudomonas putida* were suitably enhanced for heterologous gene expression (Martinez *et al.,* 2004). Direct selection of positive clones that has acquired resistance to an antibiotic or heavy metal can be achieved by excluding microorganisms that are incapable of growth in the presence of these selective compounds (Riesenfeld *et al.,* 2004; Mirete *et al.,* 2007; Parsley *et al.,* 2010).

Metagenomic expression libraries are powerful tools to identify novel natural products or metabolic activities from the yet-to-be-cultured organisms. It is often limited by a number of obstacles that can be overcome by suitability of vector and host systems. Table 2.1 summarizes the different enzymes discovered via function-based screening using different vector systems and from different samples.

Table 2.1: Enzymes obtained from function-based metagenomic libraries

Enzyme	Source	Vector	Reference
Amylase	Soil	Lambda	Richardson *et al.*, 2002
Amylase	Soil	Cosmid	Voget *et al.*, 2003
Amylase	Garden soil	Plasmid	Yun *et al.*, 2004
Amylase	Fecal sample	Fosmid	Xu *et al.*, 2014b
Cellulase	Water sample	Lambda	Rees *et al.*, 2003
Cellulase	Soil	Cosmid	Grant *et al.*, 2004
Cellulase	Soil	Cosmid	Voget *et al.*, 2006
Esterase	Water sample	Cosmid	Elend *et al.*, 2006
Esterase	Rumen microflora	Lambda	Ferrer *et al.*, 2005
Esterase	Soil	Fosmid	Kim *et al.*, 2006
Endoglucanase	Rice straw compost	Lambda	Yeh *et al.*, 2013
Lipase	Marine sediment	Plasmid	Peng *et al.*, 2014
Protease	Soil	Plasmid	Neveu *et al.*, 2011
Protease	Marine sediment	Fosmid	Biver *et al.*, 2013
Pectinase	Soil	Plasmid	Wang *et al.*, 2014
Xylanase	Insect gut	Lambda	Brennan *et al.*, 2004
Xylanase	Manure waste	Lambda	Lee *et al.*, 2006
Xylanase	Soil	BAC	Gong *et al.*, 2013

2.6.1.1 Screening of metagenomic libraries

Success of function-based approach depends greatly on the screening methods because gene expression may not be easily detectable if the screening method is not highly sensitive.

Screening of phenotypic traits can be achieved by assaying individually for a particular trait. Supplementing growth media with specific substrates will allow the screening and identification of enzymatic activity encoded by a metagenomic clone. Examples include the identification of amylase by screening in medium supplemented with soluble starch (Yun *et al.*, 2004; Sharma *et al.*, 2010; Vidya *et al.*, 2011; Xu *et al.*, 2014b) esterases by formation of a clear halo around a colony on the medium containing tributyrin (Elend *et al.*, 2006; Chu *et*

al., 2008). Antimicrobial activity exhibited by metagenomic clones may be detected by growth inhibition assays of suitable tester organisms using soft agar overlays over the colonies or a microtiter plate assay using the culture extracts from the clone cultures (Courtois *et al.,* 2003; Brady *et al.,* 2004; Craig *et al.,* 2009).

Another method adopted for functional screening is substrate-induced gene expression screening (SIGEX), in which the metagenomic DNA is cloned upstream of a promoterless Green Fluorescent Protein (GFP). This method allows for detection of promoters induced by the conditions applied after which, cells can be sorted employing fluorescence-activated cell sorting (FACS). SIGEX has been used to identify genes induced by aromatic-hydrocarbon compounds in a groundwater metagenome library constructed in *E. coli* (Uchiyama *et al.*, 2005). Alternatively, product-induced gene expression screening (PIGEX) can also be adopted, where a reporter strain with a product-sensitive promoter coupled to GFP is co-cultivated with the metagenomic library which facilitates detection of product formation by fluorescence. PIGEX has been used to identify amidase activities in activated sludge metagenome library in *E. coli* (Uchiyama and Miyazaki 2010).

2.6.2 Sequence-based screening (Metagenome sequencing)

Sequence-based approach involves direct sequencing of metagenomic DNA, either with or without cloning prior to sequencing, followed by bioinformatics based analyses (Kunin *et al.,* 2008; Sleator *et al.,* 2008). Clones containing phylogenetic anchors can be completely sequenced which indicates the taxonomic group contained within a sample. Alternatively, random sequencing can be employed, and once a gene of interest is recognized, phylogenetic markers can be sought in the flanking region to provide phylogenetic link with the functional gene. Sequence analysis based on the identification of phylogenetic anchors is a powerful approach first proposed in mid 90s, which produced the first

genomic sequence linked to a 16S rRNA gene of an uncultured archaeon (Stein *et al.,* 1996).

An alternative to a phylogenetic marker-driven approach is random sequencing of clones, which will produce remarkable insights when conducted on large scales. Community distribution, genomic organization, linkage of traits and horizontal gene transfer can all be inferred from sequence-based analysis. Pioneering sequencing efforts, which include reconstruction of the genomes of uncultured organisms in the Sargasso Sea (Venter *et al.,* 2004) and in acid mine drainage (Tyson *et al.,* 2004) illustrated the power of large-scale sequencing efforts to enrich our understanding about uncultured bacterial world. These studies have made new linkages between functions and phylogeny, specifying the unexpected abundance of certain types of genes and reconstructed the genomes of organisms that have not been cultured.

The limitation of this approach based on phylogenetic markers is that smaller number of markers available that provide reliable placement in the Tree of Life. If a DNA fragment with a desired trait for other reasons does not carry a dependable marker, its organism of origin remains unknown. The collection of phylogenetic markers is growing, and as the diversity of markers increases, the power of this approach will also increase, making it possible to assign more fragments of anonymous DNA to the organisms from which they were isolated (Handelsman, 2004).

Several methods including shotgun sequencing have been extensively used in metagenomic studies (Metzker, 2010) in which metagenomic DNA is randomly sheared, sequenced in short fragments, and reconstructed into consensus sequences. Over the past 10 years metagenomic shotgun sequencing has progressively shifted from classical Sanger sequencing technology to next-generation sequencing (NGS). However, Sanger sequencing is still considered to

be a gold standard for sequencing because of its high read length exceeding 700 bp and low error rate. All of these aspects improved outcomes for shotgun data, and hence Sanger sequencing can still be applicable in generating nearly complete genomes in low-diversity environments (Goltsman *et al.,* 2009). A drawback of Sanger sequencing is the labor intensive cloning process and overall cost. Recent advancements in sequencing technologies have made available a number of techniques that can be used for sequencing, although with varying costs and capabilities.

2.6.2.1 Next-Generation sequencing (NGS)

Microbial communities are known to be extremely diverse. This knowledge is primarily based on recent advances in DNA sequencing technology, which has made possible the generation of millions of sequence reads rapidly and cost-effectively. Initially, high-throughput sequencing technology employed polymerase chain reaction (PCR)-based approach, that focused on phylogenetically informative ribosomal gene that can be used to infer the taxonomic diversity and composition of soil microbial communities (Roesch *et al.*, 2007). This method known as pyrotagged sequencing which incorporates barcoded primers or tags (Hamady *et al.,* 2008), has rapidly became the standard approach for understanding the soil microbiome. Apart form phylogetic anchors, pyrotagging approach has subsequently been extended to include targeted functional genes, such as *nif*H, *amo*A etc. (Lovell *et al.,* 2011). Targeted pyrosequencing approach can focus on genes or group of genes of specific interest so that interactions among complex community members or their collective response to environmental changes can be studied.

Among NGS technologies, both the 454/Roche and the Illumina/Solexa systems have now been widely applied to metagenomic samples. 454/Roche sequencing method employs emulsion polymerase chain reaction (ePCR) to clonally amplify random DNA fragments that are then attached to microscopic

beads. The beads along with the attached DNA fragments are deposited onto picotitre plates followed by individual and parallel pyrosequencing. The pyrosequencing process is accomplished by the sequential addition of four deoxynucleoside triphosphates (dNTP), which is incorporated by DNA polymerase. This polymerization reaction releases pyrophosphate, which is converted via luciferase-based light reaction to produce light. Light produced is detected by charge-coupled device (CCD) camera and converted to the sequence of template DNA. The 454/Roche technology produces an average read length between 600-800 bp which makes it a popular choice for shotgun-sequencing metagenomics. Furthermore, the 454/Roche sequencing platform allows multiplexing so that up to 12 samples can be analyzed in a single run.

The Illumina/Solexa technology immobilizes random DNA fragments on a surface and then solid-surface PCR amplification is performed, resulting in clusters of identical DNA fragments. These are then sequenced using reversible terminators in a sequencing-by-synthesis process (Bentley *et al.,* 2008). Illumina technology can now sequence 250 bp paired ends, thereby examine 500 bases at a time employing both Illumina HiSeq and MiSeq platforms. The lower costs of this technology and recent success in its application to metagenomics makes the Illumina technology an increasingly popular choice.

The ideal region to be sequenced in metagenomic surveys of environmental bacterial communities should be whole 16S ribosomal RNA gene (16S rRNA) which is a component of the 30S small subunit of prokaryotic ribosomes and is 1542 bp in length. However, due to the read length limitation in NGS procedures, most analysis are aimed at selected hyper-variable regions of the 16S rRNA gene as they constitute optimal species molecular markers (Carpi *et al.,* 2011; Shah *et al.,* 2011; Lazarevic *et al.,* 2012). Analysis of the primary and secondary structures of the 16S rRNA gene demonstrates nine hyper-variable regions flanked by relatively conserved regions (Figure 2.1). This property makes

hyper variable regions of 16S rRNA gene the finest species molecular marker (Jeraldo *et al.,* 2011). Comparitive studies of sequences of hyper-variable regions recommended designing oligonucleotide primers for V1-V4 region for bacteria and V4-V7 region for Archaea (Chevreux *et al.*, 1999; Kim *et al.*, 2011)

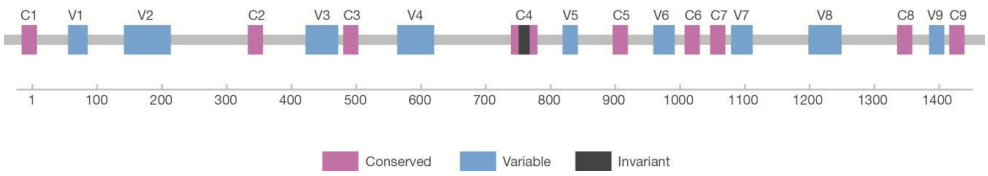

Fig 2.1: Schematic representation of the 16S rRNA gene

Location of variable (blue) and conserved (purple) regions in a canonical bacterial 16S rRNA. The grey region is invariant in all bacteria (Adapted from Illumina, 2012).

Sequencing by synthesis method is employed in both 454/Roche and the Illumina/Solexa technologies. High quality sequences were sorted out from raw reads employing software tools. Read length is the major concern regarding NGS methods, as low read lengths decreases the possibility of distinguishing distant homologs of microbial genes (Wommack *et al.*, 2008). None of the NGS platforms allowed full length coverage of the 16S rRNA gene on a single run, so emphasis has been given in identifying hyper-variable regions which are most useful and specific in species identification (Wang *et al.*, 2007). Comparing to other NSG platforms, Roche 454 based pyrosequencing became the platform of choice in metagenomic analysis because of longer read lengths (Ley *et al*., 2005; Lauber *et al*., 2009) and is often considered the most resourceful of the Next Generation Sequencing technology, permitting sequencing of large genomes and metagenomes. Major differences among the two platforms is that Illumina platform has much more sequencing depth (upto 40 Gb), while having read length of about 200bp (Luo *et al*., 2012); comparatively Roche 454 produces long reads of greater than 400 bp (Schuster, 2007) with a sequencing depth of about 400 Mb. A reduction in the sequencing cost has made metagenomic analysis more focused

on functional diversity and phylogentics microbial community. Major challenge involves obtaining accurate identification of hundreds or thousands of species in a reasonable amount of time and cost, and also requires adequate computational power.

2.7 Power of Bioinformatics in metagenomic analysis

At the dawn of next generation sequencing, reads obtained by Sanger sequencing and the data sets were small. However, data set sizes have increased by orders of magnitude by the introduction of new sequencing technology platforms. Thus, sequence similarity searches can today only be effectively handled through computational toolkits like Mothur, QIIME (Quantitative insights into microbial ecology) and OBITools (Schloss *et al.*, 2009; Caporaso *et al.*, 2010; Boyer *et al.*, 2015). A list of commonly used tools for metagenomic data analysis is given in Table 2.2. Basic processing steps in such toolkits include sequence trimming, screening and aligning sequences against a database, clustering of sequences into operational taxonomic units (OTUs) and comparative sequence analysis between different samples. The alignment of reads to the reference database is the most important step of the workflow. Different programs can be selected for this task, such as UCLUST (Edgar, 2010) CD-HIT (Huang *et al.*, 2010) and BLAST (Altschul *et al.*, 1990). Taxonomy is assigned to the aligned sequence using a predefined taxonomy map against a reference sequence set. Methods such as obiclean from OBITools detect PCR sequencing errors to avoid incorrect taxonomic assignations by the use of clustering algorithms. Based on the results of the reference database comparison, taxonomy assignation can be performed using alignment-based methods employing MEGAN (Huson *et al.*, 2011) and QIIME. Taxonomy is assigned against specific barcode loci databases, whether single loci such as 16S rRNA gene or a set of a few phylogenetic marker loci drawn from across the genome.

Table 2.2: Bioinformatic resources employed in metagenomic data analysis

Resources	Function	Reference
ChimeraSlayer	Chimera detection	Haas *et al.,* 2011
DECIPHER	Chimera detection	Wright *et al.,* 2012
UCHIME	Chimera detection	Edgar *et al.*, 2011
Denoiser	Denoising	Reeder and Knight, 2010
DADA	Denoising	Rosen *et al.,* 2012
Pyronoise	Denoising	Quince *et al.,* 2011
UCLUST	OTU clustering	Edgar, 2010
ESPRIT-Tree	OTU clustering	Fu *et al.,* 2012
CD-HIT-OTU	OTU clustering	Cai and Sun, 2011
RDP	16S database	Cole *et al.,* 2009
Greengenes	16S database	DeSantis *et al.,* 2006
SILVA	rRNA database	Quast *et al.,* 2013
Mothur	All in one	Schloss *et al.,* 2009
QIIME	All in one	Caporaso *et al.,* 2010
MEGAN	All in one	Huson *et al.,* 2011
MG-RAST	All in one	Meyer *et al.,* 2008

2.7.1 Ribosomal Database Project (http://rdp.cme.msu.edu/)

The Ribosomal Database Project (RDP) provides information related to ribosome including online data analysis; and aligned and annotated bacterial and archaeal small-subunit 16S rRNA sequences. RDP retrieves most of its rRNA sequences from the International Nucleotide Sequence Database Collaboration (INSDC). Many RDP tools are available as open-source stand-alone packages and is used for searching data collections for taxonomic classification and nearest neighbor search, for primer-probe testing, to examine statistical differences between a pair of sample libraries, tree builder for construction of phylogenetic tree and tools to align sequences. Moreover, RDPipeline tools are specifically designed for processing high volume of sequence data, including tools for assembly, quality filtering, taxonomy based analysis and taxonomy independent analysis tools, and tools to convert the data to formats suitable for common ecological and statistical packages. One of the most commonly used tools hosted

by RDP to classify sequence reads to taxonomic hierarchy is the Naïve Bayesian Classifier tool (Wang *et al.,* 2007).

2.7.2 MG-RAST - Metagenomics RAST (http://metagenomics.anl.gov/)

In 2007, Folker Meyer and Robert Edwards and a team at Argonne National Laboratory and the University of Chicago released the Metagenomics RAST server (MG-RAST) a community resource for metagenome data set analysis (Meyer *et al.,* 2008). MG-RAST server is an automated analysis platform for metagenomic data analysis providing quantitative insights into microbial populations based on sequence data. The pipeline is implemented in Perl by employing a number of open source components, including the SEED framework (Overbeek *et al.,* 2005), NCBI BLAST (Altschul *et al.,* 1997), and Sun Grid Engine as components. The server serves a platform for web based upload, quality control, automated annotation and analysis for sequence. The pipeline generates automated functional assignments of sequences by comparing both protein and nucleotide databases. Phylogenetic and functional summaries of the metagenomes are generated, and tools for comparative metagenomics are enabled via pre-computed abundance profiles. User access is controlled to ensure data privacy, but the collaborative environment underpinning the service provides a framework for sharing datasets between multiple users. The service is available to all users after simple registration process. All results are available for download in a variety of formats, including GFF3, GenBank, and flat text formats. The server is made available on a best-effort basis, and all underlying data and software are open source.

2.7.3 QIIME (Quantitative Insights Into Microbial Ecology) (http://qiime.org/)

QIIME is an open source bioinformatics pipeline for analyzing raw DNA sequencing data from microbial communities, generated based on high-throughput amplicon sequencing such as Illumina or other platforms for interpretation and

database deposition. This includes demultiplexing and quality filtering, sequence alignment, OTU picking, taxonomic assignment, inferring phylogenetic trees and phylogenetic and taxon-based analysis of diversity within and between samples and diversity analyses and visualizations. QIIME has been applied to studies based on billions of sequences from tens of thousands of samples thereby acting as an ideal platform for combining heterogeneous experimental datasets and for obtaining new insights about various microbial communities rapidly. Because QIIME scales to millions of sequences, it can be used on platforms from laptops to high-performance computing clusters (Caporaso *et al.*, 2010)

2.7.4 Other tools commonly employed

Clustal is a widely used computer program for multiple sequence alignment of protein and nucleic acid sequences. Three main variations are available, ClustalW, ClustalX and Clustal Omega . ClustalW is a command line interface whereas ClustalX has a graphical user interface (Larkin *et al*., 2007). Clustal Omega is the latest version which offers a significant increase in scalability over previous versions, allowing faster processing of hundreds of thousands of sequences. In addition to clustal, BioEdit Sequence Alignment Editor provides basic functions for protein and nucleic sequence editing, alignment, manipulation and analysis. BioEdit offers many quick and easy functions for sequence editing, annotation and manipulation, as well as a few links to external sequence analysis programs (Hall, 1999).

MEGA (Molecular Evolutionary Genetics Analysis) software is a commonly employed tool for statistical analyses of DNA and protein sequence data from an evolutionary perspective (Tamura *et al.*, 2011). It contains tools for sequence alignment, phylogenetic tree construction, phylogeny visualization, sequence divergences estimation, testing evolutionary hypotheses, web-based acquisition of sequence data, and expert systems to create natural language descriptions of the analysis methods and data chosen by the user (Kumar *et al*.,

1994; Kumar and Dudley 2007; Kumar *et al*., 2008). MEGA automatically infers the evolutionary tree by the Neighbour-Joining (NJ) algorithm that uses a matrix of pairwise distances estimated under the Jones-Thornton-Taylor (JTT) model for amino acid sequences or the Tajima and Nei model for nucleotide sequences (Jones *et al*., 1992; Tajima and Nei, 1982; Tajima and Nei, 1984). The principle of this method is to find pairs of operational taxonomic units (OTUs) that minimize the total branch length at each stage of clustering of OTUs starting with a star like tree (Saitou and Nei, 1987)

Several other computational tools used to analyze microbial ecology data are also available including MOTHUR (Schloss *et al*., 2009), DOTUR (Schloss and Handelsman, 2005), SONS (Schloss and Handelsman, 2006a), ARB (Ludwig *et al*., 2004), LIBSHUFF (Schloss *et al*., 2004), UniFrac (Lozupone and Knight, 2005; Lozupone *et al*., 2006), AMOVA and HOMOVA (Martin, 2002; Schloss, 2008), TreeClimber (Schloss and Handelsman, 2006b) and rRNA-specific databases (DeSantis *et al*., 2006; Pruesse *et al*., 2007).

2.8 Amylases

Amylases are enzymes which catalyze the hydrolysis of starch molecules to produce dextrins and progressively smaller polymers composed of glucose units (Windish and Mhatre 1965). They belong to glycoside hydrolase (GH) group of enzymes and can be derived from several sources, including animals, plants, and microorganisms. However, microbial amylases generally meet industrial demands. Today a large number of microbial amylases are available commercially. These enzymes account for about 30% of the world's enzyme production. Microbial amylases are classified into different types with their respective EC numbers (IUBMB Enzyme Nomenclature) based on their catalytic properties, mode of action, specificities for substrate and product etc.

2.8.1 Types of Amylase

Based on the action mechanism, amylases are mainly classified into two groups: retaining and inverting glycoside hydrolases**.** α-amylase (EC 3.2.1.1) are retaining enzymes, in which the anomeric configuration in the substrate is retained after the catalytic action, whereas it is inverted after the catalytic action in inverting enzymes. β-Amylase (EC 3.2.1.2) and glucoamylase (EC 3.2.1.3) are inverting enzymes. Currently, all glycoside hydrolases (GH) are classified into 133 GH families till date (CAZy database at http://www.cazy.org/CAZY), and all microbial amylases are classified into 5 GH families: GH 13, 14, 15, 31, and 77. There are basically four groups of starch-converting enzymes: (i) exoamylases (ii) endoamylases (iii) debranching enzymes and (iv) transferases.

2.8.1.1 Exoamylases

Exoamylases are 'saccharifying' enzymes that cleave α-1,4 glucosidic bonds in amylose, amylopectin and glycogen from the non-reducing end by successive removal of sugars in the form of maltose or glucose in a stepwise manner. The products of hydrolysis have the β-configuration at the C1 of the reducing glucose unit due to inversion of the product. In contrast to the action of endoamylases, action of exoamylases results in slow decrease in viscosity and iodine staining power of starch. Cereal and bacterial β-amylases as well as fungal glucoamylases come under this category (Banks and Greenwood, 1977).

2.8.1.2 Endoamylases

Endoamylases are α-amylases which cleave α-1,4 glycosidic bonds in amylose, amylopectin and related polysaccharides such as glycogen and they are also known as 'liquefying' enzymes. The products of hydrolysis are oligosaccharides of varying chain lengths having the α-configuration at the C1 of the reducing sugar unit, hence the name α-amylase. As the name suggests, endoamylases hydrolyze the bonds located in the inner regions of the substrate

resulting in rapid decrease in viscosity of the starch solution as well as decreased iodine staining power (Hill and McGregor, 1987)

2.8.1.3 Debranching enzymes

The branch points containing α-1,6 glycosidic linkages are resistant to attack by exo and endo amylases resulting in α/β limit dextrins. Debranching enzymes catalyze the hydrolysis of α-1,6 glycosidic bonds. They include isoamylase (EC 3.2.1.68) and pullanase type I (EC 3.2.1.41). Pullulanases catalyze the hydrolysis of α-1,6 glycosidic linkages in amylopectin and pullulan, a polysaccharide with a repeating unit of maltotriose that is α,1-6 linked (Bender *et al.,* 1959; Israilides *et al.,* 1999). Isoamylase can only hydrolyze the α-1,6 bond in amylopectin. These enzymes exclusively degrade amylopectin, thus leaving long linear polysaccharides.

Certain type of pullulanase enzymes can hydrolyze both α-1,4 and α-1,6 glycosidic linkages. They belong to group II pullulanase and are referred to as amylopullulanase or α-amylase–pullulanase. The main degradation products are maltose and maltotriose. A special enzyme belonging to this group of pullulanases is neopullulanase, which can also perform transglycosylation with the formation of a new α-1,4 or α-1,6 glycosidic bond (Takata *et al.,* 1992)

2.8.1.4 Transferases

Transferases can cleave α-1,4 glycosidic linkage of donor molecule and transfer part of the donor to a glycosidic acceptor with the formation of a new glycosidic bond. Enzymes such as amylomaltase (EC 2.4.1.25), cyclodextrin glycosyltransferase (EC 2.4.1.19) and glucan branching enzyme (EC 2.4.1.18), amylomaltase and cyclodextrin glycosyltransferase are capable of forming new α-1,4 glycosidic bond, while branching enzyme forms a new α-1,6 glycosidic bond. Amylomaltases are very similar to cyclodextrin glycosyltransferases with respect to the type of enzymatic reaction. The major difference is that amylomaltase

performs a transglycosylation reaction resulting in a linear product while cyclodextrin glycosyltransferase gives cyclic oligosaccharide products (Takaha and Smith, 1999)

2.8.2 Classification of amylase

2.8.2.1 α-amylase (EC 3.2.1.1)

They are endoamylases that catalyze the hydrolysis of internal α-1,4-O-glycosidic bonds in starch with the retention of α-anomeric configuration in the products. Most of the α-amylases are metalloenzymes, which require calcium ions (Ca^{2+}) for their stability, activity and structural integrity. They belong to family 13 (GH-13) of the glycoside hydrolase group of enzymes. Among the fourteen clans (A–N) defined for glycosidases and transglycosidases, α-amylase family (GH-13) belongs to the eighth clan, clan GH-H (MacGregor, 2005). Their structure contains a $(\beta/\alpha)_8$ barrel with glutamic acid as a proton donor and aspartic acid as a nucleophile on its catalytic sites. They are further classified into eight subfamilies. They cannot catalyze the hydrolysis of α-1,6 linkages. In the early stage of hydrolysis, dextrins with relatively higher molecular weights are produced, with a rapid decrease in the viscosity of starch solution. Final stage of hydrolysis produces large amounts of maltose, maltotriose, glucose, and oligosaccharides (α-limit dextrins) with the α-1,6 linkage, constituting the hydrolysis products.

α-Amylases are ubiquitous enzymes produced by animals, plants and microorganisms, where they play an important role in carbohydrate metabolism. In spite of the wide distribution, amylases from microbial sources are used for the industrial production due to advantages such as cost effectiveness, consistency, less time and space required for production and ease of process modification and optimization (Burhan *et al.,* 2003). Microbial α-amylase finds potential applications in a number of industrial processes such as in food, fermentation, textiles and paper industries. Among bacteria, *Bacillus* sp. is widely used for thermostable α-amylase production to meet industrial needs. *B. subtilis, B.*

stearothermophilus, B. licheniformis and *B. amyloliquefaciens* are known to be good producers of α-amylase and these have been widely used for commercial production of the enzyme for various applications. Characteristics of some bacterial α-amylases are given in Table 2.3

Table 2.3: Characteristics of some of the bacterial amylase

Microorganism	Temp optimum	pH optimum	MW (kDa)	Reference
Aeromonas veronii	10°C	4	63	Samie *et al.,* 2012
Bacillus dipsosauri	60°C	6.1	80	Deutch, 2002
Bacillus sp.	60°C	6.5	71	Sodhi *et al.,* 2005
B. amyloliquefaciens	55°C	6	52	Demirkan *et al.,* 2005
B. cereus	22°C	9	55	Roohi *et al.,* 2013
B. subtilis	55°C	9	21	Roy *et al.,* 2012
Haloarcula hispânica	50°C	6.5	43.3	Hutcheon *et al.,* 2005
Chromohalobacter sp.	65°C	7-9	72	Prakash *et al.,*2009
Pseudomonas stutzeri	55°C	8	57	Maalej *et al.,* 2013

2.8.2.2 β-amylase (EC 3.2.1.2)

β-Amylase are exoamylase that hydrolyzes the α-1,4 linkage next to the nonreducing end of α-glucan, which successively yields maltose in a β-configuration as the sole low-molecular weight product. They belong to GH 14 of the glycoside hydrolase group of enzymes. It cannot hydrolyze the α-1,6 linkages. This action completely stops at four glucose units before the branching points, which produces β-maltose and the high-molecular-weight branched dextrin (β - limit dextrin). β-amylases have generally been obtained from plant sources, however bacterial strains belonging to *Bacillus, Pseudomonas* and *Clostridium* sp, actinomycete strains belonging to *Streptomyces* sp. and fungal strains belonging to *Rhizopus* sp. have been reported to produce β-amylase (Pandey *et al.,* 2000).

Structurally, it has the same $(\beta/\alpha)_8$ barrel as α-amylase; however, two glutamic acid residues act as a proton donor and a nucleophile.

2.8.2.3 Glucoamylase (EC 3.2.1.3)

Glucoamylases are inverting type exo-acting enzymes, catalyzing the hydrolysis of α-1,4 linkages in the nonreducing ends of glucans, producing β-D-glucose as the sole hydrolysis product. As it was the third amylase discovered following α- and β-amylases, it was once called γ-amylase or amyloglucosidase. It can also catalyze the hydrolysis of α-1,6 linkage, though at a lower rate than that of the α-1,4 linkage hydrolysis making it the only enzyme that can hydrolyze starch completely into glucose This new amylase produced β-glucose from starch as the sole reaction product and was named glucoamylase. They belong to GH 15 of the glycoside hydrolase group of enzymes. Glucoamylase has an $(\alpha/\alpha)_6$ structure with Glutamic acid residues act as a proton donor and a nucleophile. Together with α-amylase, glucoamylase plays an important role in industrial starch saccharification process. Glucoamylase are produced mainly by fungi, such *Aspergillus* and *Rhizopus*.

2.8.3 The α-amylase family: characteristics and reaction mechanism

α-amylases are enzymes classified under glycosyl hydrolases family 13 (GH-13): Enzymes belonging to GH-13 have the following features in common:

i) they must act on α -glucosidic linkages and hydrolyze them to produce α-anomeric monosaccharides and oligosaccharides or form α-1,4 or 1,6 glucosidic linkages by transglycosylations

ii) have four highly conserved regions in their primary structures consisting of catalytic and important substrate-binding sites (Kuriki and Imanaka, 1999).

iii) have aspartic acid and glutamic acid residues as catalytic sites

iv) possess a $(\beta/\alpha)_8$ or TIM barrel structure containing the catalytic site residues

2.8.3.1 Catalytic mechanism of action

The α-glycosidic bond has a spontaneous rate of hydrolysis of approximately 2×10^{-15} s^{-1} at room temperature (Wolfenden *et al.*, 1998). Members of α-amylase family increase this rate and are therefore considered to be the most efficient enzymes known. The generally accepted catalytic mechanism of the α-amylase family is α-retaining double displacement (Koshland, 1953) and is shown in Fig 2.2. The mechanism involves two catalytic residues in the active site, glutamic acid as acid/base catalyst and an aspartate as the nucleophile.

Reaction is catalyzed in a five steps process

i) when substrate bound to active site, the glutamic acid donates a proton to the glycosidic bond oxygen, i.e. the oxygen between two glucose molecules at the subsites −1 and +1 and the nucleophilic aspartate attacks the C1 of glucose at subsite −1

ii) an oxocarbonium ion-like transition state is formed followed by the formation of a covalent intermediate

iii) the protonated glucose molecule at subsite +1 leaves the active site while a water molecule or a new glucose molecule moves into the active site and attacks the covalent bond between the glucose molecule at subsite −1 and the aspartate

iv) an oxocarbonium ion-like transition state is formed again

v) the base catalyst glutamate accepts a hydrogen from an incoming water or the newly entered glucose molecule at subsite +1, the oxygen of the incoming water or the newly entered glucose molecule at subsite +1 replaces the oxocarbonium bond between the glucose molecule at subsite −1 and the aspartate forming a new hydroxyl group at the C1 position of the glucose at subsite −1 (hydrolysis) or a new glycosidic bond between the glucose at subsite −1 and +1 (transglycosylation)

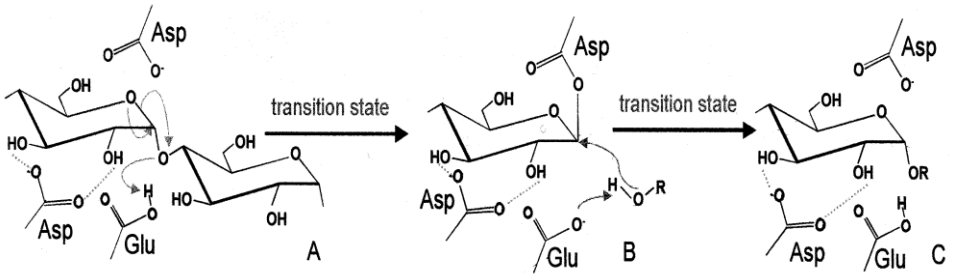

Fig 2.2: The double displacement mechanism of enzyme catalysis

Figure adapted from Van Der Maarel *et al.*, 2002.

During double displacement mechanism, only two of the three conserved catalytic residues directly play a role. The third conserved residue, a second aspartate, forms hydrogen bonds with OH-2 and OH-3 groups of the substrate thereby distorting of the substrate (Uitdehaag *et al.*, 1999). Other conserved amino acid residues can be histidine, arginine, and tyrosine. They play a role in positioning the substrate into the correct orientation into the active site, proper orientation of the nucleophile, transition state stabilization, and polarization of the electronic structure of the substrate (Nakamura *et al.*, 1993; Lawson *et al.*, 1994; Strokopytov *et al.*, 1996; Uitdehaag *et al.*, 1999). Besides the four conserved amino acid sequence regions, an additional fifth conserved region can be identified in members of the α-amylase family (Janecek, 1992, 1995). This region also contains an aspartate acting as calcium ligand.

2.8.3.2 Organization of Catalytic Domain

The two dimensional structure of α-amylase prototype consists of three domains, namely A, B and C (Fig 2.3). Domain A is the N-terminal region with a $(\beta/\alpha)_8$ barrel or TIM barrel structure consisting of eight alternate β-strands and α-helices, domain B consists of a long loop that protrudes between β-strand 3 and α-helix 3 and C domain with a β sheet structure linked to A domain. The α-amylase family have four conserved sequence regions covering the strands β_3, β_4, β_5 and β_7 of the catalytic $(\beta/\alpha)_8$-barreled domain (Kuriki *et al.*, 2005). The β-strands are

placed parallel to one another as if on a cylinder with α-helices placed outside the cylinder. Apart from the four original conserved sequence regions containing the catalytic and substrate binding residues, three additional conserved sequences are present which are located near the C-terminus of domain B, containing amino acid residues providing enzyme specificity (Janecek, 2002).

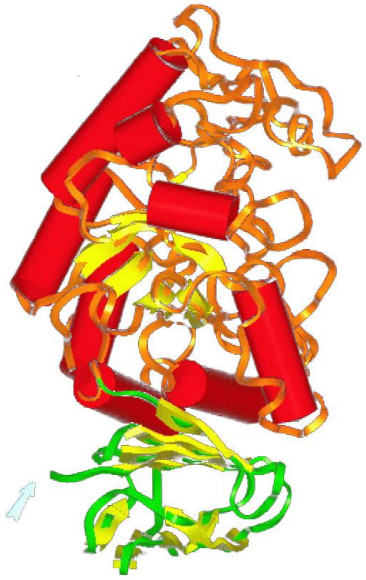

Figure 2.3: 3-D structure of α-amylase from *B. subtilis* X-23

Orange ribbon indicates domains A and B, green ribbon indicates domain C, red cylinders and yellow arrows represent α-helices and β-strands and the arrow represents the carboxyl terminus (Figure adapted from Ohdan *et al.*, 1999)

2.9 Amylases derived from metagenomics

Despite the abundance of new enzymes isolated by metagenomic approaches and their important industrial potential, there is relatively little data concerning characterization of metagenome-derived amylases. Eight clones with amylolytic activity were identified in BAC library derived from soil environments (Rondon *et al.*, 2000). A thermostable α-amylases was identified from DNA libraries originating from various environmental samples (Richardson *et al.*, 2002). In 2003, studies by Voget and colleagues identified an α-amylase gene in a

cosmid metagenomic library (Voget *et al.,* 2003). Soil-derived metagenomic library in pUC19 vector identified a novel amylase gene, and the enzyme was characterized (Yun *et al.,* 2004). Metagenomic library derived from Kargil soil of Northwestern Himalayas identified amylase clone that retained 90% activity even at low temperature with a protein size of 38kDa (Sharma *et al.,* 2010). A thermostable and calcium-dependant amylase with applications in baking and destarching was identified from a soil metagenome in fosmid vector (Vidya *et al.,* 2011). In 2014, the first α-amylase isolated from a gastrointestinal metagenomic library has been biochemically characterized (Xu *et al.,* 2014b). Cold-adapted α-amylase AmyI3C6 with a molecular mass of 56 kDa and retaining more than 70% of its activity at 1°C was identified in a metagenomic library from the cold and alkaline environment (Vester *et al.,* 2015). Metagenomics has proved to be inevitable tool for unlocking novel compounds with improved properties from nature.

2.10 Studies on α-amylase gene

The α-amylase gene from various bacterial species including *Bacillus* sp. has been cloned and expressed in *E. coli* systems for over rproduction of the enzyme. α-amylase genes from different actinomycetes species including those of *Streptomyces* have been cloned and characterized during the early 90s (Vigal *et al.,* 1991; Bahri and Ward, 1993; Chen *et al.,* 1995). A 1539 bp α-amylase gene from *Bacillus licheniformis* encoding a protein of 512 amino acids has been cloned and characterized in *E.coli* (Hmidet *et al.,* 2008). A 1650 bp gene encoding acidic, thermostable and raw starch hydrolysing α-amylase was cloned from an extreme thermophile *Geobacillus thermoleovorans* and expressed in *E.coli* suggested that it encodes a 515 amino acid protein (Mehta and Satyanarayana, 2013). A raw starch digesting α-amylase gene (amyBS-I) from *Bacillus subtilis* strain AS01, with a length of 1980 bp, consisting of 659 amino acids and molecular mass of 72kDa was cloned and expressed in *Escherichia coli* BL21 cells (Roy *et al.,* 2013). Asoodeh and colleagues (2014) cloned α-amylase gene

from *Bacillus* sp. DR90 encoding a protein of 76 kDa, which was thermoacidophilic and organic-solvent tolerant.

2.11 Bioinformatic tools for protein modeling and structure prediction

Although great progress was made in the field of experimental structure determination by X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), it is still remains a time consuming process without guaranteed success. Therefore, it is necessary to bridge this structure knowledge gap, employing computational methods for protein structure prediction. Three-dimensional (3D) structure of protein provides valuable insights into the molecular basis of protein function, structure based design of specific inhibitors, allowing an effective design of experiments, such as site-directed mutagenesis, studies of disease-related mutations etc. Various protein modelling tools are currently employed such as SWISS-MODEL, Phyre^2, ESyPred3D, 3D-JIGSAW etc are used for automated homology modelling.

2.11.1 Swiss-Model (http://swissmodel.expasy.org)

Swiss-Model is a server for automated comparative modelling of protein 3D structures (Arnold *et al.*, 2006). It pioneered the field of automated modelling starting in 1993 and is the most widely-used free web-based automated modelling facility today. It provides several levels of user interaction through its World Wide Web interface. In the first approach mode, only amino acid sequence of a protein is submitted to build a 3D model. Template selection, alignment and model building are completely automated by the server. In the alignment mode, the modelling process is based on a user-defined target-template alignment. Complex modelling tasks can be handled with the project mode employing an integrated sequence-to-structure workbench DeepView (Swiss-Pdb Viewer). All models are sent back via email with a detailed modelling report.

SWISS-MODEL employs comparative modelling which is the only method that can generate a 3D model of a protein (target) from its amino acid sequence reliably. Successful model building requires at least one experimentally solved 3D structure (template) having sequence similarity to the target sequence. All homology-modelling methods consist of the following four steps: (i) template selection; (ii) target template alignment; (iii) model building; and (iv) evaluation. These steps can be iteratively repeated, until a satisfying model structure is elucidated. Several techniques have been developed for successful model building (Sali and Blundell, 1993; Guex and Peitsch, 1997). Depending on the complexity of the modelling task and server workload, it may take a few minutes to several hours to build a protein model. The computational resources for the SWISS-MODEL server are provided in collaboration with the Swiss Institute of Bioinformatics at the Biozentrum Basel (University of Basel, Switzerland) and the Advanced Biomedical Computing Center (NCIFCRF Frederick, USA).

2.11.2 Phyre^2 (www.sbg.bio.ic.ac.uk)

Phyre^2 (Protein Homology/Analogy Recognition Engine) is a freely available web-based services used for protein structure prediction. The Phyre server employs a library of known protein structures retrieved from the Structural Classification of Proteins (SCOP) database (Murzin *et al.,* 1995) and Protein Data Bank (PDB) (Berman *et al.,* 2007). The sequence of each of these structures is scanned against a nonredundant sequence database and a profile is constructed and deposited in the 'fold library'. The known and predicted secondary structure of these proteins is also stored in the fold library. The query sequence is scanned against the nonredundant sequence database, and a profile is constructed by PSI-BLAST. Following profile construction, secondary structure of the query is predicted. Three independent secondary structure prediction programs are used in Phyre: Psi-Pred (McGuffin *et al.,* 2000), SSPro (Pollastri *et al.,* 2002) and JNet (Cole *et al.,* 2008). The predicted presence of alpha-helices, beta-strands and

disordered regions is shown graphically together with a color-coded confidence bar on the result screen.

This profile and secondary structure is then scanned against the fold library using a profile–profile alignment algorithm (Bennett-Lovsey *et al.,* 2008). This alignment process returns a score on which the alignments are ranked and the top ten highest scoring alignments are used to construct full 3D models of the query. Whereever possible, missing or inserted regions caused by insertions and deletions in the alignment are repaired using a loop library and reconstruction procedure. Finally side chains are placed on the model using a fast graph-based algorithm and side chain rotamer library. Phyre is widely used by the biological community, with >150 submissions per day, and provides a simple interface to results. Phyre takes 30 min to predict the structure of a 250-residue protein.

2.12 Application of amylases

α-amylases are important class of enzyme which finds application in variety of industrial processes such as in food, detergents, textiles and in paper industry, for the hydrolysis of starch. In this light, microbial amylases have completely replaced chemical hydrolysis in the starch processing industry. α-amylases from microbial sources are commercially available from enzyme manufactures such as Novozyme, Danisco, and Amano Enzyme Inc. for specific uses in varied industries

2.12.1 Detergent additive

Enzymes are one of the major ingredients of modern compact detergents. α-amylases are used as additives in both dishwashing detergents and powder laundry detergents. Around 90% of all liquid detergents contain α-amylase (Kottwitz *et al.,* 1994). These enzymes help to degrade the residues of starchy foods such as chocolate, potatoes, gravies, custard etc to dextrins and other smaller oligosaccharides (Olsen and Falholt, 1998). Amylases have activity at

alkaline pH and lower temperatures, maintaining the necessary stability under detergent conditions and the oxidative stability of amylases is one of the most important criteria for their use in detergents formulations. Commercial detergent enzyme suppliers Novozymes and Genencore International have improved the bleach stability of the amylases by replacing oxidation sensitive amino acids with other amino acids and introduced these new products in the market under the trade names Purafect OxAm® and Duramyl®, respectively.

2.12.2 Textile desizing

Amylases are used in textile industry as desizing agent. Desizing involves the removal of starch from the fabric which serves as the strengthening agent to prevent breaking of the warp yarn during the weaving process. Sizing agents like starch are applied to yarn before fabric production to ensure a fast and secure weaving process. Starch is the commonly used sizing agent, because it is cheap, easily available in most regions of the world, and it can be removed quite easily. Starch is later removed from the woven fabric in a wet-process in the textile finishing industry. The sizing agents in the fabric are removed by α-amylases selectively without attacking the fibres (Gupta *et al.,* 2002). It randomly cleaves the starch into dextrins that are water soluble and can be removed by washing. Amylase from *Bacillus* stain was commonly employed in textile industries.

2.12.3 Food industry

Amylases are widely employed in food industry such as preparation of digestive aids, production of cakes, baking, brewing, fruit juices and starch syrups. Amylases are used as additives in animal feeds to improve the digestibility of fiber (Van Der Maarel *et al.,* 2002). In baking industry, these enzymes are added to the bread dough to degrade the starch in the flour into smaller dextrins, which are subsequently fermented by the yeast. The addition of α-amylase to the dough enhances the rate of fermentation thereby reducing the viscosity of dough, which improves the volume and texture of the product. α-amylases also have an anti-

staleing effect in bread baking, as they improve the softness retention of baked goods thereby increasing the shelf life of these products (Gupta *et al.,* 2003). Highly thermostable α-amylase are used in starch liquefaction process by which starch is converted into high fructose corn syrups (HFCS). Because of their high sweetening property, these are used in huge quantities in the beverage industry as sweeteners for soft drinks (Nielsen and Borchert, 2000)

2.12.4 Paper industry

α-amylases are also used in paper and pulp industry for the modification of starch to produce low-viscosity, high molecular weight starch which can be used for coating paper (Bruinenberg *et al,* 1996). As for textiles, sizing of paper protects it against mechanical damage during processing. It also improves the stiffness, strength, writing quality and erasabilty of the paper. A number of amylases find use in the paper industry, including Termamyl®, Fungamyl, BAN® (Novozymes, Denmark), Amizyme® (PMP Fermentation Products, Peoria, USA), and α-amylase G9995® (Enzyme Biosystems, USA)

2.12.5 Fuel alcohol production

Ethanol is the most utilized liquid biofuel. Starch is the most commonly used substrate for ethanol production due to its low price and ease of availability. In this production process, starch is converted to fermentable sugars by enzymatic steps. The bioconversion of starch into ethanol involves liquefaction and saccharification, wherein starch is converted into sugars using α-amylase or using an amylolytic microorganism, followed by fermentation, where sugar is converted into ethanol using yeast *Saccharomyces cerevisiae*. α-amylase obtained from thermoresistant bacteria like *Bacillus licheniformis* or engineered strains of *Escherichia coli* or *Bacillus subtilis* was used during the first step of hydrolysis of starch suspensions (Sanchez and Cardona, 2008).

Chapter 3

EXTRACTION OF METAGENOMIC DNA BY DIFFERENT METHODS FROM VARIOUS ENVIRONMENTAL SAMPLES INCLUDING MARINE AND MANGROVE SEDIMENTS AND THEIR EVALUATION.

3.1 INTRODUCTION

Soil being a complex habitat for diverse microorganisms, is a rich source of novel bioactive molecules like enzymes which has numerous applications in health and other industries. Because of their complex nature, extraction of metagenomic DNA from various ecosystems pose challenges for obtaining larger quantity of pure high molecular weight DNA. Depending on the soil type, microbial cells may remain tightly bound to soil aggregates, particles or organic matter making it difficult to obtain pure DNA preparations.

Marine and mangrove ecosystems are distinctive ecological niches, with a plethora of microbes playing different, important roles in nutrient recycling and various other ecological processes. These ecosystems thereby require a thorough exploration and understanding of their microbial diversity. Similarly, Arabian Sea sediments also serve as hotspots for microbial diversity and therefore require a thorough exploration. Mangalavanam, a conserved mangrove ecosystem located in Kochi, Kerala, India, known as the “Green lungs of Kochi” could be a treasure trove of unexplored microflora. Metagenomic methods may be employed to explore the phylogenetic diversity and the potential for novel biomolecules from both sediments.

Estimations reveal that only less than 1% of the total microbial communities from the environment are readily cultivable by standard microbiological methods (Handelsman, 2004) which can now be bypassed by employing metagenomic methods. By the same rationale, community DNA isolation is a challenging process owing to the co-extraction of humic substances and other contaminants. Metagenomic DNA extraction methods are classified as direct (*in situ*) and indirect (*ex situ*) methods. In direct method, cells are initially lysed within the soil sample, followed by consequent separation of DNA from cell debris and soil matrix (Ogram *et al.,* 1987); whereas indirect method involves cell separation followed by cell lysis and DNA recovery (Holben *et al.,* 1988). Both these approaches have advantages as well as disadvantages concerning DNA yields as well as purity required for molecular analysis and unbiased representation of the entire microbiome. However as soil compositions vary greatly with regard to the organic and inorganic content, standardization of total DNA isolation protocols become a prerequisite to any analysis.

Direct lysis method provides highest DNA yields within acceptable processing time (Tsai and Olson, 1991). These methods utilizes different cell disruption protocols which is usually a combination of physical, chemical, thermal and enzymatic lysis methods. The objective of this study was to investigate the effectiveness of different direct lysis methods on yield and purity of DNA from mangrove soils to enable PCR amplification and further metagenomic analysis. In the present study five different direct lysis methods utilizing different cell disruption methods were evaluated for DNA yield, purity and PCR amenability.

3.2 MATERIALS AND METHODS

3.2.1 Extraction of metagenomic DNA from marine and mangrove sediments

3.2.1.1 Collection of mangrove sediment samples

Mangrove soils (1, 2 and 3) were collected during December, 2011 from 3 different islands located in Kochi and soil 4 i.e. from Mangalavanam mangroves were collected during January, 2012. All soils were collected by removing surface leaf litter and collecting the top soil. Samples were transferred with sterilized spatula in sterile containers and were stored at -20°C until further analysis. Sampling locations are depicted in Fig 3.1 and its coordinates are given in Table 3.1.

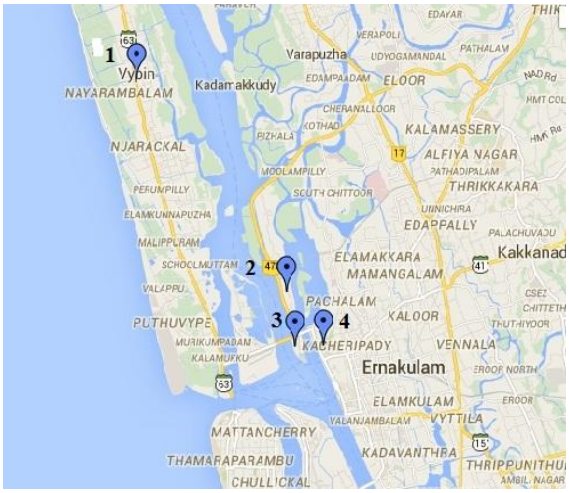

Fig 3.1: Map showing sampling locations of mangrove sediments

Table 3.1 Coordinates of sampling location.

Soil Sample	Sampling Station	Latitude	Longitude
1	Vypin	10° 4' 7.3272" N	76° 12' 47.3292" E
2	Bolgatty	10° 0' 16.2864" N	76° 15' 42.0120" E
3	Ponnarimangalam	10° 0' 0.4035" N	76° 16' 872.3201" E
4	Mangalavanam	9° 98’ 9.9424" N	76° 27’36.4992" E

3.2.1.2 Collection of marine sediment sample

Marine sediments were collected from eastern Arabian Sea onboard the research vessel FORV Sagar Sampada (Cruise No: 305) during August, 2012 using grab at a depth of 96 M. Sediment samples were transferred to sterile containers and were stored on board at -20°C until brought to the laboratory. Sampling coordinates are given in Table 3.2 and sampling station is depicted in Fig 3.2

Table 3.2 Coordinates of sampling location.

Soil Sample	Sampling Station Code	Latitude	Longitude
5	Grab 3	9°59'10.9968" N	75° 39' 26.4564" E

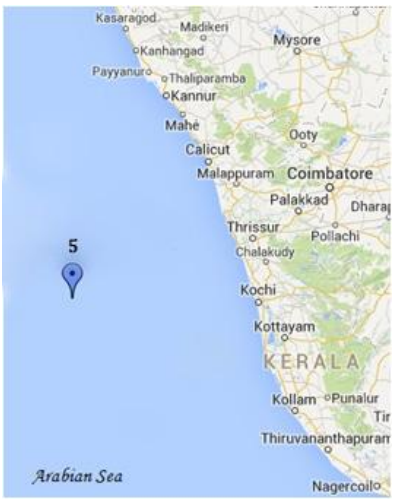

Fig 3.2: Map of sampling locations in the Arabian Sea

3.2.1.3 Extraction of metagenomic DNA from mangrove sediments using five different protocols

In order to obtain quality DNA from sediment samples with sufficient yield to carry out further analysis, five different DNA isolation protocols were compared. Metagenomic DNA was extracted from the three mangrove sediment

samples (Soil 1, 2 and 3) employing five direct lysis methods which includes the method of Zhou *et al*., (1996), slightly modified method of Volossiouk *et al*., (1995), Dong *et al*., (2006), Tsai and Olson, (1991) and that of Siddhapura *et al*., (2010). The yield and purity of DNA obtained from each protocol was compared and the most effective method among these was employed to extract metagenomic DNA from both mangrove and marine sediments (Soil 4 and 5) for phylogenetic diversity analysis and functional cloning reactions.

3.2.1.3.1 Method I (Zhou *et al*., 1996)

Mixed 5g soil with 13.5 mL DNA extraction buffer (in an Oakridge tube) (100 mM Tris-HCl (pH 8.0), 100 mM sodium EDTA (pH 8.0), 100 mM sodium phosphate (pH 8.0), 1.5 M NaCl, 1% CTAB and 100 µL of proteinase K (10 mg/mL) (Fermentas, USA) and the sample was incubated by horizontal shaking at 225 rpm for 30 minute at 37°C (Orbitek, Scigenics India). This was followed by addition of 1.5 mL of 20% SDS and incubated in a 65°C water bath for 2 h (Remi, India) with gentle end-over-end inversions every 15 to 20 min. The supernatant was transferred to new tubes after centrifugation at 6,000 x *g* for 10 minutes (Sigma, 2-16K, Germany) at room temperature. The soil pellets were further extracted twice using the same protocol. Supernatants from the three extractions were pooled, mixed with equal volume of chloroform: isoamyl alcohol (24:1, v/v), followed by recovery of the aqueous phase by centrifugation and finally precipitation with 0.6 volume of isopropanol at room temperature for 1 h. The nucleic acids obtained were pelleted by centrifugation at 16,000 x *g* for 20 min and washed with ice cold 70% ethanol, air dried and resuspended in sterile deionized water to a final volume of 500 µL.

3.2.1.3.2 Method 2 (Volossiouk *et al.,* 1995)

Soil sample (0.25 g) after adding liquid nitrogen was ground to fine powder using sterile mortar and pestle, suspended in 0.5 mL of skim milk powder solution (0.1 g skim milk in 25 mL of water), vortexed well and centrifuged for 10

min at 12,000 x *g* at 4°C. To the supernatant 2 mL of SDS extraction buffer (0.3% SDS in 0.14 M NaCl, 50 mM sodium acetate (pH 5.1) was added and vortexed to mix. An equal volume of Tris-saturated phenol solution (pH 8.0) was added and vortexed for 2 min at room temperature. Aqueous phase was collected by centrifugation at 12,000 x *g* for 10 min and the nucleic acid was precipitated with 1 volume of ice cold isopropanol at -20°C for one hour, followed by centrifugation at 12,000 x *g* for 10 min to pellet the DNA. The pellet was washed twice with ice cold 70% ethanol, with centrifugation between each rinse, air dried, dissolved in 150 µL of sterile deionized water and stored at -20^0C until further analyses.

3.2.1.3.3 Method 3 (Dong *et al.*, 2006)

In this method 0.30 g of soil sample was mixed with 0.35 g of glass beads (diameter 2.0 mm) and 300 µL of phosphate buffer (0.1 M NaH_2PO_4–$NaHPO_4$ (pH 8.0) in a microcentrifuge tube, vortexed well, followed by addition of 250 µL of SDS lysis buffer (100 mM NaCl, 500 mM Tris (pH 8.0), 10% SDS). This was vortexed horizontally for 10 min at 225 rpm. The supernatant was transferred to a new tube after centrifugation at 10,000 x *g* for 30 seconds. 250 µL of chloroform: isoamyl alcohol (24:1) was added and incubated at 4°C for 5 min, followed by centrifugation at 10,000 x *g* for 1 min. Nucleic acids were precipitated by addition of 0.5 volume of 7.5 M ammonium acetate and 1 volume of isopropanol, and incubated at -20°C for 15 min. The DNA pelleted at 12,000 x *g* for 10 min and was washed thrice with 70% ethanol and air-dried. The pellets were then dissolved in 100 µL of 10 mM Tris (pH 8.1), 100 µL of 10 mM Tris (pH 7.4), 100 µL of 10 mM Tris (pH 6.7) and 100 µL of 10 mM Tris (pH 6.0) and flocculated with 10 mM aluminium sulfate. Precipitates of humic substances was removed by centrifuging at 10,000 x *g* for 5 min.

3.2.1.3.4 Method 4 (Tsai and Olson, 1991)

One gram of soil was washed twice with 2 mL of 120 mM sodium phosphate buffer (pH 8.0), suspended in 2 mL of lysis solution (0.15 M NaCl, 0.1

M Na_2EDTA (pH 8.0) containing lysozyme (15 mg /mL), incubated for 2 h in a 37°C water bath with mixing at 20 to 30 min intervals, followed by addition of 2 mL of 0.1 M NaCl, 0.5 M Tris-HCl (pH 8.0), 10% SDS. Cells were lysed by three cycles of alternating freeze-thaw at -80°C and 65°C respectively. After phenol - choloroform extraction, the nucleic acid was precipitated with ice cold isopropanol, air dried and resuspended in 100 µL of TE buffer (20 mM Tris-HCl, 1 mM EDTA (pH 8.0).

3.2.1.3.5 Method 5 (Siddhapura *et al.*, 2010)

In this method 1g soil was mixed with 10 mL extraction buffer (100mM Tris–HCl (pH 8.2); 100mM EDTA (pH 8); 1.5M NaCl), incubated at 37°C for 10 h with shaking at 150 rpm and supernatant was collected by centrifugation at 5000 rpm for 10 min. Samples were re-extracted with 1mL of extraction buffer. To the supernatant 4 mL of lysis buffer (20%, w/v) SDS, lysozyme (20mg/mL), Protinase K (10 mg/mL), N-lauryl sarcosine (10 mg/mL), 1% (w/v) CTAB (cetyltrimethylammonium bromide) was added and incubated at 65°C for 2 h with intermittent shaking every 15 min. Centrifuged at 10,000 rpm for 10 min at 4°C to collect the supernatant. The preparation after phenol - chloroform extraction was treated with 1/10 volume of 7.5 M potassium acetate and precipitated by 2 volume of chilled ethanol. DNA was pelleted by centrifugation at 10,000 rpm for 10 min, air dried and suspended in 50 µL sterile deionized water**.**

DNA yield from all the preparations were expressed as yield in µg per gram of soil which is calculated as follows

$$\frac{\text{Concentration of DNA (µg/µL)}}{\text{Weight of soil (g)}} \times \text{Volume in which DNA suspended (µL).}$$

3.2.1.3.6 Method 6 DNA isolation using commercial kit

UltraClean™ Soil DNA isolation kit (MoBio, USA) was also used for the extraction of community DNA from both marine and mangrove sediment samples following the manufacturer's instructions.

3.2.1.4 Agarose gel electrophoresis (Sambrook *et al*., 2000)

The extracted DNA were analysed by agarose gel electrophoresis in 0.8% gel containing 10 mg/mL ethidium bromide solution prepared in 1X TAE buffer (Appendix I). The DNA sample was mixed with 6X gel loading dye (Appendix I) and loaded into the wells of the agarose gel along with the DNA markers (Lambda DNA *Eco*R I/ *Hin*d III double digest) and 1kb DNA ladder (Thermo Scientific, MA, USA). Electrophoresis was carried out at 80 V for 1 h (GeNei Mini Electrophoresis system, GeNei, India). The gel was visualized under ultraviolet illumination and gel pictures were captured using Gel documentation system (Syngene, USA).

3.2.1.5 DNA quantification (Sambrook *et al*., 2000)

The DNA was quantified using a UV-Visible spectrophotometer (Shimadzu, Kyoto, Japan). The spectrophotometric readings were taken at wavelengths of 260 nm, 280 nm and 230 nm. The absorbance at 260 nm allows calculation of the concentration of nucleic acid in the sample. An absorbance value of 1 at 260 nm corresponds to approximately 50 µg/mL for double stranded DNA.

Purity of DNA was obtained by calculating A_{260}/ A_{280} and A_{260}/A_{230} ratios for protein and humic acid contaminants respectively in the preparation. A_{260}/A_{280} ratio less than 1.8 indicates protein contamination and A_{260}/A_{230} ratio less than 2 indicates the presence of humic acid substances.

3.2.1.6 Determination of quality of DNA by Polymerase Chain Reaction (PCR)

To determine whether PCR inhibitors were present, DNA preparations isolated by all five methods were used as template to amplify the region encoding 16S rRNA gene in a thermal cycler (Applied Biosystems, USA) using universal primers. The forward and reverse primers (Sigma Aldrich, USA) used for the

amplification is as given in Table 3.3. An appropriate dilution(~80-100ng) of metagenomic DNA template was used for PCR reaction. The concentration of different ingredients used for PCR amplification is given in Table 3.4.

Table 3.3 Primers used to amplify 16S rRNA gene

Primer	Sequence	Reference
16S Forward	5' GAGTTTGATCCTGGCTCAG 3'	Shivaji *et. al.*, 2000
16S Reverse	5' ACGGCTACCTTGTTACGACTT 3'	

Table 3.4 PCR Mix composition

Ingredient	Concentration	Volume
Template DNA	50 ng/µL	1 µL
16S Forward primer	10 µM	1 µL
16S Reverse primer	10 µM	1 µL
dNTPs	2 mM each	2 µL
$MgCl_2$	25 mM	1.2 µL
PCR buffer	10X	2 µL
Taq DNA polymerase	1 U/µL	1 µL
Sterile distilled water		upto 20 µL

PCR amplification was carried out in a Thermal Cycler (Applied Biosystems, CA, USA) using the following program

Step	Temperature	Time	
Initial Denaturation	94°C	1.5 min	
Denaturation	94°C	30 sec	
Annealing	56°C	30 sec	35 cycles
Extension	72°C	2 min	
Final Extension	72°C	10 min	

The amplicons were separated electrophoretically in 1% agarose gel and gel pictures are captured using gel documentation system as described in section 3.2.1.4.

3.2.1.7 Statistical analysis

All experiments were repeated thrice and statistical analysis was done by Microsoft Excel 2007 calculating mean and standard error.

3.3 RESULTS AND DISCUSSION

3.3.1 Extraction of metagenomic DNA from marine and mangrove sediments

Five different methods of metagenomic DNA isolation using three different soil samples from mangroves (Soil 1, 2 and 3) were compared with respect to DNA yield, purity, humic acid content, and suitability for PCR. The most suitable method was employed for the extraction of metagenomic DNA from soil 4 and 5 which was used for phylogenetic diversity studies and construction of metagenomic libraries to screen for enzyme production.

3.3.1.1 Visualization of community DNA on agarose gel

Community DNA isolated using five different methods using three different soil samples from mangroves sediments were visualized on agarose gel along with the DNA marker Lambda DNA *Eco*R I/ *Hin*d III double digest and 1kb DNA ladder (Fig. 3.3).

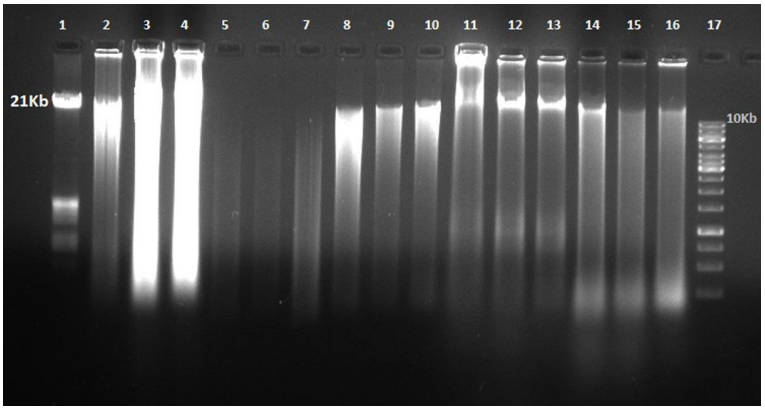

Fig 3.3: Comparison of metagenomic DNA from three Mangrove soils by five methods using Agarose Gel electrophoresis. Lane 1-21 kb DNA marker, Lane 17-1 kb ladder, Lane 2-4: DNA isolated by Method 1, Lane 5-7: DNA isolated by Method 2, Lane 8-10: DNA isolated by Method 3, Lane 11-13: DNA isolated by Method 4, Lane 14-16: DNA isolated by Method 5

From the gel picture, although same volumes were loaded, it is clear that the quantity of total DNA isolated by the different methods varied considerably; but all extracted DNA were of high molecular weight suggesting their metagenomic nature. Lower DNA concentration was obtained by method 2 which employed grinding soil particles with liquid nitrogen and this was clearly visible in the gel picture (lane 5-7). Mechanical disruption of cell wall by grinding with liquid nitrogen and bead beating as employed in method 2 and 3 resulted in increased DNA shearing, when compared to the gentle freeze-thawing employed in method 4. Similarly, method 1 which was a detergent based method, yielded high amount of DNA as is visible in the gel picture.

Physical treatments such as grinding, sonication and bead beating homogenizes soil particles and can access individual microbial cells within a sample, but with greater possibility of DNA shearing. Previous studies revealed that a combination of chemical and mechanical lysis can yield twice the amount of DNA than by any single method alone (More *et al.*, 1994). These studies suggest

that SDS based cell lysis is the most widely used DNA extraction method, whereby DNA yield is more compared to freeze thawing and use of other detergents (Trevors *et al.*, 1992).

3.3.1.2. Spectrophotometric analysis for yield and purity of isolated DNA

The isolated DNA was assessed for yield and purity by obtaining OD ratios at 260 nm/280 nm (DNA/Protein) and 260 nm/230 nm (DNA/humic acid). Comparative analysis revealed the considerable variations in yield and purity of DNA obtained by the different methods. The concentration of DNA obtained from three samples using the five different protocols was determined and is as represented in Fig. 3.4. Highest yield was obtained by method 4, giving 748.6 ± 5.7, 647.3 ± 2.1 and 353 ±1.9 µg DNA/gram of soil with soil 1, 2 and 3 respectively; while the lowest yield was obtained by method 3. Although the quantity of total DNA isolated by the different methods varied considerably, the extracted DNA were of high molecular weight despite the DNA shearing; however the DNA isolated by method 2 was considerably sheared.

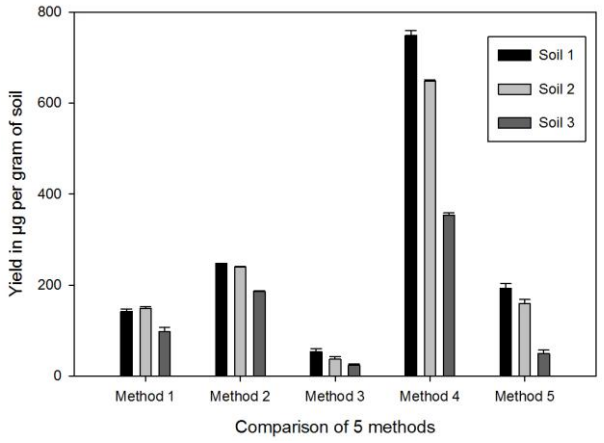

Fig 3.4: DNA yield from three different Mangrove soils by five methods.

Comparative analysis with respect to yield and purity of DNA in the present study revealed the considerable variations obtained by the different methods employed. As depicted in Fig. 3.5, method 1 gave DNA with A_{260}/A_{280} ratios close to optimum, while all the other methods tested failed to do so. But as is observed from Fig. 3.6, the A_{260}/A_{230} ratios obtained suggest that the DNA preparation contains residual humic contaminants making it unsuitable for further downstream processes. Method 1 was based on lysis with a high-salt extraction buffer followed by extended heating of the soil suspension in the presence of detergents like SDS and CTAB. A_{260}/A_{230} ratios indicating comparatively reduced humic content was obtained by method 2. Although the liquid nitrogen method yielded the shortest DNA fragments, it also had much reduced amounts of contaminants. Consequently a combination of chemical lysis along with mild physical methods can greatly influence the total DNA content in terms of quantity and quality.

Studies on soil samples collected from hot springs in Himachal Pradesh, India, employed Q-Sepharose for the purification of DNA following extraction by method 1 (Sharma *et al*., 2007). Similar method was employed in another study for extraction of metagenomic DNA from microbial communities of deep-sea methane seeps in New Zealand (Ruff *et al.*, 2013). However the purity of the isolated DNA was poor therefore it was further purified using Wizard DNA clean-up system. So depending on the soil type, DNA extracted by method 1 requires further purification steps to obtain pure, high quality DNA.

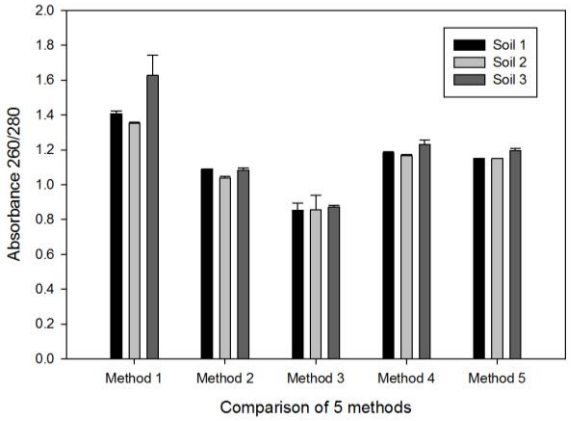

Fig 3.5: Purity of DNA (A_{260}/A_{280}) from different Mangrove soils by five methods.

Similarly in method 2, skim milk powder was added prior to addition of the extraction buffer, which may have helped to retain high quality DNA. Our results suggested that addition of skim milk helped to make the extracted DNA from the three soil samples tested amenable to PCR (section 3.3.1.3).

The addition of carriers and polyvalent polymers helps to reduce DNA loss due to adsorption and degradation (Denhardt, 1966). The use of skim milk to improve the quality of the DNA was previously reported (Volossiouk *et al.*, 1995; Garcia-Pedrajas *et al.,* 1999; Takada-Hoshino and Matsumoto, 2004; Ikeda *et al.*, 2008, Nair *et al.,* 2014), where skim milk by acting as a carrier could reduce the adsorption and degradation of nucleic acids.

On the other hand precipitating DNA with isopropanol improved DNA yield compared to the original study which used absolute alcohol instead. Observations from the present study suggest that starting with a low gram weight of soil for DNA isolation as seen in method 2 and addition of skim milk during extraction can possibly help to reduce the humic contaminants, which would otherwise interfere with all other downstream processing of DNA, like amplification and cloning, to name a few.

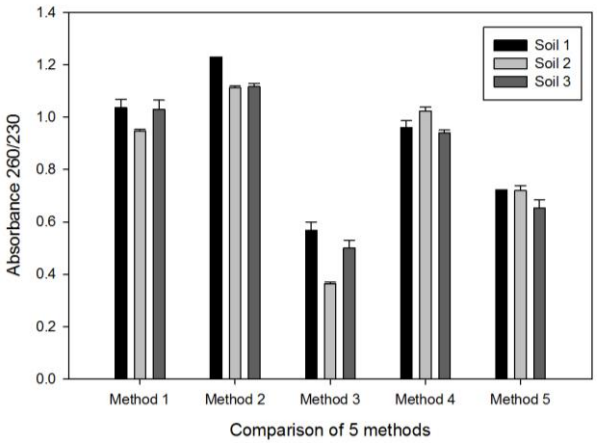

Fig. 3.6: Purity of DNA (A_{260}/A_{230}) from different Mangrove soils by five methods.

Method 3 additionally used Aluminium sulfate which resulted in flocculation of the humic substances by the excessive Al^{3+}, followed by removal of superfluous Al^{3+} by pH adjustment and finally released the soil microbial DNA by SDS lysis. But the disadvantage of the method was low DNA yield as was observed in the present study using the three soils tested. This may be due to the co-precipitation of DNA along with humic acids, as DNA and humic acid has similar physical and chemical characteristics. This method yielded DNA A_{260}/A_{230} ratios ≤ 0.6, suggesting that the DNA isolated by method 3 had higher amounts of humic contaminants present in the three soil samples tested.

In a study on South African deep mine biofilm samples, the sample was treated with aluminium sulfate to remove inhibitors that are present, followed by DNA extraction by alternate methods (Dong *et al.*, 2006; Litthauer *et al.*, 2010)

Method 4 employed lysozyme-SDS treatment followed by three cycles of freeze-thawing, and consequentially high DNA yields were obtained in all the three soil samples analyzed. This method was specifically developed to extract community DNA from subsurface soil samples from manufactured gas site and

sediment samples from a settling pond (Tsai and Olson, 1991). DNA obtained was less sheared when compared to other methods giving a thick band of DNA in the 21 kb range as shown in Fig 3.3. But the purity ratios were not in the optimum range suggesting that further purification could be required for removal of inhibitors.

Similarly method 5 was developed to extract total environmental DNA from two different habitats - saline soil near the salt pan of Gujarat and Sambhar Soda Lake, Rajasthan. This method relies on soft lysis employing a combination of lysozyme and proteinase K along with various detergents like SDS, CTAB and N-lauroyl sarcosine. Soft lysis is based on the disruption of microbial cells by enzymatic and chemical means. The enzymatic lysis treatment is based on enzymatic digestion of microbial cells to release DNA followed by the treatment of soil with surfactants and chelating agents resulting inhibitors free good quantity DNA.

Various other environmental DNA isolation protocols have been previously compared and studied (Purohit and Singh, 2009; Delmont *et al.,* 2011). Extracting pure DNA from environmental samples is practically as important as the yield; however it is also one of the most complex problems associated with the application of molecular techniques on environmental samples. Heterogeneous nature of the environmental samples requires each extraction procedure to be precise and optimized for every soil sample. Most DNA extraction procedures co-extract humic acids, pigments, heavy metals, and other contaminants. Humic contaminants due to their three dimensional structure and functional reactive groups bind with organic compounds (Stevenson, 1976) and are therefore one of the major problems associated with any soil community DNA isolation. Depending on soil types, crude DNA extracts can be contaminated by approximately 0.7-3.3 µg/µL of humic acid (Tebbe and Vahjen, 1993). In addition, due to similar physicochemical properties with nucleic acid they easily

co-precipitate with nucleic acid. These contaminants may not only hinder PCR reactions acting as inhibitor, but also can degrade the DNA during storage. Humic acid may through specific binding to DNA inhibit amplification in PCR reactions by limiting the amount of available template (Opel *et al.*, 2010). Purification of DNA employing polyvinylpolypyrrolidone, embedding DNA in agarose blocks followed by successive washing steps or by using sephadex columns can help improve quality of soil DNA and subsequent PCR amplification (Moreira, 1998; Frostegard *et al.*, 1999; Miller *et al.*, 1999). The aim of any extraction protocol is to succeed in obtaining genomic DNA which is a representative of the microbial diversity present within a soil. However different extraction procedures target only specified group of microbiota present within a soil which results in biased estimates of DNA quantity, evidently due to differences in individual component steps, sorption of DNA to soil particles, DNA degradation or co-extraction of inhibitors (More *et al.*, 1994; Madsen *et al.*, 1996; Feinstein *et al.*, 2009) suggesting that additional measures should be considered when divergent soil types are compared or when comprehensive community analysis is required.

3.3.1.3 16S rRNA gene amplification

To check for the purity of isolated metagenomic DNA obtained from the three different mangrove sediments, PCR amplification of 16S rRNA gene was performed. PCR amplification was successful only with DNA obtained by method 2 (Fig 3.7) which had comparatively reduced humic acid contaminants as indicated by the A_{260}/A_{230} ratio. Despite the shearing of DNA in all 3 soil samples employing liquid nitrogen extraction technique, they yielded 16S rRNA gene amplification using a single set of primer without the addition of any PCR enhancers or additives, thereby suggesting the suitability of the method in diverse soils and also in diversity studies. In the present study maximum DNA yield was obtained in lysozyme-freeze-thawing protocol (method 4), although the presence of residual amounts of humic and protein contaminants hindered PCR reaction.

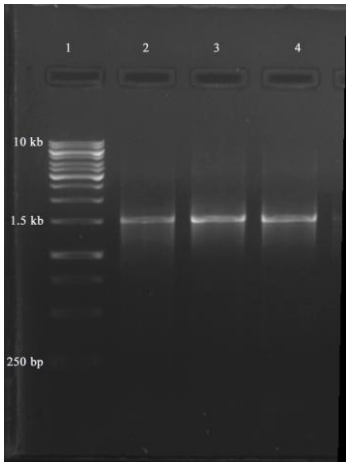

Fig. 3.7: 16S rDNA amplification profile of DNA obtained by method 2

Lane 1: 1 kb ladder, Lane 2–4: 16S rDNA amplicon.

Previous studies suggested that successful PCR amplification is generally used as an indicator of DNA purity (Cullen and Hirsch 1998; Moreira 1998; Burgmann *et al.*, 2001).

From the DNA isolation comparison studies it was understood that all methods yielded an acceptable amount of DNA, but all were not suitable for further downstream processing, except that obtained by method 2. As method 2 employed liquid nitrogen for grinding soil particles, it yielded sheared DNA making it unsuitable for shot gun cloning purposes. And from the 16S rRNA gene amplification profile, it can be concluded that the method can be employed for diversity studies following PCR amplification. For phylogenetic diversity studies, metagenomic DNA from Mangalavanam mangrove sediment (soil 4) and Arabian Sea sediment (soil 5) was isolated employing liquid nitrogen method (Method 2) and was analyzed on agarose gel (Fig. 3.8).

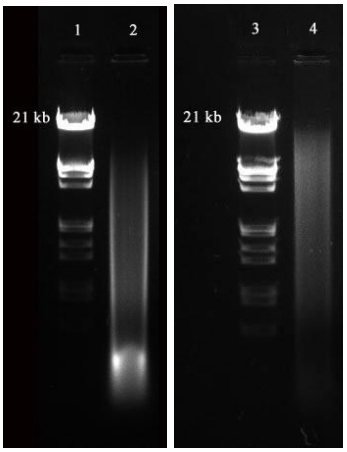

Fig 3.8: Agarose gel electrophoresis of metagenomic DNA.

Lane 1 and 3: 21 kb ladder, Lane 2 and 4: DNA isolated by Method 2 from marine and mangalavanam mangrove sediment respectively.

3.3.1.4 Metagenomic DNA isolation using kit

All the classical metagenomic methods used for comparison yielded DNA with residual humic contaminants, except that employing liquid nitrogen which yielded DNA that was considerably sheared, making it unusable for shot gun library construction. Metagenomic DNA isolation of Mangalavanam mangrove sediment (Soil 4) and Arabian Sea sediment (Soil 5) was therefore done using the commercially available kit (MoBio UltraClean^{TM} soil DNA isolation kit, USA) and this was visualized on agarose gel (Fig. 3.9). DNA isolation following kit yielded high molecular DNA without shearing in both soil 4 and 5, making it suitable for functional library construction.

Commercial DNA extraction kits are now commonly used for extraction of high molecular weight DNA from complex habitats. Studies evaluating various commercial kits to other methods have shown that DNA yield and purity vary based on methodology and soil type.

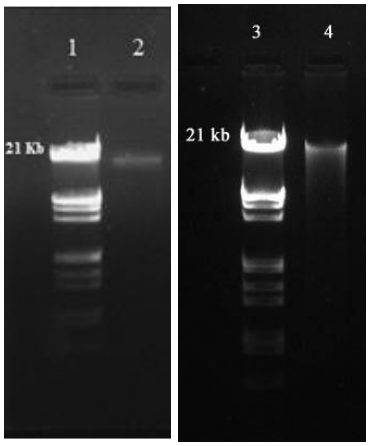

Fig 3.9: Agarose gel electrophoresis of metagenomic DNA.

Lane 1 and 3: 21 kb ladder, Lane 2 and 4: DNA isolated by commercial kit from marine and mangalavanam mangrove sediment respectively.

The mechanism of DNA purification in kits is based on adsorption and desorption of the nucleic acids in presence of chaotropic salts (Gray and Herwig, 1996) which results in contaminants- free DNA, but the quantity of DNA obtained will be less compared to classical methods of DNA extraction. Previous studies recommended that slight modification of protocols employing commercial kits or a combination of classical isolation methods followed by purification of DNA using commercial kits can greatly affect the quantity and quality of the isolated DNA (Martin-Laurent *et al.,* 2001; Lakay *et al.,* 2007; Gutiérrez-Lucas *et al.,* 2014).

Graph showing DNA yield from soil 4 and 5 employing method 2 and kit based method is shown in Fig. 3.10. As depicted in the graph method 2 yielded 251.4 ± 1.7 and 284.9 ± 2.6 µg DNA/gram of soil in comparison to 173.9 ± 1.9 and 184.7 ± 3.3µg DNA/gram of soil employing kit based method from soil 4 and 5 respectively. Despite the higher DNA yield by method 2 from both the soils, the DNA was greatly sheared due to the mechanical grinding with liquid nitrogen making it inappropriate for shot gun library construction.

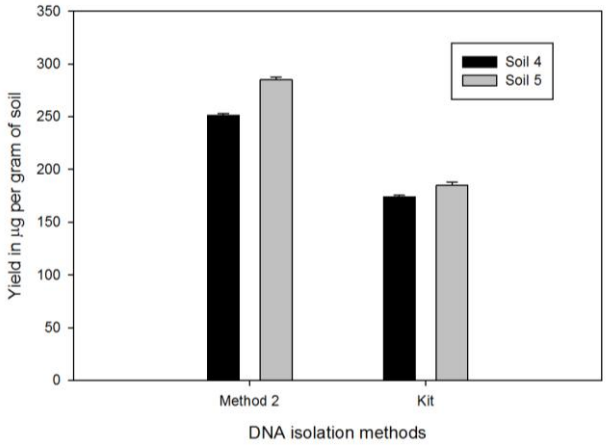

Fig 3.10: DNA yield from soil 4 and 5 by method 2 and kit

The purity ratios of A_{260}/A_{230} and A_{260}/A_{280} were much better in the kit based DNA preparation as shown in Fig. 3.11 and Fig 3.12. Comparing to the classical liquid nitrogen method, kit based method yielded less sheared DNA which is therefore amenable for shot gun cloning processes.

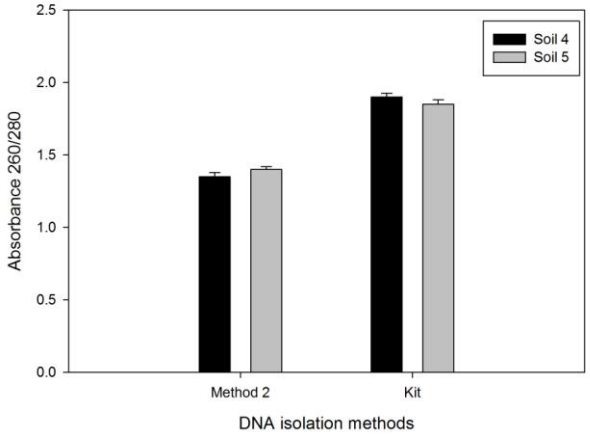

Fig 3.11: Purity of DNA (A_{260}/A_{280}) from soil 4 and 5 by method 2 and kit based method

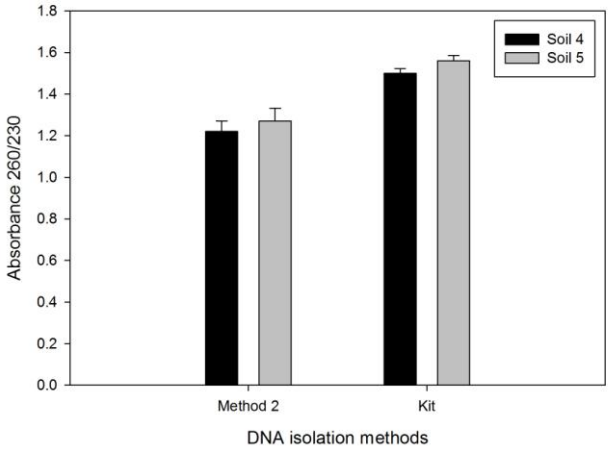

Fig 3.12: Purity of DNA (A_{260}/A_{230}) from soil 4 and 5 by method 2 and kit based method

From the comparative evaluation studies of five classical metagenomic DNA isolation methods, only method 2 (Volossiouk *et al.*, 1995) yielded PCR amenable DNA from all the soil samples tested. Since DNA isolated by method 2 from Mangalavnam mangrove sediment and Arabian Sea sediment was highly sheared, it can be used only for phylogenetic diversity studies. Due to increased DNA shearing a kit based method was also tried, which yielded high molecular weight DNA essential for cloning reactions. So for construction of metagenomic libraries to screen for enzyme production, kit based method proved appropriate to extract unsheared, pure DNA from marine and mangrove sediments.

Chapter 2

PHYOLOGENETIC DIVERSITY OF MARINE AND MANGROVE SEDIMENTS BASED ON ANALYSIS OF 16S rRNA GENE SEQUENCES OBTAINED BY SANGER AND NEXT GENERATION SEQUENCING.

4.1 INTRODUCTION

The marine environment is the largest habitat on Earth, as oceans cover ~70% of the planet surface; and microorganisms survive and grow throughout this environmental niche. These microorganisms are responsible for more than 98% of marine primary productivity (Sogin *et al.,* 2006), playing key roles in marine food webs and in carbon and energy cycles. Arabian Sea, located in the northwestern part of the Indian Ocean, covers an area of about 1,491,000 square miles. However, bacterial composition of the sediments of Arabian Sea and their contribution to the various underlying biogeochemical cycles are largely unknown. Diversity and composition of the microbial community in the sea floor is greatly influenced by seasonal variations, which cause changes in environmental factors like salinity, temperature, nutrient availability etc. All these factors make marine sediments a rich resource of novel and rare bacterial phylotypes, necessitating a thorough exploration of their diversity to understand the contribution of the unknown ones.

Similarly mangrove ecosystems represent a large area (60-70%) of the coastlines in the tropical and subtropical regions of Earth. The mangrove ecosystems are necessary for maintenance of sea level and for protection of the coast (Duke *et al.,* 2007). The Mangrove forests are by far one of the most productive ecosystems on the planet and due to the availability of rich sources of

nutrients, mangrove ecosystems are called the homeland of microorganisms (Sahoo and Dhal, 2009). Bacterial communities play essential roles in the functioning and maintenance of mangrove ecosystem. Environmental parameters specific to this ecosystem are salinity, anaerobic condition caused by tidal variation which in turn results in a redox potential that ranges from –200 mV to +150 mV (Clark *et al.*, 1998; Holguin *et al.*, 2001). These environmental conditions make mangroves hotspots of microbial diversity.

Traditional culture dependent methods could hardly estimate microbial diversity as well as the structure and function of microbial ecosystems comprehensively and accurately. Culture-independent or metagenomic analysis provides a relatively unbiased view about the microbial diversity, potential metabolic pathways and novel biomolecules present in various environments. Among the various culture-independent tools, diversity analysis targeting the 16S rRNA gene is nevertheless, a widely employed technique to describe the composition of complex microbial community. 16S rDNA clone library analysis based on Sanger sequencing technique is a widely employed method for bacterial community analysis. Even though it is labor-intensive and time consuming, these methods still provide an accurate picture of the bacterial community prevailing in a particular locale at various taxonomic levels. However recent advances in next-generation sequencing techniques have facilitated large-scale exploration of the geographic distribution and taxonomic diversity of bacterial communities from wide range of environments. Similarly, sequencing variable regions within the 16S rRNA gene employing Illumina sequencing platform provides a quick, affordable and comprehensive estimation of bacterial communities (Barriuso *et al.*, 2011).

The objective of this study was to investigate the phylogenetic diversity of bacterial communities present in marine and mangrove sediments. This is one of the first study undertaken to understand bacterial populations in the sediments of the Arabian Sea and Mangalavnam mangrove ecosystem based on Sanger and

Illumina sequencing. This has contributed to our understanding of sediment bacterial populations and to some extent their ecological functions.

4.2 MATERIALS AND METHODS

All protocols explained in this chapter were applicable to both Arabian Sea and Mangalavanam mangrove sediments unless otherwise mentioned.

4.2.1 Analysis of bacterial diversity based on 16S rRNA gene employing Sanger sequencing method

Phyologenetic diversity studies of marine and Mangalavanam mangrove metagenome based on Sanger sequencing method was done by amplification of the ~1.5 kb size 16S rRNA genes in the metagenome and construction of the phylogenetic clone library, followed by sequencing and *in silico* analysis.

4.2.1.1 PCR amplification of 16S rRNA gene from isolated marine and mangrove metagenomic DNA

The 16S rRNA gene was amplified from the Mangalavanam mangrove and marine metagenomic DNA using universal primers as described in table 3.3, following appropriate dilutions (~80-100 ng) of metagenomic DNA template, with a PCR mix composition as given in table 3.4. PCR amplification was carried out in a Thermal Cycler (Applied Biosystems, USA) using the program described in section 3.2.1.6. Agarose gel electrophoresis was carried out for visualization of PCR products and gel pictures were captured as described in section 3.2.1.4. The PCR products were purified using the GeneiPure™ Gel Extraction (GeNei, India.) according to manufacturer's instructions and the purified PCR products were used for 16S rDNA library construction of both marine and mangrove sediments.

4.2.1.2 Construction of 16S rDNA library

The purified PCR products were ligated into a pGEM®-T vector systems using TA cloning kit (Promega, USA) following the manufacturer's

protocol and transformed onto *E. coli* JM109 host cells using Transform Aid Bacterial Transformation Kit (Thermo Scientific, USA). The transformed cells were plated on Luria Bertani agar (HiMedia, India) plates containing Ampicillin (SRL, India), X-Gal (Thermo Scientific) and IPTG (Thermo Scientific) (Appendix I) and incubated at 37ºC overnight with appropriate control plates. The clones which appeared white on the plates were selected as recombinants and patched onto LB agar plates with ampicillin (Appendix I), and constituted the two phylogenetic library of marine and mangrove sediments; they were also maintained as stock in 20% glycerol at -80^{o}C.

4.2.1.3 Glycerol stocking

The clones were inoculated in LB broth containing ampicillin. After 24 h growth at 37°C, 800 µL of culture broth was mixed with 200 µL of 100% glycerol (HiMedia), giving a final concentration of 20% and stored at -80°C for long term storage; and at -20°C for use as working stock.

4.2.1.4 Isolation of plasmids from phylogenetic clones

Plasmids from selected recombinant clones from marine and mangrove library were isolated by alkaline lysis method which is a modification of the methods of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981). The method relies on bacterial lysis by sodium hydroxide and sodium dodecyl sulfate (SDS), followed by neutralization with a high concentration of low-pH potassium acetate. This gives selective precipitation of the bacterial chromosomal DNA and other high molecular-weight cellular components. The plasmid DNA remains in suspension and is precipitated with ethanol.

A single isolated colony was picked from an LB agar plate and inoculated in a test tube containing 3 mL of LB broth containing ampicillin (60 µg/mL) and incubated overnight at 37ºC with shaking. 1.5 mL of the culture was centrifuged (Sigma) in a microfuge tube for 5 min at 6000 rpm at 4ºC, the supernatant was

discarded and the pellet was air dried. The bacterial pellet was resuspended in 100 µL of ice cold Solution I (Appendix I) by vortexing and kept on ice for 5 min. About 200 µL of freshly prepared Solution II (Appendix I) was added, mixed by inverting the tube gently and the tubes were stored on ice for 10 min. About 150 µL of ice cold Solution III (Appendix I) was added, mixed by inverting the tube gently and the tube stored for 15 min on ice. The tubes were centrifuged for 10 min at 12,000 rpm and the supernatant was transferred carefully to a fresh tube avoiding the white pellet.

Equal volume of phenol: chloroform (24:1 v/v) mixture was added to the supernatant solution, mixed gently, centrifuged at 10000 rpm for 5 min and the upper aqueous layer was transferred to a new microfuge tube. The plasmids was precipitated by adding two volumes of ethanol to the supernatant, mixed well by inverting the tube several times and allowed to stand for 30 min in ice. The precipitated plasmids were collected by centrifuging (Sigma) at 12,000 rpm for 10 min and the supernatant was discarded. To the DNA pellet, 1 mL of ice cold 70% ethanol was added, centrifuged for 30 sec and the supernatant was discarded. The pellet was air dried and resuspended in 50 µL of sterile deionised H_2O and stored at -20ºC. Plasmids were visualized by agarose gel and electrophoresis was carried out as described in section 3.2.1.4.

4.2.1.5 Confirmation of recombinants containing inserts

Confirmation of recombinants was performed by reamplification of 16S rDNA inserts from the plasmids employing 16S rRNA gene universal primers. PCR amplification of the 16S rDNA insert was done with 50 ng of the plasmid DNA as template following the methodology as described in section 3.2.1.6. Reamplified 16S rDNA amplicons were visualized using agarose gel electrophoresis as described in section section 3.2.1.4.

4.2.1.6 Sequencing and *in silico* analysis for phylogenteic diversity studies

16S rDNA inserts within the recombinant plasmids were partially sequenced employing Sanger's Dideoxy method using ABI 3730 Excel (Applied Biosystems, USA) at Scigenom Labs, Kochi, Kerala. The identity of the sequences was determined by comparing the sequences obtained with those available in the GenBank database using Basic Local Alignment Search Tool (BLAST) (Altschul *et al*., 1990) at NCBI website. (http://blast.ncbi. nlm.nih.gov). The sequences were compiled and multiple sequence alignment was done using ClustalX 2.1 program (Larkin *et al*., 2007). Taxonomical hierarchy was assigned to the sequence using Ribosomal Database Project (RDP) Naive Bayesian rRNA Classifier Version 2.5 (Wang *et al*., 2007) and the distribution of various phyla was represented as pie diagram.

Raw sequences were processed through the RDP pipeline (http://wildpigeon.cme.msu.edu/) (Cole *et al*., 2013). Aligned sequences were clustered into operational taxonomic units (OTUs) defined at 3%, 5%, 10 % and 20% genetic distance level using complete-linkage clustering and rarefaction curve were constructed by plotting the number of OTUs observed against the number of sequences sampled to estimate the species richness in the samples. Phylogenetic tree were constructed using MEGA software version 5.0 (Tamura *et al*., 2007). Tree topology was deduced by Neighbour-Joining method (Saitou and Nei, 1987) using 1000 bootstrap iterations. The sequences were deposited as Sequin file to GenBank database and accession numbers were obtained for the submissions.

4.2.2 Analysis of bacterial diversity based onV3 regions of 16S rRNA gene employing Next Generation Sequencing (NGS) method

Phyologenetic diversity studies of marine and mangalavanam mangrove metagenome based on NGS method was done by amplification of the

hypervariable region 3 (V3) of 16S rRNA gene followed by Illumina sequencing and *in silico* analysis.

4.2.2.1 PCR amplification of V3 regions of 16S rRNA gene from marine and mangrove metagenomic DNA

Appropriate dilution of metagenomic DNA from marine and mangrove sediments were used as template to amplify the ~ 200 bp hypervariable region 3 (V3) of the 16S rRNA gene. The forward and reverse primers used for the amplification is given in Table 4.1.

Table 4.1: Primers used to amplify V3 region of 16S rRNA gene

Primer	Sequence	Reference
341 Forward	5' CCTACGGGAGGCAGCAG 3'	Klindworth *et. al.*, 2012
518 Reverse	5' ATTACCGCGGCTGCTGG 3'	

The concentration of different ingredients used for PCR amplification is given in Table 4.2.

Table 4.2: PCR Mix composition for V3 amplification

Ingredient	Concentration	Volume
Template DNA	50 ng/μL	1 μL
341 Forward primer	10 pmol/μL	2 μL
518 Reverse primer	10 pmol/μL	2 μL
dNTPs	40 mM	0.5 μL
Phusion HF reaction buffer	5X	5 μL
F-540Special Phusion HS DNA Polymerase	2 U/μL	0.2 μL
Sterile distilled water		upto 25 μL

PCR amplification was carried out using the following program

Step	Temperature	Time	
Initial Denaturation	98°C	30 sec	
Denaturation	98°C	10 sec	
Annealing	72°C	30 sec	30 cycles
Extension	72°C	5 sec	
Final Extension	72°C	2 min	

The PCR product was bound by adapter sequences on either side which is compatible with the Illumina index and sequencing adapters. The agarose gel electrophoresis was carried out for the visualization of PCR products as described in section 3.2.1.4. The PCR product with the bound adapter was gel extracted using the Gel and PCR Clean-up kit (Macherey-Nagel, Germany) according to manufacturer's instructions and the purified PCR product was used for a second PCR reaction which attached Illumina sequencing adapters and dual-index barcodes to the amplicon target.

151bp x 2 paired end Illumina sequencing was performed using MiSeq Genome Analyser at Scigenom Labs, Kochi, Kerala. Second set PCR primers with illumina bar code sequences used for the analysis are the proprietary sequences of the company hence the primer sequences are not provided.

4.2.2.2 *In silico* analysis for phylogenetic diversity

Sequences were checked for quality parameters such as base quality score distributions, average base content and GC distribution in the reads, quality filtering, operational taxonomic units (OTUs) picking and annotation followed by diversity analysis.

4.2.2.2.1 Identification of V3 region from paired-end reads

The paired-end sequence reads contain some portion of conserved region, spacer and V3 region. As a first step the spacer and conserved regions were removed from paired-end reads. After trimming the unwanted sequences from original paired-end data, a consensus V3 region sequence was constructed using ClustalW program. Quantitative Insights into Microbial Ecology (QIIME) program (Caporaso *et al.,* 2010) was employed to filter and analyze the raw tags. Multiple filters such as conserved region filter, spacer filter, read quality filter and mismatch filter were employed to obtain high quality V3 region sequences for

various downstream analyses. Reads were quality filtered using an average phred score of 20 (Q20) during demultiplexing. Sequences with a mean quality score <20 were excluded from analysis, and chimeras were also excluded using ChimeraSlayer detection method available in QIIME software.

4.2.2.2.2 Identification of Operational Taxonomy Units (OTU)

V3 sequences from the metagenome were combined and the combined reads are clustered into Operational Taxonomic Units (OTUs) based on their sequence similarity using UCLUST program (Edgar, 2010) (similarity cutoff = 0.97). OTUs with low read count (< 10 reads) were filtered out from further analysis. The reads from filtered OTUs were processed using QIIME program (Caporaso *et al.,* 2010) to construct a representative sequence for each OTU. The representative sequence was aligned to the Greengenes core set reference databases using PyNAST program. Further, taxonomy classifications to phylum, class, order, family, genus and species of bacteria were performed using RDP classifier (Wang *et al.,* 2007) and Greengenes OTUs database (DeSantis *et al.,* 2006). Heatmap were generated using QIIME pipeline. Shannon, Chao1 and observed species metrices were calculated using QIIME software.

4.2.2.2.3 Diversity analysis

Alpha diversity (within samples) and beta diversity (among samples) (Lozupone *et al.,* 2007) were calculated with QIIME software package. In alpha diversity analysis, Shannon, Chao1 and observed species metrices were calculated. The Shannon metric is the measure to estimate observed OTU abundances, and accounts for both richness and evenness thereby estimating the diversity of species. The Chao1 metric estimates the species richness and observed species metric is the count of unique OTUs identified in the sample. Richness is a measure of the number of different kinds of species present in a particular sample or area, whereas evenness is a measure of the relative abundance of the different species making up the richness of a sample or area. Rarefaction curves were generated

based on these three metrices. Beta diversity represents the explicit comparison of microbial communities based on their comparison. In this study the comparison was performed between the Arabian Sea and the Mangalavanam mangrove metagenome. Distance matrixes were calculated using UniFrac approach in which both weighted and unweighted unifrac distances were calculated.

4.3 RESULTS AND DISCUSSION

4.3.1 Analysis of bacterial diversity of Arabian Sea sediment based on 16S rRNA gene employing Sanger sequencing method

To study phylogenetic diversity of marine metagenome based on Sanger sequencing method, 16S rRNA genes were amplified and used for the construction of the phylogenetic clone library followed by sequencing and *in silico* analysis.

4.3.1.1 PCR amplification of 16S rRNA gene from isolated marine metagenomic DNA

The 16S rRNA gene was amplified from the marine metagenome and the1.5 kb amplicon was visualized on agarose gel (Fig. 4.1). The16S rRNA-based bacterial diversity study of the oxygen minimum zone of Arabian Sea targeted the 1,099 bp fragment of 16S rRNA gene (Divya *et al.*, 2011).

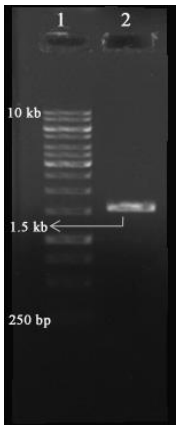

Fig 4.1: Agarose gel electrophoresis of amplified 16S rRNA gene
Lane 1: 1 kb ladder, Lane 2: 16S rDNA amplicon

4.3.1.2 Construction of 16S rDNA phylogenetic library

The 1.5 kb 16S rDNA amplicons from the marine metagenome were ligated into pGEM®-T vector systems and transformed onto competent *E. coli* JM109 host cells. The transformed cells were plated on Luria Bertani agar plates containing ampicillin, X-Gal and IPTG, followed by blue-white screening. The clones that appeared white on the plates were selected as recombinants and constituted the marine phylogenetic library.

4.3.1.3 Plasmid isolation and reamplification of 16S rDNA inserts from phylogenetic clones

Recombinant plasmids from phylogenetic clones were isolated and used as template for reamplification of 16S rDNA inserts. The presence of the 16S rDNA inserts in the recombinant plasmids was confirmed by reamplification of inserts. Marine phylogenetic library with 105 clones were selected for further analysis. Inserts within the 105 recombinant plasmids were sequenced and analysed.

4.3.1.4 *In silico* analysis of 16S rDNA inserts from marine library for phylogenetic diversity studies

The identity of the partial 16S rDNA sequences was determined using nucleotide BLAST of the NCBI database. The 105 sequences were submitted to GenBank and accession numbers were obtained. The accession numbers obtained were KF453864 to KF453950, KF569952 to KF569958, KF569962 to KF569964, KF569966 to KF569972 and KF569960.

The sequences were classified into taxonomic hierarchy using RDP classifier, with the 105 sequences falling into seven phyla of bacterial domain. Sequences were distributed amoung the phylum *Proteobacteria, Bacteriodetes, Actinobacteria, Firmicutes, Chloroflexi, Deferribacter* and *Acidobacterium*. The distribution of the clones among different bacterial phyla is depicted in Fig. 4.2.

The molecular phylogenetic analysis in the present study revealed the occurrence of the bacterial 16S rRNA gene sequences that were unique, as well as sequences that were previously reported in other marine sediments, but which are phylogenetically distinct from those in terrestrial environments.

The predominant phyla identified in sediments of Arabian Sea were *Proteobacteia*, with the 66 clones representing the phylum constituting 62.8% of the library. The taxonomic positions of the clones as identified by RDP classifier with 80% confidence threshold is tabulated in Table 1 (Appendix II) with few clones classified upto genus level and the remaining clones classified upto family level with the selected cut-off value.

Proteobacteria are reported in various environments and play essential roles in nutrient cycling. As suggested by previous studies, *Proteobacteria* can dominate the bacterial community even in deep-sea environments (Liao *et al.*, 2009). Previous diversity analysis of marine sediments from the South China Sea also identified the dominance of phylum *Proteobacteria* (Xu *et al*., 2004; Lai *et al*., 2006; Li *et al*., 2008a; 2008b).

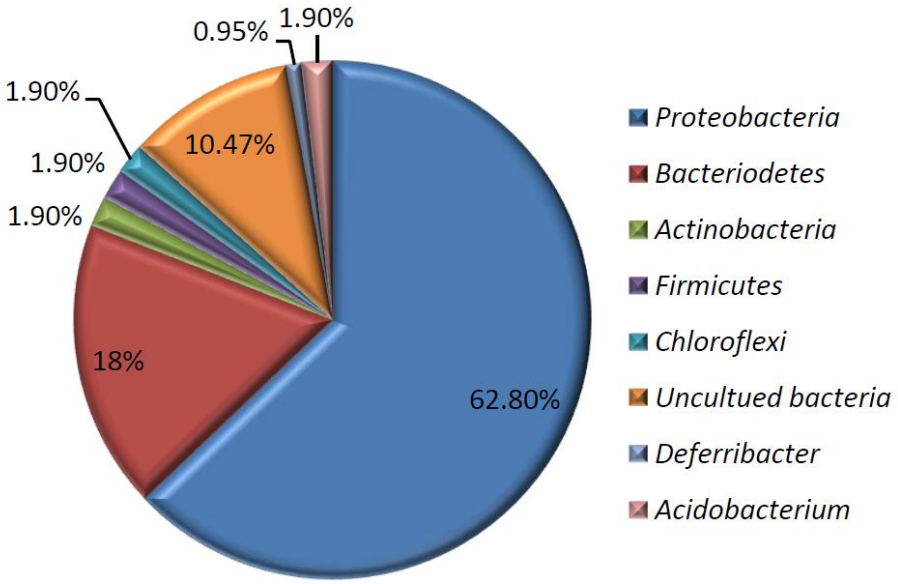

Fig. 4.2: The phylogenetic diversity of the Arabian sea sediment metagenome based on 16S rDNA sequence analysis

Marine waters contain large amounts of organic and inorganic materials which will in turn accumulate in the sea floor making it a suitable ecosystem for highly diverse microbial populations. Marine microbial communities mediate biogeochemical cycles such as carbon, nitrogen and sulphur cycle, thereby playing pivotal roles in preventing environmental changes such as warming and ocean acidification and maintaining the balance of marine ecosystem (Fuhrman, 2009; Graham *et al.*, 2012).

Phylum *Proteobacteria* encompass a large number of morphologically distinct, physiologically and metabolically diverse microbial population that are of great significance to global carbon, nitrogen and sulfur cycling. Despite the abundance and identification of large number of cultured isolates in this phylum in than any other, the vast majority of soil *Proteobacteria* are yet to be cultured (Kersters *et al.*, 2006). It is suggested that specific isolation and enrichment methods like acidic, hypoxic or anoxic conditions may prove useful in obtaining

these lineages in pure cultures (Spain *et al.*, 2009). In marine sediments, bacteria gain energy and carbon either by oxidizing organic compounds or by chemical energy, using hydrogen, methane, hydrogen sulphide and iron (Jørgensen & Boetius, 2007). Since oxygen penetrates less into the organic-rich sediments, anaerobic mineralization is of prime importance. So anaerobic or facultative aerobic bacteria mainly oxidize these organic compounds by reducing inorganic electron acceptors in a sequence of Mn oxides, nitrate, Fe oxides, sulphate, and ultimately CO_2 (Canfield *et al.*, 1993). In marine ecosystems, sulphur cycle has been proved to be the important biogeochemical factor that controls the flow of electrons in prevailing anaerobic condition (Pester *et al.*, 2012).

In the present study, the dominant phylum *Proteobacteria* was represented by three classes namely, *Alphaproteobacteria, Gammaproteobacteria* and *Deltaproteobacteria*, and their distribution is represented in Fig. 4.3. In the Arabian Sea sediment metagenome, the class *Alphaproteobacteria* comprising 58.46% of the clones was the major group, followed by *Gammaproteobacteria* (38.46%) and *Deltaproteobacteria* (3.07%). The class *Alphaproteobacteria* comprising 38 clones was represented by the order *Rhodospirillales, Rhodobacterales* and *Rhizobiales,* which included members of the genus *Pelagicola, Citreicella, Aurantimonas, Filomicrobium* and *Oceanibulbus.* Similar results were obtained in bacterial diversity studies on sediments of Kusuura Bay, Japan where class *Alphaproteobacteria* consistently dominated the microbial community throughout the study period (Kunihiro *et al.*, 2011).

Bacterial diversity study on sediments of Sagami Bay, Japan identified the predominance of bacteria within the phylum *Proteobacteria* representing (86%) of the clone libraries. Of the clones within *Proteobacteria*, 52% were related to the *Gammaproteobacteria* followed by *Deltaproteobacteria* representing 24% of the total bacterial clones and three clones were related to the *Cytophaga-Flavobacterium-Bacteriodes* (CFB) group (Fang *et al.*, 2006).

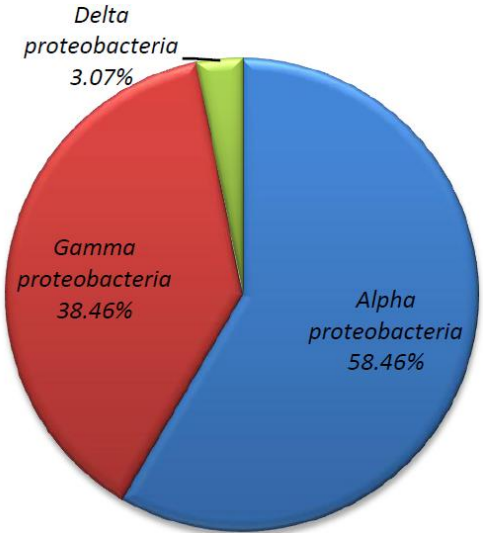

Fig. 4.3: Distribution of phylum *Proteobacteria* in the Arabian Sea sediment

Alphaproteobacteria are diverse class of organisms within the phylum *Proteobacteria* playing important roles in organic matter degradation in marine sediments. Members of the class may participate in a variety of metabolic strategies, including photosynthesis, ammonia oxidation, nitrogen fixation and methylotrophy. *Alphaproteobacteria* relative to other phylogenetic groups contributes to uptake of low-molecular weight dissolved organic matter such as amino acids, protein, glucose and N-acetyl-glucosamine in various marine environments (Cottrell and Kirchman, 2000; Malmstrom *et al.*, 2005; Elifantz *et al.*, 2007; Yokokawa and Nagata, 2010).

Therefore the dominant presence of *Alphaproteobacteria* in the benthic sediments may be attributed to the significant role that these bacteria in decomposing and assimilating the organic matter in the organically enriched sediment.

The Class *Gammaproteobacteria* with 25 clones was represented by orders *Vibrionales*, *Alteromonadales* and *Oceanospirillales*, including members of the genus *Halomonas*, *Aliidiomarina* and *Marinobacter* in the

Arabian Sea metagenome. A number of gammaproteobacterial clones identified showed similarity towards bacteria involved in sulfur cycling. Class *Deltaproteobacteria*, with two clones was represented by the order *Desulfobacterales*.

Studies on sediments of Sagami Bay, Japan showed that high GC - Gram-positive bacteria and *Gammaproteobacteria* were dominant species at the location (Urakawa *et al.*, 1999). Preliminary research on microbial diversity of Parece Vela Basin of Pacific Ocean by culture-independent method identified the dominance of alpha and gamma proteobacterial species (Xie *et al.*, 2005). Studies at a deep-sea station of the Pacific nodule province showed that *Gammaproteobacteria* was the most primary bacterial group in the sediments (Xu *et al.*, 2005). Similarly *Gammaproteobacteria* dominated the sediments of Nankai trough (Li *et al.*, 1999).

The oxidation of hydrogen sulfide is essential to sulfur cycling in marine habitat. Members of class *Gammaproteobacteria* are associated with bioconversion of sulphur containing organic molecules- sulfur-oxidizing bacteria (S-oxidisers), while most of the *Deltaproteobacteria* were sulphate reducing bacteria. Sulfur-oxidising bacterial strains play key roles in detoxification of sulphide in marine sediments whereas the sulfur-reducing bacterial community is vital in organic carbon oxidation in marine sediments. This observation is supported by the fact that sulphate is one of the major electron acceptors present in these environments and that the reduction of sulphate may be an important pathway of organic matter mineralization in organic rich sediments (Pester *et al.*, 2012).

Among the 105 clones, *Bacteroidetes* was the second prominent phylum, with 19 clones representing the phylum and constituting 18% of the library, making this a prominent phylum amid the non-proteobacterial group. Members of

Bacteroidetes was represented by the orders *Flavobacteriales* and *Sphingobacteriales* with members in the genus *Formosa, Salegentibacter* and *Aurantimonas.*

At the phylum level 11 clones, representing 10.47% of the constructed library did not fit into any taxonomic hierarchy, hence they were grouped as uncultured bacteria, which indicated the possibility of novel yet- to-be- cultured bacteria present in marine sediments which awaits discovery. Minor representatives from phylum *Actinobacteria, Firmicutes, Chloroflexi, Acidobacterium* were also obtained representing 1.90% each. The phyla *Deferribacteres* was represented by a single clone in the library belonging to the orders *Deferribacterales* and to the genus *Caldithrix.* In the present analysis, clones representing *Epsilonproteobacteria* were absent; which was also observed in previous findings, wherein members of this phylum were absent or scarce in other clone libraries of coastal marine sediments (Gray and Herwig, 1996; Urakawa *et al*., 1999; Asami *et al.,* 2005).

In a similar study on bacterial diversity in deep-sea sediment from northeastern Pacific Ocean based on bacterial 16S rRNA gene library of 79 clones, 11 phylotypes were identified in which *Gammaproteobacteria* (22.8%) and *Alphaproteobacteria* (16.5%) were the dominant components of the sediment bacterial community, followed by *Planctomycetacia* (7.6%), *Deltaproteobacteria* (6.3%), *Nitrospira* (6.3%), *Actinobacteria* (6.3%), *Betaproteobacteria* (5%), *Acidobacteria* (5.1%), *Sphingobacteria* (3.8%), *Firmicutes* (2.5%) and 17. 7 % were identified as uncultured bacteria (Hongxiang *et al.*, 2008). Studies on microbial community distribution in the Baltic Sea sediments of coast of Sweden, showed that most abundant active bacteria belonged to phylum *Proteobacteria* which is distributed among the Alpha, Beta, Delta and Gammaproteobacterial classes, followed by representatives of the phylum *Bacteroidetes, Chloroflexi, Actinobacteria and Planctomycetes* (Edlund *et al.,* 2008)

Previous studies suggested that *Chloroflexi* (or green non-sulfur bacteria) dominates organic-rich marine sediments. A study on sediments from the eastern Mediterranean Sea had shown that more than 70% of the total bacteria belong to uncultured green non-sulfur bacteria, and identified presence of organic-rich sediments in the location (Coolen *et al*., 2002). Chloroflexi was also shown to be abundant in organic-rich, methane hydrate-free sites of the Pacific Ocean margins, but show very small representation in hydrate-rich sediment (Inagaki *et al*., 2006). Similarly *Planctomycetes* were also abundant in these methane hydrate-rich sediments.

However, in the present study since only two clones with identity to green non-sulfur bacteria were observed, while none showed affiliation to phylum *Planctomycetes, it* can be indication to the prevalence of organic rich, methane hydrate-free environment in the area of study. Although phylum *Acidobacteria* with a few culturable representatives are rarely encountered in marine habitats (DeLong *et al.*, 2006; Quaiser *et al.*, 2008), two clones affiliated to this phylum were obtained in the present study.

Phylogenetic tree of partial 16S rRNA gene sequences of proteobacterial and non- proteobacterial clones were constructed separately and depicted in Fig. 4.4 and 4.5. The analysis involved 66 proteobacterial nucleotide sequences (N=66) and 39 non-proteobacterial nucleotide sequences (N=39). *Pseudoalteromonas lipolytica* (Accession number JQ905098) was used as outgroup. Accession numbers of clones are given in parentheses. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches and scale bar shows sequence divergence.

Fig 4.4: Phylogenetic tree of partial 16S rRNA gene sequences of proteobacterial clones (N=66) obtained from Arabian Sea sediments

Out of 66 proteobacterial clones, the predominant class *Alphaproteobacteria* clustered together and formed a separate clade except three clones belonging to order *Rhizobiales* and family *Hyphomicrobiaceae* which were grouped in between the gammaproteobacterial clade. The second largest group *Gammaproteobacteria* represented by 25 clones formed another major clade. *Deltaproteobacteria* (2 clones) also stood distinctly based on their phylogenetic relationship. The outgroup *Pseudoalteromonas lipolytica* was noticeably separated from all these clones.

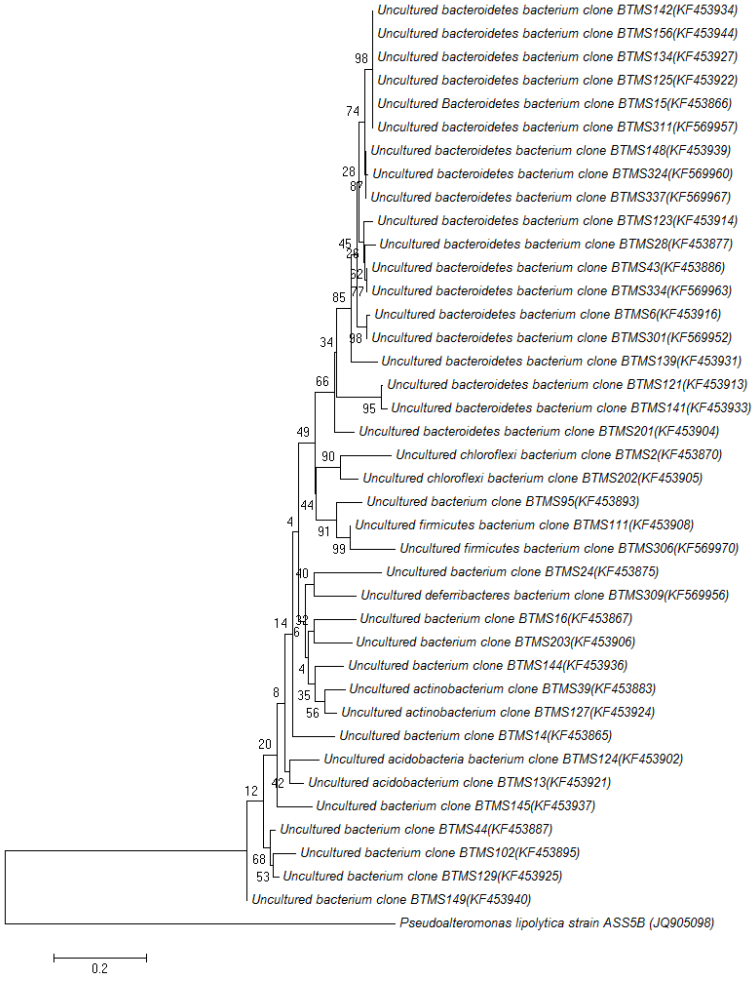

Fig 4.5: Phylogenetic tree of partial 16S rRNA gene sequences of non-proteobacterial clones (N=39) obtained from Arabian Sea sediments

In the phylogenetic tree of partial 16S rRNA gene sequences of non-proteobacterial clones, *Bacteroidetes* the prominent phylum among the non-proteobacterial group clustered together and formed a separate clade. The second abundant group, the uncultured bacterial group also claded separately from the *Bacteroidetes*; with the minor representatives grouped in between the uncultured bacterial group indicating the phylogenetic similarity of the minor representatives to the unclassified ones.

4.3.1.5 Determination of species richness by rarefaction curve

The sequences were aligned and clustered into Operational Taxonomic Units (OTUs) based on the genetic distance. OTUs were identified at genetic distances of 3%, 5%, 10% and 20% by using 105 sequences and rarefaction curve was plotted using the RDP Pipeline (Fig. 4.6). At 3% sequence divergence, 62 different OTUs were observed within the 105 sequences sampled indicating the species richness of the sample.

Ninety OTUs were identified among the 115 sequences from the Arabian sea oxygen minimum zone clone library (Divya *et al*., 2011) whereas 74 OTUs were detected out of the 87 sequences from the anoxic sediment clone library of South China Sea (Liao *et al*., 2009). Forty seven OTUs were detected from the 60 sequences from the sediment clone library of North Sea (Wegener *et al*., 2008) while 27 OTUs were identified from the clone library of cold seep sediments from the Gulf of Mexico out of the 28sequences analyzed (Orcutt *et al.,* 2010).

At 20% sequence divergence, the rarefaction curve reached saturation, indicating that the sampling effort covered almost the full extent of taxonomic diversity at these genetic distances at the phylum level. At 3% and 5% and 10% genetic distance, the rarefaction curves were not saturated indicating that full extent of taxonomic diversity at these genetic distances were not sufficiently sampled. However, a substantial fraction of the bacterial diversity within the

Arabian Sea sediment samples was assessed at phylum level by the sampling effort targeting the 16S rRNA gene. Maximal diversity can be accessed by targeting more specific phylogenic anchor regions within the 16S rRNA gene like the variable regions V3-V6 which can infer a more accurate picture about the bacterial diversity prevailing in that particular locale.

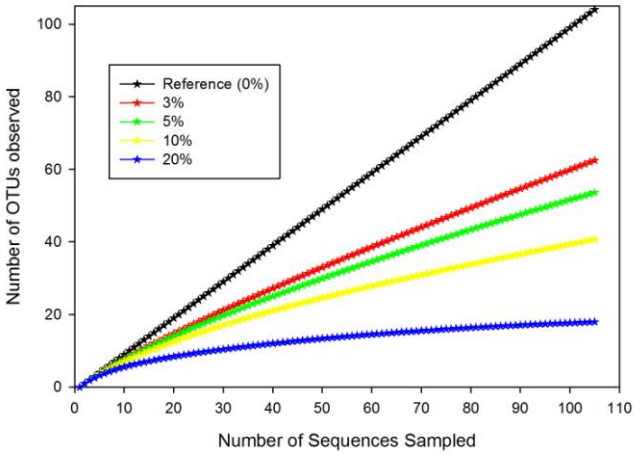

Fig. 4.6: Rarefaction curves indicating the observed number of operational taxonomic units (OTUs) at genetic distances of 3, 5, 10 and 20%. The reference line represents the OTUs at 0% genetic distance.

This study provides a preliminary insight into the microbial diversity in sediments from the eastern region of Arabian Sea based on partial sequencing of 16S rRNA gene, indicating the diverse microbial community prevailing in the location. Further studies based on next generation sequencing, targeting V3 regions of 16S rRNA gene can infer more about the microbial diversity in the study area.

4.3.2 Analysis of bacterial diversity of Mangalavanam mangrove sediment based on 16S rRNA gene employing Sanger sequencing method

The aim was to study phyologenetic diversity of Mangalavanam mangrove metagenome based on Sanger sequencing method. Here 16S rRNA gene was amplified and used for construction of phylogentic clone library followed by sequencing and *in silico* analysis.

4.3.2.1 PCR amplification of 16S rRNA gene from mangrove metagenomic DNA

The 16S rRNA gene was amplified from the Mangalavanam mangrove metagenome and the 1.5 kb amplicon was visualized on agarose gel (Fig. 4.7).

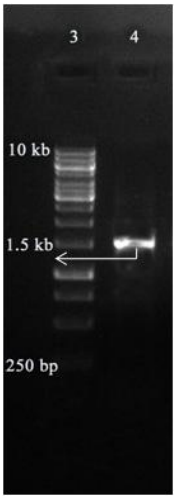

Fig 4.7: Agarose gel electrophoresis of amplified 16S rRNA gene from Mangalavanam mangrove metagenomic DNA

Lane 3: 1 kb ladder, Lane 4: 16S rDNA amplicon

4.3.2.2 Construction of 16S rDNA phylogenetic library

The 1.5 kb 16S rDNA amplicons from the mangrove metagenome were ligated into pGEM®-T vector systems and transformed onto competent *E. coli* JM109 host cells. The transformed cells were plated on Luria Bertani agar plates containing ampicillin, X-Gal and IPTG, followed by blue-white screening. The

clones that appeared white on the plates were selected as recombinants and constituted the mangrove phylogenetic library.

4.3.2.3 Plasmid isolation and reamplification of of 16S rDNA inserts from phylogenetic clones

Recombinant plasmids from phylogenetic clones were isolated and used as template for reamplification of 16S rDNA inserts. The presence of the 16S rDNA inserts in the recombinant plasmids was confirmed by reamplification of inserts. Mangalavanam mangrove phylogenetic library with 34 clones were selected for further analysis. Inserts within the 34 recombinant plasmids were sequenced and analysed.

4.3.2.4 *In silico* analysis of 16S rDNA inserts from mangalavanam mangrove library for phylogenteic diversity studies

The identity of the partial 16S rDNA sequences was determined using nucleotide BLAST of the NCBI database. The 34 sequences were submitted to GenBank and accession numbers were obtained. The accession numbers obtained were JX465646 to JX465653, JX852421 to JX852429 and KC143083 to KC143099.

The sequences were classified into taxonomic hierarchy using RDP classifier with the 34 sequences falling into 7 major phyla of bacterial domain, namely *Proteobacteria, Firmicutes, Acidobacteria, Actinobacteria,* and *Planctomycetes, Chloroflexi*, Candidate phylum WS3 and 5.88% of the clones were found similar to uncultured bacterium, i.e. which doesn't fall into any taxonomic hierarchy. Even this small scale study with only 34 clones revealed the species richness in the study area with bacteria falling in 7 phyla of bacterial domain, further indicating the species diversity in this mangrove sample. At the same time for the Arabian Sea sediment, 105 clone sequences were analyzed to obtained taxonomic diversity, so the study on Mangalavanm mangrove sediment

was limited with only 34 clones. The distribution of the clones among different bacterial phyla is as depicted in Fig. 4.8. The dominant phylum in the mangrove sediments was *Proteobacteria*, representing 44.11% of the total clones in the library. The taxonomic position of the clones as identified by RDP classifier with 80% confidence threshold is tabulated in Table 2 (Appendix II) with few clones classified upto genus level, and the remaining clones classified upto class or family level with the selected cut-off value.

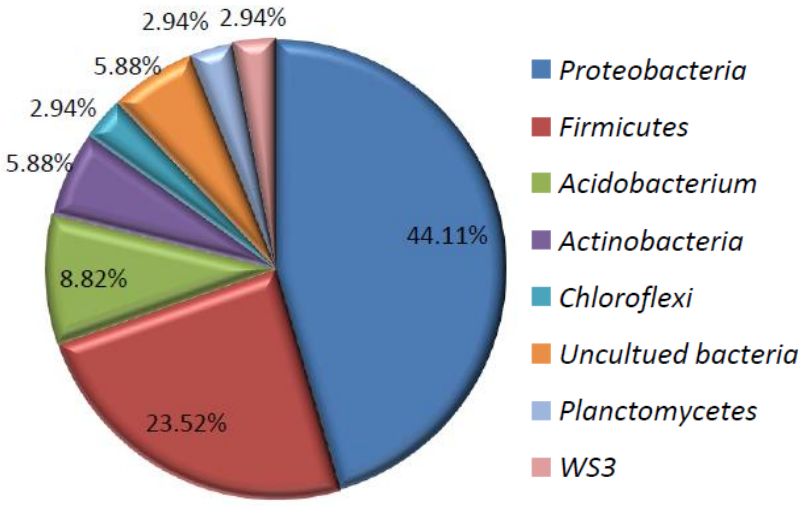

Fig. 4.8: The phylogenetic diversity of the Mangalavanam mangrove sediment metagenome based on 16S rDNA sequence analysis

Mangrove ecosystems are generally nutrient rich with high microbial diversity. Except the thin aerobic surface layer, mangrove sediments are mostly anaerobic in which anaerobic biochemical processes are catalyzed by sediment microbial communities especially members of *Deltaproteobacteria* such as *Desulfosarcina/Desulfococcus* group (Lyimo *et al.*, 2009). Microbes are important in controlling the chemical environment of the mangrove sediments and sulfate-reducing bacteria are the most important group in the prevailing anaerobic environment. They acts as primary decomposers of organic residues, playing key

roles in carbon cycle (Loka Bharathi *et al.,* 1991). Microbes in mangrove sediments are also key players in nitrogen cycle involved in all its transformation from fixation, ammonification, nitrification and denitrification (Abraham *et al.,* 2004). Microbes can control the nutrient availability in mangrove sediments, thereby a link can be made between microbial communities in sediments and vegetation patterns of mangroves (Sherman *et al.,* 1998).

Among the 34 clones analyzed from the mangrove sediment study, 15 clones showed maximum similarity to phylum *Proteobacteria* which were taxonomically identified as four classes namely, *Alphaproteobacteria* (20%), *Betaproteobacteria* (20%), *Gammaproteobacteria* (20%) and *Deltaproteobacteria* (40%). The class *Deltaproteobacteria* represented the dominant group comprising of 40% of the clones, comprising the order *Syntrophobacterales, Desulfobacterales, Desulfuromonadales* and *Myxococcales* including members of the genus *Desulfomonile, Geobacter* and *Sorangium.* Class *Alphaproteobacteria*, with 3 clones, was represented by order *Sphingomonadales* and *Rhizobiales* including members of the genus *Altererythrobacter* and *Pseudolabry.* The class *Betaproteobacteria,* with 3 clones, was represented by order *Burkholderiales* and *Gallionellales* including members of the genus *Sideroxydans* and *Thiobacter.* The class *Gammaproteobacteria*, with 3 clones was represented by order *Alteromonadales* and the genus *Haliea.*

Similarly, in a study from Sundarban mangrove sediments (Ghosh *et al*., 2010) on 16S rDNA gene libraries constructed, members belonging to 8 different bacterial phyla like *Proteobacteria, Cytophaga-Flexibacter-Bacteroides* (CFB) group, *Actinobacteria, Chloroflexi, Firmicutes, Gemmatimonadetes, Acidobacteria* and *Planctomycetes* were detected, with the major divisions of detected bacterial phyla being *Proteobacteria;* in which the clones were distributed among the Alpha, Beta, Gamma, and Deltaproteobacterial class which

supports the finding in the present study suggesting that mangrove sediments are dominated with proteobacterial phylum distributed among these four classes.

In the Mangalavanam mangrove sediment diversity analysis, some of the identified deltaproteobacterial clones belonging to order *Syntrophobacterales* and *Desulfobacterales* showed similarity to the sulfur and sulphate reducing bacteria isolated from marine sediments. Similarly Gammaproteobacterial clones are also involved in sulfur cycling contributing to oxidation of sulphur compounds. Sulfur-reducing and oxidizing bacterial strains are important in organic carbon oxidation thereby suggesting that sulphate is one of the main electron acceptors present in these environments (Li *et al.,* 2009). Similar results were reported previously in which sulfur-oxidizing bacterial strains are found to play an important role in detoxification of sulphide in mangrove sediments. It is identified that sulphate reduction may be an important pathway of organic matter mineralization in organic rich deposits typical of mangrove forests (Asami *et al.,* 2005).

Firmicutes the second prominent phylum, with 8 clones representing the phylum constituting 23.52% of the mangrove sediment 16S rDNA clone library. All the members of *Firmicutes* was represented by the class *Bacilli*, order *Bacillales* and genus *Staphylococcus.* Members of phylum *Firmicutes* represented by the class *Bacilli* were also obtained in studies on Amazonian mangrove ecosystem (Pureza *et al.,* 2012). Similar results were obtained from sediments of northern slope of the South China Sea**,** in which 8 clones affiliated to *Firmicutes,* representing 7% of the total sequences analyzed (Liao *et al,* 2009).

In the Mangrove Library, members of phylum *Acidobacteria* was represented by the class *Acidobacteria Gp17* and *Acidobacteria Gp1*.Two clones representing the phylum *Actinobacteria* were identified as order *Coriobacteriales* and *Acidimicrobiales.* Single clone representative from the phylum *Planctomycetes, Chloroflexi* were identified as order *Planctomycetales* and

Anaerolineales respectively. Single clone representing the phylum *WS3* were also obtained.

Representatives of the *Acidobacteria* can be found in a wide range of environments, even though most of the recognized taxa are still unclassified and the ecology of this phylum is not well understood (Pureza *et al.,* 2012).

Bacterial diversity of Amazonian mangrove ecosystem by culture-independent revealed the abundant groups such as *Proteobacteria* and unclassified bacteria, followed by representatives from other phyla including *Bacteroidetes, Cyanobacteria, Acidobacteria, Firmicutes, Actinobacteria, Chloroflexi, Siprochaetes and TM7* (Pureza *et al.,* 2012). In a study about microbial populations in a non-disturbed Brazilian mangrove sediment, *Alphaproteobacteria* dominated the bacterial community (Dias *et al.,* 2010). Microbial community response to a simulated crude oil exposure in mangrove sediments indicated that bacterial groups belonging to Deltaproteobacteria increased with a subsequent decrease in Alphaproteobacterial bacterial groups such as Anaerolinea (Taketani *et al.,* 2010). Under anoxic conditions, several microorganisms are capable of degrading aliphatic and aromatic hydrocarbons (Widdel *et al.,* 2010).

Predominance of sulphate reducing bacterial clones point towards the anaerobic conditions prevailing in the mangrove sediments and at the possible maintenance of the biogeochemical cycle in Mangalavanam mangrove sediments.

Phylogenetic tree of partial 16S rRNA gene sequences obtained in the present study was constructed and depicted in Fig. 4.9. The analysis involved 34 nucleotide sequences (N=34) *Pseudoalteromonas lipolytica* (Accession number

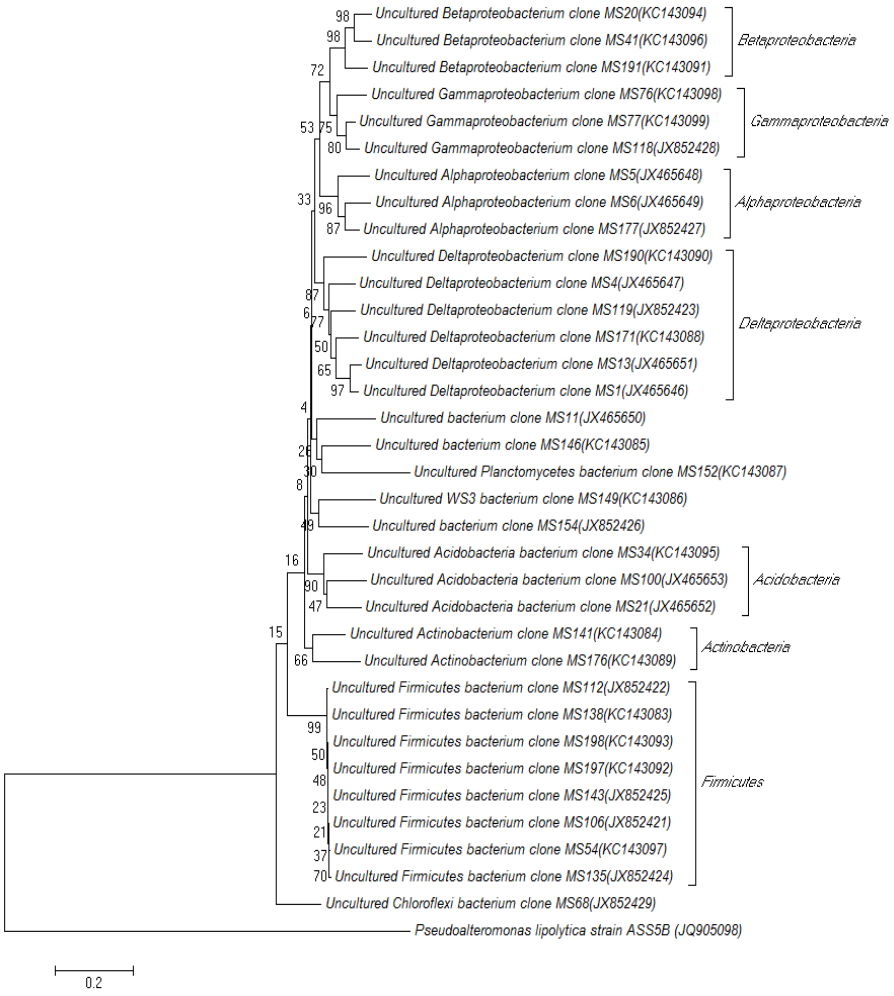

Fig 4.9: Phylogenetic relationship based on partial 16SrDNA sequences of selected clones (N=34)

JQ905098) was used as outgroup. Accession numbers of clones are given in parentheses. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches and scale bar shows sequence divergence.

In the phylogenetic tree of partial 16S rRNA gene sequences of 34 clones, *Deltaproteobacteria* the prominent class among the Proteobacterial

phylum were clustered together forming a separate clade. All other similar phyla were grouped together forming separate clades. Clone sequences showing similarity to phylum *Planctomycetes* and candidate phylum *WS3* grouped within the uncultured bacterium clone signifying that they may be new sequences identified, without having any known cultured representatives. The outgroup *Pseudoalteromonas lipolytica* was distinctly separated from all these clones.

4.3.2.5 Determination of species richness by rarefaction curve

The sequences were aligned and clustered into Operational Taxonomic Units (OTUs) based on the genetic distance. OTUs were identified at genetic distances of 3%, 5%, 10% and 20% by using 34 sequences and rarefaction curve was plotted using the RDP Pipeline (Fig. 4.10). At 3% genetic distances, 26 different OTUs were observed and at 5% distance, 25 OTUs were observed with the 34 sequences sampled indicating the species richness of the sample. In a similar study on three different sediments from Amazonian mangrove ecosystem, 29, 34, and 52 OTUs were identified (Pureza *et al.*, 2012). At 3% and 5% genetic distance, the rarefaction curves were not saturated indicating that full extent of taxonomic diversity at these genetic distances were not sampled. At 10% distance, the curve is partially saturated indicating that by increasing the sampling more taxonomic diversity can be obtained at the class level. At 20% sequence divergence, the rarefaction curves reached saturation, indicating that the sampling effort covered almost the full extent of taxonomic diversity at these genetic distances at the phylum level. However, a substantial fraction of the bacterial diversity within the Mangalavanam mangrove sediment samples was assessed at the phylum level by the sampling effort targeting the 16S rRNA gene. Maximal diversity can be accessed by targeting more specific variable regions within the 16S rRNA gene.

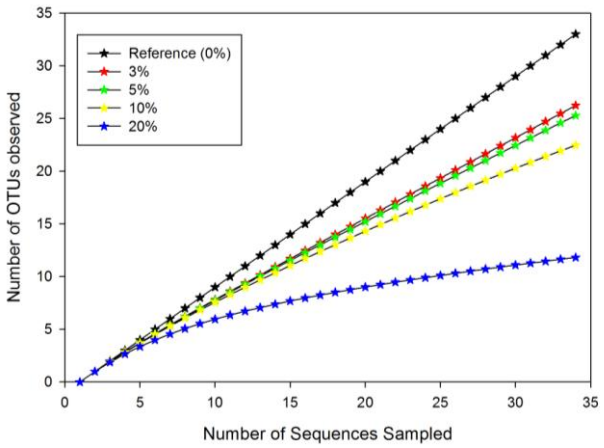

Fig. 4.10: Rarefaction curves indicating the observed number of operational taxonomic units (OTUs) at genetic distances of 3, 5, 10 and 20%. The reference line represents the OTUs at 0% genetic distance.

In the present study, 16S rRNA gene clone library based bacterial diversity analysis was performed on the sediment from the conserved Mangalavanam mangrove ecosystem for the first time. The study suggests that additional studies are needed to explore the full extent of taxonomic diversity to provide additional ecological and biological perspective of mangrove sediments that will emphasize bacterial community diversity studies into the future.

Sanger sequencing based bacterial diversity studies on Arabian Sea and Mangalavanam mangrove sediments revealed the phylogenetic diversity these sediments with the small number of clones sequenced. These studies identified that both these sediments harbors a huge microbial population which requires more thorough studies to explore the taxonomic diversity to full extent. In order to study the bacterial community composition in details, the study was extended to include next generation sequencing based on V3 regions of 16S rRNA gene.

4.3.3 Analysis of bacterial diversity of Arabian Sea sediment based onV3 regions of 16S rRNA gene employing Next Generation Sequencing (NGS) method

Phylogenetic diversity studies of marine metagenome was carried out based on NGS method by amplification of the hypervariable region 3 (V3) of 16S rRNA gene, followed by Illumina sequencing and *in silico* analysis.

4.3.3.1 PCR amplification of V3 regions of 16S rRNA gene from isolated marine metagenomic DNA

The V3 region of 16S rRNA gene was amplified from the marine metagenome and the ~250 bp amplicon was visualized on agarose gel (Fig. 4.11).

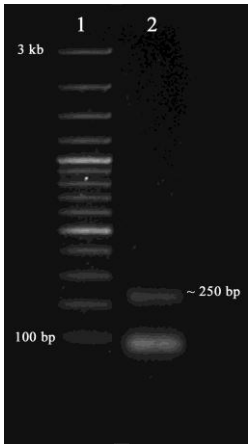

Fig 4.11: Agarose gel electrophoresis of amplified V3 region

Lane 1: 100 bp ladder, Lane 2: V3 amplicon

4.3.3.2 *In silico* analysis for microbial diversity analysis of marine metagenome

An output of 295.35 Mb data with a total of 978,007 raw reads having 54.01% GC content was obtained. Average base quality (Phred score>= Q20) obtained was 95.95. Base composition distribution of amplified V3 of Arabian Sea metagenome is shown in Table 4.3.

In a similar Illumina based study of bacterial community the average GC content of Indo-Burman Biodiversity hotspot was 56.48% (Panda *et al.*,2015); while in Jakrem hot spring, Meghalaya, it was 56.35% (Mandal *et al.,* 2015). The GC content in bacterial genomes range from about 25% to 75% (Fleischmann *et al.,* 1995; Zhang and Zhang, 2004; Zhou *et al.,* 2014). Previous studies suggested that the genomic GC content of bacteria is related to their phylogeny (Gupta, 2000).

Table 4.3: Base composition distribution of Arabain Sea metagenome

Sample Name	Base Composition (%)			
	A	C	G	T
Arabian Sea Metagenome	23.33	25.17	28.84	22.64

4.3.3.2.1 Quality filtering and counting of marine metagenome sequencing data

A total of 9,78,007 raw reads were obtained. After quality filtering a total of 8,52,121 reads were obtained After filtering and removing potential erroneous sequences, a total of 6,97,074 pre-processed reads were obtained.

4.3.3.2.2 Identification of Operational Taxonomy Units (OTU) and relative abundance of marine metagenome

Total reads obtained were clustered into Operational Taxonomic Units (OTUs) based on their sequence similarity using Uclust program (similarity cut off = 0.97). A total of 6, 309 OTUs were identified out of the 6, 97,074 pre-processed reads. Community metagenomics of Arabian Sea sediment revealed the distribution of the OTUs into different taxonomic level of bacterial domain. At the phylum level, all OTUs were classified into 43 bacterial phyla including 18 formally described bacterial phyla (Fig:4.12) and 25 candidate phyla (Fig:4.13) (Appendix II). Therefore this study presents the first report on the microbial diversity of Arabian Sea based on Illumina sequencing technique.

This study also revealed that the overall diversity in the Arabain Sea with 43 different bacterial phyla was higher, than the reported 40 phyla from South China Sea (Zhu *et al.*, 2013) and 35 and 32 different phyla obtained from other marine habitats, including the Western English Channel and Arctic Ocean (Gilbert *et al.*, 2009; Galand *et al.*,2009).

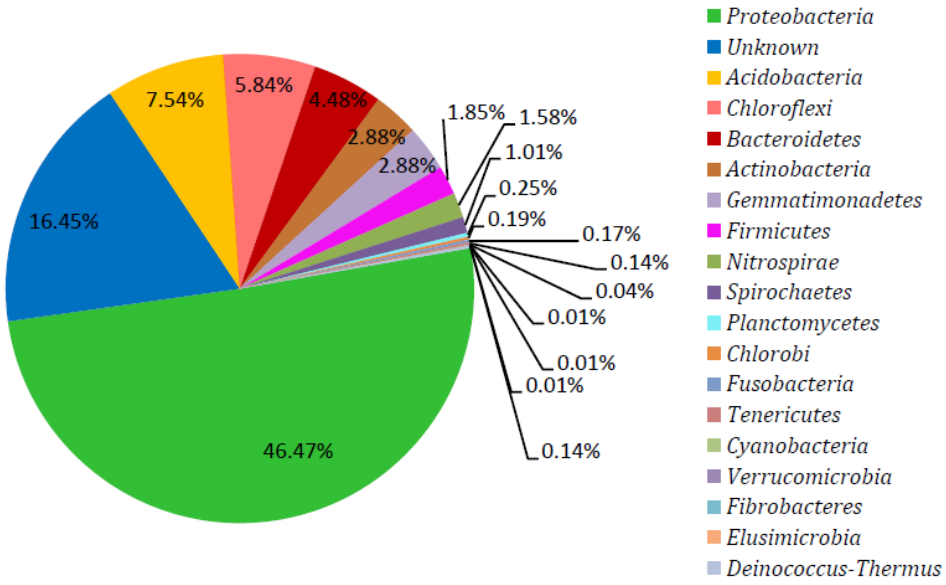

Fig 4.12: Phylogenetic distribution of OTUs at phylum level

Taxonomic classification identified that *Proteobacteria* was the most abundant phylum. Predominance of *Proteobacteria* supports the study on Arabian Sea microbial diversity based on Sanger sequencing. A total of 2932 OTUs belonging to proteobacterial phylum representing 46.47% of the total diversity were obtained. 1038 OTUs were unknown representing 16.45% indicating the possibilities for the identification of novel yet to be cultured organisms in Arabian Sea sediments awaiting discovery. 476 OTUs belonging to *Acidobacteria*, 369 OTUs to *Chloroflexi*, 283 OTUs to *Bacteroidetes*, 182 OTUs from *Actinobacteria* and *Gemmatimonadetes* were also identified. *Firmicutes*,

Nitrospirae, Spirochaetes, Planctomycetes, Chlorobi, Fusobacteria, Tenericutes, Cyanobacteria, Verrucomicrobia, Fibrobacteres, Deinococcus-Thermus and *Elusimicrobia* contributed less than 2% of the total identified OTUs.

Recent studies on microbial communities of sediments of South Sea of Korea identified 13 described bacterial phyla, the prevalence of *Proteobacteria, Bacteroidetes, Cyanobacteria, Actinobacteria, Verrucomicrobia* and *Tenericutes* with *Proteobacteria* representing 63% of the population. (Suh *et al.*, 2015), whereas studies on sediments of South China Sea identified the occurrence of *Proteobacteria, Firmicutes, Planctomycetes, Actinobacteria Chloroflexi and Bacteroidetes* with the predominance of protobacterial phylum representing 44.7% (Zhu *et al.*, 2013). Similarly a culture independent study from the sea mouth of Chilika Lake, India has revealed the phylogenetic diversity of bacterial phyla including *Proteobacteria, Spirochaetes, Firmicutes, Chlamydiae, Tenericutes and Planctomycetes* (Parag *et al.*, 2013). So the analysis revealed that members of phylum *Proteobacteria* predominates in global oceans, while the minor representatives may vary considerably based on the nutrient availability and growth conditions prevailing in the location.

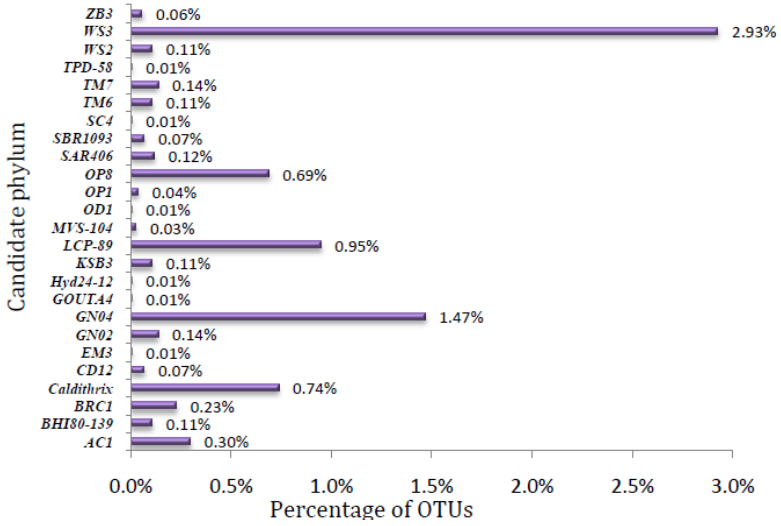

Fig 4.13: Phylogenetic distribution of OTUs at candidate phylum level

Remaining 503 OTUs clustered among 25 candidate phylum for which no cultured representatives have been isolated yet. The concept of a waiting position for putative taxa in a category called Candidatus or Candidate phyla was first describes by Murray and Schleifer in 1994 signifying that this can be used for describing prokaryotic entities for which more than a mere sequence is available; but for which characteristics required for higher level bacterial hierarchical classification are lacking. These are usually derived from metagenomic sequencing studies, *in situ* hybridization or other similar techniques for cell identification. In the present study Candidate phylum WS3 is represented by 185 OTUs contributing 2.93% of the identified bacterial population of Arabain Sea sediment. GN04 is represented by 93 OTUs constituting 1.47% of the total OTUs. All the other members of candidate phyla contribute less than 1%.

Pyrosequencing based diversity analysis of microbial communities on deep-sea sediments of the eastern Mediterranean Sea identified 23 candidate phylum, in which OD1 is the most dominating phylum among the candidate division representing 1.1% and WS3 constitute only less than 1% of the total population (Polymenakou *et al*., 2015), while OD1 is representing only 0.01% of the total population in the present study. Previous studies on deep sediments of Pacific Ocean identified candidate divisions such as OP1, OP3, OP8, OP10, OP11, WS1, and WS3 which accounts for a few percent of the total sequences (Inagaki *et al*., 2006), while in the present study representatives from OP1, OP8 and WS3 is only obtained. From the analysis it is evident that members of candidate phyla only contribute a few percent of microbial population.

Among the *Proteobacterial* OTUs, *Deltaproteobacteria* was the most dominant class, with 1426 OTUs accounting for 22.60% in total; *Gammaproteobacteria* was the second most dominant class with 1293 OTUs falling in the class representing 20.49% and *Alphaproteobacteria* with 172 OTUs falling in the class contributes the third most abundant class among proteobacterial

phylum. However, *Betaproteobacteria* and *Epsilonproteobacteria* were represented only less that 1% of the total class. Distribution of proteobacterial phylum is shown in figure 4.14.

Comparable results were reported from sediments of Pacific Ocean (Liao *et al.,* 2011), Mid-Okinawa Trough (Yanagawa *et al.,* 2014) and eastern Pacific nodule province (Xu *et al.,* 2007) in which high abundance of Gamma, and *Deltaproteobacteria* were reported.

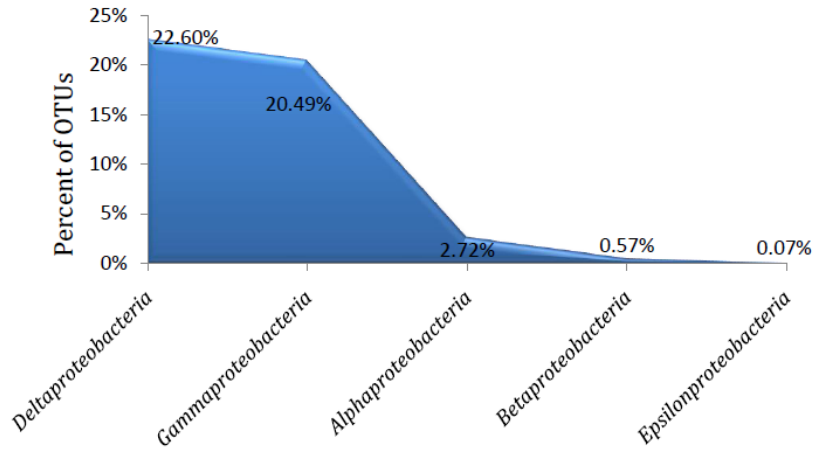

Fig 4.14: Distribution of different classes of proteobacterial phylum (n=2932 OTUs)

In class level taxonomic identification 1521 OTUs remained unknown, while it was 1038 OTU in phylum level identification, suggesting that the number of unknown sequences increased with higher taxonomic hierarchy. At order level identification, 3615 OTUs remained unknown, while at genus level it was 5777 OTUs, but at species level 6252 out of 6309 OTUs remained unidentified in the present study. These results revealed the likelihood of unknown bacteria in marine sediments. Since they are unidentified, the role they play in ecosystem maintenance also remain unknown, implying that many more prokaryotic lineages await discovery methods. Apart from members of proteobacterial phylum, OTUs

falling into 91 classes were obtained in the study and the distribution of major classes is shown in Fig 4.15.

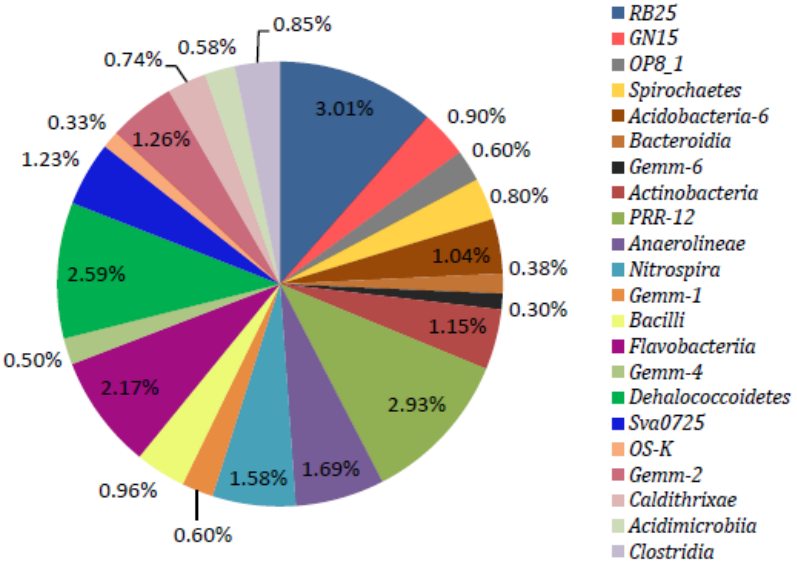

Fig 4.15: Taxonomy classification of OTUs at class level

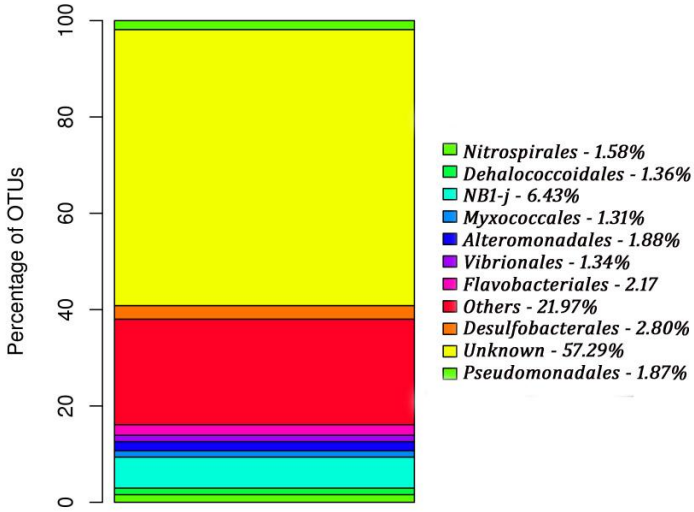

Fig 4.16: Taxonomy classification of OTUs at order level

Taxonomic hierarchy of OTUs in the level of order identified a total of 114 orders in which members of *Deltaproteobacteria* were distributed among 7 orders including *Bdellovibrionales, Desulfobacterales, Desulfovibrionales, Desulfarculales, Desulfuromonadales, Syntrophobacterales* and *Myxococcales* and the members of *Gammaproteobacteria* were distributed among 9 orders including *Alteromonadales, Chromatiales, Enterobacteriales, Oceanospirillales, Pasteurellales, Pseudomonadales, Thiotrichales, Vibrionales and Xanthomonadales.* Heatmap generated using QIIME pipeline with top ten enriched orders is represented in Fig 4.16.

Classification of OTU at genus level had recognized 5777 OTUs, all of which remained unidentified. Remaining 532 OTUs were distributed among 168 genera in which *Acinetobacter* was the dominated genus and was represented by 61 OTUs, followed by *Pseudoaltremonas* represented by 46 OTUs. Distribution of 17 most abundant genera is shown in figure 4.17. All the remaining 151 genus contains only less than 5 OTUs as representatives. In species level identification 99% of the OTUs are unknown.

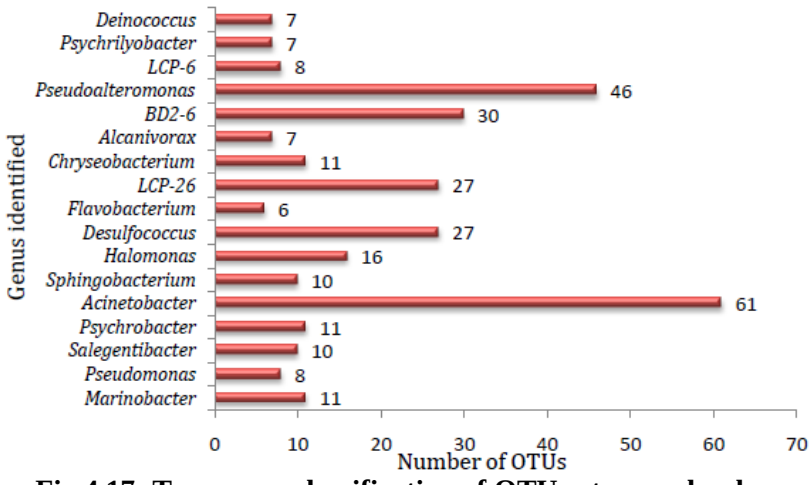

Fig 4.17: Taxonomy classification of OTUs at genus level

Oceans represent a major reservoir of sulfur on Earth and microbial transformation of sulfur compounds has proved to have profound effect on the properties of the biosphere and in turn affect the geochemistry (Klotz *et al*., 2011). Sulfide oxidation and reduction is one of the important microbial chemosynthetic pathways in marine ecosystems. Members of *Deltaproteobacteria* and *Gammaproteobacteria* are identified to be the key players in sulfur cycle in marine habitats. Sulfur-metabolizing microorganisms of *Deltaproteobacteria* include sulfate reducers and organic sulfur utilizers (Sievert *et al*., 2007). Anoxic ecosystems displayed higher proportions of *Deltaproteobacteria*, comprising many anaerobes such as the sulfate-reducing bacteria, and a large dominance of Gammaproteobacteria which is also in accordance with earlier studies on anoxic zone of the Cariaco Basin (Madrid *et al*., 2001).Major bacterial members known to contribute for sulfide oxidation and reduction includes members of Delta, Epsilon, and *Gammaproteobacteria* (Nakagawa *et al.,* 2004; Campbell *et al.,* 2006; Muyzer and Stams, 2008). In the present study, potential sulfide oxidation/reduction microbes comprised over 40.31% of the total OTUs identified. Major members identified in the present study known to participate in sulfide oxidation and reductions were *Chromatiales, Desulfobacterales, Desulfovibrionales, Desulfuromonadales* and *Syntrophobacterales*. Previous studies on bacterial diversity profiles of various marine sediments indicated that Delta and Gammaproteobacterial members contributing to sulphur cycle are prevalent in global oceans (Inagaki *et al.,* 2003; Nakagawa *et al.,* 2004; Campbell *et al.,* 2006; Muyzer and Stams, 2008; Liao *et al.,* 2011).

Previous studies have documented that *Epsilonproteobacteria* is dominant in microbial habitats coupled with hydrothermal vent fluid and chimney structure, and plays important role in carbon and sulfur cycling (Takai *et al.,* 2004; Huber *et al.,* 2010; Flores *et al.,* 2011). So it is understood that *Epsilonproteobacteria* prefer high temperature environments such as hydrothermal vent fluid and chimney, while Alpha, Beta, Delta and *Gammaproteobacteria* prefer low-

temperature habitats away from active hydrothermal regions. Although there was low abundance of Alpha and *Betaproteobacteria*, the predominance of *Deltaproteobacteria* and *Gammaproteobacteria* in the present study point towards the low temperatures prevailing in the sampling depth. Previous studies demonstrated that *Gammaproteobacteria*, prevail over other taxa in several deep-sea investigations, including the Eastern Mediterranean Sea (Polymenakou *et al*., 2015) and Northeastern Pacific Ocean (Kouridaki *et al*., 2010). In sediments of Eastern Mediterranean Sea *Gammaproteobacteria* contributes to 20.9% of total microbial sequences while in sediments of Northeastern Pacific Ocean *Gammaproteobacteria* represents the dominant class contributing 23.3%, followed by *Deltaproteobacteria* contributing 13.6% of the total OTUs. Studies based on global survey of microbial distribution in seafloor and seawater indicated that *Gammaproteobacteria* was found to dominate benthic communities which is followed by Delta and *Alphaproteobacteria, Actinobacteria, Flavobacteria, Planctomycetes* and *Acidobacteria* (Zinger *et al.,* 2011)

In this study, *Chloroflexi* contributes to 5.84% of the total OTU and contained members belonging to three classes, *Anaerolineae, Dehalococcoidetes, and Thermomicrobia.* Previous studies reported that *Chloroflexi* is widely distributed in many deep-sea hydrothermal sediments (Fry *et al.,* 2008; Flores *et al.,* 2012) and are identified as one of the most abundant phylum in the sediments of Western Pacific Ocean and sub-seafloor sediments of Okinawa Trough (Liao *et al.,* 2011;Yanagawa *et al.,* 2014). Since no deep-sea *Chloroflexi* have been cultured previously, the metabolic pathways of these bacteria remain unidentified. Recent studies with single cell genomic approach have revealed that *Dehalococcoidetes* from marine subsurface sediments are strictly anaerobic organotrophs or lithotrophs (Kaster *et al.,* 2014; Wasmund *et al.,* 2014).

In the Arabian Sea sediment, members of this phylum *Acidobacteria* was represented by 476 OTUs contributing to 5.84% of the total identified OTUs.

Members of the phylum *Acidobacteria* dominate the sediments of southern Ionian Sea (Polymenakou *et al.,* 2006) and deep-sea sediments of Cretan margin of eastern Mediterranean Sea (Polymenakou *et al.,* 2009). Interestingly, the high abundance of members of *Acidobacteria* in soils and sediments was reported to be associated with their ability to withstand metal contaminated, acidic and other extreme environments (Barns *et al.,* 2007). Genome based analysis of selected *Acidobacteria* strains suggested that they are suited to survive in low-nutrient conditions (Ward *et al.,* 2009). Members of this phylum possess the lower-specificity sugar transport system, which is assumed to be helpful to survive in low-nutrient environments (Paulsen *et al.,* 1998).

Other dominant groups of bacteria identified in Arabian sea sediment includes members of the phylum *Bacteroidetes, Actinobacteria, Gemmatimonadetes, Firmicutes, Nitrospirae* and *Spirochaetes* contributing to 14.68% of the total identified OTUs. Members of these phyla are known to be widely distributed in deep-sea sediments, crustal fluids and inactive hydrothermal chimneys, and rarely found in active hydrothermal regions (Suzuki *et al.,* 2004; Liao *et al.,* 2011; Sylvan *et al.,* 2012; Kato *et al.,* 2013).

Seafloor serves as valuable source metallic mineral resources. They occur in many forms including polymetallic nodules, metallic oozes, massive sulfide deposits, cobalt (Co) rich crusts etc. Co-rich crusts can be found globally on the ocean floor, and are also known as ferromanganese crusts, iron-manganese crusts, or Co-rich ferromanganese crusts. These resources have immense economic potentials due to their high content of metals with commercial values, e.g. manganese (Mn), copper (Cu), nickel (Ni) and platinum (Pt). Although the mechanism of formation of deep-sea metallic mineral is not completely understood, evidences suggest that microbes play a key role in the process (Wang *et al.,* 2009). Cobalt cycling is considered to be coupled with Mn-cycling, in which Mn-oxidizing or reducing bacteria possibly will also participate in Co

metabolism (Murray *et al.,* 2007). Mn-oxidizing bacteria produce Mn oxides, which promotes the adsorbtion of trace metals such as Co, Cu, Ni and Mn, which is helpful in the enrichment of such metals (Nealson, 2006).

Five genera of Co-metabolizing and Mn-oxidizing bacteria were detected benthic sediments of the Arabian sea, including *Acinetobacter*, *Flavobacterium*, *Pseudoaltremonas*, *Pseudomonas* and *Bacillus*. Microbial species that are involved in metallic crust formation are expected in the examined ocean environments (Liao *et al.,* 2011). In the genus level taxonomic identification, apart from unknown OTUs, bacteria belonging to *Acinetobacter* and *Pseudoaltremonas* were dominating suggesting the metallic mineral resources in the seafloor of Arabian Sea.

The dominant representative of the bacterial genera obtained include *Pseudoalteromonas* and *Acinetobacter, which* are identified to be involved in metal oxidation (Tebo *et al.,* 1997; Templeton *et al.,* 2005; Krishnan *et al.,* 2006). There are reports suggesting that bacteria belonging to the genera *Flavobacterium*, *Pseudomonas* and *Acinetobacter* are responsible for the immobilization of metals such as Co in marine environments (Krishnan *et al.,* 2006). In addition, many other phylogenetically distinct bacteria such as *Shewanella* and *Caulobacter* within the class *Alphaproteobacteria* have the abilities of biosorption and metabolism of metals such as Mn, Co, Fe and Ur (Konishi *et al.,*1997; Krishnan *et al.,* 2006).

However, most of the sequences identified at the genus level were unknown with unidentified physiological function. Either cultivation experiments should be conducted to isolate potential bacteria responsible for metal cycling, or investigations targeting functional genes that are potentially involved in metal cycling are required to understand the metabolic processes of seafloor in details.

The results of this study demonstrated that the bacterial community compositions of sediments of eastern Arabian Sea are diverse at higher taxonomic levels. The study revealed the anoxic, nutrient rich environment prevailing in the location in which bacterial communities were abundant with microbes potentially involved in metal and sulfur cycling.

4.3.3.2.3 Richness and diversity analysis of OTUs

A total of 6,309 OTUs were obtained based on 97% similarity. In alpha diversity analysis, rarefaction curves, Chao1, and Shannon's index were generated based on a species level of 97% similarity. Summary of diversity indices are shown in Table 4.4.

Table 4.4: Summary of the richness and diversity of microbial communities

Sample	**Shannon (97%)**	**Chao1 (97%)**	**Observed species (97%)**
Arabian Sea sediment	6.62	4332	3951

Rarefaction curve of the Shannon index is shown in figure 4.16. The curve approached plateau from less than 15,000 tags sampled indicating that the sampling depths were sufficient to capture the overall microbial diversities in the sample. Furthermore, the analysis is an indicative of the diversity present in the sediment.

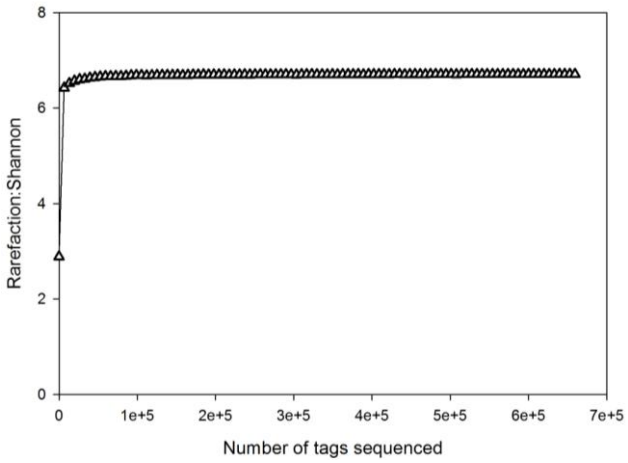

Fig 4.18: Shannon's diversity curves

This was confirmed by Chao1 index curves (Figure 4.18), in which a high value of 3951 was obtained which is indicative of the species richness in the sample. The curve is also saturated and reached plateau, suggesting that the sampling depths were adequate to infer the microbial diversity in the sample. Similarly rarefaction curve of observed species (Figure 4.19) reached plateau, indicating that sampling depth and sequencing coverage were sufficient to assess the diversity of Arabian Sea sediment.

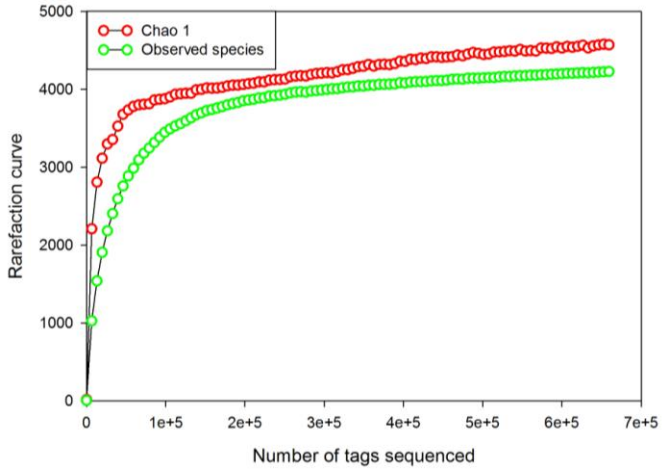

Fig4.19: Rarefaction analysis for Chao1 and observed species.

In the diversity analysis of sediments of Okinawa Trough (Yanagawa *et al.,* 2014), Shannon, Chao1 and observed species metrices obtained were 7.44, 1102.31 and 1543 respectively, whereas in a similar study on sediments of South Sea of Korea (Suh *et al.,* 2015), Shannon index of 4.726 was reported during sampling at summer, whereas diversity index increased to 5.255 during sediment sampling at spring, while Chao 1 index was 1228.18 during summer and 1705.00 during spring, suggesting that the richness of the entire bacterial community was highest during spring. Shannon index was 4.30 for sediments of Palk Bay (Aravindraja *et al.,* 2013), whereas it was 3.79 in sediments of Mediterranean Sea (Polymenakou *et al*., 2015). Based on these previous reports it is understood that the bacterial diversity of Arabian Sea is much higher which can be interpreted by the alpha diversity analysis. The richness and diversity analysis in the present study identified the microbial diversity in the Arabian Sea sediment.

4.3.3.2.4 Accession numbers

All of the sequencing data obtained in the present study was submitted to MG-RAST database with the MG-RAST ID: 4634001.3

4.3.4 Analysis of bacterial diversity of Mangalavanam mangrove sediment based on V3 regions of 16S rRNA gene employing Next Generation Sequencing (NGS) method

Phyologenetic diversity studies of mangrove metagenome based on NGS method was done by amplification of the hypervariable region 3 (V3) of 16S rRNA gene followed by Illumina sequencing and *in silico* analysis.

4.3.4.1 PCR amplification of V3 regions of 16S rRNA gene from isolated mangrove metagenomic DNA

The V3 region of 16S rRNA gene was amplified from the mangrove metagenome and the ~250 bp amplicon was visualized on agarose gel (Fig. 4.20).

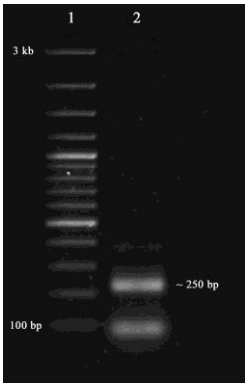

Fig 4.20: Agarose gel electrophoresis of amplified V3 region from mangrove metagenome

Lane 1: 100 bp ladder, Lane 2: V3 amplicon

4.3.4.2 *In silico* analysis for microbial diversity analysis of Mangalavanam mangrove metagenome

An output of 223.46 Mb data with a total of 7,39,964 raw reads having 58.02 % GC content was obtained. Average base quality (Phred score >= Q20) obtained is 96.22. Base composition distribution of amplified V3 of mangrove metagenome is shown in Table 4.5. G + C content of 55.75% was reported for bacterial diversity study of Brazilian Mangrove Sediments (Andreote *et al.,* 2012).

Table 4.5: Base composition distribution of mangrove metagenome

Sample Name	Base Composition (%)			
	A	C	G	T
Mangalavanam mangrove metagenome	21.02	27.47	30.55	20.93

4.3.4.2.1 Quality filtering and counting of mangrove metagenome sequencing data

A total of 7,39,964 raw reads were obtained . After quality filtering a total of 6,68,141 reads were obtained. After filtering and removing potential erroneous sequences, a 4,37,766 pre-processed reads were obtained.

4.3.4.2.2 Identification of Operational Taxonomy Units (OTU) and relative abundance of mangrove metagenome

Total read obtained were clustered into Operational Taxonomic Units (OTUs) based on their sequence similarity using Uclust program (similarity cutoff = 0.97). A total of 9,362 OTUs were identified out of the 4,37,766 pre-processed reads. Community metagenomics of Mangalavanam mangrove sediment revealed the distribution of the OTUs into different taxonomic level of bacterial domain. At the phylum level, all OTUs were classified into 52 bacterial phyla including 21 formally described bacterial phyla (Fig: 4.21) and 31 candidate phyla (Fig: 4.22) (Appendix II). Therefore this study presents the first report on the microbial diversity of unexplored sediments of Mangalavanam mangrove ecosystem based on Illumina sequencing technique.

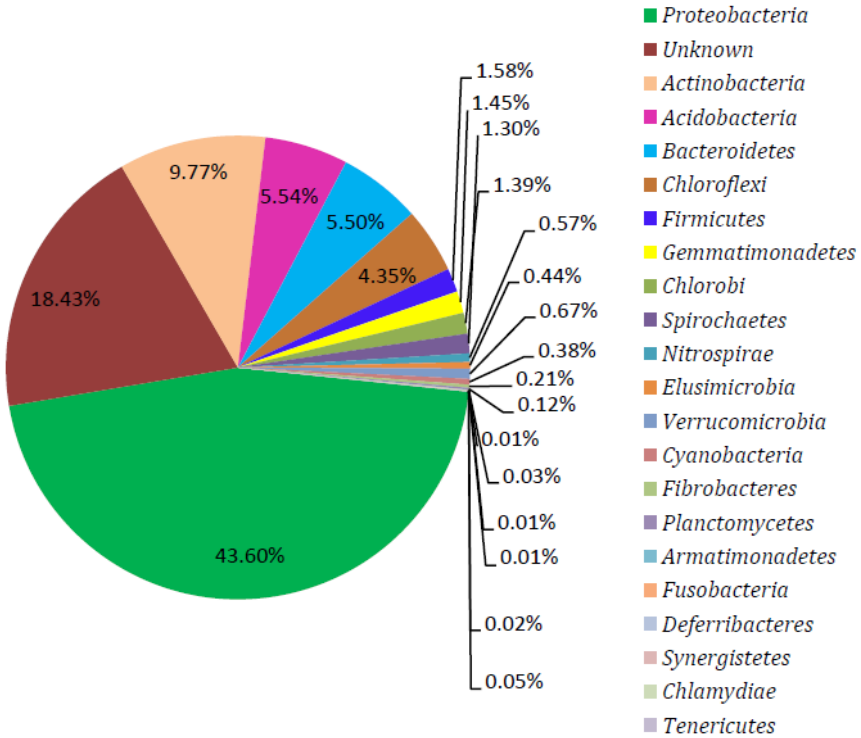

Fig 4.21: Phylogenetic distribution of OTUs at phylum level

Bacterial community profile of mangrove sediments of Sundarbans, India identified 33 different bacterial phyla with the dominance of *Proteobacteria, Firmicutes, Chloroflexi, Bacteroidetes, Acidobacteria, Nitrospirae* and *Actinobacteria* respectively. (Basak *et al.*, 2015). Similarly 44 bacterial phyla were identified by pooling the sequences from 12 mangrove sediment samples collected from Mai Po Ramsar Wetland in Hong Kong (Jiang *et al.,* 2013). A pyrosequencing based diversity analysis on oil contaminated mangrove wetland documented the presence of 22 bacterial phyla (Dos Santos *et al.,* 2011). Therefore, the present study revealed that the overall diversity in the Mangalavanam mangrove sediment was higher, with 52 different bacterial phyla, than the previous reports on different mangrove ecosystems.

Taxonomic analysis of sequences identified that *Proteobacteria* was the most abundant phylum present in the mangrove sediment and is supported by Sanger sequencing based sediment microbial diversity data. A total of 4082 OTUs belonging to proteobacterial phylum representing 43.60% of the total diversity were obtained by NGS. 1726 OTUs were unknown, representing 18.43% of the total identified OTUS, suggesting the uniqueness of Mangalavanam mangrove sediment with many unidentified bacterial community. 915 OTUs belonging to *Actinobacteria,* 519 OTUs belonging to *Acidobacteria,* 515 OTUs belonging to *Bacteroidetes,* 408 OTUs from *Chloroflexi,* 148 OTUs from *Firmicutes,* 136 OTUs from *Gemmatimonadetes*, 131 OTUs belonging to *Chlorobi*, 122 OTUs from *Spirochaetes* were also identified. Less than 1% representatives from *Nitrospirae, Elusimicrobia, Verrucomicrobia, Cyanobacteria, Fibrobacteres, Planctomycetes, Armatimonadetes, Fusobacteria, Deferribacteres, Synergistetes, Chlamydiae* and *Tenericutes* were also obtained making a total of 22 previously described phyla.

Pyrosequencing based microbial diversity analysis of sediments of Sundarbans mangrove ecosystem identified the dominance of *Proteobacteria,*

Chloroflexi, *Firmicutes*, *Bacteroidetes*, *Acidobacteria*, *Nitrospirae* and *Actinobacteria* respectively (Basak *et al.*, 2015). Whereas in studies on mangrove sediments of Hong Kong identified *Proteobacteria*, *Planctomycetes*, *Chloroflexi*, *Acidobacteria*, *Firmicutes*, *Verrucomicrobia*, *Thermomicrobia*, *Actinobacteria*, *Nitrospirae* and *Bacteroidetes* as the dominating representatives (Dos Santos *et al.*, 2011). Bacterial diversity analysis in Amazonian mangrove ecosystem identified dominance of *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Acidobacteria*, *Firmicutes*, *Actinobacteria*, *Chloroflexi* and *Siprochaetes* (Pureza *et al.*, 2012). Bacterial community analysis in mangrove sediments located in Southeastern Brazil identified bacterial phylum belonging to *Proteobacteria*, *Firmicutes*, *Verrucomicrobia*, *Chloroflexi* *Bacteroidetes* and *Acidobacteria* (Mendes and Tsai, 2014). It is therefore evident that *Proteobacteria* are the predominant phylum and contributed to approximately 50% of the total identified phylum in most of the mangrove sediments studied.

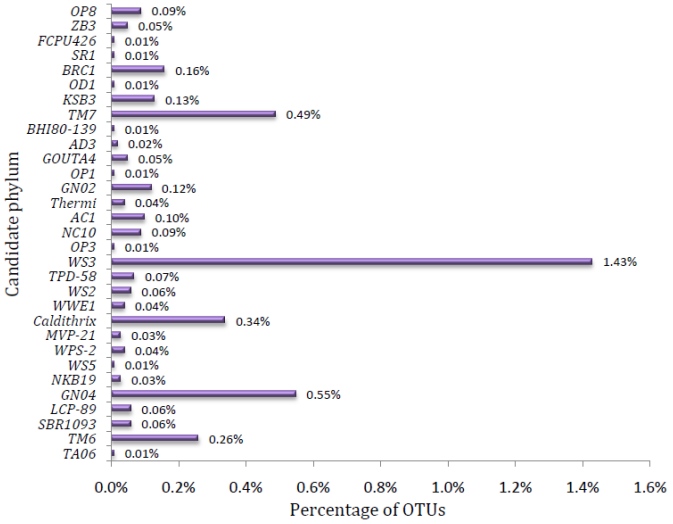

Fig 4.22: Phylogenetic distribution of OTUs at candidate phylum level

Remaining 420 OTUs were clustered among 31 candidate phylum for which no cultured representatives have been isolated yet. Candidate phylum WS3 was dominant and represented by 134 OTUs contributing 1.43% of the total

identified OTUs. All the remaining 30 candidate phyla had less than 1% representatives.

Members of candidate division WS3 were detected from various environments (Nesbo *et al.*, 2005; Tringe *et al.*, 2005; Ley *et al.*, 2006; Wilms *et al.*, 2006). Pyrosequencing based diversity analysis of microbial communities on tidal flat sediments of Ganghwa Island, Korea identified 37 candidate phylum, in which WS3 dominated among the candidate division (Kim *et al*., 2008). This was also observed in the Mangalavanam mangrove sediment suggesting that members of candidate division WS3 might be the dominant group in mangrove sediments.

At the class level, members of phylum *Proteobacteria* are distributed among six classes including *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria*, *Epsilonproteobacteria* and *Zetaproteobacteria* (Fig: 4.23). *Deltaproteobacteria* was the most dominant class, with 2092 OTUs accounting for 22.34% in total and *Gammaproteobacteria* was the second most dominant class with 1062 OTUs falling in the class representing 11.34%. *Betaproteobacteria*, *Alphaproteobacteria Epsilonproteobacteria* and *Zetaproteobacteria* are represented by 467, 446, 11 and 4 OTUs respectively.

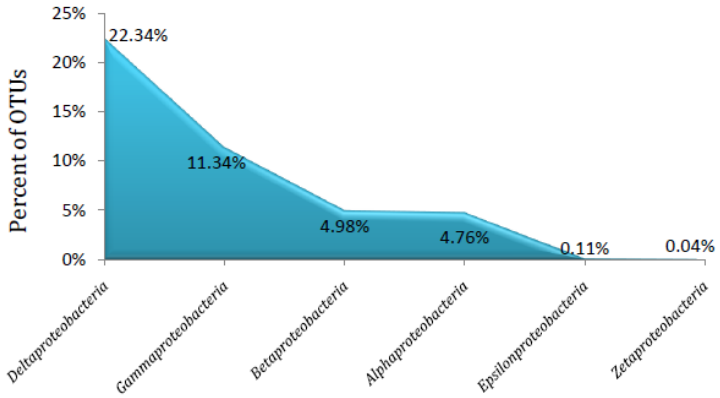

Fig 4.23: Distribution of different classes of proteobacterial phylum in mangrove sediments (n=4082 OTUs)

Similar reports are available on bacterial community structure and composition in mangrove sediment of different depths in southeastern Brazil (Mendes and Tsai, 2014), and in Brazilian mangrove sediments of Sao Paulo state (Andreote *et al.*, 2012) in which members of proteobacterial phylum are distributed among the 5 classes except *Zetaproteobacteria*. This indicative of the predominace of members of different classes of proteobacterial phylum in mangrove ecosystems.

Apart from preoteobacterial classes, OTUs falling into 132 classes of non-proteobacterial phylum were obtained in the study and their distribution is shown in figure 4.24.

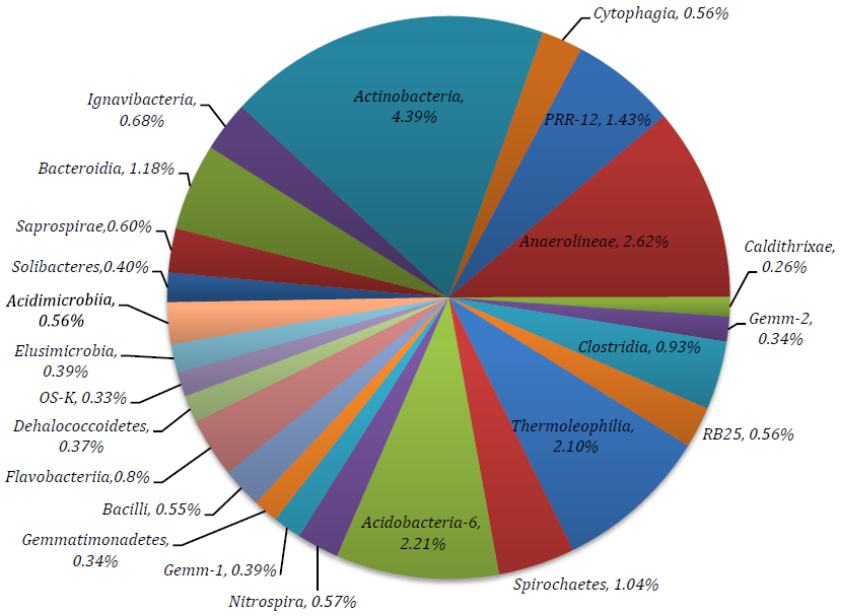

Fig 4.24: Taxonomy classification of OTUs at class level

In class level taxonomic identification 2890 OTUs remained unknown, representing 30.86% of the total identified OTUs, while it is 1726 OTU (18.43%) in phylum level identification. At order level identification 5080 OTUs remained

unknown while in genus and species level 8754 and 9313OTUs respectively remained unidentified against the total of 9362 OTUs. This suggests that the number of unknown sequences increases with higher taxonomic hierarchy. This study implies that Mangalavanam mangrove sediment houses phylogenetically diverse population of bacterial domain that are be unexplored and hence may be a source of novel biomolecules.

Taxonomic hierarchy of OTUs in the level of order identified a total of 194 orders and the heatmap generated using QIIME pipeline with top ten enriched orders were represented in Fig 4.25. *Deltaproteobacteria* the dominant class identified in the present study are distributed among 7 orders including *Bdellovibrionales, Myxococcales, Desulfobacterales, Desulfovibrionales, Desulfarculales, Desulfuromonadales* and *Syntrophobacterales. Gammaproteobacteria* the second dominated class obtained are distributed among 11 orders including *Aeromonadales, Alteromonadales, Chromatiales, Enterobacteriales, Legionellales, Methylococcales, Oceanospirillales, Pseudomonadales, Thiotrichales, Vibrionales and Xanthomonadales.* The third dominant class *Actinobacteria* is represented by three orders, *Acidimicrobiales, Actinomycetales* and *Coriobacteriales*.

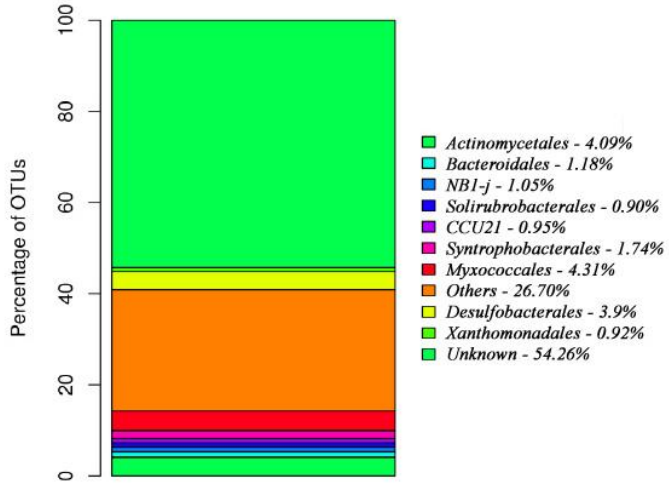

Fig 4.25: Taxonomy classification of OTUs at order level

Classification of OTU at genus level had identified 8754 OTUs which were unidentified. Remaining 608 OTUs were distributed among 227 genus in which genus *Desulfococcus* of family *Desulfobacteraceae*, order *Desulfobacterales* of class *Deltaproteobacteria* dominated being represented by 42 OTUs; followed by *Desulfobacca* of order *Syntrophobacterales*, class *Deltaproteobacteria* and represented by 29 OTUs. Distribution of 22 most abundant genera is shown in Fig 4.26. All the remaining 205 genera contained only less than 5 OTUs as representatives. In species level identification 99% of the OTUs wereunknown.

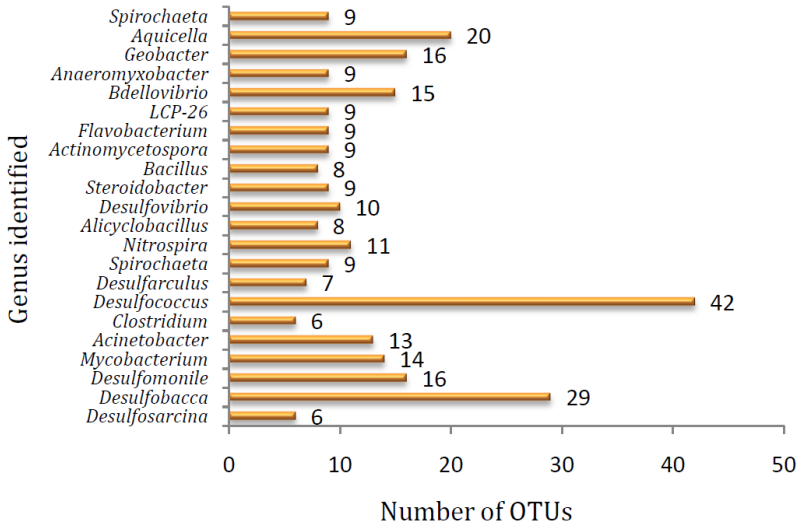

Fig 4.26: Taxonomy classification of OTUs at genus level

Mangrove ecosystem is a unique ecological niche, with a variety of microbes playing important roles in nutrient recycling and various other ecological processes, thereby requiring a thorough exploration of their microflora. Mangrove ecosystems are able to accumulate large amounts of organic carbon and in some mangrove ecosystems organic-rich sediments of several meters depth have been identified (Fujimoto *et al.*, 1999, Bouillon *et al.*, 2002), with high microbial diversity. Mangrove sediments are mainly anaerobic with an aerobic surface layer. Organic matter decomposition of the top most zones occurs mainly through

aerobic respiration where as sulfate-reduction is the key process in anaerobic layers (Nedwell *et al.,* 1994; Sherman *et al.,* 1998). In mangrove sediments, the dominant electron acceptor in anaerobic biodegradation is sulfate (Li *et al.,* 2009). *Deltaproteobacteria* was the dominant class identified representing 20.74% of the total identified OTUs in Mangalavanam mangroves. Members of this class are sulphate- reducers and they participate in mineralization of organic matter by reducing sulphate; members such as *Desulfovibrio* are linked to Fe reduction as they can reduce Fe and possibly Mn, and consequently participate in phosphorus cycling (Park *et al.*, 2008; Byrne *et al.*, 2010). Studies on Brazilian mangroves reported that sulphur metabolism occurring in mangrove sediments generates the reductive form of this compound such as sulfite, which is then reduced to H_2S (Andreote *et al.*, 2012). Further transformation of H_2S does not occurin mangrove sediment and hence released by volatilization, thus producing the typical smell of mangroves (Lyimo *et al.*, 2009). Similarly sulphate-reducers are capable of degrading complex substrates, such as long-chain and aromatic hydrocarbons (Muyzer and Stams, 2008).

93% of the sequences obtained in this study were unknown at the Genus level, hence the roles they play in mangrove ecosystem also remained unidentified; while among the taxonomically identified members, genus *Desulfococcu*s was predominant with 42 OTUs. They are Chemoautotrophs, anaerobic, thermophilic and mesophilic bacteria. Under anoxic condition they can completely oxidize acetate using sulfate as electron acceptor (Das *et al*., 2006).

Actinobacteria was the second dominant phylum obtained in Mangalavanam study, contributing 9.77% of the total identified OTUs. Members of this phylum are polyphosphate-accumulating bacteria (Kong *et al*., 2005). They grow on top of the nutrient and sulphide-rich sediments (Schulz *et al.,* 1999), and under anoxic conditions they can accumulate phosphate and release phosphate

under oxic conditions, thereby can significantly affect phosphorus effluxes (Hupfer *et al.*, 2007; Dale *et al.*, 2013).

Acidobacteria are omnipresent and are abundance among soil and sediment bacterial communities with a few cultured representatives, but their role in ecosystem maintenance has not been well documented (Jones *et al.*, 2009). At the same time studies suggested that the abundance of the *Acidobacteria* is correlated with pH of the location and more specifically, abundance increases when the pH is lower than 5.5 (Lauber *et al.*, 2009) and could possibly used as an indicator stain.

In the present study the dominant group represented was phylum *Proteobacteria*, and only 5.54% was represented by *Acidobacteria* which could not point towards the acidic pH of the sediment. In an Illumina based bacterial community study on mangrove wetland in in Hong Kong, proportion of *Acidobacteria* were higher while *Proteobacteria* was lower in one of the sediment sample analyzed (Jiang *et al.*, 2013). In addition, members of the phylum *Bacteroidetes* which is represented by 5.50% in the present study are supposed to be initial degraders of organic matter. Some members of this group are anaerobes or facultative anaerobes while others are aerobes so the species distribution of members of this class within a soil may depend on the availability of oxygen levels. Members of class *Sphingobacteria* and *Flavobacteria* are central to the initial biopolymer degradation of sedimentary organic matter (Kirchman, 2002; Bissett *et al.*, 2008; Gomez-Pereira *et al.*, 2012), such as high-molecular-weight organics (Cottrell & Kirchman, 2000). Lineages without cultured representatives have also detected (Lipson and Schmidt, 2004).

Studies indicate that members of phylum *Chloroflexi* can be abundant in sediments and involved in carbon cycling in the subsurface (Kindaichi *et al.*, 2012). Few members of the phylum perform anaerobic respiration of halogenated

hydrocarbons. Previous studies reported that representatives of class *Anaerolinea* are anaerobic and grows chemo-organotrophically on amino acids and a variety of carbohydrates (Sekiguchi, 2003), similar genome based study revealed that uncultured *Anaerolinea* sp. scavenge organic compounds from decaying debris (Kindaichi *et al.*, 2012). In Mangalavanam sediment, this phylum was represented by 4.35% and members of the class *Anaerolinea* by 2.62%, suggesting the prevalence of anaerobic respiration and possible degradation of hydrocarbons in this region. The proximity to the oil pumping station and the shipping harbor off the mangrove may be one of the contributing factor to this.

Bacterial community compositions of Mangalavanam mangrove sediments were identified by the present study revealing the roles of different bacteria in nutrient cycling and organic matter degradation. The study also revealed the prevalence of unknown bacteria suggesting that the location may be a possible source of novel biomolecules.

4.3.4.2.3 Richness and diversity analysis of OTUs

A total of 9,362 OTUs were obtained based on 97% similarity. In alpha diversity analysis, rarefaction curves, Chao1, and Shannon's index were generated based on a species level of 97% similarity. Summary of diversity indices are shown in Table 4.6.

Table 4.6: Summary of the richness and diversity of microbial communities

Sample	Shannon (97%)	Chao1 (97%)	Observed species (97%)
Mangrove sediment	9.82	5648	5661

Rarefaction curve of the Shannon index is shown in figure 4.27. High Shannon index value of 9.82 was obtained and approached plateau with minimum samples i.e. less than 15,000 tags sampled, indicating that the sampling depths

were sufficient to capture the overall microbial diversities in the Mangalavanam sample. Furthermore, the analysis is indicative of the high bacterial diversity present in the sediment.

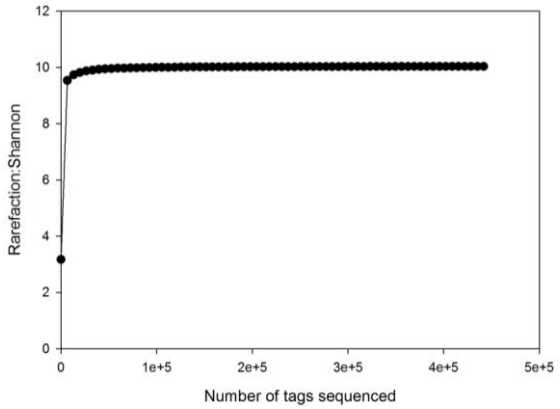

Fig 4.27: Shannon's diversity curves

This was further confirmed by rarefaction curve Chao1 index and observed species curves (Fig 4.28), in which Chao1 value of 5648 was obtained,also indicative of the species richness in the sample; the curve is also saturated and reached plateau, suggesting that the sampling depths were adequate to infer the microbial diversity in the sample. Similarly curve relating observed species also reached plateau, demonstrating that the sampling depth and sequencing coverage were sufficient to assess the bacterial diversity of Mangalavanam mangrove sediments.

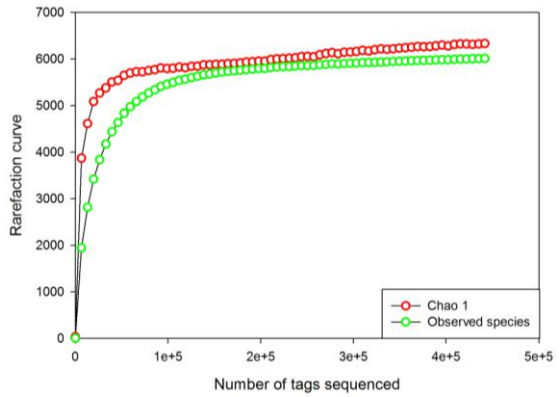

Fig 4.28: Rarefaction analysis for Chao1 and observed species.

In bacterial diversity analysis of mangrove sediments of Hong Kong (Jiang *et al.,* 2013), in one of the sample analyzed Shannon index was 7.47, while it was 6.79 in another from the same location. In a similar Illumina based bacterial diversity study on another mangrove sediment (Wang *et al.,* 2012), Shannon index value of 7.26 was obtained. In comparison, the Shannon index of 9.82 obtained in the present study is much higher than the previous reports which suggest richness and diversity of Mangalavanam mangrove sediment.

4.3.4.2.4 Accession numbers

All of the sequencing data obtained in the present study was submitted to MG-RAST database with the MG-RAST ID: 4652499.3

4.3.4.2.5 Beta diversity analysis

Beta diversity analysis was performed by comparing sequencing data of Arabian Sea and Mangalavanam mangrove metagenome. Distance matrix using Unweighted UniFrac and Weighted UniFrac distances were calculated for both the samples and shown in Table 4.7 and 4.8. From this analysis values shown in the table it is clear that bacterial diversity in Mangalavanam mangrove sediment is high compared to that of the Arabian Sea sediment.

Table 4.7: Unweighted UniFrac distance matrix for two samples

UnweightedUniFrac	Marine	Mangrove
Marine	0	0.629353261
Mangrove	0.629353261	0

Table 4.8: Weighted UniFrac distance matrix for two samples

Weighted UniFrac	Marine	Mangrove
Marine	0	0.396991744
Mangrove	0.396991744	0

From the beta diversity analysis based on Unweighted and Weighted UniFrac distance matrix, it was recognized that the bacterial diversity of Mangalavanam mangrove sediment was higher than that of the Arabian Sea sediment, which is in total agreement with the alpha diversity analysis and taxonomic profiles obtained. Shannon index value of 9.82 was obtained in the mangrove sediment against the Shannon index of 6.62 in the marine sediment. Similarly, all OTUs were classified into 52 bacterial phyla including 21 formally described bacterial phyla and 31 candidate phyla, while the OTUs of marine metagenome clustered among 43 bacterial phyla including 18 formally described bacterial phyla and 25 candidate phyla.

Bacterial diversity profiles of marine and mangrove metagenome are brought to light by the present studies based on Sanger and Next generation sequencing, revealing the roles of different bacteria in nutrient cycling and organic matter degradation. The analysis also revealed the existence of many unknown sequences, indicating a large untapped bacterial diversity in these areas.

Chapter 5

CONSTRUCTION OF METAGENOMIC LIBRARIES TO SCREEN FOR AMYLASE ENZYME PRODUCTION AND CHARACTERIZATION OF AMYLASE GENE OBTAINED FROM METAGENOMIC LIBRARY UTILIZING BIOINFORMATIC APPROACHES.

5.1 INTRODUCTION

Metagenomics is a powerful technique that can provide new insights into microbial ecology and has proved to be an efficient tool for recovery of novel genes and biomolecules (Daniel, 2005). Metagenomic methods can be focused on gene cassettes or genes encoding enzymes, and can lead to the discovery of biocatalysts for production and synthesis of secondary metabolites with bioactivity.

Microorganisms serve as a potential source for new biocatalysts, having adapted to a wide range of environmental situations which enable them to produce a variety of novel biomolecules, helping to thrive in the prevailing conditions. The resulting biochemical and physiological versatilities are considered to be a major resource for unique biotechnological products and processes (Bull *et al.*, 1992). By employing metagenomic methods small and large insert libraries can be developed and screened for novel biocatalysts. The average size of the structural genes encoding most enzymes are around 1 to 2 kb, so metagenomic library construction by means of high-copy-number plasmid vector has been widely employed to search for novel enzymes, regardless of the relatively short lengths of DNA that can be obtained by this method (Henne *et al.*, 1999).

In the past few years, there was a mounting demand from industry for novel enzymes to improve upon existing ones or to establish novel bioprocess (Steele and Streit, 2005). Among different classes of enzymes, hydrolases are of great importance due to their broad substrate spectrum, stability and activity in organic solvents and independence from cofactors (Schmid *et al.,* 2001). Amylases are among the most important hydrolytic enzymes, belonging to family 13 (GH-13) of the glycoside hydrolase group of enzymes (Bordbar *et al.,* 2005). Even though amylases can be obtained from several sources, generally microbial amylases rather than plants and animals sourced, meet industrial demands (Pandey *et al.,* 2000).

Metagenomic approach have proved to be a powerful tool in mining novel enzymes with improved properties from highly diverse bacterial communities harboring in variety of environmental samples. Regardless of the abundance of new enzymes obtained by metagenomic methods, reports regarding metagenome-derived amylases are relatively few, especially from marine and mangrove metagenome. As marine and mangrove sediments are a unique ecological niche for a variety of microorganism with virtually limitless variation among strains and representing an enormous unexplored reservoir of genetic and metabolic diversity, serve as hotspot for novel biomoleclules. So the present study focused on the potential of marine and mangrove metagenomic library for amylase production and the characterization of the resulting amylase gene employing bioinformatic methods.

5.2 MATERIALS AND METHODS

5.2.1 Construction of marine and mangrove metagenomic libraries to screen for enzyme production

5.2.1.1 Bacterial strain and vector used for functional library construction

Electrocompetent *Escherichia coli* DH10B bacterial cell (Invitrogen, USA) and *Bam*H1 digested dephosphorylated pUC19 (Thermo Scientific, USA)

were used as host and vector respectively. Characteristics of the host and the vector are given in Table 5.1. Vector map of pUC19 is shown in Appendix II.

Table 5.1: Characteristics of the host and vector

Strain/plasmid	Genotype/Description
E. coli DH10B	F⁻*end*A1 *rec*A1 *gal*E15 *gal*K16 *nup*G *rps*L Δ*lac*X74 *ϕ80lacZ*ΔM15 *ara*D139 Δ(*ara, leu*)7697 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) λ⁻
pUC19	Ampicillin resistance

5.2.1.2 Partial digestion of metagenomic DNA

Marine and mangrove metagenomic DNA isolated using UltraClean™ Soil DNA isolation kit was digested with restriction enzyme *Sau*3A1 (Fermentas, USA) and incubated at 37ºC for different time scales ranging from 5 to 30 min. The composition of the reaction mixture is given in Table 5.2. Enzyme reaction was terminated by heat inactivated at 65ºC for 10 min. Digested samples were run on 0.8% agarose gel as described in section 3.2.1.4. Optimized time fractions containing DNA fragment in the range of ~2-10 kb was noted and multiple digestions were carried out in that specific time. After digestion for appropriate time, DNA fragments were extracted from agarose gel using gel extraction kit (GeNei, India) following manufacturer's instructions.

Table 5.2: Ingredients of restriction digestion reaction mixture

Ingredient	**Quantity**
DNA (750 ng/µL)	4 µL
*Sau*3A I (1U/ µL)	3 µL
10X assay buffer	1 µL
100X BSA	0.1 µL
Sterile H_2O	1.9 µL
Total volume	10 µL

5.2.1.3 Ligation of size-fractionated metagenomic DNA with pUC19 vector

Gel eluted, partially digested metagenomic DNA (~2-10kb) were ligated into the *Bam*HI digested and dephosphorylated pUC19 vector using rapid ligation kit (Promega, USA) following manufacturer's instructions and incubated at room temperature for 30 min. The insert to vector ratio was 3:1. The reaction mixture for DNA ligation is given in Table 5.3. Appropriate positive and negative controls were also included.

Table 5.3: Ingredients of ligation reaction mixture

Ingredient	Quantity
Insert DNA (750 ng/µL)	4 µL
Vector DNA (0.5ng/ µL)	2 µL
10X assay buffer	1 µL
T4 DNA ligase	0.1 µL
Sterile H_2O	2.9 µL
Total volume	10 µL

5.2.1.4 Transformation of *E. coli* DH10B

E. coli DH10B was transformed with ligated vectors by electroporation using Micropulser II (BioRad, USA). Five micro liter of ligation reaction mix was mixed with 45µL of electrocompetent *E. coli* DH10B and incubated on ice for 5 min. The mixture was transferred to 0.1 cm chilled electroporation cuvettes (Bio Rad). An electric pulse of strength of 12.5 KV/cm was applied and cuvette was transferred back in ice and 950 µL of SOC medium (Appendix I) was transferred to cuvette to collect transformed *E. coli* DH10B cells. Transformed *E. coli* DH10B were grown at 37°C for 1 h with constant shaking at 200 rpm. Appropriate positive and negative controls were also included.

5.2.1.5 Selection of recombinant clones

Blue-white screening was employed to determine the recombinant colonies in the metagenomic library. Transformed *E. coli* DH10B was plated onto

LB agar medium supplemented with ampicillin (60 mg/mL), X-gal (20 mg/mL) and IPTG (100 mg/mL) to screen for the recombinants. The white colonies representing the recombinants were picked and stored as functional library and maintained as glycerol stock as described in section 4.2.1.3. Clones in the marine sediment metagenomic library were prefixed with BTM and those in the mangrove library were prefixed with MS followed by the numeric clone number.

5.2.1.6 Screening of the library for amylase enzyme production by plate assay

The clones were spot inoculated on LB agar plates containing 1% soluble starch. Starch substrates used for enzyme screening were sterilized under 10 lbs for 10 min to prevent charring and mixed with sterilized LB agar. Plates used for enzyme screening were supplemented with ampicillin (60 mg/mL) to reduce contaminants. Plates were incubated for 1-2 days. The amylolytic activity was determined as zone of clearance surrounding the colonies after flooding with Iodine-Potassium iodide solution (I-KI) (Appendix I) over the plate (Skerman, 1969).

5.2.2 Characterization of amylase gene from amylolytic metagenomic clone BTM109

5.2.2.1 Plasmid DNA isolation from clone BTM109

The amylase positive clone was inoculated in 5mL LB broth containing ampicillin and plasmid isolation was carried out employing alkaline lysis method as described in section section 4.2.1.4

5.2.2.2 PCR amplification of DNA insert

The recombinant plasmids were subjected to PCR amplification using vector specific primers (Fermentas, USA), which is the M13 forward and reverse primers flanking the multiple cloning site of pUC19 vector, and the primer sequences are given in Table 5.4. PCR program is as given in Table 5.5. All the

PCR components are same as described in Table 3.4 except that Long *Taq* DNA polymerase (Fermentas, USA) was used instead.

Table 5.4: M13 sequencing primers

Primer	Primer sequence
M13 F	5'-TGTAAAACGACGGCCAGT-3'
M13 R	5'-CAGGAAACAGCTATGACC-3'

Table 5.5: Program for PCR amplification of amylase gene

Step	Temperature	Time	
Initial Denaturation	94°C	1.5 min	
Denaturation	94°C	30 sec	
Annealing	58°C	30 sec	30 cycles
Extension	72°C	2 min	
Final Extension	72°C	10 min	

The PCR products were analysed on agarose gel as described under section 3.2.1.4 for the presence of amplicons. Appropriate DNA ladders were also included.

5.2.2.3 Sequencing of DNA insert

The recombinant plasmid was sequenced from both ends using vector specific M13 sequencing primers employing Sanger's Dideoxy method using ABI 3730 Excel (Applied Biosystems, USA) at Scigenom Labs, Kochi, Kerala.

5.2.2.4 *In silico* analysis and structure prediction of amylase gene

Nucleotide sequence was compared with the sequences in the GenBank database using blastn (Altschul *et al.,* 1997) at NCBI website. (http://blast.ncbi. nlm.nih.gov). Sequence manipulation was conducted using sequence manipulation suite (www.bioinformatics.org/sms2). Sequences were screened for vector contamination using VecScreen Tool of NCBI. Open reading frame (ORF) in the

nucleotide sequences was determined using the ORF finder (Wheeler *et al*., 2003) of NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).

The deduced amino acid sequence was obtained by translation of the nucleotide sequence using online tool ExPASy (Gasteiger *et al*., 2003). The predicted function of ORFs was annotated using protein-blast (blastp) against the NCBI non-redundant protein database (Altschul *et al.,* 1997) to search for identical proteins in the database. Nucleotide and amino acid sequences were aligned using ClustalW2 (Larkin *et al*., 2007). Phylogenetic tree were constructed using MEGA software version 5.0 (Tamura *et al*., 2007). Tree topology was deduced by Neighbour-Joining method (Saitou and Nei, 1987) using 1,000 bootstrap iterations. Signal peptides in deduced amino acid sequence were analyzed with SignalP version 4.1 (Center for Biological Sequence Analysis, Technical University of Denmark [http://www.cbs.dtu.dk]) (Petersen *et al.,* 2011). ProtParam tool of ExPASy was used for the computation of various physical and chemical parameters of the sequence based on its amino acid sequence (http://web.expasy.org/protparam).

Conserved domains in the amino acid sequence were analyzed using conserved protein domain database of NCBI (Marchler-Bauer and Bryant, 2004). The secondary structures encoded by the amino acid sequences were also predicted with deduced amino acid sequences using Phyre^2 software (Kelley and Sternberg, 2009). Based on amino acid sequence homologies, protein model was built using Swiss Model (Arnold *et al*., 2006). The sequences were deposited as Sequin file to GenBank database as accession number were obtained for the submission.

5.3 RESULTS AND DISCUSSION

5.3.1 Construction of marine and mangrove metagenomic library for screening of amylase producer

5.3.1.1 Restriction digestion of marine and mangrove metagenomic DNA

The restriction digestion of marine and mangrove metagenomic DNA is shown in Fig 5.1. Fragments of ~2-10 kb size were extracted from the gel and used for ligation into the pUC19 vector.

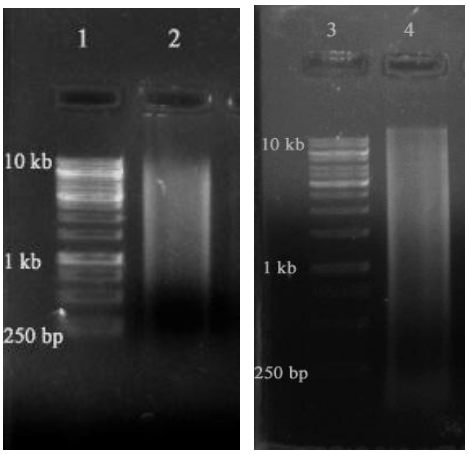

Fig 5.1: Agarose gel showing of restriction digestion of metagenomic DNA

Lane 1 and 3- 1 kb ladder, Lane 2 - Digested marine metagenomic DNA,
Lane 4 - Digested mangrove metagenomic DNA

5.3.1.2 Construction of marine and mangrove metagenomic libraries

Metagenomic libraries were constructed from the digested metagenomic DNA in the pUC19 vector transformed into *E. coli* DH10B hosts. Marine sediments metagenomic library consisting of 562 recombinant clones was obtained and was designated as BTM1-562. Mangrove metagenomic library with 551 recombinant clones were obtained and were designated as MS1-551.

Various techniques have been developed to extract specific genes with novel properties from environmental samples. One among the most popular

approach is the metagenomic shot gun library construction and activity based screening of libraries. Construction of large insert libraries with potential biomeolecules have also been reported previously (Brady *et al*., 2001; Gillespie *et al*., 2002), suggesting that large insert libraries are more informative because of the large genomic content within the libraries and also for allowing access to neighboring genes required for effective expression of target genes; in addition it will possibly provide phylogenetic origin of the target genes which can easily be missed in small insert libraries. Still large insert libraries have certain drawbacks associated with them wherein the heterologous transcription signals might not get recognized by the host cell. Previous studies have reported the construction of small insert DNA libraries and subsequent recovery of various novel genes like chitinase, lipase, amylase, protease and pectinase (Cottrell *et al.,* 1999; Henne *et al.,* 2000; Yun *et al.,* 2004; Neveu *et al.,* 2011; Wang *et al.,* 2014). The vector pUC18 or pUC19 has been widely used for cloning purpose as they have high copy number (Donovan *et al*., 1997; Sharma *et al*., 2005). Similarly increased numbers of transformants were obtained during *E. coli* transformation with pUC19 (Kerkhof and Goodman, 2009).

5.3.1.3 Screening of metagenomic clones for amylase production

The metagenomic library comprising a total of 1113 recombinant clones from both mangrove and marine sediments were screened for amylase activity by plate assay on starch agar plates. Out of the 1113 metagenomic clones screened, only marine metagenomic clone BTM109 produced extracellular amylase enzyme, with an observable zone of clearance on the starch agar plate. BTM109 showing zone of clearance is shown in Fig 5.2.

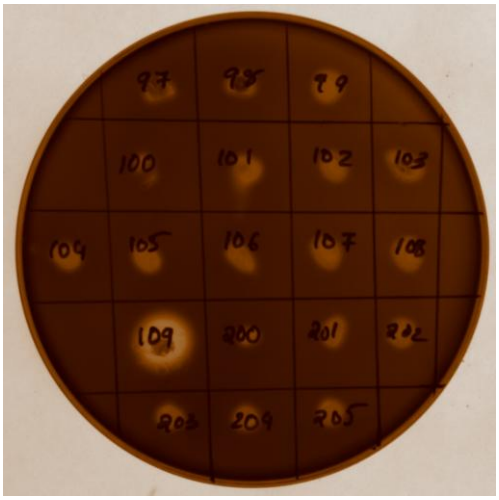

Fig 5.2: Starch agar plate showing amylase production by clone BTM109

In spite of the high abundance of novel enzymes captured by metagenomic approaches, there is comparatively less data concerning metagenome-derived amylases. However, a few α-amylases were identified through metagenomic approach and characterized which include a thermostable α-amylase identified from DNA libraries originating from environmental samples (Richardson *et al.,* 2002; Voget *et al.,* 2003) and out of 14 amylolytic clones reported from soil metagenomic library, only 4 clones are characterized (Yun *et al.,* 2004). Two soil metagenomic library derived α-amylase are reported and characterized (Sharma *et al.,* 2010; Vidya *et al.,* 2011). An α-amylase isolated from a gastrointestinal metagenomic library has been biochemically and molecularly characterized (Xu *et al.,* 2014b) and a cold-adapted α-amylase was identified in a metagenomic library from the cold and alkaline environment (Vester *et al.,* 2015). However the present study reports for the first time, information regarding an α-amylase obtained from metagenomic library from Arabian Sea sediments from 96 meter depth.

5.3.2 Characterization of amylase gene from amylolytic metagenomic clone BTM109

5.3.2.1 PCR amplification of insert DNA

The DNA insert within the plasmid was successfully amplified using the M13 vector specific primers and was visualized on agarose gel (Fig. 5.3). From the gel picture the ~1800 bp amplicon was clearly visible.

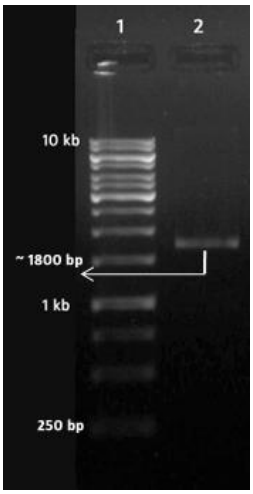

Fig 5.3: Agarose gel electrophoresis of amplified insert

Lane 1: 1 kb ladder, Lane 2: amplified DNA insert

5.3.2.2 BLAST analysis of nucleotide sequence of clone BTM109

The insert within clone BTM109 was sequenced. Nucleotide sequence of 1743 bp length was obtained. The identity of the sequence was determined by comparing with sequences in the NCBI database. Megablast algorithm (shows similarity with highly similar sequences) showed identity only to a single sequence in the database with 99% identity, i.e. to an α-amyalse gene obtained from an uncultured bacterium clone. Discontiguous megablast (shows similarity with more dissimilar sequences) was employed and the accession numbers and description of the ten hits with maximum identity with sequence from BTM109 are detailed in Table 5.6.

Table 5.6: Nucleotide BLAST analysis of amylase gene of clone BTM109

Accession No.	Description	Identity
AY383543	Uncultured bacterium AmyM genes, complete cds	99%
CP003346	*Echinicola vietnamensis* DSM 17526, complete genome	71%
CP003281	*Belliella baltica* DSM 15883, complete genome	72%
CP002349	*Marivirga tractuosa* DSM 4126, complete genome	71%
CP001656	*Paenibacillus* sp. JDR-2, complete genome	71%
CP010777	*Rufibacter* sp. DG31D, complete genome	64%
CP010429	*Spirosoma radiotolerans* strain DG5A, complete genome	68%
FP929043	*Eubacterium rectale* M104/1 draft genome	71%
FP929042	*Eubacterium rectale* DSM 17629 draft genome	70%
CP001107	*Eubacterium rectale* ATCC 33656, complete genome	70%

The gene sequence from the clone BTM109 showed 99% identity to an α-amylase gene obtained from an uncultured bacterial clone and similarities were observed to complete genome of several organisms with 71% or less identity. The sequence was analyzed by NCBI-ORF finder tool, by which an ORF consisting of a 1554 bp was identified in the +3 reading frame encoding a protein with 517 amino acid. The ORF obtained by ORF finder tool was depicted in Fig 5.4 identifying ATG as the start codon and TAA as the stop codon in the protein. The gene sequences obtained in this study were submitted to GenBank under accession number KR514290.

```
  1 atgaaaaaatccatcctaactacctccatttttgctcttgtcgct
    M  K  K  S  I  L  T  T  S  I  F  A  L  V  A
 46 ttttcttcctgcgaaaaaaagcctgctccggaagtcaaaaactat
    F  S  S  C  E  K  K  P  A  P  E  V  K  N  Y
 91 tggcctcaagcaggagtgacctatgagatttttgttcaatctttt
    W  P  Q  A  G  V  T  Y  E  I  F  V  Q  S  F
136 tatgattctaacggagacagtattggggattttaatggggtcact
    Y  D  S  N  G  D  S  I  G  D  F  N  G  V  T
181 caaaaactggactatgtgaaggagttgggggccaatgccatttgg
    Q  K  L  D  Y  V  K  E  L  G  A  N  A  I  W
226 tttatgccgattatgccttcgccaacttaccataagtacgatgtg
    F  M  P  I  M  P  S  P  T  Y  H  K  Y  D  V
271 acggactacaaggcggttcatccagattacggtacgctggatgat
```

```
    T  D  Y  K  A  V  H  P  D  Y  G  T  L  D  D
316 ttcaaaaagcttttggacgaagctcacaagcgggacatcaagatt
    F  K  K  L  L  D  E  A  H  K  R  D  I  K  I
361 gtgatcgatttgatcatcaatcacaccagcaacgaacatccgtgg
    V  I  D  L  I  I  N  H  T  S  N  E  H  P  W
406 tttttggaagcaaaatccggtagggataatccctatcgcgattac
    F  L  E  A  K  S  G  R  D  N  P  Y  R  D  Y
451 tacgtgtgggcgcaaaaggacaccattgctgatttcttgaacaaa
    Y  V  W  A  Q  K  D  T  I  A  D  F  L  N  K
496 aagaccatcacgtttgatttggataatatccgtcaatggcatgac
    K  T  I  T  F  D  L  D  N  I  R  Q  W  H  D
541 ccgggacagggagaagatttttactacgggtttttctggggtgga
    P  G  Q  G  E  D  F  Y  Y  G  F  F  W  G  G
586 atgcctgatctgaactttgacaatcctaaggtaagagaggaaatc
    M  P  D  L  N  F  D  N  P  K  V  R  E  E  I
631 tatgaaatcggacgattctggttggaagaagtgggtgtggacgga
    Y  E  I  G  R  F  W  L  E  E  V  G  V  D  G
676 tttcggttggacgctgccaagcatatttttcccgatgaccgacct
    F  R  L  D  A  A  K  H  I  F  P  D  D  R  P
721 ttggataatcatgccttttggaaagaattccgcgcaaaaatggaa
    L  D  N  H  A  F  W  K  E  F  R  A  K  M  E
766 gtcataaagccggatgtttacttggtaggagaggtgtatgacaaa
    V  I  K  P  D  V  Y  L  V  G  E  V  Y  D  K
811 aaagaagtcgtggctccttatcttcctgggttgccagccttgttc
    K  E  V  V  A  P  Y  L  P  G  L  P  A  L  F
856 aactttgattttcattacactctgcttgagaccatgaataccggc
    N  F  D  F  H  Y  T  L  L  E  T  M  N  T  G
901 gacgggatgcttttggccaagaagcagaaggagattttggacttt
    D  G  M  L  L  A  K  K  Q  K  E  I  L  D  F
946 tatcagggaatcacttcaagctttatcgatgcgaccatttcttct
    Y  Q  G  I  T  S  S  F  I  D  A  T  I  S  S
991 aaccatgatcagccccgtctgctgaatgaattgggatctgatccg
    N  H  D  Q  P  R  L  L  N  E  L  G  S  D  P
1036 gccaaatacaagcaggcgatcgcagtgatgctcagcatgccgggt
     A  K  Y  K  Q  A  I  A  V  M  L  S  M  P  G
1081 gcgccatatttgtattatggggaagagatcggcatgctgggtctc
     A  P  Y  L  Y  Y  G  E  E  I  G  M  L  G  L
1126 aagccggacgagcatatccgggagcctttcctttgggatgaaaaa
     K  P  D  E  H  I  R  E  P  F  L  W  D  E  K
1171 agcaaggatacaggtcgcaccaagtggatcaaacccaaatacagc
     S  K  D  T  G  R  T  K  W  I  K  P  K  Y  S
1216 aaagactcaacagtaacctcattggaggtccaaaagaaagattcg
     K  D  S  T  V  T  S  L  E  V  Q  K  K  D  S
1261 aacagctatttcaatcattacaaaaacttaattgctcttcgtaat
     N  S  Y  F  N  H  Y  K  N  L  I  A  L  R  N
1306 tcttatcctgctttggctattggttctttggaacttccagctgag
     S  Y  P  A  L  A  I  G  S  L  E  L  P  A  E
1351 gaattacctaaaagtgtaatggcatatttcagaaagtctggagat
     E  L  P  K  S  V  M  A  Y  F  R  K  S  G  D
```

```
1396 caggaaatctttgtcgttcacaatgtggacaaggaggaagttgac
     Q  E  I  F  V  V  H  N  V  D  K  E  E  V  D
1441 attcagcttccagagggatttgaagaagtaattttctatttgggt
     I  Q  L  P  E  G  F  E  E  V  I  F  Y  L  G
1486 gaggggaaaaacagttcaggaaaacttcagctaaaaggcaactcg
     E  G  K  N  S  S  G  K  L  Q  L  K  G  N  S
1531 agtatggtttttttgaaggattaa 1554
     S  M  V  F  L  K  D  *
```

Fig 5.4: Sequence analysis of BTM109 using ORF finder
Bold letters indicates the start and stop codons

5.3.2.3 Phylogenetic analysis of the alpha-amylase gene of clone BTM109

In order to find the phylogentic relationship of nucleotide sequence of BTM109, unrooted tree was constructed based on neighbor joining method using alpha-amylase sequences of different classes of microorganisms and uncultured domain of microbes reported in nucleotide sequence database of NCBI. Results indicated that BTM109 clustered with α-amylase of uncultured bacteria and are claded separately from alpha-amylase obtained from different microbes, suggesting that α-amylase obtained from BTM109 might be a representative from uncultured division of bacteria. Phylogenetic tree is shown in Fig 5.5.

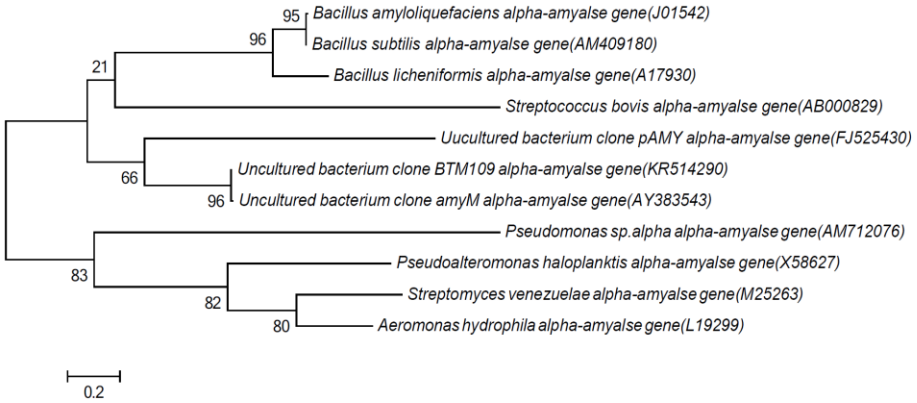

Fig 5.5: The phylogenetic affiliation of nucleotide sequences from BTM109 compared to similar sequences in the GenBank. Bootstrap values are given at the branching points and scale bar shows sequence divergence. Accession numbers are given in parentheses.

5.3.3 *In silico* analysis using deduced amino acid sequence of alpha-amylase gene BTM109

5.3.3.1 BLAST analysis of deduced amino acid sequence of amylase gene of clone BTM109

The 1554 bp nucleic acid sequence of the amylase gene of clone BTM109 was translated by ExPASy into its corresponding amino acid sequence consisting of 517 amino acids, and was compared with those available from GenBank using online protein BLAST tool-blastp. The GenBank accession numbers and description of ten hits with maximum identity after protein blast of deduced amino acid sequences of BTM109 is as detailed in Table 5.7.

Table 5.7: Protein BLAST analysis of amylase gene of clone BTM109

Accession No.	Description	Identity
AAQ89599	AmyM [uncultured bacterium]	99%
WP_026950142	α-amylase [*Algoriphagus mannitolivorans*]	85%
WP_026970068	α-amylase [*Algoriphagus terrigena*]	81%
WP_029660871	α-amylase [*Algoriphagus marincola*]	81%
WP_035483031	α-amylase [*Algoriphagus marincola*]	81%
WP_014773782	α-amylase [*Belliella baltica*]	71%
WP_015266641	Glycosidase [*Echinicola vietnamensis*]	71%
WP_009184110	α-amylase [*Cecembia lonarensis*]	73%
ERM84602	α-amylase [*Rhodonellum psychrophilum*]	71%
WP_026333543	α-amylase [*Rhodonellum psychrophilum*]	71%

The protein BLAST analysis confirmed that the amino acid sequence obtained was similar to that encoding α-amylase. The deduced amino acid sequence showed 99% identity with α-amylase gene obtained from uncultured bacterium clone (AAQ89599) and similarities with α-amylase gene of other organisms.

In a previous study, soil metagenomic subclone constructed in pUC19 vector was sequenced with M13 primers and identified putative α-amylase gene pAMY of length 909 bp encoding a protein of 302 amino acids (Sharma *et al.,* 2010). Similarly studies on cold-adapted amylase from a metagenomic library identified an α-amylase Amy_{I3C6} with 486 amino acids (Vester *et al.,* 2015). A function-driven metagenomic approach identified two non-homologous endo-acting amylases with 479 and 843 amino acids which share no sequence similarity with any known amylase or glycosidase (Delavat *et al.,* 2012). In a similar study on fecal microbial metagenome an α-amylase gene amyPL with 1,539 bp, with deduced AmyPL polypeptide consisting of 512 amino acid was identified (Xu *et al.,* 2014b).

Similarly studies on cloning and characterization of recombinant α-amylase have identified amylase gene with varying nucleotide sequences and protein chain length. Cloning and expression of α-amylase gene from *Halothermothrix orenii* identified an amylase gene amyA of 1545 bp encoding a 515 residue protein (Mijts and Patel, 2002). Likewise, cloning of α-amylase gene from *Thermotoga maritima* identified an amylase gene amyB of 1269 bp in length, encoding a protein of 422 amino acids (Lim *et al.,* 2003). Cloning and characterization of a alpha-amylase gene from *Alkalimonas amylolytica* identified 1764 bp amylase gene and was predicted to encode a 587 amino acid (Wang *et al.,* 2006). Previous reports on cloning and expression of α-amylase gene from marine bacterium *Pseudoalteromonas* sp. reported an ORF of 2,007 base pairs which encodes a protein of 669 amino acids (Tao *et al.,* 2008). All these studies indicate the considerable size differences in bacterial α-amylase gene and the encoded protein.

5.3.3.2 Phylogenetic analysis of deduced amino acid sequence of BTM109

Unrooted tree was constructed based on neighbor-joining method to determine the phylogenetic interrelationship of the deduced amino acid sequences

of the α-amylase of BTM109 with amino acid sequences of different classes of microorganisms and uncultured domain of microbes reported in nucleotide sequence database of NCBI. From the phylogenetic analysis it is clear that the deduced amino acid sequences of α-amylase of BTM109 clustered with α-amylase of uncultured bacterium and are well separated from other sequences. Their clading separately suggests that the deduced amino acid sequence of BTM109 do not show any similarity with alpha-amylases from identified microorganisms. Phylogenetic tree is shown in Fig 5.6.

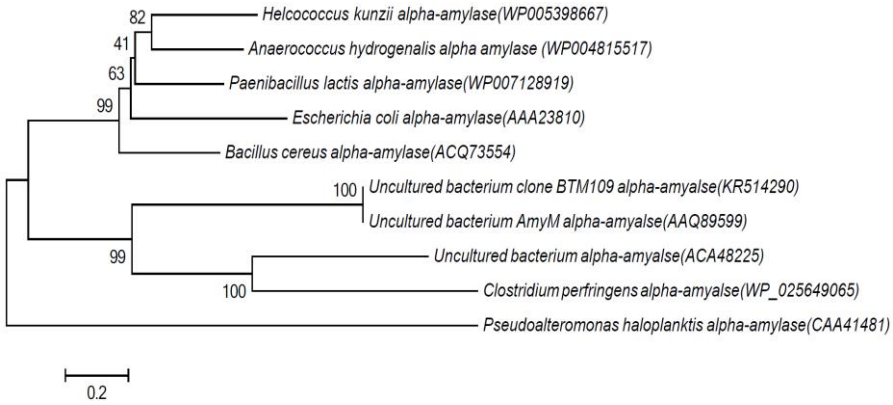

Fig 5.6: The phylogenetic affiliation of deduced amino acid sequence from BTM109 compared to similar sequences in the GenBank. Bootstrap values are given at the branching points and scale bar shows sequence divergence. Accession numbers are given in parentheses.

5.3.3.3 Prediction of signal peptide of BTM109

The signal peptide sequence present in sequence of BTM109 was predicted using SignalP 4.0 online program. The deduced amino acid sequence of BTM109 was found to contain a prokaryotic signal peptide of 24 amino acids at the N-terminus and is shown in Table 5.8.

Previous studies on a metagenomic α-amylase, the first 38 residues were predicted to be a signal peptide (Sharma *et al.,* 2010). Likewise a recombinant α-

amylase expressed in *E. coli* reportedly had a 34 amino acid signal peptide (Roy *et al.,* 2013), while that from *Geobacillus thermoleovorans* also had 34 amino acids at the N-terminus (Mehta and Satyanarayana, 2013). A 25 amino acid putative signal peptide was identified in a recombinant α-amylase from *Halothermothrix orenii* (Mijts and Patel, 2002). Similarly the first 24 amino acids were predicted to be signal peptide in a recombinant α-amylase obtained from *Pseudoalteromonas* sp. (Tao *et al.,* 2008).

Table 5.8 depicts the signal peptide obtained in the present study and the previous reports discussed above. Sequence comparison identified that the signal peptides differ significantly in length and sequence relatedness.

Table 5.8: Signal peptide sequences of α-amylase

Identity	Signal peptide sequence
BTM 109 α-amylase	MKKSITTSIFALVAFSSCEKKP
Uncultured bacterium clone α-amylase (Sharma *et al.,* 2010)	MDAVPGGGAGIVDRSKIPSWEYEAGASAIPW ISAPKGM
Bacillus subtilis α-amylase (Roy *et al.,* 2013)	MFAKRFKTSLLPLFAGFLLLFHLVLAGPAAA SAE
Geobacillus thermoleovorans α-amylase (Mehta and Satyanarayana, 2013)	MLTFHRIIRKGWMFLLAFLLTASLFCPTGQPA KA
Halothermothrix orenii α-amylase (Mijts and Patel, 2002)	MVKLKRLSFFMFVTLLVFISVFPVY
Pseudoalteromonas sp. α-amylase (Tao *et al.,* 2008).	MKLSKMITTAGFSLGLTLPSLVSA

5.3.3.4 Prediction of protein parameters using ProtParam

The computed parameters employing ProtParam tool of ExPASy included the molecular weight, theoretical pI and amino acid composition (Table 5.9) of the sequence and as given below.

Number of amino acids:	517
Molecular weight:	59kDa
Theoretical pI:	5.18
Total number of negatively charged residues (Asp + Glu):	80
Total number of positively charged residues (Arg + Lys):	60

Table 5.9: Amino acid composition of BTM109

Amino acid	No. of residues	Percentage
Ala (A)	28	5.40%
Arg (R)	14	2.70%
Asn (N)	23	4.40%
Asp (D)	44	8.50%
Cys (C)	1	0.20%
Gln (Q)	14	2.70%
Glu (E)	36	7.00%
Gly (G)	33	6.40%
His (H)	13	2.50%
Ile (I)	31	6.00%
Leu (L)	44	8.50%
Lys (K)	46	8.90%
Met (M)	12	2.30%
Phe (F)	31	6.00%
Pro (P)	29	5.60%
Ser (S)	30	5.80%
Thr (T)	20	3.90%
Trp (W)	10	1.90%
Tyr (Y)	29	5.60%
Val (V)	29	5.60%
Pyl(O)	0	0.00%
Sec (U)	0	0.00%

The analysis indicated that the dominant amino acid present in the sequence was Lysine (8.90%). Lysine is positively charged, polar amino acid and is quite frequent in protein active or binding sites. They bond with negatively charged amino acids (such as Asp, Glu) to create stabilizing hydrogen bonds, that can be important for protein stability. Asp is represented by 8.50% of the total amino acids in BTM109. This might be the possible reason of salt tolerance, stability of the enzyme with solvents and other chemicals tested and are explained in chapter 6.

5.3.3.5 Prediction of conserved domain of BTM109

The conserved domain search identified that BTM109 belongs to alpha amylase catalytic domain family which comprises the largest family of glycoside hydrolases (GH), with the majority of enzymes acting on starch, glycogen, and

related oligo- and polysaccharides. These proteins catalyze the transformation of α-1,4 and α-1,6 glycosidic linkages with retention of the anomeric center. The protein is described as having 3 domains: A, B, C. A is a (beta/alpha) 8-barrel; B is a loop between the beta 3 strand and alpha 3 helix of A; C is the C-terminal extension. The majority of the enzymes have an active site cleft found between domains A and B where a triad of catalytic residues (Asp, Glu and Asp) performs catalysis. Catalytic domains identified are shown in Fig. 5.7.

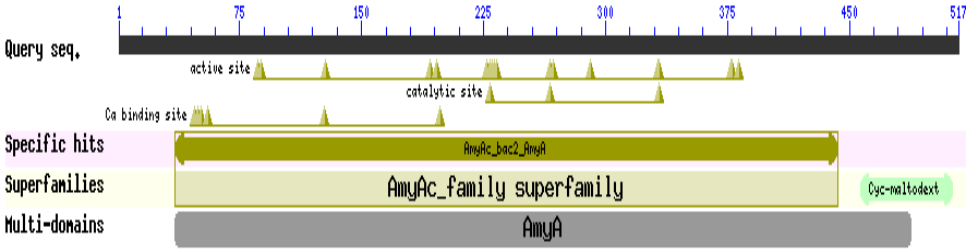

Fig 5.7: Conserved domains of BTM109

It was identified that the active site stretched from amino acid 90 to 375 in which the catalytic residues reside between the amino acids from 225 to 345, while catalytic triad was formed by three residues corresponding to Asp229, Glu266 and Asp333; where Glu acts as a proton donor and Asp as a nucleophile during enzyme catalysis. It was also noted that the calcium binding site stretched from amino acid 46 to 198. Multiple sequence alignment of deduced amino acid sequences with the top ten BLAST hits obtained showing the residues of catalytic triad and calcium binding site and the conserved active site residues are depicted in Fig. 5.8, 5.9 and 5.10 respectively. Complete multiple sequence alignment of deduced amino acid sequences with the top ten BLAST hits were shown in Appendix II.

Fig 5.8: Multiple sequence alignment of deduced amino acid sequences showing residues of catalytic triad, Underlined residues represents catalytic triad and accession numbers of the sequences are given in the left

Previous studies reported that conserved domains present in recombinant α-amylase gene from *Bacillus subtilis* strain AS01a expressed in *Escherichia coli* was identified with CDD of NCBI and predicted to be a member of family glycoside hydrolase (Roy *et al.,* 2013). Similarly the catalytic triad residue in a fecal metagenomic α-amylase was identified to be Asp233, Glu265 and Asp336 (Xu *et al.,* 2014b). Another study on a recombinant α-amylase gene from a *Bacillus licheniformis* reported the residues involved in the active site as Asp231, Asp328 and Glu261 (Hmidet *et al.,* 2008), while it was Asp268, Glu298 and Asp365 (Mehta and Satyanarayana, 2013), whereas it was identified as Asp417, Glu446 and Asp511 in a recombinant α-amylase from *Petrotoga* sp. (Le *et al.,* 2012). A recombinant α-amylase gene from a *Bacillus* strain reportedly had Asp217, Glu249 and Asp310 as catalytic residues (Asoodeh *et al.,* 2014).

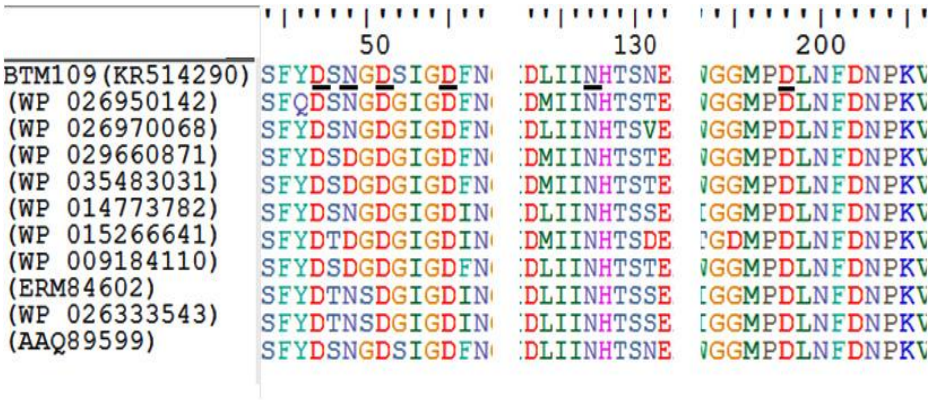

Fig 5.9: Multiple sequence alignment of deduced amino acid sequences showing residues involved in calcium binding. Underlined residues represents calcium binding residues and accession numbers of the sequences are given in the left

It was identified that calcium binding site present in BTM109 was composed of 6 residues that comprised the conserved feature identified by the multiple sequence alignment pattern. It was noted that the residues were located from amino acid 46 to 198.

Calcium ions are essential for the action of most members of α-amylase family and studies have indicated that the number of calcium binding residues within an enzyme may vary considerably. Four calcium binding residues were identified in a recombinant α-amylase obtained from ruminal bacterium *Butyrivibrio fibrisolvens* (Rumbak *et al.*, 1991), whereas it was 17 different secondary binding sites for calcium in an α-amylase obtained from *B. amyloliquefaciens* (Saboury, 2002) .

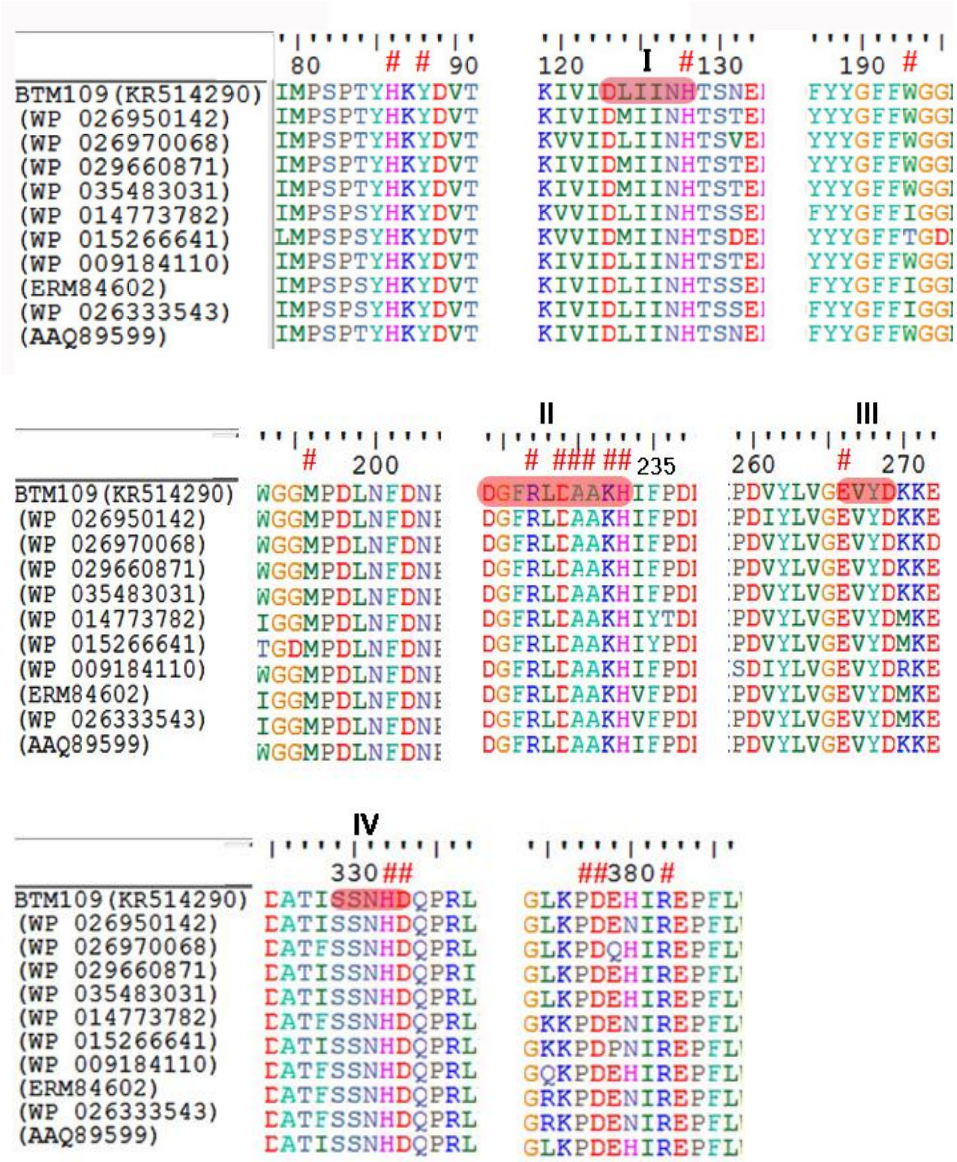

Fig 5.10: Multiple sequence alignment of deduced amino acid sequences showing conserved residues of active site (Four conserved regions are indicated by colored boxes with the active site residues indicated by #)

The multiple sequence alignment also allowed the determination of the four highly conserved regions among the amylolytic enzymes, designated as I, II,

III and IV in the deduced amino acid sequence of BTM109; and they were found to be conserved in all the sequences. 18 residues present in the active site of BTM109 identified by NCBI conserved domain search was also indicated in the alignment pattern. The four conserved regions contained residues that are involved in the key catalytic process of the enzyme. It was identified that the catalytically active residues (2 Asp and 1 Glu) were located in regions II, III and IV, while region I contained a calcium binding site.

Previous study on cloned α-amylase gene from *Bacillus subtilis* had reported four conserved regions (Roy *et al.,* 2013), but the sequences were dissimilar from those in BTM109 except for the region II. Similarly, in a recombinant α-amylase gene from *Geobacillus thermoleovorans* (Mehta and Satyanarayana, 2013), only region II was identical with that of BTM109. So it can be inferred that among the four regions, second region was identified to be highly conserved among all different alpha-amylases.

5.3.3.6 Secondary structure prediction using Phyre^2 software

The secondary structure was predicted using Phyre^2 software. The amino acid sequence was aligned with the predicted secondary structures and this is as shown in Fig 5.11. The predicted structure consisted of 33% α-helix and 18% β-strand and the model was predicted with 10% disorder. Secondary structure prediction indicated that calcium binding sites and active sites are present in the deduced amino acid sequence of BTM109 which is in agreement with the conserved domain elucidated using conserved domain database of NCBI.

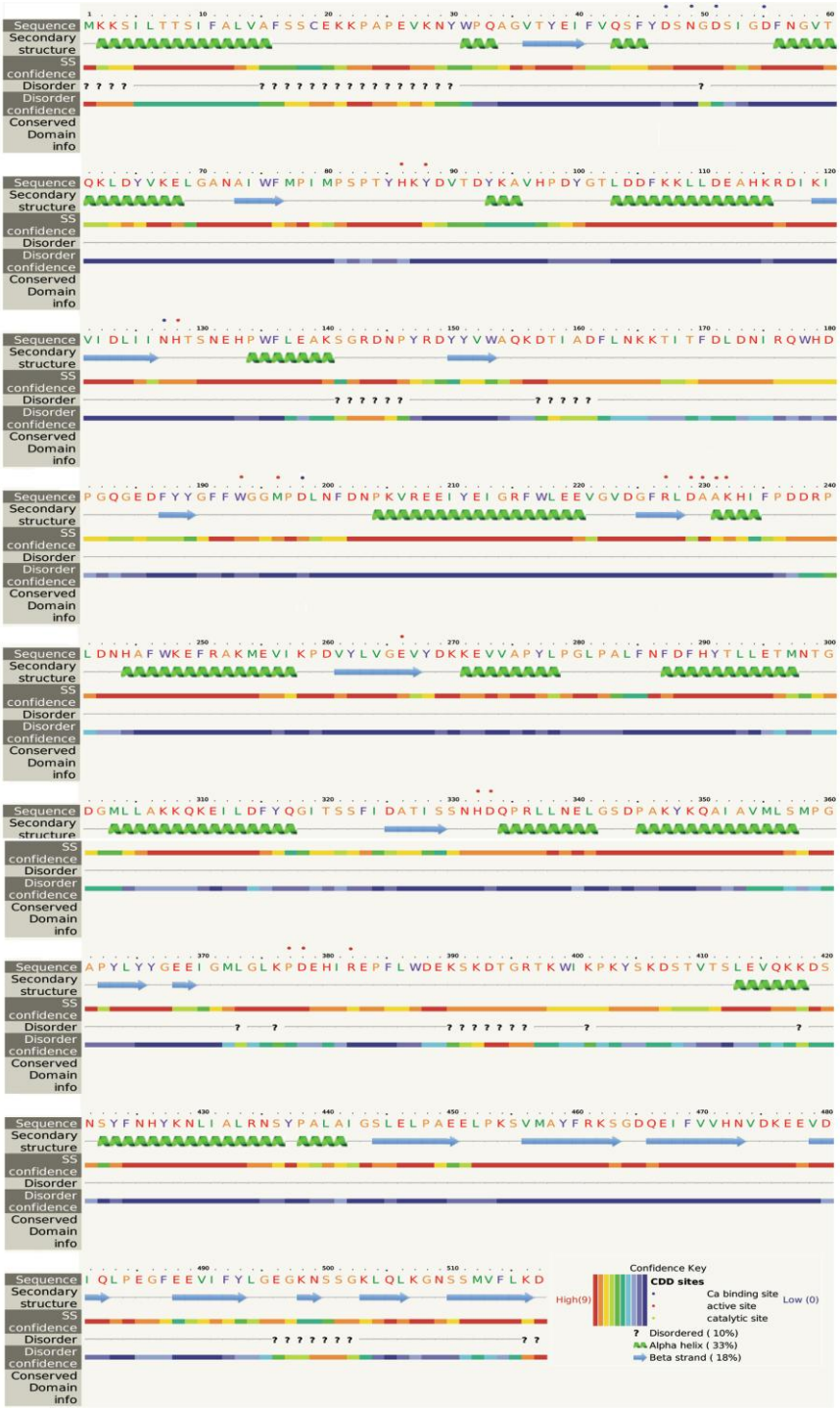

Fig 5.11: Secondary structures of deduced amino acid sequence of BTM109

5.3.3.7 Tertiary structure prediction using Swiss model

Based on amino acid sequence homologies, a protein model was built using Swiss Model (Fig 5.12), using automated computer algorithms, based on the crystal structure of alpha-amylase of *Halothermothrix orenii* (1wza.1.A) as template (Sivakumar *et al.*,2006) Models were built based on the target-template alignment using Promod-II. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions were remodeled using a fragment library. The structure was predicted to be an 8-stranded β/α barrel, typical of GH family-13 proteins.

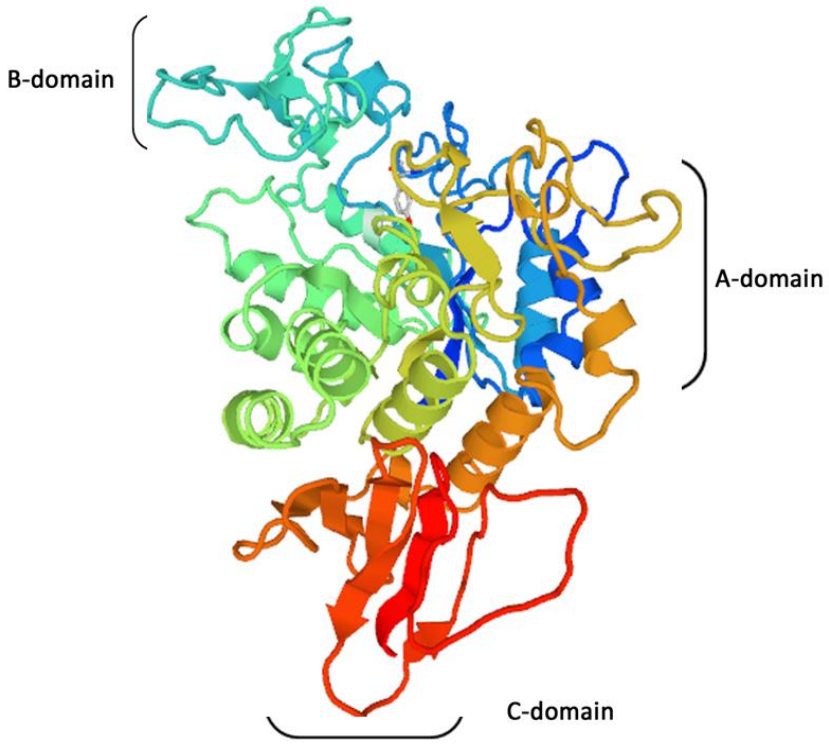

Fig 5.12: Predicted structure of BTM109

The overall 3D structure of BTM109 appeared to be similar to those of other α-amylases, with three domains: domain A, B and C. It was identified that the active site residues and catalytic triad are located in the A-domain. It was also noted that the calcium binding domains are present in the B-domain, and it protruded between the helix and strands of A-domain. Domain-C contained antiparallel β-sheets and forms the C-terminal region of the protein; and the function of domain C is not fully understood.

Studies suggested that the most conserved domain found in all α-amylase family enzymes are A, B and C-domain (Svensson and Sogaard, 1991, Takata *et al.,* 1992). The A-domain consists of a symmetrical fold of eight parallel β-strands arranged in a barrel encircled by eight α-helices. The highly conserved amino acid residues of the α-amylase family that are involved in catalysis and substrate binding are located in loops at the C-termini of β-strands in this domain. The (β/α)8 barrel has first been observed in chicken muscle triose phosphate isomerase (Banner *et al.,* 1975) and is therefore also called the TIM barrel. Apart from these, additional domains have been identified in members of the α-amylase family. In few maltogenic α-amylase and cyclodextrin glycosyltransferase, the C-domain is followed by a D-domain of unknown function (Penninga *et al.,* 1996). Studies have identified that a number of α-amylase family enzymes have additional raw starch binding domain or E-domain that facilitates interaction with the substrate (Knegtel *et al.,* 1995; Penninga *et al.,* 1996). Similarly other additional N-terminal F, H, and G-domains have also been reported in the α-amylase family enzymes that have an endo action or those that hydrolyze α,1-6 glycosidic linkages of branched substrates (Dalmia *et al.,* 1995).

The present study has identified an amylolytic clone from the Arabian Sea sediment marine metagenomic library. The nucleotide sequence of the α-amylase BTM109 was similar to an uncultured bacterium clone. From the database several reports have pointed that the genomic information generated in these genomic

studies is staggering, with many of the open reading frames identified having no homologous identity to known proteins in GenBank (Kerkhof and Goodman, 2009). Therefore, it is necessary to develop an approach to characterize these hypothetical ORFs that comprise nearly 50% of the genomic data and will eventually help to assign functions to genes encoding previously uncharacterized predicted proteins. The power of functional metagenomic approach to explore the unexplored ones can be utilized.

Chapter 6

PURIFICATION AND CHARACTERIZATION OF AMYLASE ENZYME OBTAINED FROM METAGENOMIC CLONE AND ITS APPLICATION STUDIES.

6.1 INTRODUCTION

Soil microorganisms represent a reservoir of innumerable genes with potential for application in health and other industries. This search for biomolecules among cultivable microbiota, which represent a small fraction of the total microbes in nature is due to lack of knowledge of culture requirements of these unculturable microbes. Metagenomic methods can be used to bypass this impediment, wherein libraries of the total community DNA is constructed and screened for novel biomolecules; there by offering an efficient way for rapid screening hitherto unknown enzyme candidates with improved properties, which may find application in various industrial processes. Several enzymes with such as nitrilase, lyase, protease, esterase and amylase have been reported from metagenomic libraries (Robertson *et al*., 2004; Solbak *et al.,* 2005; Jin *et al.,* 2012; Neveu *et al.,* 2011; Vidya *et al.,* 2011).

Amylases find applications in food, fermentation, textile, paper industries, breweries and in detergent industries. They are also useful for biopharmaceutical, medicinal and clinical applications (Becks *et al.,* 1995). In detergent formulations, amylases occupy second place after proteases, for effective removal of starchy stains from cloths (Niyonzima and More, 2014). Similarly the major market share for α-amylases lies in starch liquefaction for the production of starch hydrolysates

such as glucose and fructose (Van Der Maarel *et al.,* 2002). α-amylase (E.C. 3.2.1.1.) catalyze the hydrolysis of α-1,4-glucosidic linkages between adjacent glucose units in starch, and belong to the glycoside hydrolase family of enzymes (Bordbar *et al.,* 2005). Most α-amylases are metalloenymes requiring the presence of calcium ions (Ca^{2+}) for their stability and activity.

In this study, we report the characterization of the α-amylase obtained through metagenomic approach from sediments of Arabian Sea. The physicochemical properties of the amylase and its applications were also investigated and discussed.

6.2 MATERIALS AND METHODS

6.2.1. Purification and characterization of the amylase enzyme obtained from clone BTM109

6.2.1.1 Extraction of crude amylase

The amylase positive clone BTM109 was inoculated in LB broth containing ampicillin (60 µg/mL) and incubated overnight at 37ºC. The cell free supernatant was collected after centrifugation at 6000 rpm for 10 min. Extracted amylase protein from clone BTM109 was designated as P109 and were used as crude amylase enzyme and used for further purification.

6.2.1.2 Purification of amylase

Purification of amylase was done by standard protein purification methods which included acetone precipitation, followed by Sephadex G-75 gel filtration chromatography. All purification steps were carried out at 4°C unless otherwise mentioned.

6.2.1.2.1 Acetone precipitation of amylase

Protein samples commonly contain substances that interfere with downstream applications. Several strategies exist for eliminating these substances from samples. One strategy for removing undesirable substances is to add a compound that causes protein to precipitate. After pelleting the precipitated protein by centrifugation, the supernatant containing the interfering substance is removed and the protein pellet is resuspended in buffer solution. Acetone precipitation (Hamilton *et al.*, 1999) was employed in this study and the procedure is as detailed below.

One volume of ice cold extra pure acetone (Sigma-Aldrich, USA) was added to crude enzyme and vortexed thoroughly followed by incubation at -20°C for 1hour. The precipitated protein was collected by centrifugation at 16,000 rpm for 15 min at 4°C. Decanted and properly disposed the supernatant, the protein pellet was air dried and suspended in minimum volume of 0.1M phosphate buffer (pH 7) (Appendix I) and checked for amylase activity.

6.2.1.3 Gel filtration chromatography by Sephadex G-75

Gel filtration chromatography was performed using the concentrated acetone precipitated fraction.

6.2.1.3.1 Preparation of column

Four grams of sephadex G-75 (Sigma-Aldrich) was suspended in 50 mL MilliQ water and allowed to hydrate for 3 h at 90°C in a water bath, and fine particles were removed by decantation. Fairly thick slurry was degassed under vacuum to remove the air bubbles. Gel suspension was carefully poured into the column (50 X 1cm) without air bubbles and allowed to settle under gravity while maintaining a slow flow rate through the column. Column was equilibrated with three times the bed volume of eluent (0.1M phosphate buffer, pH 7) passed through the column bed in a descending eluent flow.

6.2.1.3.2 Sample preparation and application to Sephadex G-75 column

Ten milliliters of acetone fraction was concentrated using 30 kDa cutoff Amicon Ultra filter (Millipore, USA). This concentrated protein (1 mL) was applied to the column. Care was taken to ensure that the sample was completely free of undissolved substances. After the complete entry of sample to the column, the proteins were eluted using 0.1M phosphate buffer (pH-7) with a flow rate of 0.3 mL/min. Five millilitre fractions were collected. Twelve fractions were collected and each fraction was assayed for amylase activity. Active fractions were pooled and concentrated using Amicon-30 kDa filter and used as the purified enzyme.

6.2.1.4 Analytical Methods

6.2.1.4.1 Amylase activity assay

Amylase activity was determined according to 3, 5-di nitro salicylic acid (DNS) method (Miller, 1959) using maltose as standard. Maltose liberated by hydrolytic activity of amylase reduces 3, 5 dinitro salicylate to orange red 5-nitro 3-amino salicylate, which can be measured at 540 nm. Appropriate blanks for the enzyme and the substrate were also included in the assay along with the test.

The reaction mixture containing 0.5 mL of 1% starch substrate (w/v) in 0.1 M Phosphate buffer (pH 7) and 0.5 mL of appropriately diluted enzyme, the mixtures were incubated at 40°C for 15min. The reaction was stopped by adding 1 mL of DNS reagent (Appendix I) and incubated in a boiling water bath for 5 min followed by cooling to room temperature. Assay mixture was made upto 10 mL with milliQ water and absorbance was measured at 540 nm in UV-Visible spectrophotometer (Shimadzu) against appropriate blanks.

One unit of amylase activity was defined as the amount of enzyme that liberates 1 µmol of reducing sugar as maltose per min under the conditions of the assay and was expressed as U/mL.

6.2.1.4.2 Protein estimation

Protein content was estimated according to the method of Bradford (1976) using bovine serum albumin (BSA) as the standard and the concentration was expressed in mg/mL.

About 10 µL of the protein sample to be estimated was made upto 100 µL volume. To the standards as well as test samples, 1 mL Bradford reagent (Appendix I) was added and incubated for 5 min at room temperature. The absorbance was measured at 595 nm in a UV-Visible spectrophotometer. The concentration of the protein sample was determined from the slope of the standard curve.

6.2.1.5 Calculations

6.2.1.5.1 Specific Activity

Specific activity of the sample was calculated by dividing the enzyme activity (Units) with the protein content (mg) and expressed as U/mg protein.

$$\text{Specific activity} = \frac{\text{Enzyme activity (U/mL)}}{\text{Protein (mg/mL)}}$$

6.2.1.5.2 Fold of purification

Fold of purification of the amylase in each step was calculated by dividing the specific activity of the respective fraction with that of the crude extract.

$$\text{Fold of purification} = \frac{\text{Specific activity of the purified fraction}}{\text{Specific activity of the crude extract}}$$

6.2.1.6 Characterization of amylase

Purified amylase was further characterized for its biophysical and physicochemical properties like molecular weight, isoelectric point, stability at different temperature and pH, effect of detergents, metal ions, oxidizing and

reducing agents, organic solvents, halotolerance and substrate specificity as described in the following sections.

6.2.1.6.1 Electrophoretic methods

The molecular weight was determined by electrophoretic methods followed by zymogram analysis.

6.2.1.6.1.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The active amylase fraction after acetone precipitation and gel filtration chromatography was subjected to electrophoretic analysis by non denaturing SDS-PAGE in a vertical slab electrophoresis (BioRad Mini-PROTEAN Tetra cell, USA). Electrophoresis was carried out in a 10% polyacrylamide gel according to the method described by Laemmli (1970). The reagents used in the gel preparation (Appendix I) and their composition is given in Table 6.1

Table 6.1 Gel preparation for SDS-PAGE

Reagents	Stacking gel (5%) (2mL)	Resolving gel (10%) (5mL)
Stock acrylamide: bis-acrylamide (mL)	0.336	2.5
Stacking gel buffer stock (mL)	0.5	-
Resolving gel buffer stock (mL)	-	1.25
10% SDS (µL)	20	50
Ammonium persulphate (µL)	40	75
TEMED (µL)	4	4
Water (mL)	1.13	1.16

6.2.1.6.1.2 Sample preparation

The samples were mixed with 1X SDS-PAGE sample buffer (Appendix I) and 20 µL sample was loaded in the wells of the gel.

6.2.1.6.1.2 Protein marker for SDS-PAGE

Broad range molecular weight protein marker mix (New England BioLabs, USA) was used for detecting the approximate size of the purified protein. It is a ready-to-load marker. 7 µL of protein marker was taken in a tube and heated for 5 min at 100ºC. After a quick microcentrifuge spin (Tarsons Spinwin, India) the marker was loaded directly onto the gel.

6.2.1.6.1.3 Procedure

The gel plates were cleaned and assembled. 10% resolving gel solution was prepared without APS and TEMED. Mixed gently, APS and TEMED were added and mixed well before immediately pouring into the cast; a layer of water was poured over the gel and allowed to polymerize at least for 30 min. The components of stacking gel were added into a beaker except APS and TEMED, mixed gently, followed by APS and TEMED. The contents were mixed well and poured into the cast above the resolving gel after removing the layered water and immediately inserted the comb between the glass plates. It was allowed to polymerize for at least for 30 min.

The gel was placed in the electrophoresis apparatus, and the reservoir was filled with running buffer (Appendix I) for SDS-PAGE. Protein samples were loaded on to the gel. The gel was run at 4°C at 80 V till the sample entered the resolving gel. When the dye front entered the resolving gel, the current was increased to 120 V. The run was stopped when the dye front reached 1 cm above the lower end of the glass plate. The gel was removed from the cast and one portion of the gel were stained using Coomassie staining solution (Appendix 1) for 1 h, followed by washing in destaining solution (Appendix 1) until the gel became clear, with blue colored protein bands and other half of the gel was used for zymogram analysis.

6.2.1.6.1.4 Zymogram analysis

The approximate size of the active amylase protein band was determined by zymogram analysis wherein, a zone of clearance in substrate incorporated polyacrylamide gels indicated the presence of the active protein at that site.

After SDS-PAGE the gel was washed with deionized water to remove SDS. The gel was soaked in 1% starch solution at 40°C for 1h followed by staining with I-KI solution for 5 min and observed for the zone of clearance.

6.2.1.6.2 Intact mass determination by MALDI-TOF MS

The purified amylase was subjected to intact mass analysis by MALDI-TOF mass spectrometry. The intact mass was determined by MALDI TOF Mass spectrometer (Bruker Daltonics, Germany), at Indian Institute of Science (IISc) Bengaluru, India.

6.2.1.6.3 Determination of isoelectric point

Isoelectric point (pI) of the purified amylase was determined by isoelectric focusing, performed using the isoelectric focusing unit (Bio-Rad PROTEAN IEF cell, USA). Immobilized pH gradient (IPG) strip of pH 3-10 (Bio-Rad) was used for the purpose. Isoelectric focusing (IEF) was followed by two dimensional (2D) electrophoresis. For the purpose IEF is performed using two IPG strips. One strip was Coomassie stained for visualization of isoelectric point and the other strip was subjected to 2D electrophoresis. The detailed procedure is as given below.

6.2.1.6.3.1 Rehydration of sample with IPG strip

The lyophilized sample (1mg/mL) was resuspended in 125 µL rehydration buffer and loaded on to the equilibration tray. Immobilized pH gradient (IPG) strip of pH 3-10 was gently placed gel side down in the equilibration tray and air bubbles if any were removed. The strips were overlaid with 2.5 mL of mineral oil

to prevent evaporation during rehydration process. Covered the equilibration tray and left the tray overnight to rehydrate the IPG strips.

6.2.1.6.3.2 Isoelectric focusing

Paper wicks were placed at both ends of the clean, dry IEF focusing tray covering the wire electrodes. IPG strips were taken out from the rehydration tray and drained the mineral oil by holding the strip vertically for some time. Paper wicks were wet with 8 µL Nanopure water and placed the IPG strips in the focusing tray. Placed in the PROTEAN IEF cell, overlaid with mineral oil and closed the cover.

Programmed the IEF cell as given below and run the electrophoresis (Table 6.2).

Table 6.2 Steps involved in isoelectric focusing

	Voltage	**Time**	**Volt-Hours**	**Ramp**
Step 1	250	20 min	---	Linear
Step 2	4000	2 h	---	Linear
Step 3	4000	---	10,000 V-h	Rapid

The cell temperature was maintained at 20°C with maximum current of 50µA/strip and no dehydration in all steps.

6.2.1.6.3.3 Staining of IPG strips after IEF

The gel was subjected to Coomassie staining for 1 hour followed by destaining to visualize the isoelectric point of P109.

6.2.1.6.4 2-D Electrophoresis

IPG strips were removed from the focusing tray after electrophoresis. Mineral oil was drained by pressing the strip against a wet blotting paper. IPG strip was then transferred to the equilibration tray and 2.5mL equilibration buffer I

(Appendix 1) was added and kept at orbital shaker for 10 min. Decanted the tray to discard the equilibration buffer I completely from the strip. 2.5mL of equilibration buffer II (Appendix 1) was added and kept at shaking condition for 10 min. IPG strip was taken and rinsed in a 1X Tris-glycine buffer and placed on top of precast 10% SDS-PAGE gel (A portion of the gel was left at the top, according to the dimension of IPG strip during the casting itself of SDS-PAGE gel). IPG strip was overlaid with molten agarose and allowed to set. The electrophoresis was carried out at 200 V. The gel was Coomassie stained to visualize the protein spots.

6.2.1.6.5 Effect of physico-chemical parameters on amylase activity

The effect of various physicochemical factors like pH and temperature, metal ions, inhibitors, detergents, oxidizing agent and reducing agent, organic solvents on enzyme activity was studied. Substrate specificity, salt tolerance, pH and temperature stability of amylase were also tested. The relative activity and residual activity were also determined.

6.2.1.6.5.1 Relative activity

Relative activity is the percent enzyme activity of the sample with respect to the activity of the sample for which maximum activity was obtained.

$$\text{Relative activity} = \frac{\text{Activity of sample (U/mL)}}{\text{Activity of sample with maximal enzyme activity (U/mL)}} \times 100$$

6.2.1.6.5.2 Residual activity

Residual activity is the percent enzyme activity of the sample with respect to the activity of the control sample.

$$\text{Residual activity} = \frac{\text{Activity of the sample (U/mL)}}{\text{Activity of the control (U/mL)}} \times 100$$

6.2.1.6.5.3 Determination of pH optima and pH stability of amylase

The optimum pH for activity of the amylase was determined by conducting enzyme assay at various pH levels in the range of 1-13. The enzyme assay was essentially the same as described in section 6.2.1.4.1 except the preparation of the substrate in buffer systems of different pH.

The starch substrate was prepared in different buffer systems with pH ranging from 1-13. The buffer systems used included hydrochloric acid/potassium chloride buffer (pH 1–2), citric acid/ sodium citrate buffer (pH 3–5), phosphate buffer (pH 6 –7), Tris amino methane/hydrochloric acid buffer (pH 8–9), sodium bicarbonate/sodium hydroxide buffer (pH 10), sodium phosphate dibasic/sodium hydroxide buffer (pH 11–12) (Vincent and John, 2009). Preparation of buffers is charted in Appendix I. The enzyme activity and relative activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.1 respectively.

To determine the pH stability of amylase, the enzyme was pre-incubated in different buffers of pH 1-13 for 1h. After incubation the sample was assayed for amylase activity as explained in section 6.2.1.4.1 and the enzyme activity and relative activity was determined as described in sections 6.2.1.4.1 and 6.2.1.6.5.1 respectively.

6.2.1.6.5.4 Determination of optimum temperature and thermal stability of amylase

Optimum temperature for maximum enzyme activity was determined by assaying enzyme activity by varying incubation temperature ranging from 4°C-100°C. As the enzyme is derived from a marine sediment metagenomic clone, temperature ranging from 4°C was considered as the temperatures at the seafloor is usually low. All other assay procedure was essentially the same as described in section 6.2.1.4.1. The enzyme activity and relative activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.1 respectively

Thermal stability of the enzyme was determined by pre-incubating the enzyme for 1h at different temperatures ranging from 4°C-100°C. After pre incubation the sample was assayed for amylase activity as explained in section 6.2.1.4.1 and the enzyme activity and relative activity was determined as described in sections 6.2.1.4.1 and 6.2.1.6.5.1 respectively.

6.2.1.6.5.5 Effect of metal ions on amylase activity

Effect of various metal ions on activity of amylase was assessed by pre-incubating the enzyme in buffer solution containing 5 mM concentrations of various metals ions for 30 min followed by measuring the amylase activity as described under section 6.2.1.4.1. The metal salts studied included barium chloride, ferric chloride, aluminum sulphate, cadmium sulphate, copper sulphate, calcium chloride, cobalt chloride, zinc sulphate, sodium carbonate, lead nitrate, magnesium sulphate and manganese chloride which contributes the metal ions Ba^{2+}, Fe^{3+}, Al^{3+}, Cd^{2+}, Cu^{2+}, Ca^{2+}, Co^{2+}, Zn^{2+}, Na^{+}, Pb^{2+}, Mg^{2+} and Mn^{2+} respectively. The enzyme activity and residual activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.2 respectively.

6.2.1.6.5.6 Effect of inhibitors on amylase activity

The influence of inhibitors on the activity profile of amylase was studied using 5mM concentration of different protein inhibitor compounds such as dithiothreitol (DTT), Ethylenediaminetetraacetic acid (EDTA), N-bromosuccinimide (NBS), Phytic acid, Phenyl methyl sulfonyl fluoride (PMSF), and Urea. Enzyme activity was determined as explained under section 6.2.1.4.1 after pre-incubating the enzyme with inhibitors at room temperature for 30min. The enzyme activity and residual activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.2 respectively.

6.2.1.6.5.7 Effect of various detergents on amylase activity

Effect of various non-ionic and ionic detergents such as Triton X-100, Tween 80, Tween 20, SDS and CTAB (0.5% each w/v) on amylase activity was determined by pre-incubating the enzyme with detergents at room temperature for 30 min followed by enzyme assay as explained under section 6.2.1.4.1 The enzyme activity and residual activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.2 respectively.

6.2.1.6.5.8 Effect of oxidizing and reducing agent on enzyme activity

Effect of hydrogen peroxide (H_2O_2) and sodium hypochlorite (NaClO) as oxidizing agents and β-mercaptoethanol (β-ME) as reducing agent on amylase activity was determined by pre-incubating the enzyme with oxidizing and reducing agent at different concentrations (5 mM, 10 mM, 15 mM, 20 mM and 25 mM) at room temperature for 30 min followed by enzyme assay as explained under section 6.2.1.4.The enzyme activity and residual activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.2 respectively

6.2.1.6.5.9 Effect of various solvents on amylase activity

Effect of various polar and non-polar solvents like Hexane, Benzene, Toluene, Chloroform, Dimethyl Formamide, Dimethyl Sulfoxide, Acetonitrile, Isopropanol, Ethanol, and Methanol (10%, 25% and 50% each v/v) on amylase activity was determined by pre-incubating the enzyme with solvents at room temperature for 30min under shaking condition (150 rpm) followed by enzyme assay as explained under section 6.2.1.4.1. The enzyme activity and residual activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.2 respectively.

6.2.1.6.5.10 Halotolerance of amylase

As P109 is derived from a marine metagenomic clone library, the salt tolerance of the amylase was studied at a range of 0.5–2.5 M NaCl. The enzyme

reactions were performed in buffer with different salt gradient of 0.5, 1, 1.5, 2, 2.5 M concentration. The enzyme activity and residual activity were calculated as described in sections 6.2.1.4.1 and 6.2.1.6.5.2 respectively.

6.2.1.6.5.11 Determination of substrate specificity

The substrate specificity of the purified amylase was determined based on their ability to hydrolyze various starch substrates like rice starch, potato starch, wheat starch, corn starch, amylopectin from maize and amylose from potato by conducting enzyme assay as explained under section 6.2.1.4.1 with 1% of the respective substrates prepared in phosphate buffer (pH 7). Relative activity was calculated as described in sections 6.2.1.6.5.2.

6.2.1.6.5.12 Determination of kinetic parameters - Km and Vmax

The kinetic studies using the purified enzyme was conducted for determining the Km and Vmax. Km is the substrate concentration at which the reaction velocity is half maximum and Vmax is the maximum velocity of the enzyme reaction. The enzyme assay was conducted as explained under section 6.2.1.4.1 by varying substrate [S] concentrations (1 mg – 20 mg/mL) and enzyme activity [V] was calculated. Kinetic parameters, such as Km (mg/mL) and Vmax (U/mL) were obtained using Line-weaver Burk plot.

6.2.1.6.5.13 Determination of starch grain degradation by Scanning Electron Microscopy (SEM)

Structure of native and digested starch was analyzed by SEM by incubating 1% starch grains (Rice starch and potato starch) with amylase for 2 and 4 h at 40°C. Following incubation, grains were pelleted and rinsed with pure ethanol and air dried for visualization. The hydrolyzed starches were photographed using SEM (JSM-6390) at Sophisticated Test Instrumentation Centre, CUSAT.

6.2.1.6.5.14 End product analysis by thin layer chromatography (TLC)

End products liberated from the hydrolysis of starch were analyzed by conducting enzyme assay for 2 and 4 h at 40°C followed by spotting the samples onto a TLC silica gel 60 F plate (Merck, Germany). The plates were developed in a saturated chromatographic chamber with 2-propanol: n-butanol: water (12:3:4) as the solvent system, and the spots of sugars were visualized by spraying aniline-diphenylamine reagent (Appendix 1) followed by incubation in a hot-air oven at 110 °C. End products liberated by starch hydrolysis were identified using a mixture of glucose (G1) and malto-oligosaccharides standards such as maltose (G2), maltotriose (G3), maltotetrose (G4) and maltopentose (G5)

6.2.2 Application studies of amylase

6.2.2.1 Application of amylase as a detergent additive

6.2.2.1.1 Commercial detergent compatibility of the enzymes

The stability of the enzymes in the presence of commercial detergents was determined using 7 detergents which includes Ariel®, Surf Excel®, Sunlight®, Tide® and Wheel®, Ujala® at 1% (w/v). The enzymes already present in the detergent were first heat inactivated by boiling for 15 minutes. The detergent solutions were incubated with purified amylase enzyme for 1 h at 40ºC and enzyme assay was carried out as explained under section 6.2.1.4.1 and residual activity were calculated. The enzyme activity of a control sample (without detergent) was taken as 100 percent.

6.2.2.1.2 Wash performance studies

The wash performance analysis of the purified amylase was studied on white cotton cloth pieces (5 cm x 5 cm) stained with chocolate (Hmidet *et al.*, 2009). The cloth pieces were stained with 200 µL of liquefied chocolate and dried using hot air oven. The stained cloth pieces were taken in separate flasks and subjected to the following wash treatments as shown below.

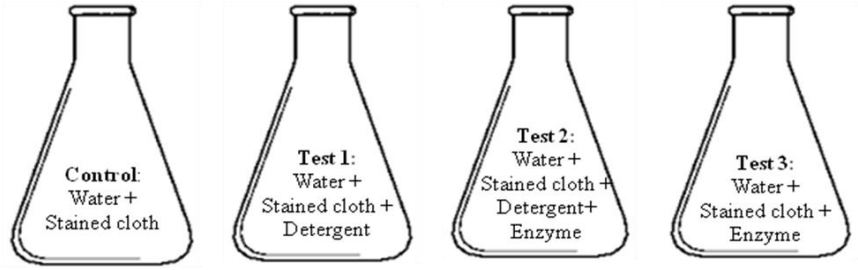

Figure 6.1: Wash treatment protocols

Flask 1: 100 mL distilled water + stained cloth piece

Flask 2: 100 mL detergent solution + stained cloth piece

Flask 3: 100 mL detergent solution + stained cloth piece + 1 mL purified enzyme

Flask 4: 100 mL distilled water + stained cloth piece + 1 mL purified enzyme

After 30 min incubation at room temperature with shaking at 200 rpm, the cloth pieces were taken out, rinsed with tap water, dried and visually examined to check the effectiveness of stain removal, and photographed.

6.2.2.2 Application of amylase as a desizing agent

Desizing efficiency of the purified amylase was studied on (5 cm x 5 cm) white cotton cloth pieces (Hmidet *et al.,* 2009). Equal sizes of pieces were used for the study and the cloth strip was then starched using 5% starch solution and dried. The starched cloth strip was then dipped into a flask containing (a) 25 mL of tap water (control) and (b) 24 mL of tap water and 1 mL of purified amylase. The flasks were then incubated for 1 h at room temperature with shaking of 200 rpm. After the time interval, the cloth strips were oven dried and stained with I-KI solution to visualize effectiveness of starch removal and photographed.

6.2.2.3 Application of amylase as a feed supplement

To check the suitability of amylase as a feed supplement, *in vitro* cytotoxic effect of P109 on cultured L929 murine fibroblast cells were tested.

6.2.2.3.1 Cell culture maintenance and treatment with P109

L929 fibroblast cell lines were maintained in Dulbecco's modified Eagles medium (Himedia, India) supplemented with 10% fetal bovine serum (Invitrogen, USA). Cells were grown to confluency at 37°C in 5 % CO_2 in a humidified atmosphere in a CO_2 incubator (Eppendorf, Germany). The cells were trypsinized (500µL of 0.025% Trypsin in phosphate buffer saline (PBS)/ 0.5mM EDTA solution) for 2 minutes and passaged to T flasks in complete aseptic conditions. P109 were added to grown cells at a final concentration of 6.25, 12.5, 25, 50 and 100µg/mL from a stock of 1mg/mL and incubated for 24 hours. The percentage difference in viability was determined by standard MTT assay (Arung *et al.,* 2010) after 24 hours of incubation. The morphological characteristics of cells were imaged using inverted phase contrast microscope (Olympus CKX41, Japan)

6.2.2.3.2 MTT Assay

MTT assay measures the reduction of yellow 3-(4, 5dimethythiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) by mitochondrial succinate dehydrogenase. The MTT enters the cells and passes into the mitochondria where it is reduced to an insoluble, coloured (dark purple) formazan product. The cells are then solubilised with Dimethyl sulfoxide and the released, solubilised formazan product was measured at 540nm. Since reduction of MTT can only occur in metabolically active cells the level of activity is a measure of the viability of the cells. The cells were washed with 1X PBS followed by the addition of 30 µL of MTT solution to the culture (MTT -5mg/mL dissolved in PBS). It was then incubated at 37°C for 3 hrs. MTT was removed by washing with 1X PBS and 200µL of DMSO was added to the culture. Incubation was at room temperature for 30 min until the cell lysed and colour was obtained. The solution was then centrifuged at top speed for 2 min to precipitate cell debris. Optical density was read at 540 nm using DMSO as blank. The percent viability was calculated by the equation:

% viability = (OD of Test/ OD of Control) X 100

6.2.3 Statistical analysis

All experiments were repeated thrice wherever required; and statistical analysis was done using Microsoft Excel 2007 by calculating mean and standard error.

6.3 RESULTS AND DISCUSSION

6.3.1 Purification and characterization of P109

6.3.1.1 Fold of purification of P0109

Amylase protein P109 produced by clone BTM109 was purified from the crude extract employing gel filtration chromatography using sephadex G-75 column. There was much increase in enzyme activity after purification and the specific activity increased gradually after each purification process. The details of purification steps including total activity, specific activity and fold of purification of P109 are as depicted in Table 6.3.

Table 6.3: Fold of purification of P109

Purification step	Total Activity (U)	Total Protein (mg)	Specific activity (U/mg)	Fold of purification
Crude extract	1175	3200	0.36	1
Acetone precipitation	744	254	2.92	8.1
Gel filtration	67	6.3	10.36	28.7

From the table it may be noted that acetone precipitation of P109 resulted in an approximately eight fold increase in specific activity of the enzyme. This fraction upon further purification by gel filtration chromatography yielded a 28.7 fold increase in specific activity.

Various methods are followed for the purification of extracellular enzymes. The first step in any purification process is the extraction of proteins as well as the removal of cellular components which is usually aided by

centrifugation or ultrafiltration (Bell *et al.*, 1983). In another study, an α-amylase obtained from a soil metagenomic library of Western Ghats of Kerala was purified by acetone precipitation followed by ion exchange chromatography using Q-Sepharose anion exchange column. Upon purification the specific activity increased to 18.2 U/mg from the initial 0.204 U/mg activity of crude enzyme (Vidya *et al.*, 2011). An α-amylase from Northwestern Himalayas soil metagenomic library purified by ammonium sulphate fractionation, yielded an specific activity of 2.4U/mg (Sharma *et al.*, 2010). Studies on recombinant α-amylase from *Halothermothrix orenii* purification by affinity chromatography reported 4.87 fold purification (Mijts and Patel, 2002).

6.3.1.2 SDS-PAGE of P109 and zymogram analysis

SDS-PAGE was performed to check purity of the enzyme P109 after each stage of purification; and was also used to determine the molecular mass of the purified protein followed by zymogram analysis to confirm amylase activity of P109. Protein profile obtained by SDS-PAGE and the zymogram are shown in Fig. 6.2.

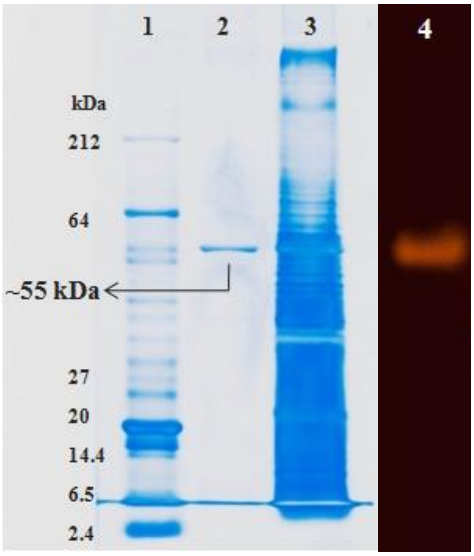

Fig: 6.2: SDS-PAGE and zymogram of P109

Lane 1- Protein marker (NEB); Lane 2- Purified P109; Lane 3- Acetone precipitated fraction; Lane 4-Zone of clearance obtained by zymogram analysis

SDS- PAGE followed by Coomassie staining identified multiple protein bands in the crude acetone fraction of P109, while a single band was visible in the gel filtered purified fraction. The apparent molecular weight of the this amylase P109 was observed to be about~55 kDa. After purification by gel filtration, single band of pure protein was obtained, which inferred complete purification of P109. Single protein band obtained after gel filtration chromatography was responsible for the amylase activity based clearing on the zymogram (Fig. 6.1), in the starch incorporated gel after staining with I-KI solution. Clearance zone is also in line with the purified protein fraction which also confirmed the size of the protein as ~55 kDa protein.

6.3.1.3 Intact mass determination by MALDI-TOF Mass spectrometry

The intact mass of amylase was determined by MALDI-TOF MS. The mass of P109 was determined as 55.7 kDa from the mass spectrum as depicted in Fig. 6.3

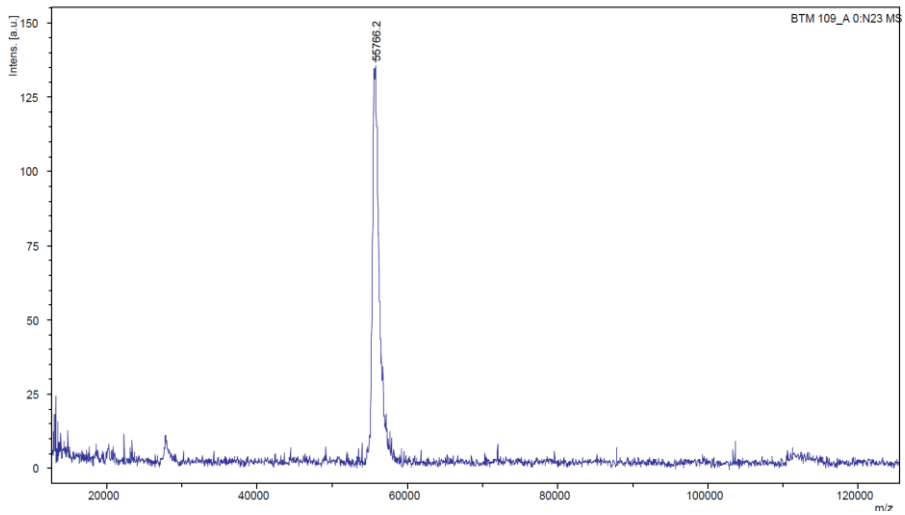

Fig: 6.3 Mass spectrum of P109 obtained by MALDI-TOF MS

From cloned α-amylase genes and deduced amino acid sequences, molecular mass of microbial α-amylases were reported to usually range from 50-

60 kDa (Gupta *et al.,* 2003). However, an α-amylase from *Bacillus caldolyticus* with a molecular weight of 10 kDa has been reported (Schwab *et al.,* 2009) , as also a large size α-amylase of 210 kDa from *Chloroflexus aurantiacus* (Ratanakhanokchai *et al.,* 1992).

Metagenomic studies identified a variety of α-amylases from various environments. A fecal microbial community library derived α-amylase had a molecular mass of 55.4 kDa (Xu *et al.,* 2014b), while a 56.07 kDa amylase was identified from a cold and alkaline environment library (Vester *et al.,* 2015). These reports and that from this study imply a 50-60kDa molecular mass for α-amylases.

Characterization of α-amylase enzyme from different bacterial groups surviving in extreme environments, wherein α-amylase from thermophilic *Anoxybacillus flavithermus* with molecular weight of 60 kDa (Fincan *et al.,* 2014); extracellular 55 kDa α-amylase from an extremely alkalophilic *Bacillus licheniformis* (Roy and Mukherjee, 2013) have also been conducted; besides the recombinant α-amylase from extreme thermophile *Geobacillus thermoleovorans* with a molecular mass of 59 kDa (Mehta and Satyanarayana, 2013).

6.3.1.4 Isoelectric Focusing (IEF) and 2D-electrophoresis

Isoelectric focusing was carried out in pre-casted IPG strips. After IEF, the strip was Coomassie stained and visualized as blue coloured bands on the strip. The 7cm IPG strip are equally divided into seven points of pH. From the position of the bands, the isoelectric point (pI) was calculated. The pI of P109 was determined as 6.9 and shown in Fig. 6.4 (a) and the arrow mark indicated the pI.

IEF followed by 2D further confirmed the purity of P109 and it is identified that the enzyme did not show any isoforms. Gel showing a single protein spot indicated by an arrow is shown in Fig. 6.4 (b)

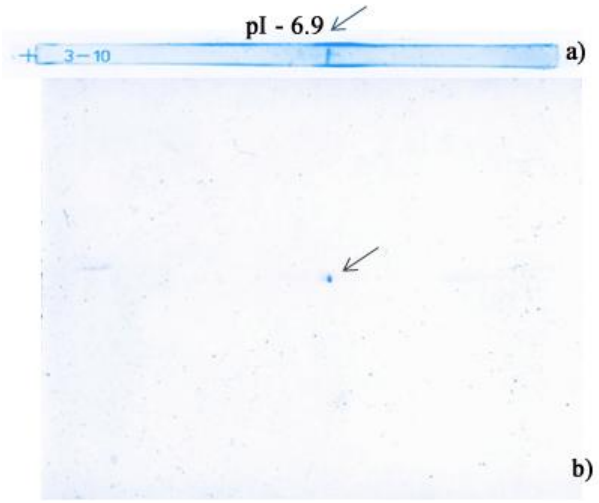

Fig 6.4: Isoelectric focusing and 2D electrophoresis of P109

Isoelectric point of a soil metagenome derived α-amylase was reported to be 8.5 (Vidya *et al.,* 2011), while that for a metagenomic amylase derived from acid mine drainage was pI of 6.14 (Delavat *et al.,* 2012). The pI of enzyme from *Halomonas meridian* was 4.65 (Coronado *et al.,* 2000) and that from *Aeromonas hydrophila* was 6.43 (Chang *et al.,* 1993). Similarly α-amylase isoforms have been reported in *Bacillus* sp. (Ohdan *et al.,* 1999; Božić *et al.,* 2014). It is clear therefore that the isoelectric points of amylases vary considerably, which in turn is based on the amino acid composition of the protein.

6.3.2 Effect of physico-chemical parameters on amylase activity

6.3.2.1 Determination of pH optima and pH stability of P109

In order to study the effect of pH on activity and stability of P109, buffers having different pH in the range of 1-13 were used. Relative activity was determined by comparing the activity of the enzymes at different pH with the sample having maximum activity. The effect of pH on the activity and stability of P109 is depicted in Fig 6.5 and Fig 6.6.

It was observed that the P109 enzyme activity increased as the pH increased from the acidic range, peaking at pH 7 and maintaining more than 80% relative activity at pH 8 and 9 and declining thereafter; followed by complete loss of activity at higher alkaline pH. It was also noted that the pH maxima for enzyme activity was P109 was pH 7; so it was neither an acidic nor an alkaline enzyme, but could tolerate different pH to a certain extent.

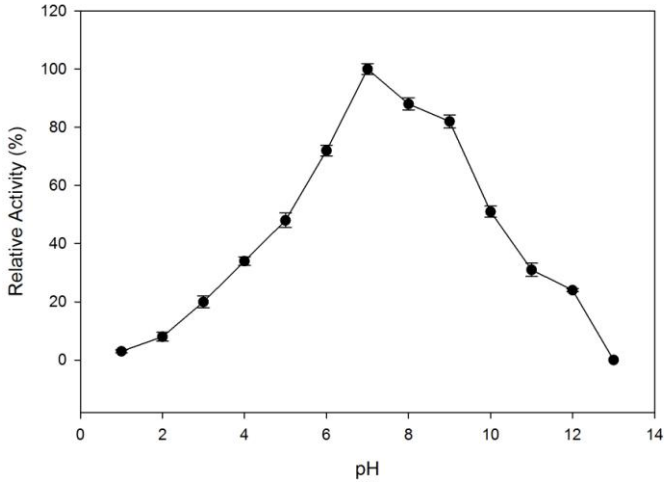

Fig 6.5: Effect of pH on enzyme activity of P109

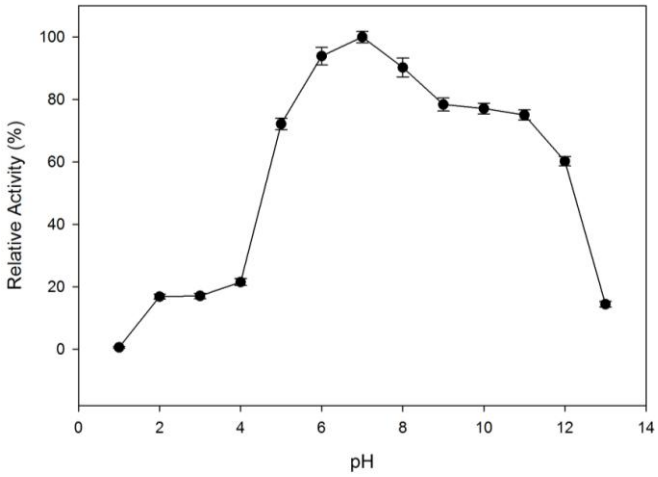

Fig 6.6: Stability of P109 at different pH

Similar patterns were observed when the pH stability of P109 was analysed, with maximal stability at pH 7; while maintaining more that 75% activity between pH 6 -11. At the same time, P109 lost approximately 80% of its activity at pH 1, 2, 3, 4 and 13. So from the present analysis, it identified that P109 was highly active at neutral pH and maintained more than 75% relative activity at near alkaline pH, thereby suggesting its potential as a detergent additive for stain removal.

The pH optima of most microbial amylases are in the acidic to neutral range (Pandey *et al.,* 2000). However, amylases with extreme acidic and alkaline pH optima have also reported from acidophilic and alkalophilic microbes (Schwermann *et al.,* 1994; Lee *et al.,* 1994). Amylases in general are stable over a wide range of pH from 4 to 11 (Roy *et al.,* 2012; Maalej *et al.,* 2013; Roohi *et al.,* 2013) however, α-amylases with stability in a narrow range have also been reported (Ghorbel *et al.,* 2009; Shafiei *et al.,* 2010).

The pH optimum of a metagenome derived amylase was between pH 8-9, while enzyme was stable between pH 6-10 (Vester *et al.,* 2015). An amylase AmyM from soil metagenomic library showed highest activity at pH 9.0 with soluble starch substrate (Yun *et al.,* 2004). The influence of pH on the enzyme derived from Northwestern Himalayas soil metagenomic library showed activity from pH 5.5 to 7.5, with maximum activity at 6.5 and enzyme stability between pH 5.5 and 7.0 (Sharma *et al.,* 2010).

6.3.2.2 Determination of optimum temperature and thermal stability of P109

The effect of temperature from 4°C-100°C on activity and stability of P109 was studied. Fig 6.7 depicts the effect of temperature on the activity of P109 and Fig 6.8 depicts the thermal stability profile of P109.

The optimum temperature for P109 enzyme activity was 60°C, with the enzyme maintaining 35-55% relative activity at temperature range from 4-30°C and more than 85% activity at 40 and 50°C. Enzyme activity declined after 60°C, reaching a relative activity of 34% at 100°C.

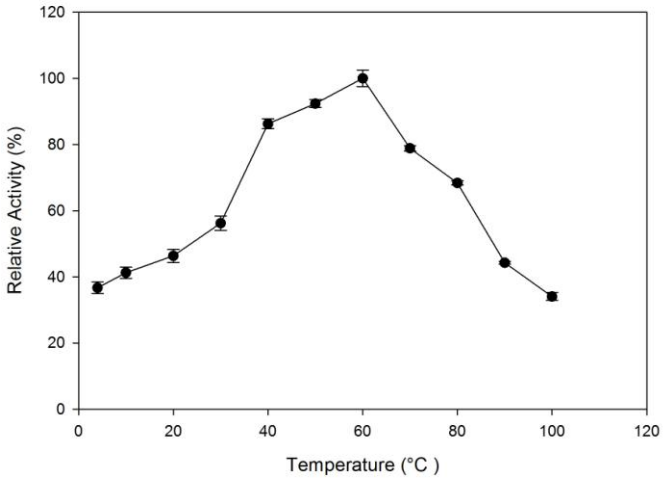

Fig 6.7: Effect of temperature on enzyme activity of P109

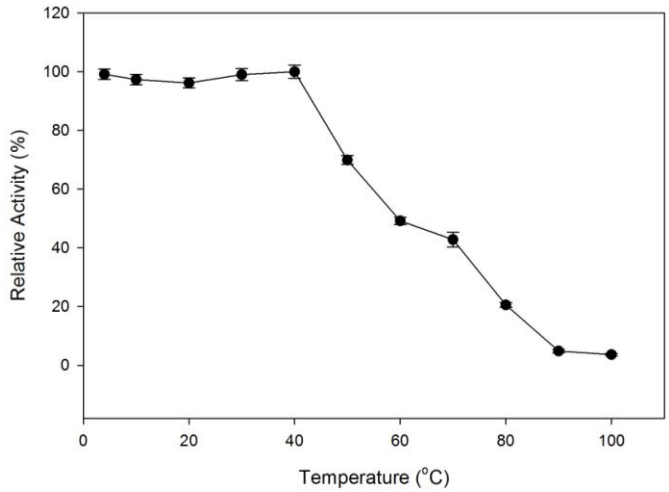

Fig 6.8: Effect of temperature on stability of P109

Based on the temperature stability studies, it was noted that the enzyme was highly stable from 4-40°C, and even incubating P109 up to 40°C for 1h did

not cause loss of enzyme activity. Thermal stability however declined from 40°C with a complete loss of activity at 100°C. Nevertheless P109 was observed to be stable up to 80ºC, although relative activity was reduced to 20%. From Fig 6.7 the optimum reaction temperature was 60°C, but based on the thermostability studies it was clear that the enzyme could retain only 50% activity at 60°C after 1h. For this reason the amylase assay was conducted at 40°C for maximal activity.

Temperature optimum for activity of amylase vary considerably depending on the isolation source, while thermostability is affected by factors like calcium, substrate and other stabilizers (Vihinen and Mantsala, 1989). The optimal temperature for an amylase from soil metagenome was 42°C with starch substrate (Yun *et al.,* 2004). The temperature profile of a metagenome derived amylase Amy_{I3C6} showed an optimum at 10–15 °C with more than 70 % of the activity retained at 1°C, while temperature stability studies indicated that it to be a heat-labile enzyme unable to withstand more than 45°C (Vester *et al.,* 2015). The optimal reaction temperature for pAMY a soil metagenome derived amylase was 35°C; pAMY was highly stable at 10°C to 50°C, retaining almost 90% activity in 10°C to 30°C range (Sharma *et al.*, 2010).

6.3.2.3 Effect of metal ions on the activity of P109

The effect of twelve different metal ions on enzyme activity was studied, residual activity was calculated and is as depicted in Fig 6.9. Corresponding percentage inhibition on activity of P109 is tabulated in table 6.4 and the most prominent effect is highlighted.

Metals ions are known to stabilize or inhibit enzyme activity. In this study interaction with Ca^{2+} ions enhanced activity of P109, while that with other ions did not have any stimulatory effect. The activity enhancement in the presence of calcium ions was 110% of its initial activity. This activity enhancement in the presence of Ca^{2+} supports our finding that P109 is an α-amylase. On the other

hand, presence of Cu^{2+} and Zn^{2+} decreased the residual activity of P109 to 69 and 67% respectively. While P109 maintained around 75% residual activity in the presence of all the other metal ions tested.

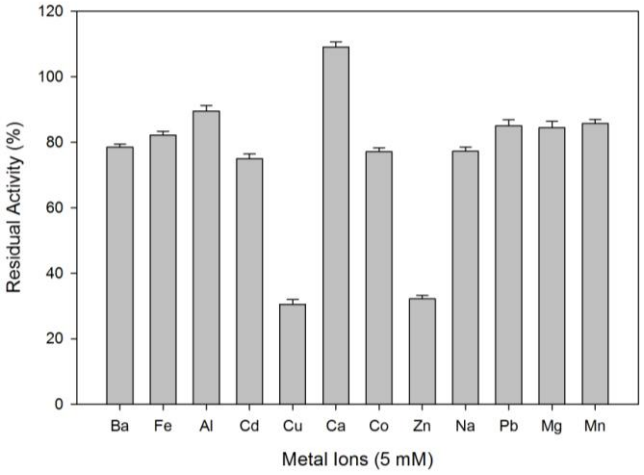

Fig. 6.9: Effect of metal ions on enzyme activity of P109

Table 6.4: Inhibitory action of metal ions

Metal ions	Percentage Inhibition (%)
Ba^{2+}	21.56%
Fe^{3+}	17.87%
Al^{3+}	10.48%
Cd^{2+}	25.03%
Cu^{2+}	69.51%
Ca^{2+}	**0%**
Co^{2+}	22.92%
Zn^{2+}	67.86%
Na^{+}	22.77%
Pb^{2+}	15%
Mg^{2+}	14.51%
Mn^{2+}	14.25%

Most α-amylases are metalloenzymes that requires metal ions for stability and catalytic activity. The amylolytic activity of most bacterial amylases is enhanced by Ca^{2+}. The affinity of Ca^{2+} to α-amylases is much stronger than that of

other ions. The amount of bound calcium varies from one to ten (Heinen and Lauwers, 1975; Arikan, 2007; Kiran and Chandra, 2008). Enhancement of amylase activity in the presence of calcium ions is due to the formation of a calcium-sodium-calcium metal triad in the main Ca^{2+} binding site, bridging domains A and B of the enzyme (Machius *et al.,* 1998).

In this study amylase activity was hiked 110%, while that from soil metagenome increased to 197% in the presence of 1mM Ca^{2+} (Yun *et al.,* 2004). In another α-amylase from a metagenomic library, the initial activity was enhanced to 119% in the presence of 10mM Ca^{2+}(Vidya *et al.,* 2011).

Amylase activity of P109 was inhibited more than 60% by Cu^{2+} and Zn^{2+}, whereas with another metagenomic amylase, there was complete inhibition (Vester *et al.,* 2015). The inhibitory effect of Cu^{2+} and Zn^{2+} was also observed in the enzyme from a thermophilic *Geobacillus* and *Bacillus* sp. (Lin *et al,* 1998; Fincan and Enez 2014). The analysis of deduced amino acid sequence of P109 indicated that calcium binding sites contained 6 residues, that are conserved (section 5.3.3.6 in chapter 5). The inhibitory effect of Cu^{2+} and Zn^{2+}can be attributed either to binding of these ions to the catalytic residues or because they replace the required Ca^{2+} ions from the enzyme (Nirmala and Muralikrishna 2003).

6.3.2.4 Effect of inhibitors on activity of P109

The effect of inhibitors on the activity profile of amylase P109 was studied using different protein inhibitor such as Dithiothreitol (DTT), Ethylenediaminetetraacetic acid (EDTA), N-bromosuccinimide (NBS), Urea, Phenyl methyl sulfonyl fluoride (PMSF) and Phytic acid; these will either modify the amino acids residues in the enzyme molecule or acts as metal ion chelators. The residual activity calculated by comparing the enzyme activity with that of

control is depicted in Fig 6.10. Percentage inhibition exhibited by the inhibitors on P109 activity is tabulated in table 6.5 and the most prominent effect is highlighted.

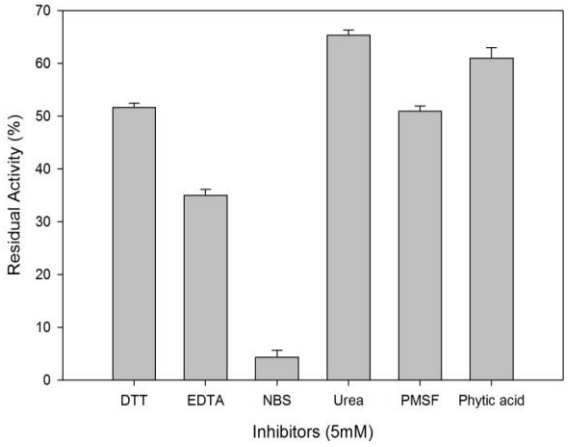

Fig. 6.10: Effect of inhibitors on enzyme activity of P109

Table 6.5: Percentage inhibition by various inhibitors

Inhibitors	Percentage Inhibition (%)
DTT	49%
EDTA	65%
NBS	**96%**
Urea	35%
PMSF	50%
Phytic acid	40%

Among the six different inhibitors tested, NBS treatment caused 96% inhibition of amylase activity suggesting the possible role of tryptophan in maintaining the structure and activity of P109. It was identified by conserved domain search of deduced amino acid sequence that tryptophan was present as an active site residue in P109. There was loss of activity due to the action of EDTA suggesting that P109 is a metal-dependent amylase, which was also confirmed by activity enhancement in the presence of Ca^{2+} ions. Similarly PMSF caused 50% inhibition of amylase activity, implying the role for serine in enzyme catalysis.

The loss of activity due to DTT, phytic acid and urea was 49%, 40% and 35% respectively.

The aromatic amino acid tryptophan plays a major role not only in active site of amylases, but also in maintaining the secondary structure thereby coordinating amylase activity (Rao and Satyanarayana, 2008). N-bromosuccinimide acts by oxidizing tryptophan residues, resulting in loss of enzyme activity and amylases are inhibited by NBS at minimum concentration (Kawaminami *et al.*, 1994; Igarashi *et al.*, 1998). In P109, 96% inhibition by NBS is indicative of tryptophan in the active site. NBS at 5mM concentration can also acts as a stabilizer by increasing the activity amylase (Vidya *et al.*, 2011). 1mM NBS completely inhibited activity of a recombinant α-amylase from an extreme thermophile *Geobacillus thermoleovorans* ,although the enzyme retained 97 and 93% activity in the presence of 5mM EDTA and phytic acid (Mehta and Satyanarayana, 2013). EDTA and phytic acid are metal ion chelators and act by binding to divalent ions. Reduction in enzyme activity in the presence of these compounds further confirmed that P109 is a metalloenzyme. In a previous study, 1mM EDTA treatment resulted in complete loss of activity of metagenomic α-amylase (Vester *et al.*, 2015), but another from an alkalophilic *Bacillus licheniformis* exhibited 52% inhibition in the presence of 2mM EDTA (Roy and Mukherjee, 2013). Moreover, α-amylases insensitive to phytic acid have also been previously reported in *Geobacillus* sp. (Ezeji and Bahl, 2006; Mollania *et al.*, 2010).

P109 enzyme was 50% inhibited in the presence of 5mM PMSF. PMSF inhibits enzyme activity by sulfonating serine residues. Hence, enzymes with serine playing a catalytic role may get inhibited by varying concentrations of PMSF; indicating the participation of seryl hydroxyl groups in enzyme catalysis (Das *et al.*, 2004). An α-amylase from *Aeromonas veronii* exhibited 98% activity in 3mM PMSF (Samie *et al.*, 2012). A recombinant α -amylase from a *Bacillus*

strain expressed in *E. coli* was 70% inhibited in 5mM PMSF (Asoodeha *et al.*, 2014), whereas another recombinant retained 97% activity even in 10mM PMSF (Lin *et al.*, 2002).

DTT acts on thiol groups present in the proteins and can possibly alkylate lysine residues. Inhibitory action of urea is ascribed to its denaturing character that unfolds the amylases by acting on the hydrophobic amino acids in the enzyme polypeptide chain (Chakraborty *et al.*, 2011). Due to the action of DTT, P109 was inhibited by 49%. The amino acid composition of tP109 protein shows that lysine was the dominant amino acid (8.90%), hence the inhibition of P109 may be attributed to alkylation of these predominant lysine resides. Amylase activity was reported to be stabilized by 5mM DTT (Vidya *et al.*, 2011), however 30% inhibition was also observed with 10mM DTT (Mehta and Satyanarayana, 2013), while 5mM DTT and urea caused 39 and 14% inhibition in a recombinant α amylase (Asoodeha *et al.*, 2014). An α-amylase from *Bacillus subtilis* was inhibited 21% in 2M urea (Roy *et al.*, 2012), while that from *Aeromonas veronii* exhibited 80% inhibition (Samie *et al.*,2012).

6.3.2.5 Effect of various detergents on activity of P109

Effect of non-ionic detergents like Tween 20, Tween 80 and Triton X 100 and ionic detergents like SDS and CTAB were tested on activity of P109; residual activity was calculated and is as depicted in Fig 6.11. Percentage inhibition exhibited by the detergents on activity of P109 is tabulated in table 6.6 and the most prominent effect is highlighted.

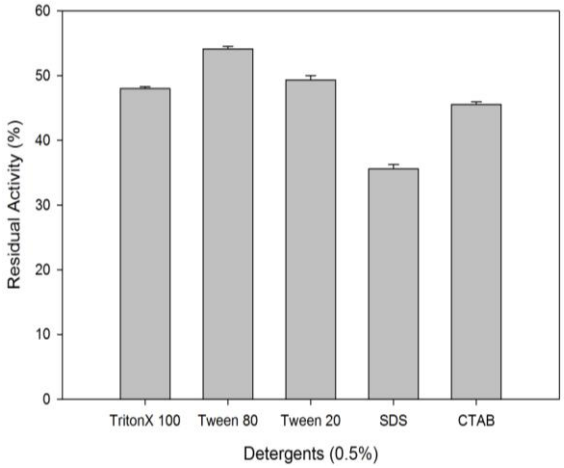

Fig. 6.11: Effect of detergents on enzyme activity of P109

Table 6.6: Inhibitory action of detergents

Detergents	Percentage Inhibition (%)
TritonX 100	52%
Tween 80	45%
Tween 20	50%
SDS	**64%**
CTAB	54%

Five detergents were tested and all except Tween 80 caused more than 50% inhibition of P109 activity. It was noted that ionic detergents were more inhibitory than non-ionic detergents on activity of P109. SDS reduced enzyme activity to 36% residual activity, whereas CTAB and TritonX 100 reduced it to 46% and 48% respectively. The partial inhibition of amylase activity by detergents may be due to conformational changes caused in the enzyme, thereby preventing enzyme-substrate binding and leading to amylolytic activity loss (Niyonzima and More, 2014).

Non-ionic surfactants Tween 20 and Triton X-100 upto 10 % concentration had a moderate effect on the activity of α-amylase from a

metegneomic library, whereas incubation with 0.1% anionic surfactant SDS resulted in almost complete loss (Vester *et al.*, 2015). Another study on metagenomic α-amylase reported 30% inhibition with SDS, whereas activity increased at 0.5% Tween20, Tween 80 and Triton X 100 (Sharma *et al.*, 2010). Likewise, an alkaline amylase of *B. cereus* treated with 1% SDS was 50 % inhibited (Maalej *et al.*, 2013). However, only partial inhibition was observed for amylase of *B. licheniformis* in 1 % Tween 80 and Triton X-100 (Roohi *et al.*, 2013), while 90 % of the initial amylolytic activity was maintained for an amylase from *Bacillus* sp. (Kiran and Chandra, 2008).

6.3.2.6 Effect of oxidizing and reducing agent on activity of P109

Effect of various concentrations (5-25mM) of hydrogen peroxide (H_2O_2) and sodium hypochlorite (NaClO) as oxidizing agents and β-mercaptoethanol (β-ME) as reducing agent on activity of P109 was analysed and depicted in Fig 6.12, and percentage inhibition is tabulated in table 6.7.

It was noted that varying concentration of oxidizing and reducing agents did not have a profound inhibitory effect on activity of P109. More than 50% residual activity was retained at all the concentrations of β-mercaptoethanol, NaClO and H_2O_2. Maximal inhibitory effect was at 25mM NaClO, with inhibition of 43% enzyme activity.

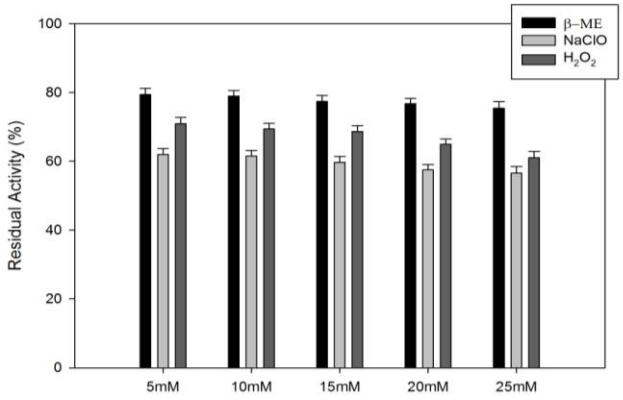

Fig. 6.12: Effect of oxidizing and reducing agent on activity of P109

Table 6.7: Inhibitory action of oxidizing and reducing agents

Conc.	β-ME	NaClO	H_2O_2
5mM	20%	38%	29%
10mM	21%	38%	30%
15mM	22%	40%	31%
20mM	23%	42%	35%
25mM	**23%**	**43%**	**38%**

P109 maintained 80% and 77% activity in the presence of 5 and 25mM β-ME. β-mercaptoethanol acts by disrupting the structure of proteins by breaking disulfide bonds within it. Disulfide bonds are commonly found in extracellular proteins and it is widely accepted that they contribute to the stabilization of the native conformation of proteins (Pons *et al.,* 1995). Atleast two cysteine residues are required in the formation of a disulfide bridge, in the amino acid composition of P109 deduced from its nucleic acid sequence, only one cysteine residue was identified, suggesting that the P109 protein may not have disulfide bridges. Hence this could be why enzyme P109 did not lose its activity in the presence of β-ME.

A soil metagenome derived α-amylase maintained around 80% residual activity in 5mM β-ME (Vidya *et al.,* 2011; Vester *et al.,* 2015), while it was reduced to 30% when concentration was increased to 10mM (Vester *et al.,* 2015). β-ME also enhanced activity of α-amylase from fecal microbial metagenome (Xu *et al.,* 2014b), but the enzyme from *Geobacillus thermoleovorans* exhibited 98% activity in 10mM β-ME (Mehta and Satyanarayana, 2013).

Amylase activity of *Streptomyces* strain and *Saccharopolyspora* sp. was not altered by 0.2 % H_2O_2 (Chakraborty *et al.,* 2011; Roy and Mukherjee, 2013), while that from alkaliphilic *Bacillus* isolate maintained > 80% of its original activity with 1.8 M H_2O_2 (Hagihara *et al.,* 2001); 15-30 % of initial activity of alkaline amylase of *Bacillus* sp. in 2–5 mM H_2O_2 was lost, but completely inhibited in 5mM NaOCl (Saxena *et al.,* 2007).

6.3.2.7 Effect of solvents on activity of P019

Effect of various polar and non-polar solvents on activity of P109 tested at different concentrations ranging from 10, 25, 50% (v/v) was analysed and depicted in Fig 6.13; percentage inhibition is tabulated in table 6.8.

Among the solvents tested, P109 activity was affected maximally by dimethyl formamide and chloroform. At 50% (v/v) concentration these two solvents inhibited the enzyme completely, while the other solvents did not. It was also observed that P109 maintained more than 50% residual activity at 10% concentration of all the solvents tested except dimethyl formamide. Therefore it is suggested that P109 is a moderately solvent tolerant enzyme. Increasing solvent concentration to 25% (v/v) also did not have a prominent effect on enzyme activity, except those with dimethyl formamide, chloroform and toluene which caused 89, 84 and 59% enzyme inhibition respectively.

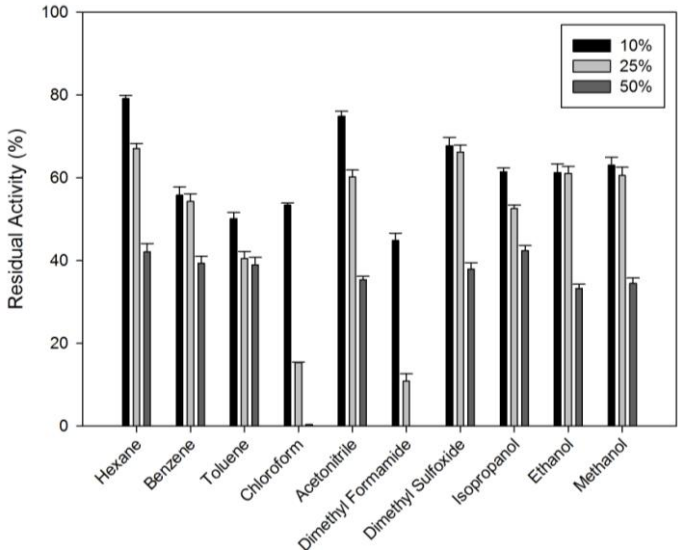

Fig. 6.13: Effect of solvents on activity of P109

Table 6.8: Inhibitory action of solvents

Solvents	10%	25%	50%
Hexane	20%	32%	57%
Benzene	44%	45%	60%
Toluene	49%	59%	61%
Chloroform	46%	84%	99%
Acetonitrile	25%	39%	64%
Dimethyl Formamide	**55%**	**89%**	**100%**
Dimethyl Sulfoxide	32%	33%	62%
Isopropanol	38%	47%	57%
Ethanol	38%	38%	66%
Methanol	36%	39%	65%

Enzymes with resistance toward organic solvents have advantageous over others due to the suppression of undesirable water-dependent side reactions, as well as decrease in microbial contaminations (Vieille and Zeikus, 2001). Polarity or hydrophobicity of organic solvents is measured by a parameter termed "log $P_{o/w}$" value. Solvents with low log $P_{o/w}$ values (water miscible solvents) exhibit more biological toxicity and result in more inhibition of biocatalysts compared to water immiscible solvents with high $P_{o/w}$ solvents (Laane *et al.*, 1987). Water immiscible solvents such as chloroform, toluene and hexane increased amylase activity of a recombinant α-amylase from a *Bacillus* sp. expressed in *E. coli*, while water miscible solvent isopropanol had destabilizing effect on amylase activity with 10 and 20% (v/v) concentration (Asoodeha *et al.*, 2014). Similarly α-amylase from *Bacillus agaradhaerens* was active in hexane, methanol and propanol (Pandey and Singh, 2012), while that from *Nesterenkonia* sp. was dramatically inhibited by organic solvents with low log $P_{o/w}$ values, but activated by organic solvents with higher log $P_{o/w}$ (Shafiei *et al.*, 2011). However no such activation with organic solvents was observed in the present study, although P109 exhibited solvent tolerance moderately at 25% solvent concentration. Another α-amylase retained 97, 99 and 88% residual activity in 30 % ethanol, methanol and acetonitrile respectively (Roy and Mukherjee, 2013).

6.3.2.8 Halotolerence studies on P109

Halotolerance of P109 was tested at different concentrations of NaCl ranging from 0.5-2.5M, residual activity was calculated and depicted in Fig 6.14; and percentage inhibition is tabulated in table 6.9.

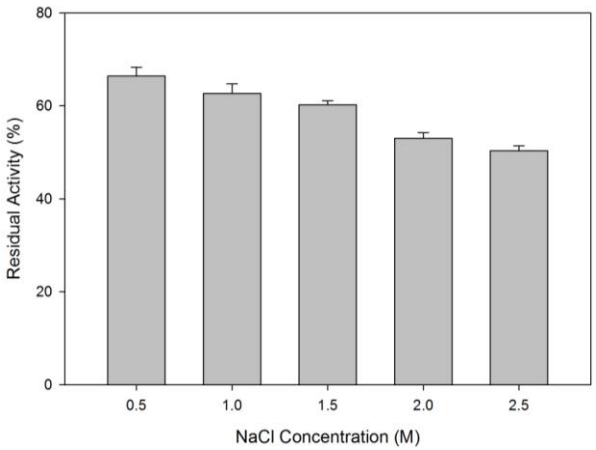

Fig. 6.14: Halotolerence of P109

Table 6.9: Inhibitory action of NaCl

NaCl	Percentage Inhibition (%)
0.5M	33%
1M	37%
1.5M	39%
2M	46%
2.5M	**49%**

In this study, P109 retained 63% activity at 1 M NaCl and almost 67% at 0.5 M NaCl, but residual activity was 51% at 2.5M NaCl. At 1.5M NaCl only 37% inhibition in P109 activity was noted. All these outputs indicate that enzyme P109 was halotolerant at the NaCl concentrations tested.

Extracellular enzymes produced by halophilic microorganisms are adapted to high salinity. At least part of this adaptation might be due to abundance

of acidic residues (Lanyi, 1974). Halotolerant amylases retaining >80% at different concentrations of NaCl have been reported; including metagenomics based amylases as well as those from other sources. A metagenomic α-amylase exhibited 90% activity in 1 M NaCl, but retained only 20% of its initial activity at 2 M NaCl. (Vidya *et al.*, 2011). The α-amylase from *Bacillus* sp. was unaffected by 3% NaCl, however there was a gradual decrease in the amylase activity with increasing salt concentration. Nevertheless the remaining activity was more than 80% even up to 15% NaCl (Aygan *et al.*, 2008). Similarly an α-amylase from *Pseudoalteromonas haloplanktis* retained around 80% activity at 4.5 M NaCl (Srimathi *et al.*, 2007).

6.3.2.9 Substrate specificity of P109

The hydrolytic activity of P109 on different carbohydrate polymers including rice starch, potato starch, wheat starch, corn starch, amylopectin from maize and amylose from potato was determined and depicted in Fig 6.15.

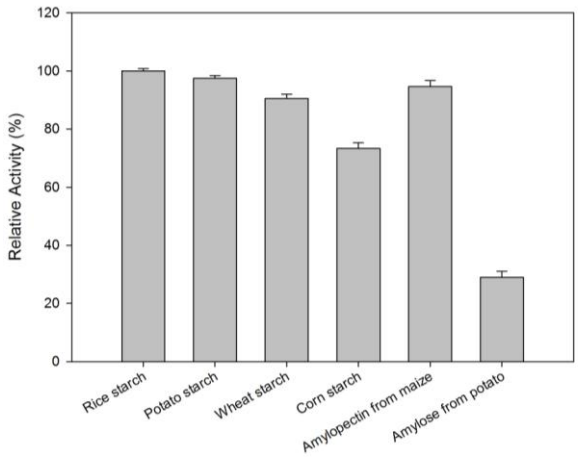

Fig 6.15: Substrate specificity of P109

Substrate specificity studies identified rice starch as the best substrate for P109 followed by potato starch and amylopectin from maize, with 97 and 94% relative activity respectively. Action of P109 on corn starch resulted in 73% relative activity, while the lowest activity was obtained with amylose from potato

resulting in 23% relative activity, suggesting that P109 was more specific towards starch and cannot hydrolyze amylose effectively.

In general, amylases display highest activity towards starch followed by amylose (Gupta *et al.*, 2003), although pAMY showed highest activity with amylose rather than starch (Sharma *et al.*, 2010). Alpha amylase from *Bacillus* sp. exhibited maximum activity with corn starch, maintaining only 74 and 65% activity with amylopectin and amylase respectively (Hagihara *et al.*, 2001). A recombinant α-amylase exhibited maximal hydrolysis of soluble potato starch, while maintaining 68% activity with wheat starch and 61% activity with corn starch (Asoodeha *et al.*, 2014).

6.3.2.10 Determination of kinetic parameters - Km and Vmax

The kinetic parameters of P109 was determined. Km and Vmax were estimated by plotting the initial velocity as the function of the concentration of substrate. The Lineweaver-Burk plot was constructed and shown in Fig 6.16.The Km and Vmax of P109 were determined as 2.7 mg/mL and 454 U/mL respectively.

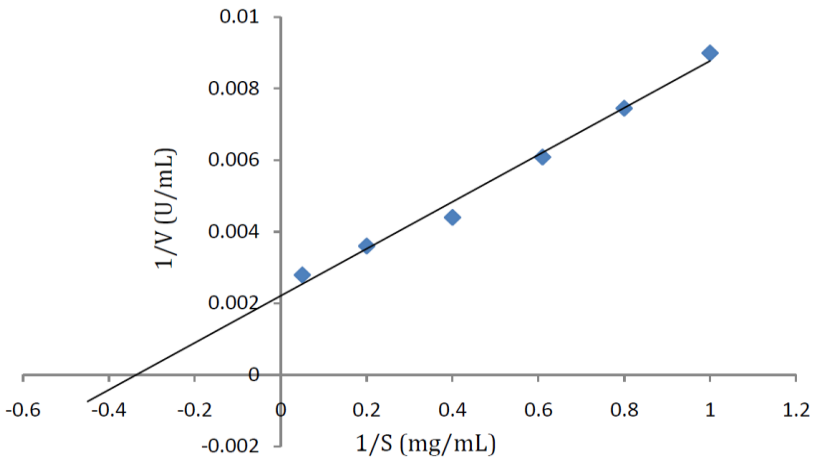

Fig 6.16: Lineweaver–Burk plot for estimation of Km and Vmax of P109

The Michaelis constant (Km) is the substrate concentration yielding half maximal velocity and a low Km value indicates high affinity of the enzyme for the substrate (Hamilton *et al.*, 1999). The Km of most detergent compatible amylases is in the range 0.1 to 5.0 mg/mL using soluble starch as substrate (Niyonzima and More, 2014). A Km value of 3.28 mg/mL has been reported from a metagenomic α-amylase using amylopectin as substrate (Vester et al., 2015). Km and Vmax of *B. cereus* α-amylase was 0.27 mg/mL and 2,600 U/mL respectively (Maalej *et al.*, 2013), while it was 0.7 mg/mL and 2.2 U/mL for that of *B.* US147 (Ghorbel *et al.*, 2009) using starch substrate. A Km of 2.9 mg/mL and Vmax of 7.936 U/mL was recorded for enzyme from *Alicyclobacillus acidocaldarius* employing starch substrate (Chakraborty *et al.*, 2012).

6.3.1.4.11 Determination of starch grain degradation by Scanning Electron Microscopy (SEM)

The effect of P109 on starch grains, structure of native and enzyme treated starch granules were examined by scanning electron microscope. Substrate specificity studies identified that rice starch as the best substrate for the enzyme, so the effect of P109 on rice starch was studied after 2 and 4 h of enzyme treatment. Effect of P109 on potato starch after 2 h of enzyme treatment was also studied and shown in Fig 6.17.

Rice starch was hydrolyzed to a greater extent than the potato starch which was also supported by the substrate specificity studies. Comparing to the control, distortions and holes were observed in P109 treated starch granules. Starch granules after 4 h treatment was maximally distorted suggesting the hydrolyzing property of P109 on its substrate.

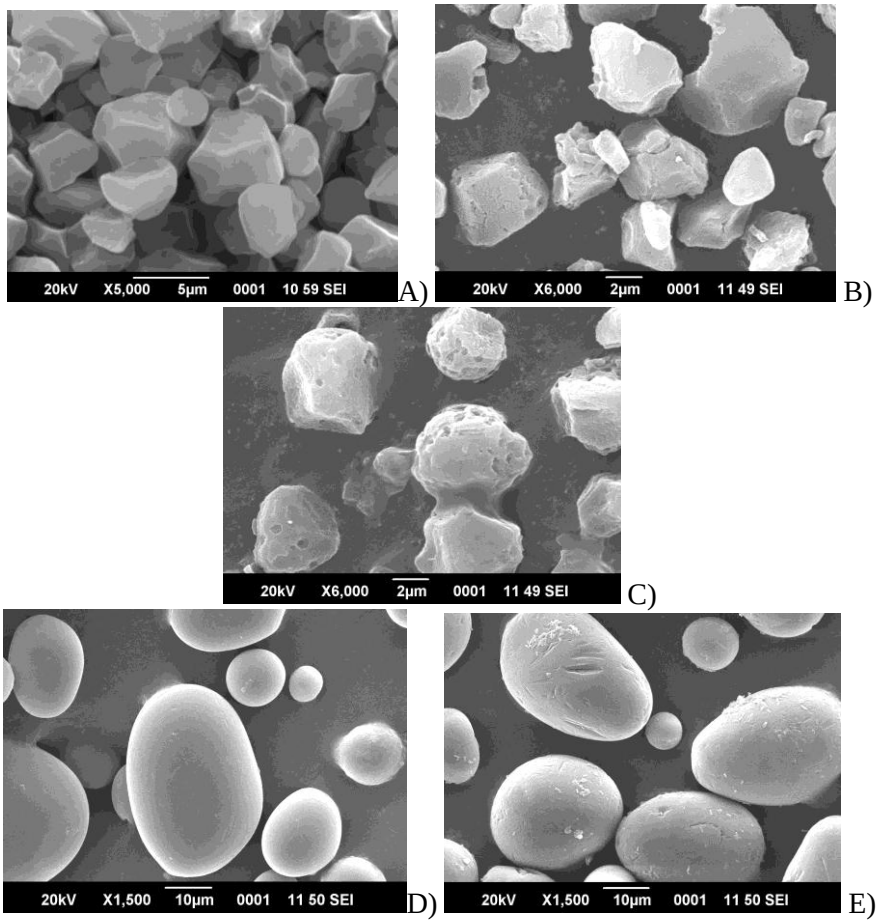

Fig 6.17: SEM images of starch granules before and after hydrolysis

A) Untreated rice starch, B) 2 h treated rice starch, C) 4 h treated rice starch, D) Untreated potato starch, E) 4 h treated potato starch

Prior reports have suggested that starch hydrolysis is due to the interplay of many other factors like granule shape, amylose-lipid complexes, amylose to amylopectin ratio, amylose chain length, phosphorus content, degree of crystallinity, botanical origin and cultivar/variety (Tester and Karkalas, 2006). The method of adsorption of enzymes on starch granules is still unclear, but binding probably occurs through a C-terminal binding domain (Jespersen *et al.*, 1991). Treatment of rice and potato starch granules by α-amylase from *Bacillus subtilis* (Roy *et al.*, 2012; Roy *et al.*, 2013) and *B. amyloliquefaciens* (Demirkan *et al.*,

2005) resulted in holes and crevices in the granules due to the hydrolytic action of the enzyme.

6.3.2.12 End product analysis by thin layer chromatography (TLC)

The nature of enzyme was characterized by TLC based on the degradation products obtained after starch hydrolysis (Fig 6.18).

Starch hydrolysis pattern by P109 indicated that the main hydrolysis products post 2h and 4 h of incubation were glucose(G1), maltose (G2) and malto oligosaccharides such as maltotriose (G3) and maltopentose (G5). The degradation patterns indicted that P109 was an α-amylase which was further confirmed by activity enhancement in the presence of Ca^{2+}.

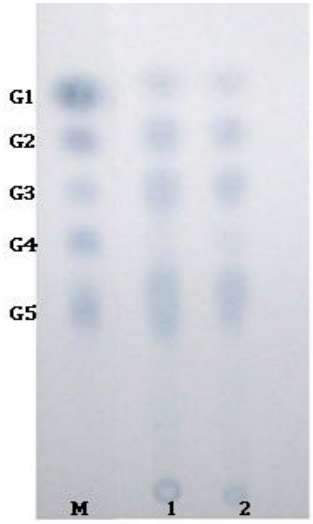

Fig 6.18: Thin-layer chromatographic analysis of hydrolysis products

M represents standards, 1 and 2 represents end products formed after 2 and 4 h of hydrolysis

A metagenomic amylase liberated more of glucose and maltose as hydrolysis products and was identified as α-amylase (Vidya *et al.,* 2011). Degradation products of an α-amylase from *B. licheniformis* resulted in the formation of G2, G3 and G5 in the starch hydrolysate (Hmidet *et al.,* 2008). A

recombinant α-amylase expressed in *E. coli* produced maltose as major product, small malto oligosaccharides and a minor amount of glucose upon action on soluble starch (Demirkan *et al.*, 2005). The main hydrolysis products of soluble starch by α –amylase from *Nocardiopsis* sp. were mainly G1, G2 and G3 including a little G4 (Zhang and Zeng, 2008)

6.3.3 Application studies of amylase

Applications of P109 as a detergent additive, as desizing agent and *invitro* cytotoxic effect of P109 were also tested to check for the suitability as a feed supplement.

6.3.3.1 Commercial detergent compatibility of the P109

The compatibility of P109 with seven different commercial detergents like Ariel®, Surf Excel®, Sunlight®, Tide®, Wheel®, Ujala® at 1% (w/v) was studied, the residual enzyme activity in the presence of the detergents was determined, and represented in Fig. 6.19.

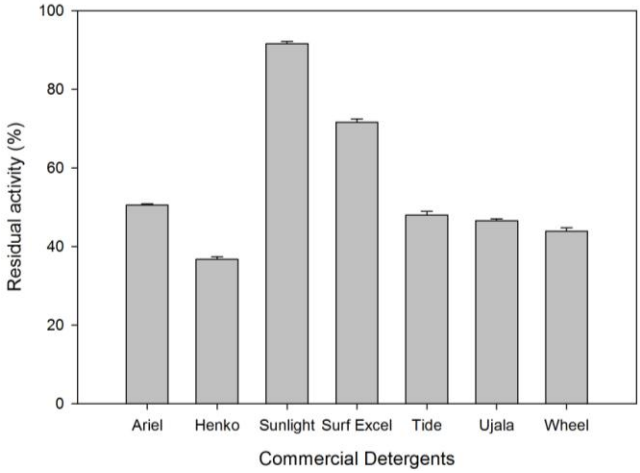

Fig 6.19: Commercial detergent compatibility of P109

It was noted that the α-amylase P109 showed highest compatibility with Sunlight detergent retaining 91.58% activity after 1 hour incubation. However, it retained 71% activity with Surf Excel and 50% activity with Ariel. With all other

detergents tested the enzyme was maintaining less that 50% residual activity. Therefore, the effectiveness of P109 along with Sunlight detergent in starchy stain removal was tested. The partial loss of α-amylase activity in the presence of some of the laundry detergents may be attributed to inhibitory effect of other components of these detergents like bleaching agents, anionic surfactants, water softening agents etc., which may influence the stability of enzyme in detergent (Mukherjee *et al.,* 2009)

6.3.3.2 Wash performance studies

The wash performance analysis was carried out to determine the efficiency of P109 in the removal of starchy stains from fabrics. White cotton cloth stained with chocolate was used for the purpose and wash performance was conducted with appropriate controls followed by visual examination to check the effectiveness of stain removal and photographs were taken (Fig 6.20)

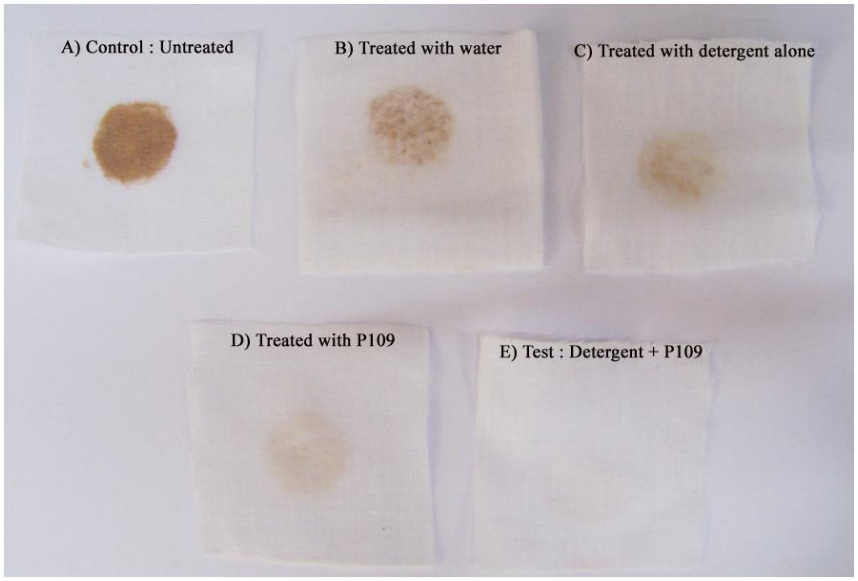

Fig 6.20: Wash performance studies of P109

Visual examination of the washed cloth pieces indicated that the amylase P109 supported the detergent in the chocolate stain removal process and this is clearly visible from Fig 6.20(E) that the stain is almost completely removed from

the cloth piece only in the presence of both the detergent and P109. Thus suggesting P109 as a suitable candidate as a detergent additive for starchy stain removal. It was identified that detergent Sunlight® supplemented with the purified P109 resulted in better stain removal from cotton fabrics as compared to that of detergent alone. Similarly the P109 was also able to remove stain from the cloth piece in the absence of the detergent, as is visible from Fig 6.20(D). When the stained cloth was washed with detergent alone, the stain was not removed completely Fig 6.20(C), with traces of stain retained on the cloth.

The suitability of an enzyme preparation for use in detergents formulations depends on its compatibility with widely used detergents. An ideal detergent enzyme should be stable and active in the detergent solution for a longer period of time and should have adequate temperature stability to be effective in a wide range of washing temperatures. Amylase P109 showed excellent stability and compatibility in the presence of the commercial laundry detergent Sunlight®. In a similar study, α-amylase from *Bacillus licheniformis* maintained around 90% of residual activity in commercial detergent Safed® out of the 8 detergents tested. The wash performance test in chocolate stained cloth piece recognized that the enzyme-detergent combination resulted in better stain removal from cotton fabrics as compared to that of detergent alone (Roy and Mukherjee, 2013). Similarly detergent Fena® supplemented with α-amylase from *B. subtilis* strain showed in increased chocolate stain removal than with detergent alone (Roy *et al.*, 2012). Likewise, the alkaline amylase of *B. cereus* also improved the washing capacity of a detergent by removing food gravy stain from white cloth (Maalej *et al.*, 2013).

6.3.3.3 Application of P109 as a desizing agent

In textile industry, amylases play an important role in desizing. Sizing agents like starch are applied to yarn before weaving, serving as strengthening agents to prevent breaking of the warp thread during the weaving process. Desizing involves the removal of starch from the fabric. To ascertain the desizing

efficiency of P109, white cotton cloth pieces were starched and treated with purified amylase enzyme and effectiveness of starch removal was visualized and photographs were shown in Fig 6.21.

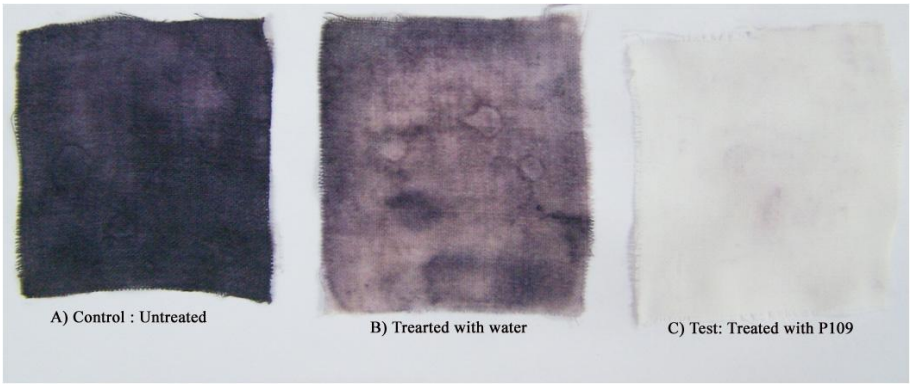

Fig 6.21: Desizing capability of P109

Studies on desizing efficiency of P109 proved that the purified enzyme could successfully remove starch from cotton fabrics and this is clearly visible from Fig 6.21 (C). Comparing to the control and the water treated cloth piece, the starch present in the enzyme treated cloth piece was almost complete removed.

Previous studies on marine *Bacillus subtilis* identified the suitability of extracellular α-amylase enzyme in desizing (Kalpana *et al.*, 2014). In another study an α-amylase from *Bacillus licheniformis* assists in successful removal of starch from cotton fabrics (Roy and Mukherjee, 2013), while that obtained from mutant strain of *Bacillus amyloliquefaciens* also proved useful as a desizing agent (Haq *et al.*, 2010).

6.3.3.4 Application of amylase as a feed supplement

Amylase preparations, together with other hydrolytic enzymes are added to animal feeds to increase the absorption efficiency of nutrients. Before being supplemented as a feed additive, it is important to study the cytotoxic effect of the protein to be incorporated. Therefore effect of P109 on cultured L929 murine

fibroblast cells was tested at concentrations ranging from 6.25-100 µg/mL, to check the suitability of amylase as a feed supplement. Percentage viability of cells treated with varying concentration of P109 is as shown in Fig 6.22 and the inverted phase contrast microscopic images are shown in Fig 6.23.

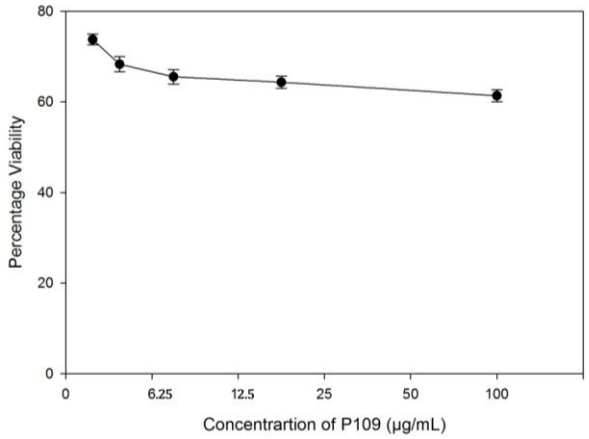

Fig 6.22: Percentage viability of cells with varying concentration of P109

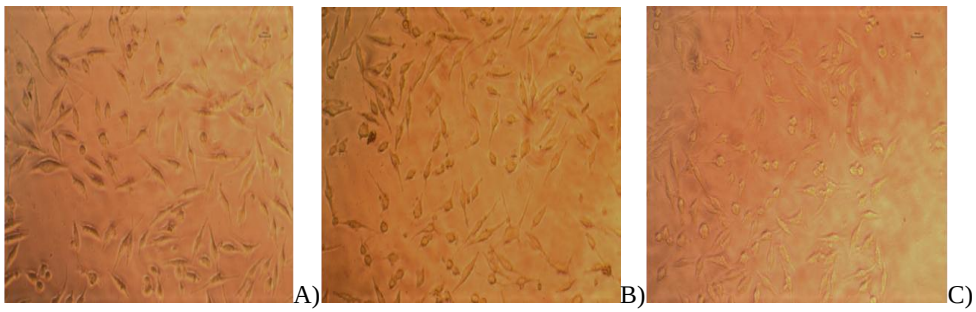

Fig 6.23: Phase contrast micrographs showing the effect of control and P109 treated cells A) Control B) 50 µg/mL treated C) 100 µg/mL treated.

From this *in vitro* test, it was observed that P109 did not exhibit cytotoxicity at the concentrations tested, with the cells maintaining more than 60% viability in all concentrations. This was further confirmed by the microscopic images in which the cell morphology was clearly visible in the treated cells.

Safety evaluation of an α-amylase enzyme from *Thermococcales* sp. was by oral administration to mice and was reported to be avirulent (Landry *et al.*, 2003). According to WHO food additives series 52, LE399 *alpha*-amylase from a genetically modified *Bacillus licheniformis* tested in rats by giving water containing the enzyme was not toxic to the animals (Pronk and Leclercq, 2004).

Using biotechnological interventions and metagenomic approach the potential of sediment microbial community of Arabian Sea was assessed. The outcome was an α-amylase, with distinct properties like halotolerance, stability in the presence of oxidizing and reducing agents and with moderate solvent tolerance. All these physico-chemical properties of P109 suggest its suitability for application as feed enzymes and in textile industries.

Chapter 7

SUMMARY AND CONCLUSION

The marine and mangrove proved to be a highly productive environment with tremendously diverse microbial community profile that were not only related to biogeochemical cycling but are also involved in bioremediation and production of substances of biotechnological interest. Particularly, a high bacterial diversity reflects a high diversity of potential biomolecules. In this milieu, culture independent metagenomic methods explored the microbial diversity of the Arabian Sea and Mangalavanam mangrove sediment and searched for a potent biocatalyst from the metagenome.

Metagenomic DNA extraction was the primary and as yet most challenging process in culture-independent studies. In order to obtain high quality metagenomic DNA for bacterial diversity studies and functional library construction, comparative evaluation of five classical metagenomic DNA isolation protocols were employed in three mangrove sediments and assessed for amenability for PCR reaction. Only method 2 employing liquid nitrogen, proved to be efficient in obtaining PCR amenable DNA, and confirmed with the quality appraisal of DNA by spectrophotomeric methods. Although method 2 extracted good quality DNA, it proved unsuitable for shot-gun library construction because of increased DNA shearing. So a kit based method was established to be useful for functional library construction.

Bacterial community composition of Arabian Sea and Mangalavanam mangrove sediment was assessed employing Sanger sequencing strategy and next generation sequencing method utilizing 16S rRNA gene as phylogenetic anchor.

16S rDNA clone libraries from Arabian Sea and Mangalavanam mangrove sediments with 105 clones and 34 clones respectively were constructed and assessed for bacterial diversity. Diversity analysis revealed the dominance of proteobacterial phylum within both marine and mangrove sediments. The study on marine phylogenetic library identified a wide representation of various taxonomic groups falling into seven major phyla of bacterial domain. It also revealed the dominance of members that plays significant roles in decomposition and assimilation of organic matter in the organically enriched sediment and also associated with bioconversion of sulphur containing organic molecules. The anaerobic conditions prevailing in the locale and sulfur cycling capabilities of the microbes was brought to light.

The mangrove phylogenetic library identified members related to six major phyla of bacterial domain and representative from one candidate phyla. The dominance of *Deltaproteobacteria* was identified, suggesting the prevalence of sulphate reduction processes in this region. Identification of species richness by rarefaction curve indicated that both sediments were enormously rich with bacterial population; and this study represents an almost complete extent of taxonomic diversity at the phylum level. 16S rDNA clone library based studies on both marine and mangrove sediments identified the wealth of uncultivable bacteria which hinted at the possibility for the presence of novel yet-to-be-cultured organisms.

An in depth and high throughput identification of the bacterial communities present in both Arabian Sea and mangrove sediments was also conducted, based on Illumina sequencing. The results exposed the enormous richness of the Arabian Sea with 6,309 OTUs falling into 43 bacterial phyla, including 18 formally described bacterial phyla and 25 candidate phyla. However, studies on Mangalavanam mangrove metagenome identified that this niche was even richer with the bacterial community falling into 9,362 OTUs, classified into 53 bacterial

phyla including 22 formally described bacterial phyla and 31 candidate phyla; was all further confirmed by the alpha and beta diversity patterns. These Illumina based studies revealed the nutrient rich anoxic conditions prevailing in both these sampling sites and further highlighted that more than 15% of the sequences remained unknown at the phylum level, while at the genus level more than 90% of the sequences did not show similarities to any known identified sequences, implying that these sediments were diverse with novel yet to be cultured organisms.

For the functional analysis, two metagenomic shotgun libraries comprising a total of 1113 clones were constructed in vector pUC19 and *E. coli* DH10B host using the metegenomic DNA of Arabian sea and Mangalavanam mangrove sediments. Only one clone BTM109, among the 1113 clones screened for hydrolytic activity exhibited amylolytic activity. The putative α-amylase gene within the clone was identified to be an ORF with 1554 bp, encoding a protein of 517amino acids. Nucleotide similarity search by megablast showed similarity only to a gene for α-amylase derived from uncultured soil metagenomic library, suggesting origin of the gene from an unclassified division of bacterial domain.

The deduced amino acid sequence of α-amylase in BTM109 identified the presence of four conserved domains in the sequence. The catalytic triad of the enzyme consisting of two Asp and one Glu residues acting as proton donor and an Asp acting as a nucleophile during enzyme catalysis. The study also revealed the seven calcium binding residues indicating the role of calcium in the stability and activity of the enzyme. P109 is a metalloenzyme requiring Ca^{2+}. Similarly 18 active site residues were also identified.

Phylogenetic analyisis of the nucleotide and deduced amino acids suggested that BTM109 is widely separated from α-amylase of cultured representatives and cladded with α-amylase obtained from the uncultured

bacterium clone which further confirmed the novelty of the sequence. 3D structure was predicted using Swiss-Model and was identified to be similar to those of other α-amylases, with three domains: domain A, B and C. It was identified that domain-A consisted of the active site residues and catalytic triad, while the calcium binding domains were present in the B-domain followed by the C-terminal residues in the domain-C.

Protein purification process for α-amylase designated as P109 was by acetone precipitation followed by gel filtration chromatography resulting a 28.7 fold increase in the specific activity of the enzyme. Molecular mass of P109 was determined as ~55 kDa by SDS-PAGE and intact mass confimed to be 55.7 kDa by MALDI-TOF MS. The isoelectric point of the enzyme was 6.9.

The enzyme was active at neutral and near alkaline pH with optimum activity at pH-7. pH stability was maximal at pH 7, while retaining > 75% activity between pH 6-11. P109 exhibited optimum activity at 60ºC, but exhibited maximum stability at 40ºC with complete loss of activity at 100 ºC. Calcium ions enhanced the activity of P109. While exposure to metal ions like Ba, Fe, Al, Cd, Co, Na, Mg and Mn, did not affect activity of P109, Cu and Zn ions decreased activity.

N-bromosuccinimide completely inhibited P109 enzyme activity, while inhibition was partial in the presence of PMSF, indicating the role of tryptophan and serine in enzyme catalysis; these two amino acids are specified in the active site of P109, as identified by the conserved domain search of the deduced amino acid sequence. EDTA exhibiting its chelating effect inhibited the enzyme confirming the metalloenzyme nature of P109. Urea, DTT and phytic acid caused 35-49% inhibition.

Ionic detergents had a more inhibitory effect on activity of P109 than non-ionic detergents at 0.5% concentration. Detergents like SDS, CTAB, Tween 20, Triton X-100 caused > 50% activity inhibition. Inhibitory action by these detergents may be due to conformational changes caused in the active site of the enzyme thereby preventing enzyme-substrate binding.

Varying concentration of oxidising agents like hydrogen peroxide and sodium hypochlorite and reducing agent β-mercaptoethanol did not have a profound inhibitory effect on activity of P109. Similarly P109 exhibited stability in various polar and non-polar solvents at 25% concentration, but was completely inhibited in 50% dimethyl formamide.

Halotolerance of P109 was also tested because of its marine nature. The enzyme P109 exhibited halotolerance and even retained 51% activity at 2.5M NaCl. The hydrolytic activity of P109 on different carbohydrate polymers identified that the enzyme acted well on rice starch followed by potato starch, but was a poor degrader of amylose. The Km and Vmax of P109 was 2.7 mg/mL and 454 U/mL respectively.

Starch grain degradation by P109 was assessed by scanning electron microscope. Comparing to the control, distortions and holes were observed in P109 treated rice starch and potato starch granules. The starch hydrolysis pattern analysed by thin layer chromatography indicated that P109 was an α-amylase, producing glucose, maltose and malto oligosaccharides.

Application of P109 included its suitability as biobuilders along with detergents to remove starchy stains, as desizing agent and its *invitro* cytotoxicity to check its suitability as a feed supplement. P109 showed excellent stability and compatibility in the presence of the laundry detergent Sunlight®. Detergent supplemented P109 proved efficient in removing chocolate stains from cloth. P109

could also remove starch from cotton fabrics, indicating its candidature in textile desizing. *Invitro* cytotoxicity tests on cultured L929 murine fibroblast cells showed that P109 was not cytotoxic at the concentrations tested, supporting its candidature role as a feed enzyme.

CONCLUSION

Environments act as reservoir for diverse groups of bacterial communities. Metagenomic methods are valuable to open a window into the unknown microbial diversity, which can be mined to throw up novel products and molecules of industrial importance. This is the first report of bacterial community profile of Arabian Sea and Mangalavanam mangrove sediments using Sanger sequencing and high-throughput sequencing method. Thus the presence of diverse bacterial lineages is reported; and this will help to understand the biological and physicochemical conditions and interactions prevailing in these sediments. Simultaneously the study highlighted the presence of unknown lineages of bacteria in the metagenome, signifying the impending application of the sediment bacteria for novel biocatalysts.

Considering the ever increasing market demand for hydrolases especially α-amylase with improved stability and activity, a successful attempt to characterize the α-amylase obtained from marine metagenomic library and checked for its suitability in various industrial applications was made. This presented the primary report of an α-amylase from Arabian Sea metagenome. The amylase, by virtue of interesting features like halotolerance, solvent stability, detergent compatibility, stability in the presence of reducing agents and efficient digestion of raw starch suggests its biotechnological application in diverse industries.

Exploring the biotechnological potential of microbial communities of Arabian Sea and Mangalavanam mangrove sediments to mine for many more novel biocatalysts for various industrial applications and other biomolecules using the metagenomic approach are strongly recommended.

REFERENCES

Abraham, T. J., Ghosh, S., Nagesh, T. S., & Sasmal, D. (2004). Distribution of bacteria involved in nitrogen and sulphur cycles in shrimp culture systems of West Bengal, India. *Aquaculture*, *239*(1), 275-288.

Allen, E. E., & Banfield, J. F. (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology*, *3*(6), 489-498.

Alongi, D. M. (2002).Present state and future of the world's mangrove forests. *Environmental conservation*, *29(03)*, 331-349.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), 3389-3402.

Amann, J. (1911). The direct count of the water bacteria using the ultra microscope. *Journal of the Royal Microscopical Society*, 29, 381-384

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation.*Microbiological reviews*, *59*(1), 143-169.

Aminin, A. L. N., Warganegara, F. M., & Aditiawati, P. (2008). Culture-Independent and culture-dependent approaches on microbial community analysis at Gedongsongo (GS-2) hot spring. *International Journal of Integrative Biology*, 2, 145-152.

Andreote, F. D., Jiménez, D. J., Chaves, D., Dias, A. C. F., Luvizotto, D. M., Dini-Andreote, F., & de Melo, I. S. (2012). The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS One*, *7*(6), e38600.

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., & Rohwer, F. (2006). The marine viromes of four oceanic regions. *PLoS Biology*,*4*(11), e368.

Antón, J., Rosselló-Mora, R., Rodríguez-Valera, F., & Amann, R. (2000). Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Applied and Environmental Microbiology*, *66*(7), 3052-3057.

Aravindraja, C., Viszwapriya, D., & Pandian, S. K. (2013). Ultradeep 16S rRNA sequencing analysis of geographically similar but diverse unexplored marine samples reveal varied bacterial community composition, PLOS One, 8(10), e76724

Arikan, B. (2007). Highly thermostable, thermophilic, alkaline, SDS and chelator resistant amylase from a thermophilic *Bacillus* sp. isolate A3-15. *Bioresource Technology*, *99*(8), 3071-3076.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.*Bioinformatics*, *22*(2), 195-201.

Arung, E. T., Yoshikawa, K., Shimizu, K., & Kondo, R. (2010). Isoprenoid-substituted flavonoids from wood of *Artocarpus heterophyllus* on B16 melanoma cells: cytotoxicity and structural criteria. *Fitoterapia*, *81*(2), 120-123.

Asami, H., Aida, M., & Watanabe, K. (2005). Accelerated sulfur cycle in coastal marine sediment beneath areas of intensive shellfish aquaculture.*Applied and Environmental Microbiology*, *71*(6), 2925-2933.

Asoodeh, A., Emtenani, S., Emtenani, S., Jalal, R., & Housaindokht, M. R. (2014). Molecular cloning and biochemical characterization of a thermoacidophilic, organic-solvent tolerant α-amylase from a *Bacillus* strain in *Escherichia coli. Journal of Molecular Catalysis B: Enzymatic*, *99*, 114-120.

Aygan, A., Arikan, B., Korkmaz, H., Dinçer, S., & Çolak, Ö. (2008). Highly thermostable and alkaline α-amylase from a halotolerant-alkaliphilic *Bacillus* sp. AB68. *Brazilian Journal of Microbiology*, *39*(3), 547-553.

Azam, F. (1998). Microbial control of oceanic carbon flux: the plot thickens. *Science*, 280, 694-695.

Bahri, S. M., & Ward, J. M. (1993). Sequence of the Streptomyces thermoviolaceus CUB74 α-amylase-encoding gene and its transcription analysis in Streptomyces lividans. *Gene*, *127*(1), 133-137.

Baker, B. J., & Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS microbiology ecology*, *44*(2), 139-152.

Bakken, L. R. (1985). Separation and purification of bacteria from soil. *Applied and Environmental Microbiology*, *49*(6), 1482-1487.

Bakken, L. R., & Lindahl, V. (1995). Recovery of bacterial cells from soil. *Nucleic Acids in the Environment* (pp. 9-27). Springer Berlin Heidelberg.

Banks, W., & Greenwood, C. T. (1977). Mathematical models for the action of alpha-amylase on amylose. *Carbohydrate Research*, *57*, 301-315.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., & Waley, S. G. (1975). Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution: using amino acid sequence data. *Nature*, *255*, 609-614.

Barns, S. M., Cain, E. C., Sommerville, L., & Kuske, C. R. (2007). Acidobacteria phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. Applied and Environmental Microbiology, 73(9), 3113-3116.

Barns, S. M., Delwiche, C. F., Palmer, J. D., & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences*, 93(17), 9188-9193.

Barns, S. M., Fundyga, R. E., Jeffries, M. W., & Pace, N. R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proceedings of the National Academy of Sciences, 91(5), 1609-1613.

Barriuso, J., Valverde, J. R., & Mellado, R. P. (2011). Estimation of bacterial diversity using next generation sequencing of 16S rDNA: a comparison of different workflows. *BMC bioinformatics*, *12*(1), 473.

Basak, P., Pramanik, A., Roy, R., Chattopadhyay, D., & Bhattacharyya, M. (2015). Cataloguing the bacterial diversity of the Sundarbans mangrove, India in the light of metagenomics. Genomics Data, 4, 90-92.

Becks, S., Bielawski, C., Henton, D., Padala, R., Burrows, K., & Slaby, R. (1995). Application of a liquid stable amylase reagent on the Ciba Corning express, *Clinical Chemistry*, 41(6), 186

Beja, O. EN Spudich, JL Spudich, M. Leclerc, and EF Delong. 2001. Prote orhodopsin phototrophy in the ocean. *Nature*, *411*, 786-789.

Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., & DeLong, E. F. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, *289*(5486), 1902-1906.

Bell, D. J., Hoare, M., & Dunnill, P. (1983). The formation of protein precipitates and their centrifugal recovery. *Downstream processing* Springer Berlin Heidelberg. pp. 1-72.

Bender, H., Lehmann, J., & Wallenfels, K. (1959). Pullulan , an extracellular glucan from *Pullularia pullulans*. *Biochimica et biophysica acta*, *36*(2), 309-316.

Benlloch, S., López-López, A., Casamayor, E. O., Øvreås, L., Goddard, V., Daae, F. L., & Rodríguez-Valera, F. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environmental Microbiology*, *4*(6), 349-360.

Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., & Kelley, L. A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: Structure, Function, and Bioinformatics*, *70*(3), 611-625.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., & Anastasi, C. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. N*ature*, *456*(7218), 53-59.

Berman, H., Henrick, K., Nakamura, H., & Markley, J. L. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, *35*(suppl 1), D301-D303.

Berry, A. E., Chiocchini, C., Selby, T., Sosio, M., & Wellington, E. M. (2003). Isolation of high molecular weight DNA from soil for cloning into BAC vectors.*FEMS Microbiology Letters*, *223*(1), 15-20.

Bhaya, D., Grossman, A. R., Steunou, A. S., Khuri, N., Cohan, F. M., Hamamura, N & Heidelberg, J. F. (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *The ISME journal*, *1*(8), 703-713.

Birnboim, H. C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, *7*(6), 1513-1523.

Bissett, A., Bowman, J. P., & Burke, C. M. (2008). Flavobacterial response to organic pollution. *Aquatic Microbial Ecology*, *51*(1), 31.

Biver, S., Portetelle, D., & Vandenbol, M. (2013). Characterization of a new oxidant-stable serine protease isolated by functional metagenomics.*Springerplus*, *2*, 410.

Blank, C. E., Cady, S. L., & Pace, N. R. (2002). Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. *Applied and Environmental Microbiology*, 68(10), 5123-5135.

Bond, P. L., Smriga, S. P., & Banfield, J. F. (2000). Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Applied and Environmental Microbiology*, *66*(9), 3842-3849.

Bordbar, A. K., Omidiyan, K., & Hosseinzadeh, R. (2005). Study on interaction of α-amylase from *Bacillus subtilis* with cetyl trimethylammonium bromide.*Colloids and Surfaces B: Biointerfaces*, *40*(1), 67-71.

Bott, T. L., & Brock, T. D. (1969). Bacterial growth rates above 90 C in Yellowstone hot springs. *Science*, *164*(3886), 1411-1412.

Bouillon, S., Raman, A. V., Dauby, P., & Dehairs, F. (2002). Carbon and nitrogen stable isotope ratios of subtidal benthic invertebrates in an estuarine mangrove ecosystem (Andhra Pradesh, India). *Estuarine, Coastal and Shelf Science*, *54*(5), 901-913.

Bowman, J. P., McCammon, S. A., Brown, M. V., Nichols, D. S., & McMeekin, T. A. (1997). Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Applied and Environmental Microbiology*, *63*(8), 3068-3078.

Bowman, J. P., McCammon, S. A., Gibson, J. A. E., Robertson, L., & Nichols, P. D. (2003). Prokaryotic metabolic activity and community structure in Antarctic continental shelf sediments. *Applied and Environmental Microbiology*,*69*(5), 2448-2462.

Bowman, J. P., Rea, S. M., McCammon, S. A., & McMeekin, T. A. (2000). Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hills, Eastern Antarctica. *Environmental Microbiology*, *2*(2), 227-237.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2015). obitools: a unix-inspired software package for DNA metabarcoding.*Molecular Ecology Resources*.

Božić, N., Slavić, M. Š., Gavrilović, A., & Vujčić, Z. (2014). Production of raw-starch-digesting α-amylase isoform from *Bacillus* sp. under solid-state fermentation and biochemical characterization. *Bioprocess and Biosystems Engineering*, *37*(7), 1353-1360.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.*Analytical biochemistry*, *72*(1), 248-254.

Brady, S. F., Chao, C. J., & Clardy, J. (2004). Long-chain N-acyltyrosine synthases from environmental DNA. *Applied and Environmental Microbiology*,*70*(11), 6865-6870.

Brady, S. F., Chao, C. J., Handelsman, J., & Clardy, J. (2001). Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Organic letters*, *3*(13), 1981-1984.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, *99*(22), 14250-14255.

Brennan, Y., Callen, W. N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., & Steer, B. A. (2004). Unusual microbial xylanases from insect guts.*Applied and Environmental Microbiology*, *70*(6), 3609-3617.

Brock, T. D. (1967). Life at high temperatures. *Science*, *158*(3804), 1012-1019.

Brosch, R., Gordon, S. V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., & Cole, S. T. (1998). Use of a Mycobacterium tuberculosisH37Rv Bacterial Artificial Chromosome Library for Genome Mapping, Sequencing, and Comparative Genomics. *Infection and immunity*, *66*(5), 2221-2229.

Brown, M. V., & Bowman, J. P. (2001). A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiology Ecology*, *35*(3), 267-275.

Bruce, K. D., Hiorns, W. D., Hobman, J. L., Osborn, A. M., Strike, P., & Ritchie, D. A. (1992). Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. *Applied and Environmental Microbiology*, 58(10), 3413-3416.

Bruinenberg, P. M., Hulst, A. C., Faber, A., & Voogd, R. H. (1996). A process for surface sizing or coating of paper. European Patent Application.

Bull, A. T., Goodfellow, M., & Slater, J. H. (1992). Biodiversity as a source of innovation in biotechnology. *Annual Reviews in Microbiology*, *46*(1), 219-246.

Burgmann, H., Pesaro, M., Widmer, F., & Zeyer, J. (2001). A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods*, *45*(1), 7-20.

Bürgmann, H., Pesaro, M., Widmer, F., & Zeyer, J. (2001). A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods*, 45(1), 7-20.

Burhan, A., Nisa, U., Gökhan, C., Ömer, C., Ashabil, A., & Osman, G. (2003). Enzymatic properties of a novel thermostable, thermophilic, alkaline and chelator resistant amylase from an alkaliphilic *Bacillus* sp. isolate ANT-6.*Process Biochemistry*, *38*(10), 1397-1403.

Byrne, M. E., Ball, D. A., Guerquin-Kern, J. L., Rouiller, I., Wu, T. D., Downing, K. H., ... & Komeili, A. (2010). Desulfovibrio magneticus RS-1 contains an iron-and phosphorus-rich organelle distinct from its bullet-shaped magnetosomes.*Proceedings of the National Academy of Sciences*, *107*(27), 12263-12268.

Cai, Y., & Sun, Y. (2011). ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. *Nucleic Acids Research*, gkr349.

Campbell, B. J., Engel, A. S., Porter, M. L., & Takai, K. (2006). The versatile ε-*proteobacteria*: key players in sulphidic habitats. *Nature Reviews Microbiology*, *4*(6), 458-468.

Campbell, B. J., Jeanthon, C., Kostka, J. E., Luther, G. W., & Cary, S. C. (2001). Growth and Phylogenetic Properties of Novel Bacteria Belonging to the Epsilon Subdivision of the Proteobacteria Enriched fromAlvinella pompejana and Deep-Sea Hydrothermal Vents. *Applied and Environmental Microbiology*,*67*(10), 4566-4572.

Canfield, D. E., Jørgensen, B. B., Fossing, H., Glud, R., Gundersen, J., Ramsing, N. B., & Hall, P. O. (1993). Pathways of organic carbon oxidation in three continental margin sediments. *Marine Geology*, *113*(1), 27-40.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335-336.

Carpi, G., Cagnacci, F., Wittekindt, N. E., Zhao, F., Qi, J., Tomsho, L. P., & Schuster, S. C. (2011). Metagenomic profile of the bacterial communities associated with Ixodes ricinus ticks. *PloS One*, *6*(10), e25604-e25604.

Carrigg, C., Rice, O., Kavanagh, S., Collins, G., & O'Flaherty, V. (2007). DNA extraction method affects microbial community profiles from soils and sediment. *Applied Microbiology and Biotechnology*, *77*(4), 955-964.

Chakraborty, S., Khopade, A., Biao, R., Jian, W., Liu, X. Y., Mahadik, K., & Kokare, C. (2011). Characterization and stability studies on surfactant, detergent and oxidant stable α-amylase from marine haloalkaliphilic *Saccharopolyspora* sp. A9. *Journal of Molecular Catalysis B: Enzymatic*, *68*(1), 52-58.

Chakraborty, S., Raut, G., Khopade, A., Mahadik, K., & Kokare, C. (2012). Study on calcium ion independent alpha-amylase from haloalkaliphilic marine Streptomyces strain A 3. *Indian Journal of Biotechnology*, *11*(4), 427-437.

Chang, M. C., Chang, J. C., & Chen, J. P. (1993). Cloning and nucleotide sequence of an extracellular α-amylase gene from *Aeromonas hydrophila* MCC-1. *Journal of General Microbiology*, *139*(12), 3215-3223.

Chen, I., Marcos, A. T., da Costa, S. O., Martin, J. F., & Padilla, G. (1995). Cloning and characterization of an alpha-amylase gene from Streptomyces sp WL6. *Biochemistry and Molecular biology International*, *35*(5), 1059-1067.

Chen, K., & Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Computational Biology*, *1*(2), e24.

Chevreux, B., Wetter, T., & Suhai, S. (1999, October). Genome sequence assembly using trace signals and additional sequence information. In *German Conference on Bioinformatics* (pp. 45-56).

Christner, B. C., Kvitko II, B. H., & Reeve, J. N. (2003). Molecular identification of bacteria and eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles*,*7*(3), 177-183.

Christner, B. C., Mosley-Thompson, E., Thompson, L. G., & Reeve, J. N. (2001). Isolation of bacteria and 16S rDNAs from Lake Vostok accretion ice, *Environmental Microbiology*, *3*(9), 570-577.

Chu, X., He, H., Guo, C., & Sun, B. (2008). Identification of two novel esterases from a marine metagenomic library derived from South China Sea.*Applied Microbiology and Biotechnology*, *80*(4), 615-625.

Cifuentes, A., Antón, J., Benlloch, S., Donnelly, A., Herbert, R. A., & Rodríguez-Valera, F. (2000). Prokaryotic diversity in Zostera noltii-colonized marine sediments. *Applied and Environmental Microbiology*, *66*(4), 1715-1719.

Clark, M. W., McConchie, D., Lewis, D. W., & Saenger, P. (1998). Redox stratification and heavy metal partitioning in Avicennia-dominated mangrove sediments: a geochemical model. *Chemical Geology*, *149*(3), 147-171.

Cole, C., Barber, J. D., & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Research*, *36*(suppl 2), W197-W201.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., & Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, *37*(suppl 1), D141-D145.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., & Tiedje, J. M. (2013). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, gkt1244.

Colwell, R. R., Tamplin, M. L., Brayton, P. R., Gauzens, A. L., Tall, B. D., Herrington, D., & Sack, D. A. (1990). Environmental aspects of *Vibrio cholerae* in transmission of cholera. *Advances in research on cholera and related diarrhoea, 7th ed. Tokyo, KTK Scientific Publishers*, 327-343.

Coolen, M. J., Cypionka, H., Sass, A. M., Sass, H., & Overmann, J. (2002). Ongoing modification of Mediterranean Pleistocene sapropels mediated by prokaryotes. *Science*, *296*(5577), 2407-2410.

Corliss, J. B., Dymond, J., Gordon, L. I., & Edmond, J. M. (1979). on the Galapagos Rift. *Science*, *203*, 16.

Coronado, M. J., Vargas, C., Mellado, E., Tegos, G., Drainas, C., Nieto, J. J., & Ventosa, A. (2000). The α-amylase gene amyH of the moderate halophile *Halomonas meridiana:* cloning and molecular characterization. *Microbiology*,*146*(4), 861-868.

Cottrell, M. T., & Kirchman, D. L. (2000). Natural assemblages of marine *Proteobacteria* and members of the *Cytophaga-Flavobacter* cluster consuming low-and high-molecular-weight dissolved organic matter. *Applied and Environmental Microbiology*, *66*(4), 1692-1697.

Cottrell, M. T., Moore, J. A., & Kirchman, D. L. (1999). Chitinases from uncultured marine microorganisms. *Applied and Environmental microbiology*,*65*(6), 2553-2557.

Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P., Helynck, G., & Pernodet, J. L. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology*, *69*(1), 49-55.

Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P., & Simonet, P. (2001). Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Applied andEnvironmental Microbiology*, 3(7), 431-439.

Craig, J. W., Chang, F. Y., & Brady, S. F. (2009). Natural products from environmental DNA hosted in Ralstonia metallidurans. *ACS Chemical Biology*,*4*(1), 23-28.

Cullen, D. W., & Hirsch, P. R. (1998). Simple and rapid method fordirect extraction of microbial DNA from soil for PCR. *Soil Biology and Biochemistry*,*30*(8), 983-993.

Curtis, T. P., Sloan, W. T., & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences*,*99*(16), 10494-10499.

Dale, A. W., Bertics, V. J., Treude, T., Sommer, S., & Wallmann, K. (2013). Modeling benthic–pelagic nutrient exchange processes and porewater distributions in a seasonally hypoxic

sediment: evidence for massive phosphate release by Beggiatoa. *Biogeosciences (BG)*, *10*(2), 629-651.

Dalmia, B. K., Schütte, K., & Nikolov, Z. L. (1995). Domain E of *Bacillus macerans* cyclodextrin glucanotransferase: An independent starch-binding domain. *Biotechnology and Bioengineering*, *47*(5), 575-584.

Daniel, R. (2005). The metagenomics of soil. *Nature Reviews Microbiology*,*3*(6), 470-478.

Das, K., Doley, R., & Mukherjee, A. K. (2004). Purification and biochemical characterization of a thermostable, alkaliphilic, extracellular α-amylase from *Bacillus subtilis* DM-03, a strain isolated from the traditional fermented food of India. *Biotechnology and applied biochemistry*, *40*(3), 291-298.

Das, S., Lyla, P. S., & Khan, S. A. (2006). Marine microbial diversity and ecology: importance and future perspectives. *Current Science*, *90*(10), 1325-1335.

D'Costa, V. M., McGrann, K. M., Hughes, D. W., & Wright, G. D. (2006). Sampling the antibiotic resistome. *Science*, *311*(5759), 374-377.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.*Proceedings of the National Academy of Sciences*, *107*(33), 14691-14696.

Degrange, V., & Bardin, R. (1995). Detection and counting of Nitrobacter populations in soil by PCR. *Applied and Environmental Microbiology*, *61*(6), 2093-2098.

Delavat, F., Phalip, V., Forster, A., Plewniak, F., Lett, M. C., & Lièvremont, D. (2012). Amylases without known homologues discovered in an acid mine drainage: significance and impact. *Scientific reports*, *2*(354), 1-6

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., & Vogel, T. M. (2011a). Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology*, *77*(4), 1315-1324.

Delmont, T. O., Robe, P., Clark, I., Simonet, P., & Vogel, T. M. (2011b). Metagenomic comparison of direct and indirect soil DNA extraction approaches. *Journal of Microbiological Methods*, *86*(3), 397-400.

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., & Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, *311*(5760), 496-503.

Demirkan, E. S., Mikami, B., Adachi, M., Higasa, T., & Utsumi, S. (2005). α-Amylase from *B. amyloliquefaciens:* purification, characterization, raw starch degradation and expression in *E. coli. Process Biochemistry*, *40*(8), 2629-2636.

Denhardt, D. T. (1966). A membrane-filter technique for the detection of complementary DNA. *Biochemical and Biophysical Research Communications*,*23*(5), 641-646.

DeSantis, G., Zhu, Z., Greenberg, W. A., Wong, K., Chaplin, J., Hanson, S. R., & Burk, M. J. (2002). An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. *Journal of the American Chemical Society*, *124*(31), 9024-9025.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069-5072.

Deutch, C. E. (2002). Characterization of a salt-tolerant extracellular a-amylase from *Bacillus dipsosauri*. *Letters in Applied Microbiology*, *35*(1), 78-84.

Dias, A. C., Andreote, F. D., Rigonato, J., Fiore, M. F., Melo, I. S., & Araújo, W. L. (2010). The bacterial diversity in a Brazilian non-disturbed mangrove sediment. *Antonie van Leeuwenhoek*, *98*(4), 541-551.

Diaz-Perez, S. V., Alatriste-Mondragon, F., Hernandez, R., Birren, B., & Gunsalus, R. P. (1997). Bacterial artificial chromosome (BAC) library as a tool for physical mapping of the archaeon Methanosarcina thermophila TM-1. *Microbial & comparative genomics*, *2*(4), 275-286.

Divya, B., Parvathi, A., Bharathi, P. L., & Nair, S. (2011). 16S rRNA-based bacterial diversity in the organic-rich sediments underlying oxygen-deficient waters of the eastern Arabian Sea. *World Journal of Microbiology and Biotechnology*, *27*(12), 2821-2833.

Dong, D., Yan, A., Liu, H., Zhang, X., & Xu, Y. (2006). Removal of humic substances from soil DNA using aluminium sulfate. *Journal of Microbiological Methods*, *66*(2), 217-222.

Donovan, W. P., Tan, Y., & Slaney, A. C. (1997). Cloning of the nprA gene for neutral protease A of *Bacillus thuringiensis* and effect of in vivo deletion of nprA on insecticidal crystal protein. *Applied and environmental microbiology*,*63*(6), 2311-2317.

Dos Santos, H. F., Cury, J. C., Do Carmo, F. L., Dos Santos, A. L., Tiedje, J., Van Elsas, J. D., & Peixoto, R. S. (2011). Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. PLoS One, 6(3), e16943.

Duke, N. C., Meynecke, J. O., Dittmann, S., Ellison, A. M., Anger, K., Berger, U., & Dahdouh-Guebas, F. (2007). A world without mangroves?. *Science*,*317*(5834), 41-42.

Eden, P. A., Schmidt, T. M., Blakemore, R. P., & Pace, N. R. (1991). Phylogenetic analysis of Aquaspirillum magnetotacticum using polymerase chain reaction-amplified 16S rRNA-specific DNA. *International Journal of Systematic Bacteriology*, *41*(2), 324-325.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*,*27*(16), 2194-2200.

Edgcomb, V. P., McDonald, J. H., Devereux, R., & Smith, D. W. (1999). Estimation of bacterial cell numbers in humic acid-rich salt marsh sediments with probes directed to 16S ribosomal DNA. *Applied and Environmental Microbiology*, *65*(4), 1516-1523.

Edlund, A., Hardeman, F., Jansson, J. K., & Sjoling, S. (2008). Active bacterial community structure along vertical redox gradients in Baltic Sea sediment.*Environmental Microbiology*, *10*(8), 2051-2063.

Edwards, K. J., Bond, P. L., Gihring, T. M., & Banfield, J. F. (2000). An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science*, *287*(5459), 1796-1799.

Eisen, J. A. (2007). Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. *PLoS biology*, *5*(3), e82.

Ekkers, D. M., Cretoiu, M. S., Kielak, A. M., & van Elsas, J. D. (2012). The great screen anomaly - A new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*, 93(3), 1005-1020.

Elend, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J. D., Steele, H. L., & Streit, W. R. (2006). Isolation and biochemical characterization of two novel metagenome-derived esterases. *Applied and Environmental Microbiology*, *72*(5), 3637-3645.

Elifantz, H., Dittel, A. I., Cottrell, M. T., & Kirchman, D. L. (2007). Dissolved organic matter assimilation by heterotrophic bacterial groups in the western Arctic Ocean. *Aquatic Microbial Ecology*, *50*(1), 39-49.

Eme, L., Reigstad, L. J., Spang, A., Lanzén, A., Weinmaier, T., Rattei, T., & Brochier-Armanet, C. (2013). Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper) thermophilic lineages related to Thaumarchaeota. *Research in microbiology*, *164*(5), 425-438.

Ezeji, T. C., & Bahl, H. (2006). Purification, characterization, and synergistic action of phytate-resistant α-amylase and α-glucosidase from *Geobacillus thermodenitrificans* HRO10. *Journal of Biotechnology*, *125*(1), 27-38.

Faegri, A., Torsvik, V. L., & Goksøyr, J. (1977). Bacterial and fungal activities in soil: separation of bacteria and fungi by a rapid fractionated centrifugation technique. *Soil Biology and Biochemistry*, *9*(2), 105-112.

Fang, J., Arakawa, S., Kato, C., & Schouten, S. (2006). Microbial diversity of cold-seep sediments in Sagami Bay, Japan as determined by 16S rDNA and lipid analyses. In *AGU Fall Meeting Abstracts* (Vol. 1, p. 1109).

Feinstein, L. M., Sul, W. J., & Blackwood, C. B. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Applied and Environmental Microbiology*, *75*(16), 5428-5433.

Felczykowska, A., Bloch, S. K., Nejman-Falenczyk, B., & Baranska, S. (2012). Metagenomic approach in the investigation of new bioactive compounds in the marine environment. *Acta Biochimica Polonica 59*, 501-505.

Fernandez, A. B., Ghai, R., Martin-Cuadrado, A. B., Sanchez-Porro, C., Rodriguez-Valera, F., & Ventosa, A. (2013). Metagenome sequencing of prokaryotic microbiota from two hypersaline ponds of a marine saltern in Santa Pola, Spain. *Genome announcements*, *1*(6), e00933-13

Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A., & Golyshin, P. N. (2005). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora.*Environmental Microbiology*, *7*(12), 1996-2010.

Fincan, S. A., & Enez, B. (2014). Production, purification, and characterization of thermostable α-amylase from thermophilic *Geobacillus stearothermophilus. Starch*, *66*(1-2), 182-189.

Fincan, S. A., & Enez, B. Ozdemir, S. & Matpan Bekler, F. (2014). Purification and characterization of thermostable α-amylase from thermophilic *Anoxybacillus flavithermus*. *Carbohydrate Polymers*, *102*, 144-150.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, *269*(5223), 496-512.

Flores, G. E., Campbell, J. H., Kirshtein, J. D., Meneghin, J., Podar, M., Steinberg, J. I., & Reysenbach, A. L. (2011). Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environmental Microbiology*, *13*(8), 2158-2171.

Flores, G. E., Shakya, M., Meneghin, J., Yang, Z. K., Seewald, J. S., Geoff Wheat, C., & Reysenbach, A. L. (2012). Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin.Geobiology, 10(4), 333-346.

Foulongne, V., Sauvage, V., Hebert, C., Dereure, O., Cheval, J., Gouilh, M. A., & Eloit, M. (2012). Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *PloS one*, *7*(6), e38499.

Frostegard, A.., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., & Simonet, P. (1999). Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology*, *65*(12), 5409-5420.

Fry, J. C., Parkes, R. J., Cragg, B. A., Weightman, A. J., & Webster, G. (2008). Prokaryotic biodiversity and activity in the deep subseafloor biosphere.FEMS Microbiology Ecology, 66(2), 181-196.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150-3152.

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature*, *459*(7244), 193-199.

Fujimoto, K., Imaya, A., Tabuchi, R., Kuramoto, S., Utsugi, H., & Murofushi, T. (1999). Belowground carbon storage of Micronesian mangrove forests.*Ecological Research*, *14*(4), 409-413.

Gabor, E. M., de Vries, E. J., & Janssen, D. B. (2003). Efficient recovery of environmental DNA for expression cloning by indirect extraction methods.*FEMS Microbiology Ecology*, *44*(2), 153-163.

Gabor, E. M., De Vries, E. J., & Janssen, D. B. (2004). Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology*, *6*(9), 948-958.

Galand, P. E., Casamayor, E. O., Kirchman, D. L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences*, *106*(52), 22427-22432.

Garcia-Pedrajas, M. D., Bainbridge, B. W., Heale, J. B., Pérez-Artés, E., & Jiménez-Díaz, R. M. (1999). A simple PCR-based method for the detection of the chickpea-wilt pathogen Fusarium oxysporum f. sp. ciceris in artificial and natural soils. *European Journal of Plant Pathology*, *105*(3), 251-259.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, *31*(13), 3784-3788.

Ghorbel, R. E., Maktouf, S., Massoud, E. B., Bejar, S., & Chaabouni, S. E. (2009). New thermostable amylase from *Bacillus cohnii* US147 with a broad pH applicability. *Applied Biochemistry and Biotechnology*, *157*(1), 50-60.

Ghosh, A., Dey, N., Bera, A., Tiwari, A., Sathyaniranjan, K., Chakrabarti, K., & Chattopadhyay, D. (2010). Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. *Saline systems*,*6*(1), 1.

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., & Joint, I. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environmental Microbiology*, *11*(12), 3132-3139.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., & Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, *312*(5778), 1355-1359.

Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., & Handelsman, J. (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology*, *68*(9), 4301-4306.

Gillespie, D. E., Rondon, M. R., Williamson, L. L., & Handelsman, J. (2005). Metagenomic libraries from uncultured microorganisms. *Molecular microbial ecology*. Taylor & Francis Group, New York, NY, 261-279.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345, 60 – 63.

Gobet, A., Böer, S. I., Huse, S. M., van Beusekom, J. E., Quince, C., Sogin, M. L., & Ramette, A. (2012). Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *The ISME journal*, *6*(3), 542-553.

Goltsman, D. S. A., Denef, V. J., Singer, S. W., VerBerkmoes, N. C., Lefsrud, M., Mueller, R. S., & Banfield, J. F. (2009). Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing "Leptospirillum rubarum"(Group II) and "Leptospirillum ferrodiazotrophum"(Group III) bacteria in acid mine drainage biofilms. *Applied and Environmental Microbiology*, *75*(13), 4599-4615.

Gomez-Pereira, P. R., Schüler, M., Fuchs, B. M., Bennke, C., Teeling, H., Waldmann, J., & Amann, R. (2012). Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean.*Environmental microbiology*, *14*(1), 52-66.

Gong, X., Gruniniger, R. J., Forster, R. J., Teather, R. M., & McAllister, T. A. (2013). Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Applied Microbiology and Biotechnology*, *97*(6), 2423-2431.

Gonzalez-Toril, E., Llobet-Brossa, E., Casamayor, E. O., Amann, R., & Amils, R. (2003). Microbial ecology of an extreme acidic environment, the Tinto River. *Applied and Environmental Microbiology*, *69*(8), 4853-4865.

Gordonrid, D. A., Priscu, J., & Giovannoniid, S. (2000). Origin and phylogeny of microbes living in permanent Antarctic lake ice. *Microbial Ecology*, *39*(3), 197-202.

Graham, D. E., Wallenstein, M. D., Vishnivetskaya, T. A., Waldrop, M. P., Phelps, T. J., Pfiffner, S. M., & Jansson, J. K. (2012). Microbes in thawing permafrost: the unknown variable in the climate change equation. *The ISME Journal*, *6*(4), 709-712.

Grant, S., Sorokin, D. Y., Grant, W. D., Jones, B. E., & Heaphy, S. (2004). A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures.*Extremophiles*, *8*(5), 421-429.

Gray, J. P., & Herwig, R. P. (1996). Phylogenetic analysis of the bacterial communities in marine sediments. *Applied and Environmental Microbiology*,*62*(11), 4049-4059.

Grice, E. A., Kong, H. H., Renaud, G., Young, A. C., Bouffard, G. G., Blakesley, R. W., & Segre, J. A. (2008). A diversity profile of the human skin microbiota. *Genome research*, *18*(7), 1043-1050

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer:an environment for comparative protein modeling. *Electrophoresis*, *18*(15), 2714-2723.

Gupta, R. S. (2000). The phylogeny of *Proteobacteria*: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiology Reviews*, *24*(4), 367-402.

Gupta, R., Beg, Q., & Lorenz, P. (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology*, *59*(1), 15-32.

Gupta, R., Gigras, P., Mohapatra, H., Goswami, V. K., & Chauhan, B. (2003). Microbial α-amylases: a biotechnological perspective. *Process Biochemistry*,*38*(11), 1599-1616.

Gutiérrez-Lucas, L. R., Montor-Antonio, J. J., Cortés-López, N. G., & del Moral, S. (2014). Strategies for the Extraction, Purification and Amplification of Metagenomic DNA from Soil Growing Sugarcane. *Advances in Biological Chemistry*, *2014*.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., & Human Microbiome Consortium. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons.*Genome Research*, *21*(3), 494-504.

Hagihara, H., Igarashi, K., Hayashi, Y., Endo, K., Ikawa-Kitayama, K., Ozaki, K., & Ito, S. (2001). Novel α-Amylase That Is Highly Resistant to Chelating Reagents and Chemical Oxidants from the Alkaliphilic *Bacillus* Isolate KSM-K38. *Applied and Environmental Microbiology*, *67*(4), 1744-1750.

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

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., & Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature methods*, *5*(3), 235-237.

Hamilton, L. M., Kelly, C. T., & Fogarty, W. M. (1999). Purification and properties of the raw starch-degrading α-amylase of *Bacillus* sp. IMD 434.*Biotechnology Letters*, *21*(2), 111-115.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews*, *68*(4), 669-685.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry and biology*, *5*(10), R245-R249.

Haq, I., Ali, S., Javed, M. M., Hameed, U., Saleem, A., Adnan, F., & Qadeer, M. A. (2010). Production of alpha amylase from a randomly induced mutant strain of *Bacillus amyloliquefaciens* and its application as a desizer in textile industry. *Pakistan Journal of Botany 42*(1), 473-484.

Hardeman, F., & Sjoling, S. (2007). Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS microbiology ecology*, *59*(2), 524-534.

Healy, F. G., Ray, R. M., Aldrich, H. C., Wilkie, A. C., Ingram, L. O., & Shanmugam, K. T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Applied microbiology and biotechnology*, *43*(4), 667-674.

Heinen, W., & Lauwers, A. M. (1975). Amylase activity and stability at high and low temperature depending on calcium and other divalent cations. *Enzymes and Proteins from Thermophilic Microorganisms Structure and Function* .26, 77-89.

Henne, A., Daniel, R., Schmitz, R. A., & Gottschalk, G. (1999). Construction of Environmental DNA Libraries in *Escherichia coli* and Screening for the Presence of Genes Conferring Utilization of 4-Hydroxybutyrate. *Applied and Environmental Microbiology*, *65*(9), 3901-3907.

Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G., & Daniel, R. (2000). Screening of Environmental DNA Libraries for the Presence of Genes Conferring Lipolytic Activity on *Escherichia coli*. *Applied and environmental microbiology*, *66*(7), 3113-3116.

Hill, R. D., & MacGregor, A. W. (1987). Cereal alpha-amylases in grain research and technology. *Advances in Cereal Science and Technology*, 9, 217-261

Hmidet, N., Ali, N. E. H., Haddar, A., Kanoun, S., Alya, S. K., & Nasri, M. (2009). Alkaline proteases and thermostable α-amylase co-produced by *Bacillus licheniformis* NH1: Characterization and potential application as detergent additive. *Biochemical Engineering Journal*, *47*(1), 71-79.

Hmidet, N., Bayoudh, A., Berrin, J. G., Kanoun, S., Juge, N., & Nasri, M. (2008). Purification and biochemical characterization of a novel α-amylase from *Bacillus licheniformis* NH1: cloning, nucleotide sequence and expression of amyN gene in *Escherichia coli*. *Process Biochemistry*, *43*(5), 499-510.

Hoff, K. J., Tech, M., Lingner, T., Daniel, R., Morgenstern, B., & Meinicke, P. (2008). Gene prediction in metagenomic fragments: a large scale machine learning approach. *BMC bioinformatics*, *9*(1), 217.

Holben, W. E., Jansson, J. K., Chelm, B. K., & Tiedje, J. M. (1988). DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Applied and Environmental Microbiology*, 54(3), 703-711.

Holguin, G., Vazquez, P., & Bashan, Y. (2001). The role of sediment microorganisms in the productivity, conservation, and rehabilitation of mangrove ecosystems: an overview. *Biology and fertility of soils*, *33*(4), 265-278.

Hongxiang, X., Min, W., Xiaogu, W., Junyi, Y., & Chunsheng, W. (2008). Bacterial diversity in deep-sea sediment from northeastern Pacific Ocean. *Acta Ecologica Sinica*, *28*(2), 479-485.

Hu, Y., Liu, Z., Yan, J., Qi, X., Li, J., Zhong, S.,& Liu, Q. (2010). A developed DNA extraction method for different soil samples. *Journal of Basic Microbiology*, 50(4), 401-407.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bi Informatics*, *26*(5), 680-682.

Huber, J. A., Butterfield, D. A., & Baross, J. A. (2002). Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat.*Applied and Environmental Microbiology*, *68*(4), 1585-1594.

Huber, J. A., Cantin, H. V., Huse, S. M., Welch, D. B. M., Sogin, M. L., & Butterfield, D. A. (2010). Isolated communities of *Epsilonproteobacteria* in hydrothermal vent fluids of the Mariana Arc seamounts. *FEMS Microbiology Ecology*, *73*(3), 538-549.

Huber, R., Huber, H., & Stetter, K. O. (2000). Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiology Reviews*, *24*(5), 615-623.

Hug, L. A., Castelle, C. J., Wrighton, K. C., Thomas, B. C., Sharon, I., Frischkorn, K. R., ... & Banfield, J. F. (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome*, *1*(1), 22.

Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998a). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity.*Journal of Bacteriology*, *180*(18), 4765-4774.

Hugenholtz, P., Pitulle, C., Hershberger, K. L., & Pace, N. R. (1998b). Novel division level bacterial diversity in a Yellowstone hot spring. Journal of Bacteriology, 180(2), 366-376.

Humayoun, S. B., Bano, N., & Hollibaugh, J. T. (2003). Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. *Applied and Environmental Microbiology*, *69*(2), 1030-1042.

Hupfer, M., Gloess, S., & Grossart, H. P. (2007). Polyphosphate-accumulating microorganisms in aquatic sediments. *Aquatic Microbial Ecology*, *47*(3), 299.

Hurt, R. A., Qiu, X., Wu, L., Roh, Y., Palumbo, A. V., Tiedje, J. M., & Zhou, J. (2001). Simultaneous recovery of RNA and DNA from soils and sediments.*Applied and Environmental Microbiology*, *67*(10), 4495-4503.

Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N., & Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4.*Genome Research*, *21*(9), 1552-1560.

Hutcheon, G. W., Vasisht, N., & Bolhuis, A. (2005). Characterisation of a highly stable α-amylase from the halophilic archaeon Haloarcula hispanica.*Extremophiles*, *9*(6), 487-495.

Igarashi, K., Hatada, Y., Hagihara, H., Saeki, K., Takaiwa, M., Uemura, T., ... & Ito, S. (1998). Enzymatic properties of a novel liquefying α-amylase from an alkaliphilic *Bacillus* isolate and entire nucleotide and amino acid sequences. *Applied and Environmental Microbiology*, *64*(9), 3282-3289.

Ikeda, S., Tsurumaru, H., Wakai, S., Noritake, C., Fujishiro, K., Akasaka, M., & Ando, K. (2008). Evaluation of the effects of different additives in improving the DNA extraction yield and quality from Andosol. *Microbes and Environments*, *23*(2), 159-166.

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever, M., Lauer, A., & Jørgensen, B. B. (2006). Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(8), 2815-2820.

Inagaki, F., Takai, K., Hirayama, H., Yamato, Y., Nealson, K. H., & Horikoshi, K. (2003). Distribution and phylogenetic diversity of the subsurface microbial community in a Japanese epithermal gold mine. Extremophiles, 7(4), 307-317.

Ish-Horowicz, D., & Burke, J. F. (1981). Rapid and efficient cosmid cloning.*Nucleic Acids Research*, *9*(13), 2989-2898.

Israilides, C., Smith, A., Scanlon, B., & Barnett, C. (1999). Pullulan from agro-industrial wastes. *Biotechnology and Genetic Engineering Reviews*, *16*(1), 309-324.

Jackson, C. R., Harper, J. P., Willoughby, D., Roden, E. E., & Churchill, P. F. (1997). A simple, efficient method for the separation of humic substances and DNA from environmental samples. *Applied and Environmental Microbiology*,*63*(12), 4993-4995.

Jacobsen, C. S. (1995). Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and PCR amplification assay. *Applied and Environmental Microbiology*, *61*(9), 3347-3352.

Jacobsen, C. S., & Rasmussen, O. F. (1992). Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Applied and Environmental Microbiology*, *58*(8), 2458-2462.

Janecek, S. (1992). New conserved amino acid region of alpha-amylases in the third loop of their (beta/alpha) 8-barrel domains. *Biochemical Journal*, *288*(Pt 3), 1069.

Janecek, Š. (1995). Close evolutionary relatedness among functionally distantly related members of the (α/β) 8-barrel glycosyl hydrolases suggested by the similarity of their fifth conserved sequence region. *FEBS Letters*, *377*(1), 6-8.

Janecek, Š. (2002). How many conserved sequence regions are there in the α-amylase family. *Biologia*, *57*(Suppl 11), 29-41.

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*,*72*(3), 1719-1728.

Jeraldo, P., Chia, N., & Goldenfeld, N. (2011). On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys.*Environmental Microbiology*, *13*(11), 3000-3009.

Jespersen, H. M., MacGregor, E. A., Sierks, M. R., & Svensson, B. (1991). Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem. J*, *280*, 51-55.

Jiang, X. T., Peng, X., Deng, G. H., Sheng, H. F., Wang, Y., Zhou, H. W., & Tam, N. F. Y. (2013). Illumina sequencing of 16S rRNA tag revealed spatial variations of bacterial communities in a mangrove wetland. Microbial Ecology,66(1), 96-104.

Jin, P., Pei, X., Du, P., Yin, X., Xiong, X., Wu, H., & Wang, Q. (2012). Overexpression and characterization of a new organic solvent-tolerant esterase derived from soil metagenomic DNA. *Bioresource Technology*, *116*, 234-240.

Johannes, T. W., & Zhao, H. (2006). Directed evolution of enzymes and biosynthetic pathways. *Current opinion in microbiology*, *9*(3), 261-267.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, *8*(3), 275-282.

Jones, R. T., Robeson, M. S., Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *The ISME journal*, *3*(4), 442-453.

Jørgensen, B. B., & Boetius, A. (2007). Feast and famine—microbial life in the deep-sea bed. *Nature Reviews Microbiology*, *5*(10), 770-781.

Kalpana, B. J., Sindhulakshmi, M., & Pandian, S. K. (2014). Amylase enzyme from *Bacillus subtilis* S8-18: A potential desizing agent from the marine environment. *Biotechnology and applied biochemistry*, *61*(2), 134-144.

Kang, S., Denman, S. E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., & McSweeney, C. S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflammatory bowel diseases*, *16*(12), 2034-2042.

Kaster, A. K., Mayer-Blackwell, K., Pasarelli, B., & Spormann, A. M. (2014). Single cell genomic study of Dehalococcoidetes species from deep-sea sediments of the Peruvian Margin. The ISME Journal.

Kato, S., Nakawake, M., Kita, J., Yamanaka, T., Utsumi, M., Okamura, K., & Yamagishi, A. (2013). Characteristics of microbial communities in crustal fluids in a deep-sea hydrothermal field of the Suiyo Seamount. Frontiers in Microbiology, 4.

Kauffmann, I. M., Schmitt, J., & Schmid, R. D. (2004). DNA isolation from soil samples for cloning in different hosts. *Applied Microbiology and Biotechnology*, 64(5), 665-670.

Kawaminami, S., Ozaki, K., Sumitomo, N., Hayashi, Y., Ito, S., Shimada, I., & Arata, Y. (1994). A stable isotope-aided NMR study of the active site of an endoglucanase from a strain of *Bacillus*. *Journal of Biological Chemistry*, *269*(46), 28752-28756.

Kelley, L. A., & Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, *4*(3), 363-371.

Kerkhof, L. J., & Goodman, R. M. (2009). Ocean microbial metagenomics.*Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(19), 1824-1829.

Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P., & Stackebrandt, E. (2006). Introduction to the *Proteobacteria*. *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses*, 3-37

Kim, B. S., Kim, B. K., Lee, J. H., Kim, M., Lim, Y. W., & Chun, J. (2008). Rapid phylogenetic dissection of prokaryotic community structure in tidal flat using pyrosequencing. *The Journal of Microbiology*, *46*(4), 357-363.

Kim, M., Morrison, M., & Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods*, *84*(1), 81-87.

Kim, Y. J., Choi, G. S., Kim, S. B., Yoon, G. S., Kim, Y. S., & Ryu, Y. W. (2006). Screening and characterization of a novel esterase from a metagenomic library. *Protein Expression and Purification*, *45*(2), 315-323.

Kindaichi, T., Yuri, S., Ozaki, N., & Ohashi, A. (2012). Ecophysiological role and function of uncultured Chloroflexi in an anammox reactor. *Water Science & Technology*, *66*(12), 2556-2561.

Kiran, K. K., & Chandra, T. S. (2008). Production of surfactant and detergent-stable, halophilic, and alkalitolerant alpha-amylase by a moderately halophilic *Bacillus* sp. strain TSCVKK. *Applied Microbiology and Biotechnology*, *77*(5), 1023-1031.

Kirchman, D. L. (2002). The ecology of Cytophaga–Flavobacteria in aquatic environments. *FEMS Microbiology Ecology*, *39*(2), 91-100.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.*Nucleic Acids Research*, gks808.

Klotz, M. G., Bryant, D. A., & Hanson, T. E. (2011). The microbial sulfur cycle.*Frontiers in Microbiology*, *2*, 241

Knegtel, R. M., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., & Dijkstra, B. W. (1995). Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 with natural substrates and products. *Journal of Biological Chemistry*,*270*(49), 29256-29264.

Kong, H. H., Oh, J., Deming, C., Conlan, S., Grice, E. A., Beatson, M. A., & Segre, J. A. (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome research*, *22*(5), 850-859.

Kong, Y., Nielsen, J. L., & Nielsen, P. H. (2005). Identity and ecophysiology of uncultured actinobacterial polyphosphate-accumulating organisms in full-scale enhanced biological phosphorus removal plants. *Applied and environmental microbiology*, *71*(7), 4076-4085.

Konishi, Y., Asai, S., & Sawada, Y. (1997). Leaching of marine manganese nodules by acidophilic bacteria growing on elemental sulfur. Metallurgical and Materials Transactions B, 28(1), 25-32.

Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews*, *28*(4), 416-436.

Kottwitz, B., Upadek, H., & Carrer, G. (1994). Application and benefits of enzymes in detergents. *Chimica oggi*, *12*(11-12), 21-24.

Kouridaki, I., Polymenakou, P. N., Tselepides, A., Mandalakis, M., & Smith Jr, K. L. (2010). Phylogenetic diversity of sediment bacteria from the deep Northeastern Pacific Ocean: a comparison with the deep Eastern Mediterranean Sea. *International Microbiology*, *13*(3), 143-150.

Kozdrój, J. (2010). Izolacja kwasów nukleinowych ze środowiska-pierwszy krok w analizie metagenomu. *Kosmos*, *59*(1-2), 141-149.

Krishnan, K. P., Fernandes, C. E., Fernandes, S. O., & Bharathi, P. L. (2006). Tolerance and immobilization of cobalt by some bacteria from ferromanganese crusts of the Afanasiy Nikitin Seamounts. *Geomicrobiology Journal*, *23*(1), 31-36.

Krsek, M.,and Wellington, E. M. H. (1999). Comparison of different methods for the isolation and purification of total community DNA from soil. *Journal of Microbiological Methods*, 39(1), 1-16.

Kumar, S., & Dudley, J. (2007). Bioinformatics software for biologists in the genomics era. *Bioinformatics*, *23*(14), 1713-1717.

Kumar, S., Nei, M., Dudley, J., & Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*, *9*(4), 299-306.

Kumar, S., Tamura, K., & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Computer Applications in the Biosciences: CABIOS*, *10*(2), 189-191.

Kunihiro, T., Takasu, H., Miyazaki, T., Uramoto, Y., Kinoshita, K., Yodnarasri, S., & Tsutsumi, H. (2011). Increase in *Alphaproteobacteria* in association with a polychaete, *Capitella* sp. I, in the organically enriched sediment. *The ISME Journal*, *5*(11), 1818-1831.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews*, *72*(4), 557-578.

Kuriki, T., & Imanaka, T. (1999). The concept of the α-amylase family: structural similarity and common catalytic mechanism. *Journal of Bioscience and Bioengineering*, *87*(5), 557-565.

Kuriki, T., Hondoh, H., & Matsuura, Y. (2005). The conclusive proof that supports the concept of the α-amylase family: structural similarity and common catalytic mechanism. *Biologia Bratislava*, *60*, 13-16.

Kuske, C. R., Banton, K. L., Adorada, D. L., Stark, P. C., Hill, K. K., & Jackson, P. J. (1998). Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil, *Applied and Environmental Microbiology*, *64*(7), 2463-2472.

Laane, C., Boeren, S., Vos, K., & Veeger, C. (1987). Rules for optimization of biocatalysis in organic solvents. *Biotechnology and Bioengineering*, *30*(1), 81-87.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, *227*(5259), 680-685.

Lai, X., Zeng, X., Fang, S., Huang, Y., Cao, L., & Zhou, S. (2006). Denaturing gradient gel electrophoresis (DGGE) analysis of bacterial community composition in deep-sea sediments of the South China Sea. *World Journal of Microbiology and Biotechnology*, *22*(12), 1337-1345.

Lakay, F. M., Botha, A., & Prior, B. A. (2007). Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *Journal of Applied Microbiology*, 102(1), 265-273.

LaMontagne, M. G., Michel, F. C., Holden, P. A., & Reddy, C. A. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *Journal of Microbiological Methods*, *49*(3), 255-264.

Landry, T. D., Chew, L., Davis, J. W., Frawley, N., Foley, H. H., Stelman, S. J., & Hanselman, D. S. (2003). Safety evaluation of an α-amylase enzyme preparation derived from the archaeal order *Thermococcales* as expressed in *Pseudomonas fluorescens* biovar I. *Regulatory Toxicology and Pharmacology*,*37*(1), 149-168.

Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriological Reviews*, *38*(3), 272.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0.*Bioinformatics*, *23*(21), 2947-2948.

Lasken, R. (2009). Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochemical Society Transactions*, *37*(2), 450.

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*,*75*(15), 5111-5120.

Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., & Dijkstra, B. W. (1994). Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. *Journal of Molecular Biology*, *236*(2), 590-600.

Lazarevic, V., Whiteson, K., Gaïa, N., Gizard, Y., Hernandez, D., Farinelli, L., & Schrenzel, J. (2012). Analysis of the salivary microbiome using culture-independent techniques. *J. Clinical Bioinformatics*, *2*(4).

Le, H. C., Truong, N. H., & Schweder, T. (2012). Characterization of the Thermophilic Starch Degrading *Petrotoga* Strain 64G3 and the Expression of its [alpha]-amylase Gene. *Biotechnology*, *11*(4), 199.

Lee, C. C., Kibblewhite-Accinelli, R. E., Wagschal, K., Robertson, G. H., & Wong, D. W. (2006). Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. *Extremophiles*, *10*(4), 295-300.

Lee, S. P., Morikawa, M., Takagi, M., & Imanaka, T. (1994). Cloning of the aapT gene and characterization of its product, alpha-amylase-pullulanase (AapT), from thermophilic and alkaliphilic *Bacillus* sp. strain XAL601. *Applied and Environmental Microbiology*, *60*(10), 3764-3773.

Lee, S. Y., Bollinger, J., Bezdicek, D., & Ogram, A. (1996). Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. *Applied and Environmental Microbiology*, *62*(10), 3787-3793.

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 11070-11075.

Ley, R. E., Harris, J. K., Wilcox, J., Spear, J. R., Miller, S. R., Bebout, B. M., & Pace, N. R. (2006). Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Applied and Environmental Microbiology,72(5), 3685-3695.

Li, C. H., Zhou, H. W., Wong, Y. S., & Tam, N. F. Y. (2009). Vertical distribution and anaerobic biodegradation of polycyclic aromatic hydrocarbons in mangrove sediments in Hong Kong, South China. *Science of the Total Environment*, *407*(21), 5772-5779.

Li, L., Guenzennec, J., Nichols, P., Henry, P., Yanagibayashi, M., & Kato, C. (1999). Microbial diversity in Nankai Trough sediments at a depth of 3,843 m.*Journal of Oceanography*, *55*(5), 635-642.

Li, T., & Wang, P. & Wang, P.X., (2008a). Bacterial and archaeal diversity in surface sediment from the south slope of the South China Sea. *Acta Microbiologica Sinica*, *48*(3), 323-329.

Li, T., Peng, W., & Pinxian, W. (2008b). Microbial diversity in surface sediments of the Xisha Trough, the South China Sea. *Acta Ecologica Sinica*,*28*(3), 1166-1173.

Liao, L., Xu, X. W., Jiang, X. W., Wang, C. S., Zhang, D. S., Ni, J. Y., & Wu, M. (2011). Microbial diversity in deep-sea sediment from the cobalt-rich crust deposit region in the Pacific Ocean. *FEMS Microbiology Ecology*, *78*(3), 565-585.

Liao, L., Xu, X. W., Wang, C. S., Zhang, D. S., & Wu, M. (2009). Bacterial and archaeal communities in the surface sediment from the northern slope of the South China Sea. *Journal of Zhejiang University Science B*, *10*(12), 890-901.

Lim, H. K., Chung, E. J., Kim, J. C., Choi, G. J., Jang, K. S., Chung, Y. R., & Lee, S. W. (2005). Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Applied and Environmental Microbiology*, *71*(12), 7768-7777.

Lim, W. J., Park, S. R., An, C. L., Lee, J. Y., Hong, S. Y., Shin, E. C., & Yun, H. D. (2003). Cloning and characterization of a thermostable intracellular α-amylase gene from the hyperthermophilic bacterium *Thermotoga maritima* MSB8. *Research in microbiology*, *154*(10), 681-687.

Lin, L. L., Chyau, C. C., & Hsu, W. H. (1998). Production and properties of a raw-starch-degrading amylase from the thermophilic and alkaliphilic *Bacillus* sp. TS-23. *Biotechnology and Applied Biochemistry*, *28*(1), 61-68.

Lin, L. L., Lo, H. F., Chen, J. N., Ku, K. L., & Hsu, W. H. (2002). Isolation of a Recombinant *Bacillus* sp. TS-23 α-Amylase by Adsorption-Elution on Raw Starch. *Starch-Stärke*, *54*(8), 338-342.

Lindahl, V., & Bakken, L. R. (1995). Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiology Ecology*, *16*(2), 135-142.

Lipson, D. A., & Schmidt, S. K. (2004). Seasonal changes in an alpine soil bacterial community in the Colorado Rocky Mountains. *Applied and environmental microbiology*, *70*(5), 2867-2879.

Litthauer, D., Abbai, N. S., Piater, L. A., & van Heerden, E. (2010). Pitfalls using tributyrin agar screening to detect lipolytic activity in metagenomic studies. *African Journal of Biotechnology*, *9*(27), 4282-4285.

Loka Bharathi, P. A., Oak, S., & Chandramohan, D. (1991). Sulfate-reducing bacteria from mangrove swamps. II, Their ecology and physiology.*Oceanologica acta*, *14*(2), 163-171.

Lovell, C. R., Friez, M. J., Longshore, J. W., & Bagwell, C. E. (2001). Recovery and Phylogenetic Analysis ofnifH Sequences from Diazotrophic Bacteria Associated with Dead

Aboveground Biomass of *Spartina alterniflora*. *Applied and Environmental Microbiology*, *67*(11), 5308-5314.

Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology,73(5), 1576-1585.

Lozupone, C., & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*,*71*(12), 8228-8235.

Lozupone, C., Hamady, M., & Knight, R. (2006). UniFrac–an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*, *7*(1), 371.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., & Schleifer, K. H. (2004). ARB: a software environment for sequence data.*Nucleic Acids Research*, *32*(4), 1363-1371.

Luo, C., Tsementzi, D., Kyrpides, N. C., & Konstantinidis, K. T. (2012). Individual genome assembly from complex community short-read metagenomic datasets. *The ISME journal*, *6*(4), 898-901.

Lyimo, T. J., Pol, A., Harhangi, H. R., Jetten, M. S., & den Camp, H. J. O. (2009). Anaerobic oxidation of dimethylsulfide and methanethiol in mangrove sediments is dominated by sulfate-reducing bacteria. *FEMS microbiology ecology*, *70*(3), 483-492.

Maalej, H., Hmidet, N., Ghorbel-Bellaaj, O., & Nasri, M. (2013). Purification and biochemical characterization of a detergent stable α-amylase from Pseudomonas stutzeri AS22. *Biotechnology and Bioprocess Engineering*, *18*(5), 878-887.

MacGregor, E. A. (2005). An overview of clan GH-H and distantly related families. *Biologia*, *60*(Suppl 16), 5-12.

Machius, M., Declerck, N., Huber, R., & Wiegand, G. (1998). Activation of *Bacillus licheniformis* α-amylase through a disorder order transition of the substrate-binding site mediated by a calcium–sodium–calcium metal triad. *Structure*, *6*(3), 281-292.

MacNeil, I. A., Tiong, C. L., Minor, C., August, P. R., Grossman, T. H., Loiacono, K. A., & Osburne, M. S. (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. *Journal of Molecular Microbiology and Biotechnology*, *3*(2), 301-308.

Madrid, V. M., Taylor, G. T., Scranton, M. I., & Chistoserdov, A. Y. (2001). Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. *Applied and Environmental Microbiology*, *67*(4), 1663-1674.

Madsen, E. L., Stotzky, G., & Bollag, J. M. (1996). A critical analysis of methods for determining the composition and biogeochemical activities of soil microbial communities in situ. *Soil biochemistry. Volume 9.*, 287-370.

Malmstrom, R. R., Cottrell, M. T., Elifantz, H., & Kirchman, D. L. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Applied and Environmental Microbiology*, *71*(6), 2979-2986.

Mandal, S., Lalremsanga, H. T., & Kumar, N. S. (2015). Bacterial diversity of Murlen National Park located in Indo-Burman Biodiversity hotspot region: A metagenomic approach. *Genomics Data*, *5*, 25-26.

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*, *32*, W327-W331.

Martin, A. P. (2002). Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Applied and Environmental Microbiology*,*68*(8), 3673-3682.

Martin, C., Diaz, N. N., Ontrup, J., & Nattkemper, T. W. (2008). Hyperbolic SOM-based clustering of DNA fragment features for taxonomic visualization and classification. *Bioinformatics*, *24*(14), 1568-1574.

Martinez, A., Kolvek, S. J., Yip, C. L. T., Hopke, J., Brown, K. A., MacNeil, I. A., & Osburne, M. S. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology*, *70*(4), 2452-2463.

Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J. C., Soulas, G., & Catroux, G. (2001). DNA extraction from soils: old bias for new microbial diversity analysis methods. *Applied and environmental microbiology*,*67*(5), 2354-2359.

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404-405.

Mehta, D., & Satyanarayana, T. (2013). Biochemical and molecular characterization of recombinant acidic and thermostable raw-starch hydrolysing α-amylase from an extreme thermophile *Geobacillus thermoleovorans*. *Journal of Molecular Catalysis B: Enzymatic*, *85*, 229-238.

Mendes, L. W., & Tsai, S. M. (2014). Variations of Bacterial Community Structure and Composition in Mangrove Sediment at Different Depths in Southeastern Brazil. Diversity, 6(4), 827-843.

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, *11*(1), 31-46.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., & Edwards, R. A. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*(1), 386.

Meyerdierks, A., Kube, M., Kostadinov, I., Teeling, H., Glöckner, F. O., Reinhardt, R., & Amann, R. (2010). Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environmental Microbiology*, *12*(2), 422-439.

Mijts, B. N., & Patel, B. K. (2002). Cloning, sequencing and expression of an α-amylase gene, amyA, from the thermophilic halophile *Halothermothrix orenii* and purification and biochemical characterization of the recombinant enzyme. *Microbiology*, *148*(8), 2343-2349.

Miller, D. N. (2001). Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. *Journal of Microbiological Methods*, 44(1), 49-58.

Miller, D. N., Bryant, J. E., Madsen, E. L., & Ghiorse, W. C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology*, 65(11), 4715-4724.

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical chemistry*, *31*(3), 426-428.

Milling, A., Gomes N, C, M., Oros-Sicher, M., Gotz, M., & Smalla K. (2005) Nucleic acids extraction from environmental samples. *Molecular microbial ecology*, Osborn AM, Smith CJ (Eds), Taylor & Francis, New York. pp 1–24.

Mirete, S., De Figueras, C. G., & González-Pastor, J. E. (2007). Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. *Applied and Environmental Microbiology*, *73*(19), 6001-6011.

Mojzsis, S. J., Arrhenius, G., McKeegan, K. D., Harrison, T. M., Nutman, A. P., & Friend, C. R. (1996). Evidence for life on Earth before 3,800 million years ago. *Nature*, (384), 55-9.

Mollania, N., Khajeh, K., Hosseinkhani, S., & Dabirmanesh, B. (2010). Purification and characterization of a thermostable phytate resistant α-amylase from *Geobacillus* sp. LH8. *International journal of biological macromolecules*,*46*(1), 27-36.

More, M. I., Herrick, J. B., Silva, M. C., Ghiorse, W. C., & Madsen, E. L. (1994). Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Applied and Environmental Microbiology*, 60(5), 1572-1580.

Moreira, D. (1998). Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Research*, *26*(13), 3309-3310.

Mukherjee, A. K., Borah, M., & Rai, S. K. (2009). To study the influence of different components of fermentable substrates on induction of extracellular α-amylase synthesis by *Bacillus subtilis* DM-03 in solid-state fermentation and exploration of feasibility for inclusion of α-amylase in laundry detergent formulations. *Biochemical Engineering Journal*, *43*(2), 149-156.

Murray, K. J., Webb, S. M., Bargar, J. R., & Tebo, B. M. (2007). Indirect oxidation of Co (II) in the presence of the marine Mn (II)-oxidizing bacterium *Bacillus* sp. strain SG-1. Applied and Environmental Microbiology, 73(21), 6905-6909.

Murray, R. G. E., & Schleifer, K. H. (1994). Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes. *International journal of systematic bacteriology*, *44*(1), 174-176.

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, *247*(4), 536-540.

Muyzer, G., & Stams, A. J. (2008). The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology*, *6*(6), 441-454.

Nair, H. P., Vincent, H., Bhat, S. G. (2014). Evaluation of five in situ lysis protocols for PCR amenable metagenomic DNA from mangrove soils, *Biotechnology Reports* 4: 134–138.

Nakagawa, T., Nakagawa, S., Inagaki, F., Takai, K., & Horikoshi, K. (2004). Phylogenetic diversity of sulfate-reducing prokaryotes in active deep-sea hydrothermal vent chimney structures. *FEMS Microbiology Letters*, *232*(2), 145-152.

Nakamura, A., Haga, K., & Yamane, K. (1993). Three histidine residues in the active center of cyclodextrin glucanotransferase from alkalophilic *Bacillus* sp. 1011: effects of the

replacement on pH dependence and transition-state stabilization. *Biochemistry*, *32*(26), 6624-6631.

Nealson, K. H. (2006). The manganese-oxidizing bacteria. In The Prokaryotes(pp. 222-231). Springer New York.

Nedwell, D. B., Blackburn, T. H., & Wiebe, W. J. (1994). Dynamic nature of the turnover of organic carbon, nitrogen and sulphur in the sediments of a Jamaican mangrove forest. *Marine Ecology Progress Series*, *110*, 223-231.

Nercessian, O., Reysenbach, A. L., Prieur, D., & Jeanthon, C. (2003). Archaeal diversity associated with in situ samplers deployed on hydrothermal vents on the East Pacific Rise (13 N). *Environmental Microbiology*, *5*(6), 492-502.

Nesbo, C. L., Boucher, Y., Dlutek, M., & Doolittle, W. F. (2005). Lateral gene transfer and phylogenetic assignment of environmental fosmid clones.Environmental Microbiology, 7(12), 2011-2026.

Neveu, J., Regeard, C., & DuBow, M. S. (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Applied Microbiology and Biotechnology*, *91*(3), 635-644.

Nielsen, J. E., & Borchert, T. V. (2000). Protein engineering of bacterial α-amylases. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1543*(2), 253-274.

Niemi, R. M., Heiskanen, I., Wallenius, K., & Lindström, K. (2001). Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *Journal of Microbiological Methods*, 45(3), 155-165.

Nirmala, M., & Muralikrishna, G. (2003). Three α-amylases from malted finger millet (Ragi, Eleusine coracana, Indaf-15)-purification and partial characterization. *Phytochemistry*, *62*(1), 21-30.

Niyonzima, F. N., & More, S. S. (2014). Detergent-compatible bacterial amylases. *Applied Biochemistry and Biotechnology*, *174*(4), 1215-1232.

Nold, S. C., Zhou, J., Devol, A. H., & Tiedje, J. M. (2000). Pacific Northwest Marine Sediments Contain Ammonia-Oxidizing Bacteria in the β Subdivision of theProteobacteria. *Applied and Environmental Microbiology*, *66*(10), 4532-4535.

Ochsenreiter, T., Pfeifer, F., & Schleper, C. (2002). Diversity of Archaea in hypersaline environments characterized by molecular-phylogenetic and cultivation studies. *Extremophiles*, *6*(4), 267-274.

Ogram, A., Sayler, G. S., & Barkay, T. (1987). The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*, *7*(2), 57-66.

Ohdan, K., Kuriki, T., Kaneko, H., Shimada, J., Takada, T., Fujimoto, Z., & Okada, S. (1999). Characteristics of two forms of α-amylases and structural implication. *Applied and Environmental Microbiology*, *65*(10), 4652-4658.

Olsen, H. S., & Falholt, P. (1998). The role of enzymes in modern detergency.*Journal of Surfactants and Detergents*, *1*(4), 555-567.

Opel, K. L., Chung, D., & McCord, B. R. (2010). A Study of PCR Inhibition Mechanisms Using Real Time PCR, *Journal of Forensic Sciences*, *55*(1), 25-33.

Orcutt, B. N., Joye, S. B., Kleindienst, S., Knittel, K., Ramette, A., Reitz, A., & Boetius, A. (2010). Impact of natural oil and higher hydrocarbons on microbial diversity, distribution, and activity in Gulf of Mexico cold-seep sediments. *Deep Sea Research Part II: Topical Studies in Oceanography*,*57*(21), 2008-2021.

Orsini, M., & Romano-Spica, V. (2001). A microwave-based method for nucleic acid isolation from environmental samples. *Letters in Applied Microbiology*, 33(1), 17-20.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., & Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, *33*(17), 5691-5702.

Pace, N. R., Stahl, D. A., Lane, D. J., & Olsen, G. J. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM American Society for Microbiology News*, *51*(1), 4-12.

Pace, N. R., Stahl, D. A., Lane, D. J., & Olsen, G. J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. *Advances in Microbial Ecology* (pp. 1-55). Springer US.

Panda, A. K., Bisht, S. S., Kumar, N. S., & De Mandal, S. (2015). Investigations on microbial diversity of Jakrem hot spring, Meghalaya, India using cultivation-independent approach. *Genomics Data*, *4*, 156-157.

Pandey, A., Nigam, P., Soccol, C. R., Soccol, V. T., Singh, D., & Mohan, R. (2000). Advances in microbial amylases. *Biotechnology and Applied Biochemistry*, (31 (Pt 2)), 135-52.

Pandey, S., & Singh, S. P. (2012). Organic solvent tolerance of an α-amylase from haloalkaliphilic bacteria as a function of pH, temperature, and salt concentrations. *Applied Biochemistry and Biotechnology*, *166*(7), 1747-1757.

Parag, B., Sasikala, C., & Ramana, C. V. (2013). Molecular and culture dependent characterization of endolithic bacteria in two beach sand samples and description of *Rhizobiumendolithicum* sp. nov. *Antonie van Leeuwenhoek*,*104*(6), 1235-1244.

Park, H. S., Lin, S., & Voordouw, G. (2008). Ferric iron reduction by Desulfovibrio vulgaris Hildenborough wild type and energy metabolism mutants.*Antonie van Leeuwenhoek*, *93*(1-2), 79-85.

Parsley, L. C., Consuegra, E. J., Kakirde, K. S., Land, A. M., Harper, W. F., & Liles, M. R. (2010). Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Applied and Environmental Microbiology*, *76*(11), 3753-3757.

Paulsen, I. T., Sliwinski, M. K., & Saier, M. H. (1998). Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. Journal of Molecular Biology, 277(3), 573-592.

Pel, J., Broemeling, D., Mai, L., Poon, H. L., Tropini, G., Warren, R. L., & Marziali, A. (2009). Nonlinear electrophoretic response yields a unique parameter for separation of biomolecules. *Proceedings of the National Academy of Sciences*, *106*(35), 14796-14801.

Peng, Q., Wang, X., Shang, M., Huang, J., Guan, G., Li, Y., & Shi, B. (2014). Isolation of a novel alkaline-stable lipase from a metagenomic library and its specific application for milkfat flavor production. *Microb Cell Fact*, *13*(1).

Penninga, D., van der Veen, B. A., Knegtel, R. M., van Hijum, S. A., Rozeboom, H. J., Kalk, K. H., & Dijkhuizen, L. (1996). The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. *Journal of Biological Chemistry*, *271*(51), 32777-32784.

Peršoh, D., Theuerl, S., Buscot, F., & Rambold, G. (2008). Towards a universally adaptable method for quantitative extraction of high-purity nucleic acids from soil. *Journal of Microbiological Methods*, *75*(1), 19-24.

Pester, M., Knorr, K. H., Friedrich, M. W., Wagner, M., & Loy, A. (2012). Sulfate-reducing microorganisms in wetlands–fameless actors in carbon cycling and climate change. *Frontiers in microbiology*, *72*(3), 1-19

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*,*8*(10), 785-786.

Pflughoeft, K. J., & Versalovic, J. (2012). Human microbiome in health and disease. *Annual Review of Pathology: Mechanisms of Disease*, *7*, 99-122.

Phillips, S. J. M., & Parnell, J. (2006). The detection of organic matter in terrestrial snow and ice: implications for astrobiology. *International Journal of Astrobiology*, *5*(04), 353-359.

Picard, C., Ponsonnet, C., Paget, E., Nesme, X., & Simonet, P. (1992). Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Applied and Environmental Microbiology*, 58(9), 2717-2722.

Pillai, S. D., Josephson, K. L., Bailey, R. L., Gerba, C. P., & Pepper, I. L. (1991). Rapid method for processing soil samples for polymerase chain reaction amplification of specific gene sequences. *Applied and Environmental Microbiology*, *57*(8), 2283-2286.

Pollastri, G., Przybylski, D., Rost, B., & Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins: Structure, Function, and Bioinformatics*, *47*(2), 228-235.

Polymenakou, P. N., Bertilsson, S., Tselepides, A., & Stephanou, E. G. (2006). Bacterial community composition in different sediments from the Eastern Mediterranean Sea: a comparison of four 16S ribosomal DNA clone libraries. Microbial Ecology, 50(3), 447-462.

Polymenakou, P. N., Christakis, C. A., Mandalakis, M., & Oulas, A. (2015). Pyrosequencing analysis of microbial communities reveals dominant cosmopolitan phylotypes in deep-sea sediments of the eastern Mediterranean Sea. *Research in Microbiology*, 166, 448-457

Polymenakou, P. N., Lampadariou, N., Mandalakis, M., & Tselepides, A. (2009). Phylogenetic diversity of sediment bacteria from the southern Cretan margin, Eastern Mediterranean Sea. Systematic and Applied Microbiology,32(1), 17-26.

Pons, J., Planas, A., & Querol, E. (1995). Contribution of a disulfide bridge to the stability of 1, 3-1, 4-PD-glucan 4-glucanohydrolase from *Bacillus licheniformis. Protein engineering*, *8*(9), 939-945.

Porteous, L. A., Seidler, R. J., & Watrud, L. S. (1997). An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. *Molecular Ecology*, *6*(8), 787-791.

Porteous, M. L. A., & Armstrong, J. L. (1991). Recovery of bulk DNA from soil by a rapid, small-scale extraction method. *Current Microbiology*, *22*(6), 345-348.

Prakash, B., Vidyasagar, M., Madhukumar, M. S., Muralikrishna, G., & Sreeramulu, K. (2009). Production, purification, and characterization of two extremely halotolerant, thermostable, and alkali-stable α-amylases from Chromohalobacter sp. TVSP 101. *Process Biochemistry*, *44*(2), 210-215.

Priscu, J. C., Adams, E. E., Lyons, W. B., Voytek, M. A., Mogk, D. W., Brown, R. L., & Avci, R. (1999). Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science*, *286*(5447), 2141-2144.

Pronk, M. I., & Leclercq, C. (2004) Alpha-Amylase from *Bacillus licheniformis* containing a genetically engineered alpha-amylase gene from *B. licheniformis.*Safety evaluation of certain food additives and contaminants, *WHO food additives series:* 52.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.*Nucleic Acids Research*, *35*(21), 7188-7196.

Purdy, K. J., Embley, T. M., Takii, S., & Nedwell, D. B. (1996). Rapid Extraction of DNA and rRNA from Sediments by a Novel Hydroxyapatite Spin-Column Method. *Applied and Environmental Microbiology*, *62*(10), 3905-3907.

Pureza, L. M., Graças, D. A., Gonçalves, E. C., Correa, J. A., Silva, A., & Schneider, M. P. (2012). Bacterial Diversity in an Amazonian Mangrove Ecosystem. *Aquatic Science and Technology*, *1*(1), 66-85.

Purohit, M. K., & Singh, S. P. (2009). Assessment of various methods for extraction of metagenomic DNA from saline habitats of coastal Gujarat (India) to explore molecular diversity. *Letters in Applied Microbiology*, *49*(3), 338-344.

Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H., Eck, J., & Schleper, C. (2008). Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics.*Molecular Microbiology*, *50*(2), 563-575.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research Online*,*41*(D1), D590-D596.

Quince, C., Lanzen, A., Davenport, R. J., & Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, *12*(1), 38.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., & Banfield, J. F. (2005). Community proteomics of a natural microbial biofilm. *Science*, *308*(5730), 1915-1920.

Ramsay, A. J. (1984). Extraction of bacteria from soil: efficiency of shaking or ultrasonication as indicated by direct counts and autoradiography. *Soil Biology and Biochemistry*, *16*(5), 475-481.

Ranjan, R., Grover, A., Kapardar, R. K., & Sharma, R. (2005). Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochemical and Biophysical Research Communications*, *335*(1), 57-65.

Rao, J. U. M., & Satyanarayana, T. (2008). Biophysical and biochemical characterization of a hyperthermostable and Ca2+-independent α-amylase of an extreme thermophile *Geobacillus thermoleovorans. Applied Biochemistry and Biotechnology*, *150*(2), 205-219.

Rappé, M. S., & Giovannoni, S. J. (2003). The uncultured microbial majority.*Annual Reviews in Microbiology*, *57*(1), 369-394.

Rappe, M. S., Kemp, P. F., & Giovannoni, S. J. (1997). Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnology and Oceanography*, *42*(5), 811-826.

Ratanakhanokchai, K., Kaneko, J., Kamio, Y., & Izaki, K. (1992). Purification and properties of a maltotetraose-and maltotriose-producing amylase from *Chloroflexus aurantiacus. Applied and Environmental Microbiology*, *58*(8), 2490-2494.

Ravenschlag, K., Sahm, K., & Amann, R. (2001). Quantitative molecular analysis of the microbial community in marine Arctic sediments (Svalbard).*Applied and Environmental Microbiology*, *67*(1), 387-395.

Ravenschlag, K., Sahm, K., Pernthaler, J., & Amann, R. (1999). High bacterial diversity in permanently cold marine sediments. *Applied and Environmental Microbiology*, *65*(9), 3982-3989.

Reeder, J., & Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature Methods*, *7*(9), 668-669.

Rees, H. C., Grant, S., Jones, B., Grant, W. D., & Heaphy, S. (2003). Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles*, *7*(5), 415-421.

Rees, H. C., Grant, W. D., Jones, B. E., & Heaphy, S. (2004). Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles*, *8*(1), 63-71.

Richardson, T. H., Tan, X., Frey, G., Callen, W., Cabell, M., Lam, D., & Miller, C. (2002). A novel, high performance enzyme for starch liquefaction discovery and optimization of a low pH, thermostable α-amylase. *Journal of Biological Chemistry*, *277*(29), 26501-26507.

Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics,38*, 525-552.

Robe, P., Nalin, R., Capellano, C., Vogel, T. M., & Simonet, P. (2003). Extraction of DNA from soil. *European Journal of Soil Biology*, 39(4), 183-190.

Robertson, D. E., Chaplin, J. A., DeSantis, G., Podar, M., Madden, M., Chi, E., & Short, J. M. (2004). Exploring nitrilase sequence space for enantioselective catalysis. *Applied and Environmental Microbiology*, *70*(4), 2429-2436.

Rochelle, P. A., Fry, J. C., Parkes, R. J., & Weightman, A. J. (1992). DNA extraction for 16S rRNA gene analysis to determine genetic diversity in deep sediment communities. *FEMS microbiology letters*, *100*(1-3), 59-65.

Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., & Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, *1*(4), 283-290.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., & Goodman, R. M. (2000). Cloning the soil metagenome: a strategy for accessing the genetic

and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, *66*(6), 2541-2547.

Rondon, M. R., Raffel, S. J., Goodman, R. M., & Handelsman, J. (1999). Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus. Proceedings of the National Academy of Sciences*, *96*(11), 6451-6455.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574.

Roohi, R., Kuddus, M., & Saima, S. (2013). Cold-active detergent-stable extracellular α-amylase from *Bacillus cereus* GA6: Biochemical characteristics and its perspectives in laundry detergent formulation. *Journal of Biochemical Technology*, *4*(4), 636-644.

Roose-Amsaleg, C. L., Garnier-Sillam, E., & Harry, M. (2001). Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology*, *18*(1), 47-60.

Rosen, M. J., Callahan, B. J., Fisher, D. S., & Holmes, S. P. (2012). Denoising PCR-amplified metagenome data. *BMC Bioinformatics*, *13*(1), 283.

Roy, J. K., & Mukherjee, A. K. (2013). Applications of a high maltose forming, thermo-stable α-amylase from an extremely alkalophilic *Bacillus licheniformis* strain AS08E in food and laundry detergent industries. *Biochemical Engineering Journal*, *77*, 220-230.

Roy, J. K., Borah, A., Mahanta, C. L., & Mukherjee, A. K. (2013). Cloning and overexpression of raw starch digesting α-amylase gene from *Bacillus subtilis* strain AS01a in *Escherichia coli* and application of the purified recombinant α-amylase (AmyBS-I) in raw starch digestion and baking industry. *Journal of Molecular Catalysis B: Enzymatic*, *97*, 118-129.

Roy, J. K., Rai, S. K., & Mukherjee, A. K. (2012). Characterization and application of a detergent-stable alkaline α-amylase from *Bacillus subtilis* strain AS-S01a. *International Journal of Biological Macromolecules*, *50*(1), 219-229.

Ruff, S. E., Arnds, J., Knittel, K., Amann, R., Wegener, G., Ramette, A., & Boetius, A. (2013). Microbial communities of deep-sea methane seeps at Hikurangi continental margin (New Zealand). *PloS one*, *8*(9), e72627.

Rumbak, E., Rawlings, D. E., Lindsey, G. G., & Woods, D. R. (1991). Cloning, nucleotide sequence, and enzymatic characterization of an alpha-amylase from the ruminal bacterium *Butyrivibrio fibrisolvens* H17c. *Journal of Bacteriology*,*173*(13), 4203-4211.

Saano, A., Tas, E., Pippola, S., Lidstorm, K., Van Elsas, J, D. (1995) Extraction and analysis of microbial DNA from soil, *In Nucleic Acids in the Environmenatal: Methods and Applications*, van Elsas JD, Trevors JT (Eds), Springer-Verlag, Heidelberg, Germany. pp 49–67.

Saboury, A. A. (2002). Stability, activity and binding properties study of αamylase upon interaction with Ca2+ and Co2+. *Biologia*, *57*, 221-228.

Sahoo, K., & Dhal, N. K. (2009). Potential microbial diversity in mangrove ecosystems: a review. Indian Journal of Marine Sciences, 38(2), 249.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406-425.

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, *234*(3), 779-815.

Sambrook, J., Fritish, E., and Maniatis, I (2000): Molecular Cloning, a Laboratory manual. In 3 (Ed.), C.S.H. Laboratory, Ed.

Samie, N., Noghabi, K. A., Gharegozloo, Z., Zahiri, H. S., Ahmadian, G., Sharafi, H., & Vali, H. (2012). Psychrophilic α-amylase from *Aeromonas veronii* NS07 isolated from farm soils. Process Biochemistry, 47(9), 1381-1387.

Samuel, B. S., & Gordon, J. I. (2006). A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proceedings of the National Academy of Sciences*, *103*(26), 10011-10016.

Sanchez, O. J., & Cardona, C. A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology*, *99*(13), 5270-5295.

Saxena, R. K., Dutt, K., Agarwal, L., & Nayyar, P. (2007). A highly thermostable and alkaline amylase from a *Bacillus* sp. PN5. *Bioresource Technology*, *98*(2), 260-265.

Schloss, P. D. (2008). Evaluating different approaches that test whether microbial communities have the same structure. *The ISME Journal*, *2*(3), 265-275.

Schloss, P. D., & Handelsman, J. (2003). Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, *14*(3), 303-310.

Schloss, P. D., & Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*, *71*(3), 1501-1506.

Schloss, P. D., & Handelsman, J. (2006a). Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Applied and Environmental Microbiology*, *72*(10), 6773-6779.

Schloss, P. D., & Handelsman, J. (2006b). Introducing TreeClimber, a test to compare microbial community structures. *Applied and Environmental Microbiology*, *72*(4), 2379-2384.

Schloss, P. D., Larget, B. R., & Handelsman, J. (2004). Integration of microbial ecology and statistics: a test to compare gene libraries. *Applied and Environmental Microbiology*, *70*(9), 5485-5492.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537-7541.

Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., & Witholt, B. (2001). Industrial biocatalysis today and tomorrow. *Nature*, *409*(6817), 258-268.

Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology*, *173*(14), 4371-4378.

Schrenk, M. O., Kelley, D. S., Delaney, J. R., & Baross, J. A. (2003). Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Applied and Environmental Microbiology*, *69*(6), 3580-3592.

Schulz, H. N., Brinkhoff, T., Ferdelman, T. G., Mariné, M. H., Teske, A., & Jørgensen, B. B. (1999). Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science*, *284*(5413), 493-495.

Schuster,S.C.(2007). Next-generation sequencing transforms today's biology. *Nature*, *200*(8), 16-18.

Schwab, K., Bader, J., Brokamp, C., Popović, M. K., Bajpai, R., & Berovič, M. (2009). Dual feeding strategy for the production of α-amylase by *Bacillus caldolyticus* using complex media. *New biotechnology*, *26*(1), 68-74.

Schwermann, B., Pfau, K., Liliensiek, B., Schleyer, M., Fischer, T., & Bakker, E. P. (1994). Purification, Properties and Structural Aspects of a Thermoacidophilic α-Amylase from *Alicyclobacillus ccidocaldarius* ATCC 27009. *European Journal of Biochemistry*, *226*(3), 981-991.

Sekiguchi, Y., Yamada, T., Hanada, S., Ohashi, A., Harada, H., & Kamagata, Y. (2003). *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. *International journal of systematic and evolutionary microbiology*, *53*(6), 1843-1851.

Shafiei, M., Ziaee, A. A., & Amoozegar, M. A. (2010). Purification and biochemical characterization of a novel SDS and surfactant stable, raw starch digesting, and halophilic α-amylase from a moderately halophilic bacterium, Nesterenkonia sp. strain F. *Process Biochemistry*, *45*(5), 694-699.

Shafiei, M., Ziaee, A. A., & Amoozegar, M. A. (2011). Purification and characterization of an organic-solvent-tolerant halophilic α-amylase from the moderately halophilic *Nesterenkonia* sp. strain. *Journal of Industrial Microbiology and Biotechnology*, *38*(2), 275-281.

Shah, N., Tang, H., Doak, T. G., & Ye, Y. (2011, January). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. In *Pacific Symposium on Biocomputing* (Vol. 16, pp. 165-176).

Sharma, P. K., Capalash, N., & Kaur, J. (2007). An improved method for single step purification of metagenomic DNA. Molecular biotechnology, 36(1), 61-63.

Sharma, R., Ranjan, R., Kapardar, R. K., & Grover, A. (2005). Unculturable'bacterial diversity: an untapped resource. *Current Science*, *89*(1), 72-77.

Sharma, S., Khan, F. G., & Qazi, G. N. (2010). Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. *Applied Microbiology and Biotechnology*, *86*(6), 1821-1828.

Sherman, R. E., Fahey, T. J., & Howarth, R. W. (1998). Soil-plant interactions in a neotropical mangrove forest: iron, phosphorus and sulfur dynamics.*Oecologia*, *115*(4), 553-563.

Shivaji, S., Bhanu, N. V., & Aggarwal, R. K. (2000). Identification of Yersinia pestis as the causative organism of plague in India as determined by 16S rDNA sequencing and RAPD-based genomic fingerprinting. *FEMS Microbiology Letters*, *189*(2), 247-252.

Siddhapura, P. K., Vanparia, S., Purohit, M. K., & Singh, S. P. (2010). Comparative studies on the extraction of metagenomic DNA from the saline habitats of Coastal Gujarat and Sambhar

Lake, Rajasthan (India) in prospect of molecular diversity and search for novel biocatalysts. *International Journal of Biological Macromolecules*, *47*(3), 375-379.

Sievert, S. M. Kiene r, Schulz H (2007) The sulfur cycle. *Oceanography*, *20*, 117-123.

Simonet, P., Capellano, A., Navarro, E., Bardin, R., & Moiroud, A. (1984). An improved method for lysis of Frankia with acharomopeptidase allows detection of new plasmids. *Canadian journal of microbiology*, *30*(10), 1292-1295.

Singh, A., Chauhan, N. S., Thulasiram, H. V., Taneja, V., & Sharma, R. (2010). Identification of two flavin monooxygenases from an effluent treatment plant sludge metagenomic library. *Bioresource Technology*, *101*(21), 8481-8484.

Sivakumar, N., Li, N., Tang, J. W., Patel, B. K., & Swaminathan, K. (2006). Crystal structure of AmyA lacks acidic surface and provide insights into protein stability at poly-extreme condition. *FEBS Letters*, *580*(11), 2646-2652.

Skerman, V. B. (1969). Techniques in pure and applied microbiology, *Abstracts of Microbiological Methods.* Wiley- Interscience, New York, 883.

Sleator, R. D., Shortall, C., & Hill, C. (2008). Metagenomics. *Letters in applied microbiology*, *47*(5), 361-366.

Sodhi, H. K., Sharma, K., Gupta, J. K., & Soni, S. K. (2005). Production of a thermostable α-amylase from *Bacillus* sp. PS-7 by solid state fermentation and its synergistic use in the hydrolysis of malt starch for alcohol production.*Process Biochemistry*, *40*(2), 525-534.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, *103*(32), 12115-12120.

Solbak, A. I., Richardson, T. H., McCann, R. T., Kline, K. A., Bartnek, F., Tomlinson, G., & Kerovuo, J. (2005). Discovery of pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton fabric. *Journal of Biological Chemistry*, *280*(10), 9431-9438.

Song, J. S., Jeon, J. H., Lee, J. H., Jeong, S. H., Jeong, B. C., Kim, S. J., & Lee, S. H. (2005). Molecular characterization of TEM-type beta-lactamases identified in cold-seep sediments of Edison Seamount (south of Lihir Island, Papua New Guinea). *J Microbiol*, *43*(2), 172-8.

Spain, A. M., Krumholz, L. R., & Elshahed, M. S. (2009). Abundance, composition, diversity and novelty of soil *Proteobacteria*. *The ISME Journal*,*3*(8), 992-1000.

Srimathi, S., Jayaraman, G., Feller, G., Danielsson, B., & Narayanan, P. R. (2007). Intrinsic halotolerance of the psychrophilic α-amylase from *Pseudoalteromonas haloplanktis*. *Extremophiles*, *11*(3), 505-515.

Stahl, D. A., Lane, D. J., Olsen, G. J., & Pace, N. R. (1985). Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Applied and Environmental Microbiology*, *49*(6), 1379-1384.

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology*, 39(1), 321-346.

Steele, H. L., & Streit, W. R. (2005). Metagenomics: advances in ecology and biotechnology. *FEMS Microbiology Letters*, *247*(2), 105-111.

Steffan, R. J., Goksoyr, J., Bej, A. K., & Atlas, R. M. (1988). Recovery of DNA from soils and sediments. *Applied and Environmental Microbiology*, 54(12), 2908-2915.

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology*, *178*(3), 591-599.

Stevenson, F. J. (1976). Stability constants of Cu2+, Pb2+, and Cd2+ complexes with humic acids. *Soil Science Society of America Journal*, 40(5), 665-672.

Strokopytov, B., Knegtel, R. M., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., & Dijkstra, B. W. (1996). Structure of cyclodextrin glycosyltransferase complexed with a maltononaose inhibitor at 2.6 Å resolution. Implications for product specificity. *Biochemistry*, *35*(13), 4241-4249.

Suh, S. S., Park, M., Hwang, J., Kil, E. J., Jung, S. W., Lee, S., & Lee, T. K. (2015). Seasonal Dynamics of Marine Microbial Community in the South Sea of Korea. *PloS One*, *10*(6), e0131633.

Suzuki, Y., Inagaki, F., Takai, K., Nealson, K. H., & Horikoshi, K. (2004). Microbial diversity in inactive chimney structures from deep-sea hydrothermal systems. Microbial Ecology, 47(2), 186-196.

Svensson, B., & Sogaard, M. (1991). Protein engineering of amylases. *Biochemical Society Transactions*, *20*(1), 34.

Sylvan, J. B., Toner, B. M., & Edwards, K. J. (2012). Life and death of deep-sea vents: bacterial diversity and ecosystem succession on inactive hydrothermal sulfides. MBio, 3(1), e00279-11.

Tajima, F., & Nei, M. (1982). Biases of the estimates of DNA divergence obtained by the restriction enzyme technique. *Journal of Molecular Evolution*,*18*(2), 115-120.

Tajima, F., & Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*, *1*(3), 269-285.

Takada-Hoshino, Y., & Matsumoto, N. (2004). An improved DNA extraction method using skim milk from soils that strongly adsorb DNA. *Microbes and Environments*, *19*(1), 13-19.

Takaha, T., & Smith, S. M. (1999). The functions of 4-α-glucanotransferases and their use for the production of cyclic glucans. *Biotechnology and Genetic Engineering Reviews*, *16*(1), 257-280.

Takai, K., & Horikoshi, K. (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics*, *152*(4), 1285-1297.

Takai, K., Gamo, T., Tsunogai, U., Nakayama, N., Hirayama, H., Nealson, K. H., & Horikoshi, K. (2004). Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field.*Extremophiles*, *8*(4), 269-282.

Takata, H., Kuriki, T., Okada, S., Takesada, Y., Iizuka, M., Minamiura, N., & Imanaka, T. (1992). Action of neopullulanase. Neopullulanase catalyzes both hydrolysis and transglycosylation at alpha-(1----4)-and alpha-(1----6)-glucosidic linkages. *Journal of Biological Chemistry*, *267*(26), 18447-18452.

Taketani, R. G., Franco, N. O., Rosado, A. S., & van Elsas, J. D. (2010). Microbial community response to a simulated hydrocarbon spill in mangrove sediments. *The Journal of Microbiology*, *48*(1), 7-15.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*(8), 1596-1599.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731-2739.

Tao, X., Jang, M. S., Kim, K. S., Yu, Z., & Lee, Y. C. (2008). Molecular cloning, expression and characterization of alpha-amylase gene from a marine bacterium *Pseudoalteromonas* sp. MY-1. *Indian J Biochem Biophys*, *45*(5), 305-309.

Tebbe, C. C., & Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and Environmental Microbiology*, 59(8), 2657-2665.

Tebo, B. M., Ghiorse, W. C., van Waasbergen, L. G., Siering, P. L., & Caspi, R. (1997). Bacterially mediated mineral formation; insights into manganese (II) oxidation from molecular genetic and biochemical studies. *Reviews in Mineralogy and Geochemistry*, *35*(1), 225-266.

Templeton, A. S., Staudigel, H., & Tebo, B. M. (2005). Diverse Mn (II)-oxidizing bacteria isolated from submarine basalts at Loihi Seamount. *Geomicrobiology Journal*, *22*(3-4), 127-139.

Tester, R. F.,& Karkalas, J. (2006). Hydrolysis of native starches with amylases. *Animal Feed Science and Technology*, *130*(1), 39-54.

Thomas, D. N., & Dieckmann, G. S. (2002). Antarctic sea ice--a habitat for extremophiles. *Science*, 295(5555), 641-644.

Tien, A. C., Backus, S., Kapteyn, H., Murnane, M., & Mourou, G. (1999). Short-pulse laser damage in transparent materials as a function of pulse duration. *Physical Review Letters*, *82*(19), 3883.

Torsvik, V., Goksøyr, J., & Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology*, *56*(3), 782-787.

Trevors, J. T., Lee, H., & Cook, S. (1992). Direct extraction of DNA from soil.*Microb. Releases*, *1*(2), 111-115.

Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., & Rubin, E. M. (2005). Comparative metagenomics of microbial communities. Science, 308(5721), 554-557.

Tsai, Y. L., & Olson, B. H. (1991). Rapid method for direct extraction of DNA from soil and sediments. *Applied and Environmental Microbiology*, *57*(4), 1070-1074.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027-131.

Turpin, P. E., Maycroft, K. A., Rowlands, C. L., & Wellington, E. M. H. (1993). An ion-exchange based extraction method for the detection of salmonellas in soil. *Journal of applied bacteriology*, *74*(2), 181-190.

Tyson, G. W., & Banfield, J. F. (2005). Cultivating the uncultivated: a community genomics perspective. *Trends in microbiology*, *13*(9), 411-415.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., & Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, *428*(6978), 37-43.

Uchiyama, T., & Miyazaki, K. (2010). Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Applied and Environmental Microbiology*, *76*(21), 7029-7035.

Uchiyama, T., Abe, T., Ikemura, T., & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, *23*(1), 88-93.

Uitdehaag, J. C., Mosi, R., Kalk, K. H., Van Der Veen, B. A., Dijkhuizen, L., Withers, S. G., & Dijkstra, B. W. (1999). X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nature Structural & Molecular Biology*, *6*(5), 432-436.

Urakawa, H., Kita-Tsukamoto, K., & Ohwada, K. (1999). Microbial diversity in marine sediments from Sagami Bay and Tokyo Bay, Japan, as determined by 16S rRNA gene analysis. *Microbiology*, *145*(11), 3305-3315.

Van Der Maarel, M. J., Van Der Veen, B., Uitdehaag, J. C., Leemhuis, H., & Dijkhuizen, L. (2002). Properties and applications of starch-converting enzymes of the α-amylase family. *Journal of Biotechnology*, *94*(2), 137-155.

Van Der Maarel, M. J., Van Der Veen, B., Uitdehaag, J. C., Leemhuis, H., & Dijkhuizen, L. (2002). Properties and applications of starch-converting enzymes of the α-amylase family. *Journal of biotechnology*, *94*(2), 137-155.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., & Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, *304*(5667), 66-74.

Vester, J. K., Glaring, M. A., & Stougaard, P. (2015). An exceptionally cold-adapted alpha-amylase from a metagenomic library of a cold and alkaline environment. *Applied Microbiology and Biotechnology*, *99*(2), 717-727.

Vidya, J., Swaroop, S., Singh, S., Alex, D., Sukumaran, R., & Pandey, A. (2011). Isolation and characterization of a novel α-amylase from a metagenomic library of Western Ghats of Kerala, India. *Biologia*, *66*(6), 939-944.

Vieille, C., & Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews*, *65*(1), 1-43.

Vigal, T., Gil, J. A., Daza, A., Garcia-González, M. D., & Martín, J. F. (1991). Cloning, characterization and expression of an α-amylase gene from Streptomyces griseus IMRU3570. *Molecular and General Genetics*, *225*(2), 278-288.

Vihinen, M., & Mantsiila, P. (1989). Microbial amylolytic enzyme. *Critical Reviews in biochemistry and Molecular Biology*, *24*(4), 329-418.

Vincent, S, S., &John, S, B. (2009): Buffers: principles and practice. In Richard RB, Murray PD (Eds): *Methods in Enzymology*. Elsevier Science & Technology Books, Massachusetts, pp. 43–56.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K. E., & Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology*, *69*(10), 6235-6242.

Voget, S., Steele, H. L., & Streit, W. R. (2006). Characterization of a metagenome-derived halotolerant cellulase. *Journal of Biotechnology*, *126*(1), 26-36.

Volossiouk, T., Robb, E. J., & Nazar, R. N. (1995). Direct DNA extraction for PCR-mediated assays of soil organisms. *Applied and Environmental Microbiology*, 61(11), 3972-3976.

Wang, G. Y. S., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., & Davies, J. (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Organic letters*, *2*(16), 2401-2404.

Wang, H., Li, X., Ma, Y., & Song, J. (2014). Characterization and high-level expression of a metagenome-derived alkaline pectate lyase in recombinant *Escherichia coli*. *Process Biochemistry*, *49*(1), 69-76.

Wang, N., Zhang, Y., Wang, Q., Liu, J., Wang, H., Xue, Y., & Ma, Y. (2006). Gene cloning and characterization of a novel α-amylase from alkaliphilic *Alkalimonas amylolytica*. *Biotechnology Journal*, *1*(11), 1258-1265.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261-5267.

Wang, X. H., Schloßmacher, U., Natalio, F., Schröder, H. C., Wolf, S. E., Tremel, W., & Müller, W. E. (2009). Evidence for biogenic processes during formation of ferromanganese crusts from the Pacific Ocean: implications of biologically induced mineralization. Micron, 40(5), 526-535.

Wang, Y., Sheng, H. F., He, Y., Wu, J. Y., Jiang, Y. X., Tam, N. F. Y., & Zhou, H. W. (2012). Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Applied and Environmental Microbiology*, *78*(23), 8264-8271.

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., & Kuske, C. R. (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Applied and Environmental Microbiology, 75(7), 2046-2056.

Wasmund, K., Schreiber, L., Lloyd, K. G., Petersen, D. G., Schramm, A., Stepanauskas, R., & Adrian, L. (2014). Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia, phylum Chloroflexi. The ISME Journal, 8(2), 383-397.

Weber, M., Teeling, H., Huang, S., Waldmann, J., Kassabgy, M., Fuchs, B. M., & Glöckner, F. O. (2011). Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *The ISME journal*, *5*(5), 918-928.

Wegener, G., Shovitri, M., Knittel, K., Niemann, H., Hovland, M., & Boetius, A. (2008). Biogeochemical processes and microbial diversity of the Gullfaks and Tommeliten methane seeps (Northern North Sea). *Biogeosciences Discussions*, *5*(1), 971-1015.

Weiss, R. F., Lonsdale, P., Lupton, J. E., Bainbridge, A. E., & Craig, H. (1977). Hydrothermal plumes in the Galapagos Rift.*Nature* 267,600-603.

Whang, K., & Hattori, T. (1988). Oligotrophic bacteria from rendzina forest soil. *Antonie van Leeuwenhoek*, *54*(1), 19-36.

Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., & Wagner, L. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Research*, *31*(1), 28-33.

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, *95*(12), 6578-6583

Widdel, F., Knittel, K., & Galushko, A. (2010). Anaerobic hydrocarbon-degrading microorganisms: an overview. In *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 1997-2021). Springer Berlin Heidelberg.

Williamson, K. E., Kan, J., Polson, S. W., & Williamson, S. J. (2011). Optimizing the indirect extraction of prokaryotic DNA from soils. *Soil Biology and Biochemistry*, 43(4), 736-748.

Wilms, R., Köpke, B., Sass, H., Chang, T. S., Cypionka, H., & Engelen, B. (2006). Deep biosphere-related bacteria within the subsurface of tidal flat sediments. Environmental Microbiology, 8(4), 709-719.

Windish, W. W., & Mhatre, N. S. (1965). Microbial amylases. *Advances in applied microbiology*, *7*, 273.

Winterberg, H. (1898). The methodology of bacterial count. *Journal of Hygiene and Infectious Diseases*, 29, 75-93

Wolfenden, R., Lu, X., & Young, G. (1998). Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society*, *120*(27), 6814-6815.

Wommack, K. E., Bhavsar, J., & Ravel, J. (2008). Metagenomics: read length matters. *Applied and environmental microbiology*, *74*(5), 1453-1463.

Wright, E. S., Yilmaz, L. S., & Noguera, D. R. (2012). DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Applied and Environmental Microbiology*, *78*(3), 717-725.

Xie, H., Xue, Y. F., Zhao, A. M., Li, T. G., & Ma, Y. H. (2005). [Preliminary research on bacterial diversity of Parece Vela Basin, Pacific Ocean by culture-independent method]. *Wei Sheng Wu Xue Bao= Acta Microbiologica Sinica*,*45*(1), 1-5.

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., & Xu, A. (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *The ISME journal*, *5*(3), 414-426.

Xu, B., Yang, F., Xiong, C., Li, J., Tang, X., Zhou, J., & Huang, Z. (2014b). Cloning and characterization of a novel α-amylase from a fecal microbial metagenome. *Journal of Microbiology and Biotechnology*, *24*, 447-452.

Xu, F., Dai, X., Chen, Y. Q., Zhou, H., Cai, J. H., & Qu, L. H. (2004). Phylogenetic diversity of bacteria and archaea in the Nansha marine sediments, as determined by 16S rDNA analysis. *Oceanologia et Limnologia Sinica*, *35*(1), 94-96.

Xu, M., Wang, F., Meng, J., & Xiao, X. (2007). Construction and preliminary analysis of a metagenomic library from a deep-sea sediment of east Pacific Nodule Province. *FEMS Microbiology Ecology*, *62*(3), 233-241.

Xu, M., Wang, P., Wang, F., & Xiao, X. (2005). Microbial diversity at a deep-sea station of the Pacific nodule province. *Biodiversity & Conservation*, *14*(14), 3363-3380.

Xu, Z., Hansen, M. A., Hansen, L. H., Jacquiod, S., & Sørensen, S. J. (2014a). Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PloS one*, *9*(4), e93445.

Yanagawa, K., Breuker, A., Schippers, A., Nishizawa, M., Ijiri, A., Hirai, M., & Takai, K. (2014). Microbial Community Stratification Controlled by the Subseafloor Fluid Flow and Geothermal Gradient at the Iheya North Hydrothermal Field in the Mid-Okinawa Trough (Integrated Ocean Drilling Program Expedition 331). *Applied and Environmental Microbiology*, *80*(19), 6126-6135.

Yeates, C., Gillings, M. R., Davison, A. D., Altavilla, N., & Veal, D. A. (1997). PCR amplification of crude microbial DNA extracted from soil. *Letters in Applied Microbiology*, 25(4), 303-307.

Yeh, Y. F., Chang, S. C. Y., Kuo, H. W., Tong, C. G., Yu, S. M., & Ho, T. H. D. (2013). A metagenomic approach for the identification and cloning of an endoglucanase from rice straw compost. *Gene*, *519*(2), 360-366.

Yelton, A. P., Comolli, L. R., Justice, N. B., Castelle, C., Denef, V. J., Thomas, B. C., & Banfield, J. F. (2013). Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. *BMC genomics*, *14*(1), 485.

Yokokawa, T., & Nagata, T. (2010). Linking bacterial community structure to carbon fluxes in marine environments. *Journal of Oceanography*, *66*(1), 1-12.

Young, C. C., Burghoff, R. L., Keim, L. G., Minak-Bernero, V., Lute, J. R., & Hinton, S. M. (1993). Polyvinylpyrrolidone-agarose gel electrophoresis purification of polymerase chain reaction-amplifiable DNA from soils. *Applied and Environmental Microbiology*, *59*(6), 1972-1974.

Yun, J., Kang, S., Park, S., Yoon, H., Kim, M. J., Heu, S., & Ryu, S. (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and environmental microbiology*, *70*(12), 7229-7235.

Zhang, C. T., & Zhang, R. (2004). A nucleotide composition constraint of genome sequences. *Computational biology and chemistry*, *28*(2), 149-153.

Zhang, J. W., & Zeng, R. Y. (2008). Purification and characterization of a cold-adapted α-amylase produced by *Nocardiopsis* sp. 7326 isolated from Prydz Bay, Antarctic. *Marine Biotechnology*, *10*(1), 75-82.

Zhao, X. Q. (2011). Genome-based studies of marine microorganisms to maximize the diversity of natural products discovery for medical treatments.*Evidence-Based Complementary and Alternative Medicine*, *2011*, 1-11

Zhou, H. Q., Ning, L. W., Zhang, H. X., & Guo, F. B. (2014). Analysis of the Relationship between Genomic GC Content and Patterns of Base Usage, Codon Usage and Amino Acid Usage in Prokaryotes: Similar GC Content Adopts Similar Compositional Frequencies Regardless of the Phylogenetic Lineages. *PLoS ONE 9*(9), e107319

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, 62(2), 316-322.

Zhu, D., Tanabe, S. H., Yang, C., Zhang, W., & Sun, J. (2013). Bacterial community composition of South China Sea sediments through pyrosequencing-based analysis of 16S rRNA genes. *PloS One*, *8*(10).

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. M., & Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One*, *6*(9), e24570.

Appendix I

50X TAE Buffer

Tris base	-	121 g
Glacial acetic acid	-	28.6mL
0.5M EDTA pH 8.0	-	50 mL

Deionised water added to make volume to 500mL.

1X TAE Buffer

50X TAE buffer	-	10mL
Deionised water	-	490mL

6X Gel-loading buffer

0.25% (w/v) bromophenol blue

0.25% (w/v) xylene cyanol FF

40% (w/v) sucrose in H_2O

Stored at 4°C.

Luria bertani broth (X-Gal, IPTG, A mpicillin)

Ingredients g/L

Casein enzymic hydrolysate	-	10
Yeast extract	-	5
Sodium chloride	-	10

Suspended 25 grams of media (Himedia, Mumbai, India) in 1000 mL distilled water. Heated to dissolve the medium completely. Sterilized by autoclaving at 15 lbs pressure (121°C) for 15 minutes. Final pH is 7.5 ± 0.2. Before pouring the plates, allow the medium to cool to 55°C. Then, add 1 mL of ampicillin stock solution (50 mg/ml) to a final concentration of 50 μg/mL, 40 μL of X-Gal stock solution (20 mg/mL) and 4 μL of IPTG 200 mg/mL. Mix gently and pour

into the plates. Allow the LB-ampicillin agar medium to solidify. Dry plates opened at room temperature under UV light for 30 min.

Solutions for plasmid isolation

Solution I

25mM Tris – HCl - pH 8.0

50mM glucose

10mM EDTA

Autoclaved the solution and stored at 4^{o}C.

Solution II

0.2N NaOH (freshly diluted from a 10N stock)

1%SDS

Solution II prepared as fresh and used at room temperature.

Solution III

5.0M Potassium Acetate	-	60 mL
Glacial acetic acid	-	11.5 mL
Water	-	28.5 mL

Prepared and stored at 4^{o}C. Transfered to an ice bucket just before use.

Buffer solutions

Hydrochloric acid- potassium chloride buffer (pH 1-2)

Solution A: 0.2 M KCl

Solution B: 0.2 M HCl

Mixed 50 ml of solution A with 10.6 ml of solution B and made up to 200ml with distilled water.

Citrate buffer (pH 3 - 6)

Solution A: 0.1 M Citric acid

Solution B: 0.1 M sodium citrate

Referring to the table below for desired pH, mixed the indicated volumes of solutions A and B, then diluted with distilled water to make up volume to 200 ml and then filter sterilized.

Desired pH	Solution A (ml)	Solution B (ml)
3	46.5	3.5
4	33.0	17
5	20.5	29.5
6	9.5	41.5

Phosphate buffer (pH 7)

Solution A: 0.1 M NaH_2PO_4

Solution B: 0.1 M Na_2HPO_4

Mixed 39 ml of solution A with 61 ml of solution B and the volume was made up to 200ml with distilled water, followed by filter sterilization.

Tris (hydroxymethylamino methane buffer system (pH 8-9)

Solution A: 0.2 M Tris buffer

Solution B: 0.2 M HCl

Referring to the table below for desired pH, mixed the indicated volumes of solutions A and B, then diluted with distilled water to 200 ml and filter sterilized.

Desired pH	Solution A (ml)	Solution B (ml)
8	50	26.8
9	50	5

Carbonate – bicarbonate buffer (pH 10-11)

Solution A: 0.2 M Na_2CO_3

Solution B: 0.2M $NaHCO_3$

Referring to the table below for desired pH, mixed the indicated volumes of solutions A and B, then diluted with distilled water to 200 ml and filter sterilized.

Sodium hydroxide - Potassium chloride buffer (pH 12-13)

Solution A: 0.2 M KCl

Solution B: 0.2M NaOH

Referring to the table below for desired pH, mixed the indicated volumes of solutions A and B, then diluted with distilled water to 200 ml and then filter sterilized.

Desired pH	Solution A (ml)	Solution B (ml)
12	50	12
13	50	132

SOC Medium

2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 2% Agar. Sterilize by autoclaving at 15 lbs pressure (121°C) for 15 min. Aseptically add filter sterilized 20 mM glucose. Final pH is 7.0 ± 0.2. Mix well and dispense as desired.

Iodine-Potassium iodide solution

Dissolved 0.02% (w/v) of iodine and 0.2% (w/v) potassium iodide in distilled water and stored in amber colored bottle.

DNS reagent

Dissolve by stirring at room temperature 1g of 3,5-dinitrosalicylic in 50 mL distilled water, then added 20 mL 2 M NaOH and 28.2 g sodium potassium tartarate, finally made up to 100 mL with distilled water. The reagent was stored at RT in amber colored bottle.

Bradford reagent

Hundred milligrams of Coomassie Brilliant Blue G-250 (Sigma Aldrich, USA) was dissolved in 50 mL 95% ethanol, added 100 mL of 85% (w/v) phosphoric acid and diluted to 1L. When the dye was completely dissolved, this was filtered through Whatman no:1 filter paper.

Reagents for polyacrylamide gel electrophoresis

1. Stock acrylamide - bis acrylamide solution (30: 0.8)

Acrylamide (SRL) (30%)	-	30 g
Bis-acrylamide (SRL) (0.8%)	-	0.8 g
Distilled water	-	100 mL

Filtered through Whatman No: 1 filter paper and stored at 4°C in amber coloured bottle.

2. Stacking gel buffer stock

Tris buffer (SRL) (0.5 M) - 6.05 g in 40 mL distilled water

Titrated to pH 6.8 with 1 M HCl and made up to 100 mL with distilled water. Filtered through Whatman No: 1 filter paper and stored at 4°C.

3. Resolving gel buffer stock

Tris buffer (SRL) (1.5 M) - 18.15 g

Titrated to pH 8.8 with 1M HCl and made up to 100 mL with distilled water. Filtered through Whatman No: 1 filter paper and stored at 4°C.

4. 10% SDS (SRL) - 1 g dissolved in 10 mL distilled water

5. Ammonium persulfate (10%, w/v) (SRL) - 0.1 g of ammonium persulfate was dissolved in1 mL distilled water (prepared freshly).

6. Running buffer for SDS-PAGE (pH 8.3)

Tris buffer	-	3 g
Glycine	-	14.4 g
SDS	-	1 g

Dissolved and made up to 1L with distilled water. Prepared in 10X concentration and stored at 4°C.

7. Sample buffer for Non-Reductive SDS-PAGE (2X)

Tris-HCl (pH 6.8)	-	0.0625 M
Glycerol	-	10%
SDS	-	2 %
Bromophenol blue	-	0.01%

Samples were diluted with sample buffer before loading into the gel.

Coomassie staining solutions

Protein staining solution

Coomassie brilliant blue (0.1%)	-	100 mg
Methanol (40%)	-	40 mL
Glacial acetic acid (10%)	-	10 mL
DW	-	50 mL

Destaining solution

Methanol (40%)	-	40 mL
Glacial acetic acid (10%)	-	10 mL
DW	-	50 mL

2-D Electrophoresis buffers

Equilibration Buffer I

20 mL of 6 M urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8), 20% glycerol and 2% DTT (w/v).

Equilibration Buffer II

20 mL of 6 M urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8) and 20% glycerol.

Aniline-diphenylamine reagent

Dissolve 1 g of diphenylamine in 50 ml of acetone to that add 1 ml of aniline and 5 ml of 85% phosphoric acid

Table 1: Taxonomic classification of the marine 16SrDNA clones by RDP Naive Bayesian rRNA Classifier

Sl.No	Clone No	Accession No:	Phylum	Class	Order	Family	Genus
1	BTMS10	KF453864	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
2	BTMS14	KF453865					
3	BTMS15	KF453866	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Formosa*
4	BTMS16	KF453867					
5	BTMS18	KF453868	*Proteobacteria*	*Gammaproteobacteria*	*Vibrionales*	*Vibrionacea*	
6	BTMS19	KF453869	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
7	BTMS2	KF453870	*Chloroflexi*	*Anaerolineae*	*Anaerolineales*	*Anaerolineaceae*	
8	BTMS20	KF453871	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	*Pelagicola*
9	BTMS21	KF453872	*Proteobacteria*	*Gammaproteobacteria*			
10	BTMS22	KF453873	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
11	BTMS23	KF453874	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
12	BTMS24	KF453875	*Unclassified Bacteria*				
13	BTMS27	KF453876	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	
14	BTMS28	KF453877	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	
15	BTMS33	KF453878	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
16	BTMS34	KF453879	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
17	BTMS35	KF453880	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*	*Halomonas*
18	BTMS36	KF453881	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Oceanospirillaceae*	*Amphritea*
19	BTMS38	KF453882	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Marinobacter*
20	BTMS39	KF453883	*Actinobacteria*	*Actinobacteria*			
21	BTMS4	KF453884	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	
22	BTMS41	KF453885	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Marinobacter*
23	BTMS43	KF453886	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	
24	BTMS44	KF453887	*Unclassified bacteria*				
25	BTMS45	KF453888	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillaceae*	
26	BTMS47	KF453889	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	

27	BTMS5	KF453890	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Citreicella*
28	BTMS68	KF453891	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
29	BTMS70	KF453892	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	
30	BTMS95	KF453893	Unclassified Bacteria				
31	BTMS100	KF453894	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
32	BTMS102	KF453895	*Unclassified Bacteria*				
33	BTMS103	KF453896	*Proteobacteria*	*Gammaproteobacteria*	*Unclassified*		
34	BTMS104	KF453897	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
35	BTMS105	KF453898	*Proteobacteria*	*Gammaproteobacteria*	*Unclassified*		
36	BTMS108	KF453899	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
37	BTMS109	KF453900	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
38	BTMS122	KF453901	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
39	BTMS124	KF453902	*Acidobacteria*	*Acidobacteria_Gp9*			*Gp9*
40	BTMS200	KF453903	*Proteobacteria*	*Deltaproteobacteria*	*Unclassified*		
41	BTMS201	KF453904	*Bacteroidetes*	*Unclassified*			
42	BTMS202	KF453905	*Chloroflexi*	*Unclassified*			
43	BTMS203	KF453906	*Unclassified Bacteria*				
44	BTMS204	KF453907	*Proteobacteria*	*Gammaproteobacteria*	*Unclassified*		
45	BTMS111	KF453908	*Fermicutes*	*Clostridia*	*Clostridiales*	*Clostridiales*	*Fusibacter*
46	BTMS116	KF453909	*Proteobacteria*	*Gammaproteobacteria*	*Unclassified*		
47	BTMS117	KF453910	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
48	BTMS118	KF453911	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
49	BTMS119	KF453912	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	
50	BTMS121	KF453913	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Cyclobacteriaceae*	
51	BTMS123	KF453914	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	
52	BTMS1	KF453915	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
53	BTMS6	KF453916	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	
54	BTMS8	KF453917	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	
55	BTMS9	KF453918	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
56	BTMS11	KF453919	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
57	BTMS12	KF453920	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
58	BTMS13	KF453921	*Acidobacteria*	*Acidobacteria_Gp23*			

59	BTMS125	KF453922	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	
60	BTMS126	KF453923	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
61	BTMS127	KF453924	*Actinobacteria*	*Actinobacteria*			
62	BTMS129	KF453925	*Unclassified bacteria*				
63	BTMS133	KF453926	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
64	BTMS134	KF453927	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Oceanibulbus*
65	BTMS135	KF453928	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
66	BTMS137	KF453929	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*	
67	BTMS138	KF453930	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
68	BTMS139	KF453931	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*		
69	BTMS140	KF453932	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Marinobacter*
70	BTMS141	KF453933	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Cyclobacteriaceae*	*Salegentibacter*
71	BTMS142	KF453934	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Unclassified*
72	BTMS143	KF453935	*Proteobacteria*	*Gammaproteobacteria*			*Unclassified*
73	BTMS144	KF453936	*Unclassified bacteria*				
74	BTMS145	KF453937	*Unclassified bacteria*				
75	BTMS147	KF453938	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*		*Unclassified*
76	BTMS148	KF453939	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	
77	BTMS149	KF453940					
78	BTMS151	KF453941	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*	*Unclassified*
79	BTMS154	KF453942	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*	*Caldithrix*
80	BTMS155	KF453943	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Unclassified*
81	BTMS156	KF453944	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Aurantimonas*
82	BTMS159	KF453945	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
83	BTMS107	KF453946	*Proteobacteria*	*Gammaproteobacteria*			*Salegentibacter*
84	BTMS112	KF453947	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*	
85	BTMS113	KF453948	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	
92	BTMS309	KF569956	*Deferribacteres*	*Deferribacteres*	*Deferribacterales*	*Deferribacterales*	*Caldithrix*
86	BTMS114	KF453949	*Proteobacteria*				
87	BTMS115	KF453950	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	
88	BTMS301	KF569952	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Salegentibacter*
89	BTMS303	KF569953	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Unclassified*	

90	BTMS304	KF569954	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Aurantimonadaceae*	*Aurantimonas*
91	BTMS307	KF569955	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Unclassified*
92	BTMS309	KF569956	*Deferribacteres*	*Deferribacteres*	*Deferribacterales*	*Deferribacterales*	*Caldithrix*
93	BTMS311	KF569957	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Unclassified*
94	BTMS315	KF569958	*Proteobacteria*	*Gammaproteobacteria*	*Unclassified*		
95	BTMS324	KF569960	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Unclassified*
96	BTMS328	KF569962	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Marinobacter*
97	BTMS334	KF569963	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Unclassified*
98	BTMS336	KF569964	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	*Unclassified*
99	BTMS332	KF569966	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Marinobacter*
100	BTMS337	KF569967	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Unclassified*
101	BTMS302	KF569968	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteracea*	*Oceanibulbus*
102	BTMS305	KF569969	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Unclassified*
103	BTMS306	KF569970	*Fermicutes*	*Clostridia*	*Clostridiales*	*Unclassified*	
104	BTMS310	KF569971	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*	*Unclassified*
105	BTMS313	KF569972	*Proteobacteria*	*Gammaproteobacteria*	*Unclassified*		

Table 2: Taxonomic classification of the mangrove 16SrDNA clones by RDP Naive Bayesian rRNA Classifier

Sl.No	Clone no		Phylum	Class	Order	Family	Genus
1	MS4	JX465647	*Proteobacteria*	*Deltaproteobacteria*	*Syntrophobacterales*	*Syntrophaceae*	*Desulfomonile*
2	MS5	JX465648	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Erythrobacteraceae*	*Altererythrobacter*
3	MS 6	JX465649	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Xanthobacteraceae*	*Pseudolabrys*
4	MS 11	JX465650	*Unclassified bacteria*				
5	MS 177	JX852427	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Phyllobacteriaceae*	
6	MS 118	JX852428	*Proteobacteria*	*Gammaproteobacteria*			
7	MS 21	JX465652	*Acidobacteria*	*Acidobacteria-Gp17*			
8	MS 100	JX465653	*Acidobacteria*	*Acidobacteria-Gp1*			
9	MS 154	JX852426	*Unclassified Bacteria*				

10	MS 106	JX852421	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
11	MS 1	JX465646	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	
12	MS 112	JX852422	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
13	MS 119	JX852423	*Proteobacteria*	*Deltaproteobacteria*	*Desulfuromonadales*	*Geobacteraceae*	*Geobacter*
14	MS 135	JX852424	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
15	MS 13	JX465651	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	
16	MS 143	JX852425	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
17	MS 68	JX852429	*Chloroflexi*	*Anaerolineae*	*Anaerolineales*	*Anaerolineaceae*	
18	MS 138	KC143083	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
19	MS 141	KC143084	*Actinobacteria*	*Actinobacteria*	*Coriobacteriales*	*Coriobacteriaceae*	
20	MS 146	KC143085	*Unclassified Bacteria*				
21	MS 149	KC143086	*WS3*				*Latescibacteria*
22	MS 152	KC143087	*Planctomycetes*	*Planctomycetia*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces*
23	MS 171	KC143088	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	
24	MS 176	KC143089	*Actinobacteria*	*Actinobacteria*	*Acidimicrobiales*	*Acidimicrobiaceae*	*Ilumatobacter*
25	MS 190	KC143090	*Proteobacteria*	*Deltaproteobacteria*	*Myxococcales*	*Polyangiaceae*	*Sorangium*
26	MS 191	KC143091	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiales*	*Thiobacter*
27	MS 197	KC143092	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
28	MS 198	KC143093	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
29	MS 20	KC143094	*Proteobacteria*	*Betaproteobacteria*	*Gallionellales*	*Gallionellaceae*	*Sideroxydans*
30	MS 34	KC143095	*Acidobacteria*	*Acidobacteria-Gp7*			
31	MS 41	KC143096	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiales*	
32	MS54	KC143097	*Firmicutes*	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
33	MS76	KC143098	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Haliea*
34	MS77	KC143099	*Proteobacteria*	*Gammaproteobacteria*			

Table 3: Classification of Arabian Sea Metagenome at phylum level

Sl No	Phylum	Total OTUs	Percentage OTUs
1	*Proteobacteria*	2932	46.47329
2	*Acidobacteria*	476	7.544777
3	*Chloroflexi*	369	5.848787
4	*Actinobacteria*	182	2.884768
5	*Firmicutes*	117	1.854494
6	*Chlorobi*	12	0.190204
7	*Nitrospirae*	100	1.585037
8	*Elusimicrobia*	1	0.01585
9	*Tenericutes*	9	0.142653
10	*Bacteroidetes*	283	4.485655
11	*Gemmatimonadetes*	182	2.884768
12	*Verrucomicrobia*	1	0.01585
13	*Caldithrix*	62	0.944968
14	*Spirochaetes*	64	1.014424
15	*Fibrobacteres*	1	0.01585
16	*Fusobacteria*	11	0.174354
17	*Cyanobacteria*	3	0.047551
18	*Planctomycetes*	16	0.253606
19	*WS3*	185	2.932319
20	*SBR1093*	5	0.079252
21	*SC4*	1	0.01585
22	*TM6*	7	0.110953
23	*LCP-89*	6	0.095102
24	*CD12*	5	0.079252
25	*GN04*	93	1.474085
26	*Hyd24-12*	1	0.01585
27	*WS2*	7	0.110953
28	*TPD-58*	1	0.01585
29	*SAR406*	8	0.126803
3	*AC1*	19	0.301157
31	*Thermi*	9	0.142653
32	*GN02*	9	0.142653
33	*OP1*	3	0.047551
34	*GOUTA4*	1	0.01585
35	*BHI80-139*	7	0.110953
36	*TM7*	9	0.142653
37	*KSB3*	7	0.110953
28	*OD1*	1	0.01585
39	*BRC1*	15	0.237756
40	*MVS-104*	2	0.031701

41	*EM3*	1	0.01585
42	*OP8*	44	0.697416
43	*ZB3*	4	0.063401
44	Unknown	1038	16.45269

Table 4: Classification of Mangalavanam Mangrove Metagenome at phylum level

Sl No	Phylum	Total OTUs	PercentageOTUs
1	*Proteobacteria*	4082	43.60179
2	*Chloroflexi*	408	4.358043
3	*Planctomycetes*	12	0.128178
4	*Firmicutes*	148	1.580859
5	*Chlorobi*	131	1.399274
6	*Nitrospirae*	54	0.5768
7	*Actinobacteria*	915	9.773553
8	*Elusimicrobia*	42	0.448622
9	*Acidobacteria*	519	5.543687
10	*Chlamydiae*	5	0.053407
11	*Tenericutes*	2	0.021363
12	*Bacteroidetes*	515	5.500961
13	*Verrucomicrobia*	63	0.672933
14	*Caldithrix*	32	0.367037
15	*Spirochaetes*	122	1.30314
16	*Synergistetes*	1	0.010681
17	*Gemmatimonadetes*	136	1.452681
18	*Deferribacteres*	1	0.010681
19	*Fibrobacteres*	20	0.21363
20	*Fusobacteria*	1	0.010681
21	*Cyanobacteria*	36	0.384533
22	*Armatimonadetes*	3	0.032044
23	*WS3*	134	1.431318
24	*TA06*	1	0.010681
25	*TM6*	25	0.267037
26	*SBR1093*	6	0.064089
27	*LCP-89*	6	0.064089
28	*GN04*	52	0.555437
29	*NKB19*	3	0.032044
30	*WS5*	1	0.010681
31	*WPS-2*	4	0.042726
32	*MVP-21*	3	0.032044
33	*WWE1*	4	0.042726
34	*WS2*	6	0.064089

35	*TPD-58*	7	0.07477
36	*GN02*	12	0.128178
37	*OP3*	1	0.010681
38	*NC10*	9	0.096133
39	*AC1*	10	0.106815
40	*Thermi*	4	0.042726
41	*OP1*	1	0.010681
42	*GOUTA4*	5	0.053407
43	*AD3*	2	0.021363
44	*BHI80-139*	1	0.010681
45	*TM7*	46	0.491348
46	*KSB3*	13	0.138859
47	*OD1*	1	0.010681
48	*BRC1*	15	0.160222
49	*SR1*	1	0.010681
50	*FCPU426*	1	0.010681
51	*ZB3*	5	0.053407
52	*OP8*	9	0.096133
53	*Unknown*	1726	18.43623

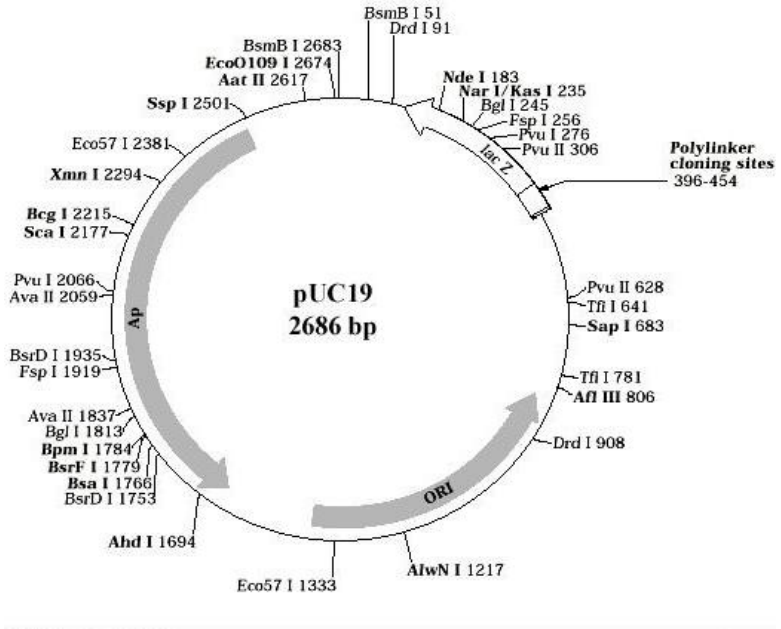

Figure 1: Vector Map of pUC19

```
(WP_014773782)        -----MKKH---LLAPILSVLTILTSCDTTS--NSKVENHWPHAGITYEIFIQSFYDSNG
(WP_009184110)        -------MKL--QLGLFISTLLLIWSCSGKG--KNEVENHWPHAGITYEIFIQSFYDSDG
(ERM84602)            MAFSFLKQKFMKKIFIYCFIFFVAVSCKNSG--DGKLENHWPNAGITYEIFIQSFYDTNS
(WP_026333543)        ----------MKKIFIYCFIFFVAVSCKNSG--DGKLENHWPNAGITYEIFIQSFYDTNS
(WP_015266641)        -----MRNLL--RSFTYCTLVLLTASCMQDANEELEVKNYWPEAGITYEIFIQSFYDTDG
(WP_026970068)        -----MKKSI--L-TASALTVAALASCEKMENEGLAVKNYWPQAGITYEIFVRSFYDSNG
(WP_029660871)        -----MKAKS--T-LIALILSLFYFSCNQEK--EPEVKNYWPNAGVTYEIFVQSFYDSDG
(WP_035483031)        -----MKAKS--T-LIALILSLFYFSCNQEK--EPEVKNYWPDAGVTYEIFVQSFYDSDG
BTM109(KR514290)      -----MKKSI--L-TTSIFALVAFSSCEKKP--APEVKNYWPQAGVTYEIFVQSFYDSNG
(AAQ89599)            -----MKKSI--L-TTSIFALVAFSSCEKKP--APEVKNYWPQAGVTYEIFVQSFYDSNG
(WP_026950142)        -----MKINL--L-LSSVLSLAILSACSK-K--EPEVKNYWPQAGVTYEIFVQSFQDSNG
                                           :*        ::*:**.**:*****::** *::.

(WP_014773782)        DGIGDINGVTKKLDHVKELGANAIWFMPIMPSPSYHKYDVTDYNGIHPDYGTMDDFKNML
(WP_009184110)        DGIGDFNGVIKKLDHVKELGANAIWFMPIMPSPTYHKYDVTDYKAVHPDYGTMDDFKQLL
(ERM84602)            DGIGDINGVTKKLDHVKELGANAIWFMPIMPSPSYHKYDVTDYKAIHPDYGTLEDFKKLM
(WP_026333543)        DGIGDINGVTKKLDHVKELGANAIWFMPIMPSPSYHKYDVTDYKAIHPDYGTLEDFKKLM
(WP_015266641)        DGIGDINGVTKKLDHVQELGANAIWFMPLMPSPSYHKYDVTDYKAIHPDYGTMDDFKQML
(WP_026970068)        DGIGDFNGVTQKLDYVKELGADAIWLMPIMPSPTYHKYDVTDYKSVHKDYGTMDDFRKLL
(WP_029660871)        DGIGDFNGVTEKLDYVQELGANAIWFMPIMPSPTYHKYDVTDYKAVHPDYGTMEDFKRLL
(WP_035483031)        DGIGDFNGVTEKLDYVQELGANAIWFMPIMPSPTYHKYDVTDYKAVHPDYGTMEDFKRLL
BTM109(KR514290)      DSIGDFNGVTQKLDYVKELGANAIWFMPIMPSPTYHKYDVTDYKAVHPDYGTLDDFKKLL
(AAQ89599)            DSIGDFNGVTQKLDYVKELGANAIWFMPIMPSPTYHKYDVTDYKAVHPDYGTLDDFKKLL
(WP_026950142)        DGIGDFNGVIQKLDHVKELGANAIWFMPIMPSPTYHKYDVTDYKAVHPDYGTMEEFKNLI
                      *.***:*** :***:*:****:***:**:****:*********:.:* ****:::*:.::

(WP_014773782)        KSAHEKNIKVVIDLIINHTSSEHPWFIESKKGRDNPYRDYYVWAQKDTIADFINKKTITL
(WP_009184110)        EEAHKRDIKIVIDLIINHTSTEHPWFLESKKSRDNPYRDYYVWAQKDTIADFINKKTITL
(ERM84602)            KEAHKRDIKIVIDLIINHTSSEHPWFLESKKGRDNPYRDYYVWAQKDTISAYLNKKTITL
(WP_026333543)        KEAHKRDIKIVIDLIINHTSSEHPWFLESKKGRDNPYRDYYVWAQKDTISAYLNKKTITL
(WP_015266641)        DEAHKRDIKVVIDMIINHTSDEHPWFQEAKKGRDNPYRDYYVWAQYDTIQDYLDKKVVTL
(WP_026970068)        DEAHKRDIKVVIDLIINHTSVEHPWFLEAKSGRDNPYRDYYIWAQRDTVAKVLNKKVLTL
(WP_029660871)        EEAHQRDIKIVIDMIINHTSTEHPWFQESKKSRDNPYRDYYVWAQKDTIADFLDKKVITL
(WP_035483031)        EEAHQRDIKIVIDMIINHTSTEHPWFQESKKSRDNAYRDYYVWAQKDTIADFLNKKVITL
BTM109(KR514290)      DEAHKRDIKIVIDLIINHTSNEHPWFLEAKSGRDNPYRDYYVWAQKDTIADFLNKKTITF
(AAQ89599)            DEAHKRDIKIVIDLIINHTSNEHPWFLEAKSGRDNPYRDYYVWAQKDTIADFLNKKTITF
(WP_026950142)        AEAHKRDIKIVIDMIINHTSTEHPWFQESKSGRDNPYRDYYVWAQKDTIADFLNKKTITL
                       .**:::**:***:****** ***** *:*..*** *****:*** **:   ::**.:*:

(WP_014773782)        DSDNIRQWHDPGNGEDFYYGFFIGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
(WP_009184110)        DSDNIRQWHDPGFGQDYYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEDIGVDGFRLDA
(ERM84602)            DSDNIRQWHNPGSGEDFYYGFFIGGMPDLNFDNPKVRKEIYDIGRFWLEEMGVDGFRLDA
(WP_026333543)        DSDNIRQWHNPGSGEDFYYGFFIGGMPDLNFDNPKVRKEIYDIGRFWLEEMGVDGFRLDA
(WP_015266641)        DSDNIRQWHDPGQGDDYYYGFFTGDMPDLNFDNPKVREEIYEIGRYWLAEVGVDGFRLDA
(WP_026970068)        DSSNRRQWHDPGRGEDYYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
(WP_029660871)        DSDNIRQWHDPGSGEDYYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
(WP_035483031)        DSDNIRQWHDPGSGEDYYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
BTM109(KR514290)      DLDNIRQWHDPGQGEDFYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
(AAQ89599)            DLDNIRQWHDPGQGEDFYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
(WP_026950142)        DSDNIRQWHDPGQGEDYYYGFFWGGMPDLNFDNPKVREEIYEIGRFWLEEVGVDGFRLDA
                      * .* ****:** *:*:***** * ***********:***:***:** ::*********
```

```
(WP_014773782)      AKHIYTDDRPKDNHAFWKEFRSEMTKIKPDVYLVGEVYDMKEVVAPYLPGLPALFNFDFH
(WP_009184110)      AKHIFPDDRPEDNHAFWKEFRDRMTAIKSDIYLVGEVYDRKEIVAPYLPGLPALFNFDFH
(ERM84602)          AKHVFPDDRPVDNHEFWIEFRSEMEKIKPDVYLVGEVYDMKEVVAPYLPGLPALFNFDFH
(WP_026333543)      AKHVFPDDRPVDNHEFWIEFRSEMEKIKPDVYLVGEVYDMKEVVAPYLPGLPALFNFDFH
(WP_015266641)      AKHIYPDDRAADSHEFWEEFRAEMEKVKPDVYLVGEVYDMKEVVAPYLTGLRALFNFDFH
(WP_026970068)      AKHIFPDDRHADNHEFWKEFRARMTAIKPDVYLVGEVYDKKDVVAPYLPGLPALFNFDFH
(WP_029660871)      AKHIFPDDRPLDNHEFWKEFRAKMEAIKPDVYLVGEVYDKKEIVAPYLPGLPALFNFDFH
(WP_035483031)      AKHIFPDDRPLDNHEFWKEFRAKMEAIKPDVYLVGEVYDKKEIVAPYLPGLPALFNFDFH
BTM109(KR514290)    AKHIFPDDRPLDNHAFWKEFRAKMEVIKPDVYLVGEVYDKKEVVAPYLPGLPALFNFDFH
(AAQ89599)          AKHIFPDDRPLDNHAFWKEFRAKMEVIKPDVYLVGEVYDKKEVVAPYLPGLPALFNFDFH
(WP_026950142)      AKHIFPDDRPLDNHAFWKEFRAKMEAIKPDIYLVGEVYDKKEIVAPYLPGLPALFNFDFH
                    ***:: ***  *.* ** *** .*  :* *:******** *::***** ** ********

(WP_014773782)      YTLLESYKKEDGMLLVRKQKEILDFYKDITDEFIDATFSSNHDQPRLLNSLDENPKKLKQ
(WP_009184110)      YTLIEAYQKEDGMLLAKKQKEVLDFYLDITKDFIDATFSSNHDQPRLLNSLDENPRKLKQ
(ERM84602)          YTLIDAFKESNGSLLVEKQKEILDYYLGITPDFIDATFSSNHDQPRLLNSLNENPRKLKQ
(WP_026333543)      YTLIDAFKESNGSLLVEKQKEILDYYLGITPDFIDATFSSNHDQPRLLNSLNENPRKLKQ
(WP_015266641)      YTLLEAYKKEDGMLLAKKQHDILAFYNGITDDFIDATISSNHDQPRLLNELGKSKDKLKQ
(WP_026970068)      YTLLTAMNSGDGLLLPEIQKETLDFYQSIAPSFIDATFSSNHDQPRLLNELGKDPARYKQ
(WP_029660871)      YTLLEAFNTADGMLLPKKQKEILDFYQNITPEFIDATISSNHDQPRIFNELGEDVEKYKQ
(WP_035483031)      YTLLEAFKSADGMLLPKKQKEILDFYQNITPEFIDATISSNHDQPRLLNELGEDVDKYKQ
BTM109(KR514290)    YTLLETMNTGDGMLLAKKQKEILDFYQGITSSFIDATISSNHDQPRLLNELGSDPAKYKQ
(AAQ89599)          YTLLETMNTGDGMLLAKKQKEILDFYQGITSSFIDATISSNHDQPRLLNELGSDPAKYKQ
(WP_026950142)      YTLLETMNTEDGMLLAKKQKDIIDFYQGITPNFIDATISSNHDQPRLLNELGSDPDKYKQ
                    ***: : :  :* ** . *:: : :*  *: .*****:********::*.* ..  : **

(WP_014773782)      AVTILMSMPGAPYLYYGEEIGMMGKKPDENIREPFLWDIRANDSGRATWIDAEYSKDDVV
(WP_009184110)      AINILMTMPGAPYLYYGEEIGMLGQKPDEHIREPFLWDIKANDKGRTSWIEPQYSIDETV
(ERM84602)          AIAILMTMPGAPYLYYGEEIGMLGRKPDENIREPFLWDIQANDKGRTTWMEPVYSIDEMV
(WP_026333543)      AIAILMTMPGAPYLYYGEEIGMLGRKPDENIREPFLWDIQANDKGRTTWMEPVYSIDEMV
(WP_015266641)      AIAILMTMPGAPYIYYGEEIGMLGKKPDPNIREPFLWDVAEQDEGRTKWITPAFSTDNTV
(WP_026970068)      AIAVLLTMPGAPYLYYGEEIGMLGLKPDQHIREPFLWDKKAKDTGRTTWIKARYSKDSTV
(WP_029660871)      AIAVMLTMPGAPYLYYGEEIGMKGLKPDEHIREPFLWDVKAKDEGRATWIEPKYSIDSEV
(WP_035483031)      AIAVMLTMPGAPYLYYGEEIGMKGLKPDEHIREPFLWDVKAKDKGRATWIEPKYSIDSEV
BTM109(KR514290)    AIAVMLSMPGAPYLYYGEEIGMLGLKPDEHIREPFLWDEKSKDTGRTKWIKPKYSKDSTV
(AAQ89599)          AIAVMLSMPGAPYLYYGEEIGMLGLKPDEHIREPFLWDEKSKDTGRTKWIKPKYSKDSTV
(WP_026950142)      AIAVLLTMPGAPYLYYGEEIGMLGLKPDENIREPFLWDVKEKDTGRPTWIDAKYSTDKTV
                    *: ::::******:******** * *** .*******   :* ** .*:   :* *. *

(WP_014773782)      TPLAIQKSMSTSYFNHYREIIKLRNTNRALALGDLEFYSEDLSKPVMAFLRKYEGQELFV
(WP_009184110)      TPLAIQKSMSTSYFNHYKEIIRLRNTNRALALGNLELYEGTLPDPIMAFFRNHKDQKVFV
(ERM84602)          TPLAIQKSMSTSYFNHYKEIIAVRNTNRGLALGNLELYSEDLPKPIMAYFRNHKDQNLFV
(WP_026333543)      TPLAIQKSMSTSYFNHYKEIIAVRNTNRGLALGNLELYSEDLPKPIMAYFRNHKDQNLFV
(WP_015266641)      TPLAIQKEDADSYFNHYKRVIQLRNTHPALAIGSLELPAEKYPKAVMAYQRKTGEQELYV
(WP_026970068)      TPLEIQRKDPNSYFNHYKTLIALRNSYPPLAIGALELSEIEFPESVMAYFRNSGEQEIFV
(WP_029660871)      TPLALQRNDPSSIFNHYKSLINLRNTYPALALGKLSLPSSEYPKPIMAYSRVHQDQEIFA
(WP_035483031)      TPLALQRNDPNSIFNHYKSLINLRNTYPALALGKLSLPSSEYSKPIMAYSRVHKGQAIFV
BTM109(KR514290)    TSLEVQKKDSNSYFNHYKNLIALRNSYPALAIGSLELPAEELPKSVMAYFRKSGDQEIFV
(AAQ89599)          TSLEVQKKDSNSYFNHYKNLIALRNSYPALAIGSLELPAEELPKSVMAYFRKSGDQEIFV
(WP_026950142)      TPLALQKNDPKSYFNHYKSLIKLRNSFPALAIGTLELPKTEYPKSIMAYLRKSGNQEIFV
                    * * :*:.   * ****: :* :**:   **:* *.:      . :**: *    * ::.
```

Figure 2: Multiple sequence alignment of BTM109

LIST OF PUBLICATIONS

Peer Reviewed Publications

1. **Harisree P. Nair,** Helvin Vincent and Sarita G. Bhat (2014), Evaluation of five in situ lysis protocols for PCR amenable metagenomic DNA from mangrove soils, *Biotechnology Reports* 4: 134–138.

2. **Harisree P. Nair**, Helvin Vincent and Sarita G. Bhat (2013), Culture independent analysis of the soil microbiome to assess microbial diversity of Mangrove Soil, *Bio-Genetics Journal*, 1(1):1-4

3. Noble K Kurian, **Harisree P Nair** and Sarita G Bhat (2015), Evaluation of anti-inflammatory property of melanin from marine *Bacillus* spp. BTCZ31, *Asian Journal of Pharmaceutical and Clinical Research* Vol 8, Issue 3

4. Noble K Kurian, **Harisree P Nair** and Sarita G Bhat (2014), Melanin producing *Pseudomonas stutzeri* BTCZ10 from marine sediment at 96 m depth (Sagar Sampada cruise #305) *International Journal of Current Biotechnology*, 2(5):6-11.

5. Helvin Vincent, **Harisree P Nair** and Sarita G Bhat(2013), Community genomics involving culture independent approach for assessing the phylogenetic diversity of mangrove sediment, *Indian Journal of Applied Research*, Vol 3(10), 29-32

6. Alphonsa Vijaya Joseph, Raghul Subin Sasidharan, **Harisree P. Nair** and Sarita G. Bhat (2013), Occurrence of potential pathogenic *Aeromonas* species in tropical seafood, aquafarms and mangroves off Cochin coast in South India, *Veterinary World*, 6(6):300-306.

Full paper in proceedings of national/international symposium/conferences/seminars

1. **Harisree P. Nair**, Helvin Vincent and Sarita G. Bhat (2012), Metagenomic approach for analysis of bacterial diversity of Mangalavanam mangrove soil, National Symposium on Emerging Trends in Biotechnology, CUSAT, December 12-13, 2012, (ISBN:978-93-80095-39-4)
2. Noble K Kurian, **Harisree P Nair** and Sarita G Bhat (2014), A novel melanin producing *Pseudoalteromonas lipolytica* BTCZ28 isolated from 96m depth marine sediments, 24th Swadeshi Science Congress (ISBN: 978-81-928129-2-2)

Oral/Poster presentations in national/international symposium/conferences/seminars

1. **Harisree P. Nair,** Rinu M. Puthusseri, Helvin Vincent and Sarita G. Bhat (2015), Analysis of microbial composition and diversity of the sediments of Arabian sea by paired-end illumina reads, International Conference on Biodiversity, Ecology and Conservation of Marine Ecosystems 2015 at School of Biological Sciences The University of Hong Kong, Hong Kong, June 1-4, 2015

2. **Harisree P. Nair,** Rinu M. Puthusseri, Helvin Vincent and Sarita G. Bhat (2015), Metagenomic analysis of bacterial diversity of marine sediments of Arabian Sea revealed proteobacterial dominance, World Ocean Science Congress, Kochi, Kerala, February 5-8, 2015

3. **Harisree P. Nair,** Rinu M. Puthusseri, Helvin Vincent and Sarita G. Bhat (2015), Fingerprinting of metagenomic clones using Amplified Ribosomal DNA Restriction Analysis (ARDRA) for assessment of bacterial diversity of Arabian Sea sediments, 102nd Indian Science Congress, University of Mumbai, Maharastra, India, January 2-7, 2015

4. **Harisree P. Nair,** Helvin Vincent and Sarita G. Bhat (2014), Culture independent analysis of the soil microbiome to assess microbial diversity, 101st Indian Science Congress, University of Jammu, Jammu and Kashmir, February 3-7, 2014

5. Noble K Kurian, **Harisree P Nair** and Sarita G Bhat (2014), Physiochemical characterization of purified melanin pigment from marine *Bacillus* sp. BTCZ31, 101st Indian Science Congress, University of Jammu, Jammu and Kashmir, February 3-7, 2014

6. Helvin Vincent, **Harisree P. Nair** and Sarita G. Bhat (2010), Novel gelatinase gene from a metagenomic library of marine sediment, International Conference on Genomic sciences, Madurai Kamaraj University, Tamil Nadu, November 12-14, 2010

Genbank submissions – 140 Numbers

KF453864 - KF453950, KF569952 - KF569958, KF569962 - KF569964

KF569966 - KF569972, JX465646 - JX465653, JX852421 - JX852429,

KC143083 - KC143099, KF569960, KR514290

Biotechnology Reports 4 (2014) 134–138

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

Evaluation of five in situ lysis protocols for PCR amenable metagenomic DNA from mangrove soils☆

Harisree P. Nair, Helvin Vincent, Sarita G. Bhat *

Department of Biotechnology, Cochin University of Science and Technology, Cochin, Kerala 682022, India

ARTICLE INFO

Article history:
Received 15 July 2014
Received in revised form 25 September 2014
Accepted 26 September 2014
Available online 30 September 2014

Keywords:
Metagenome
In situ lysis
DNA extraction

ABSTRACT

Microbes in nature are rarely amenable to growth by standard microbiological methods, with the majority being unculturable. Metagenomic methods help to bypass and overcome the limitations of traditional culturing method; wherein total community DNA is isolated, cloned into suitable vector and host systems. However, isolation of total community DNA itself remains a challenge. In this study five methods of total community DNA isolation from three different mangrove soils were evaluated to test its PCR amenability. The yield and purity of the isolated DNA was also analysed. The quantity of DNA by all 5 methods although reasonably high, contained residual humic contaminants. Of the five, the method employing liquid nitrogen yielded readily amplifiable DNA, while that by all others required further downstream processing to achieve purity and PCR amenability.

1. Introduction

Soil being a complex habitat for diverse microorganisms, is a rich source of novel enzymes and bioactive molecules for application in health and other industries. Estimations reveal that less than 1% of the total microbial communities from the environment are readily cultivable by standard microbiological methods [1]. The unculturable microbes remain uncharacterised, the deficiency of information about their culturing parameters, allowing their continuation as unexplored reservoir of metabolic and genetic diversity.

Mangrove ecosystems present at the intertidal zones of estuaries, lagoons or marshes of tropical and subtropical latitudes, are unique ecological niches, habitat to multiple microbes playing significant roles in nutrient recycling and various ecological processes; thereby necessitating a thorough exploration of these microflora. Mangrove soils are commonly nutrient rich and hence exceedingly diverse in their microbial content. By the same rationale, community DNA isolation is a challenging process owing to co-extraction of humic substances.

DNA extraction methods are classified as direct (in situ) and indirect (ex situ) methods. In direct methods, cells are lysed within the soil sample, followed by consequent separation of DNA from cell debris and soil matrix [2]; and indirect method employs cell separation followed by cell lysis and DNA recovery [3]. These approaches have advantages as well as disadvantages concerning DNA yields, purity for molecular analysis and unbiased representation of the entire microbiome.

However as soil compositions vary greatly with regard to the organic and inorganic content, standardisation of total DNA isolation protocols become a prerequisite to any analysis. The objective of this study was to investigate the effectiveness of different direct lysis methods on yield and purity of DNA from mangrove soils to enable PCR amplification and further metagenomic analysis.

2. Materials and methods

2.1. Sample collection

Mangrove soils were collected from 3 different islands located in Kochi, Kerala, India, by removing surface leaf litter and collecting the top soil. Samples were transferred with sterilised spatula in sterile containers and were stored at −20 °C until further analysis. Sampling location details are given in Table 1.

2.2. DNA extraction

The five direct lysis methods tested for isolation and purification of DNA from the three mangrove soils include the methods of Zhou et al. (1996), slightly modified method of Volossiouk et al.

* Corresponding author. Tel.: +91 484 2576267.
E-mail address: saritagbhat@gmail.com (S.G. Bhat).

http://dx.doi.org/10.1016/j.btre.2014.09.008
2215-017X/

Table 1
Coordinates of sampling location.

Soil	Sampling station	Latitude	Longitude
1	Vypin	10° 4' 7.3272"N	76° 12' 47.3292"E
2	Bolgatty	10° 0' 16.2864"N	76° 15' 42.0120"E
3	Ponnarimangalam	10° 0' 16.2864"N	76° 15' 42.0120"E

(1995), Dong et al. (1996), Tsai and Olson, (1991) and that of Siddhapura et al. (2010).

2.2.1. Method 1 [4]

Mixed 5 g soil with 13.5 mL DNA extraction buffer (in an Oakridge tube) (100 mM Tris–HCl (pH 8.0), 100 mM sodium EDTA (pH 8.0), 100 mM sodium phosphate (pH 8.0), 1.5 M NaCl, 1% CTAB) and 100 mL of proteinase K (10 mg/mL) (Fermentas, USA) and the sample was incubated by horizontal shaking at 225 rpm for 30 min at 37 °C (Orbitek, Scigenics India). This was followed by addition of 1.5 mL of 20% SDS and incubated in a 65 °C water bath for 2 h (Remi, India) with gentle end-over-end inversions every 15–20 min. The supernatant was transferred to new tubes after centrifugation at 6000 × *g* for 10 min (Sigma, 2–16 K, Germany) at room temperature. The soil pellets were further extracted twice using the same protocol. Supernatants from the three extractions were pooled, mixed with equal volume of chloroform: isoamyl alcohol (24:1, v/v), followed by recovery of the aqueous phase by centrifugation and finally precipitation with 0.6 volume of isopropanol at room temperature for 1 h. The nucleic acids obtained were pelleted by centrifugation at 16,000 × *g* for 20 min and washed with cold 70% ethanol, air dried and resuspended in sterile deionised water to a final volume of 500 μL.

2.2.2. Method 2 [5]

After adding liquid nitrogen the 0.25 g soil sample was ground to fine powder using sterile mortar and pestle, suspended in 0.5 mL of skim milk powder solution (0.1 g skim milk in 25 mL of water), vortexed well and centrifuged for 10 min at 12,000 × *g* at 4 °C. To the supernatant 2 mL of SDS extraction buffer (0.3% SDS in 0.14 M NaCl, 50 mM sodium acetate (pH 5.1) was added and vortexed to mix. An equal volume of Tris-saturated phenol solution was added and vortexed for 2 min at room temperature. Aqueous phase was collected by centrifugation at 12,000 × *g* for 10 min and the nucleic acid was precipitated with 1 volume of ice cold isopropanol at −20 °C for 1 h, followed by centrifugation at 12,000 × *g* for 10 min to pellet the DNA. The pellet was washed twice with cold 70% ethanol, with centrifugation between each rinse, air dried, dissolved in 150 μL of sterile deionised water and stored at −20 °C until further analyses.

2.2.3. Method 3 [6]

In this method 0.30 g of soil sample was mixed with 0.35 g of glass beads (diameter 2.0 mm) and 300 μL of phosphate buffer (0.1 M NaH_2PO_4–$NaHPO_4$ (pH 8.0)) in a microcentrifuge tube, vortexed well, followed by addition of 250 μL of SDS lysis buffer (100 mM NaCl, 500 mM Tris (pH 8.0), 10% SDS). This was vortexed horizontally for 10 min at 225 rpm. The supernatant was transferred to new tube after centrifugation at 10,000 × *g* for 30 s. 250 μL of chloroform: isoamyl alcohol (24:1) was added and incubated at 4 °C for 5 min, followed by centrifugation at 10,000 × *g* for 1 min. Nucleic acids were precipitated by addition of 0.5 volume of 7.5 M ammonium acetate and 1volume of isopropanol, and incubated at −20 °C for 15 min. DNA was pelleted at 12,000 x *g* for 10 min, was washed thrice with 70% ethanol and air-dried. Pellets were dissolved in 100 μL of 10 mM Tris (pH 8.1), 100 μL of 10 mM Tris [pH 7.4], 100 μL of 10 mM Tris (pH 6.7) and 100 μL of 10 mM Tris (pH 6.0) and flocculated with 10 mM aluminium sulfate. Precipitate of humic substances was removed by centrifuging at 10,000 × *g* for 5 min.

2.2.4. Method 4 [7]

One gram soil was washed twice with 2 mL of 120 mM sodium phosphate buffer (pH 8.0), suspended in 2 mL of lysis solution (0.15 M NaCl, 0.1 M Na_2EDTA [pH 8.0]) containing lysozyme [15 mg/mL], incubated for 2 h in a 37 °C water bath with mixing at 20–30 min intervals, followed by addition of 2 mL of 0.1 M NaCl, 0.5 M Tris–HCl [pH 8.0], 10% SDS. Cells were lysed by three cycles of alternating freeze-thaw at −80 °C and 65 °C respectively. After phenol–choloroform extraction, the nucleic acid was precipitated with ice cold isopropanol, dried and resuspended in 100 μL of TE buffer (20 mM Tris–HCl, 1 mM EDTA (pH 8.0)).

2.2.5. Method 5 [8]

In this method 1 g soil was mixed with 10 mL extraction buffer (100 mM Tris–HCl (pH 8.2); 100 mM EDTA (pH 8); 1.5 M NaCl), incubated at 37 °C for 10 h with shaking at 150 rpm and supernatant was collected by centrifugation at 5000 rpm for 10 min. Samples were re-extracted with 1 mL of extraction buffer. To the supernatant 4 mL of lysis buffer (20%, w/v) SDS, lysozyme (20 mg/mL), Proteinase K (10 mg/mL), *N*-lauryl sarcosine (10 mg/mL), 1% (w/v) CTAB (cetyltrimethylammonium bromide) was added and incubated at 65 °C for 2 h with intermittent shaking every 15 min. Centrifuged at 10,000 rpm for 10 min at 4 °C to collect the supernatant. The preparation after phenol–chloroform extraction was treated with 1/10 volume of 7.5 M potassium acetate and precipitated by 2 volumes of chilled absolute alcohol. DNA was pelleted by centrifugation at 10,000 rpm for 10 min, air dried and suspended in 50 μL sterile deionised water.

2.3. Determination of yield and purity of DNA

The yield and purity of DNA obtained by all the five methods was quantified using spectroscopic methods, by calculating A_{260}/A_{280} and A_{260}/A_{230} ratios for protein and humic acid contaminants in the preparation. A_{260}/A_{280} ratio less than 1.8 indicates protein contamination and A_{260}/A_{230} ratio less than 2 indicates the presence of humic acid substances.

2.4. Gel electrophoresis

The extracted DNA were analysed by agarose gel electrophoresis in 0.8% gel containing 10 mg/mL ethidium bromide solution under UV light. Gel pictures were captured using gel documentation system (Syngene, USA)

2.5. Purity of DNA by PCR

To determine whether PCR inhibitors were present, DNA preparations isolated by all protocols were used as template to amplify the region encoding 16S rRNA gene in a thermal cycler (Biorad, USA) using universal primers [9]. 50 ng template DNA was used in a 20 μL reaction with an initial denaturation for 2 min at 94 °C, 34 cycles of denaturation at 94 °C for 30 s, annealing at 54 °C for 30 s and extension at 72 °C for 2 min with a final extension for 10 min at 72 °C. The amplicons were separated electrophoretically in 1% agarose gel and visualised using ethidium bromide under ultraviolet illumination and gel pictures are captured using gel documentation system (Syngene, USA)

2.6. Statistical analysis

All experiments repeated thrice and statistical analysis was done by Microsoft Excel 2007 calculating mean and standard error.

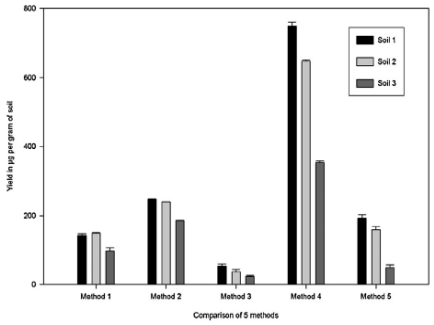

Fig. 1. DNA yield from different Mangrove soils by five methods.

3. Results

3.1. Metagenomic DNA extraction

Five different methods of metagenomic DNA isolation using three different soil samples from mangroves were compared with respect to DNA yield, purity, humic acid content, and suitability for PCR. Highest yield was obtained by method 4, giving 748.6, 647.3 and 353 µg DNA/gram of soil with soil 1, 2 and 3 respectively; while the lowest yield was obtained by method 3 which is calculated as follows: Yield in µg per gram of soil = concentration of DNA (µg/µL)/weight of soil (g) × volume in which DNA suspended (µL). The results are indicated in Fig. 1.

3.2. Spectrophotometric analysis for yield and purity of isolated DNA

The isolated DNA was assessed for yield and purity by obtaining OD ratios at 260 nm/280 nm (DNA/Protein) and 260 nm/230 nm (DNA/humic acid). Comparative analysis revealed the considerable variations in yield and purity of DNA obtained by the different methods. As depicted in Figs. 2 and 3 , method 1 gave DNA with A_{260}/A_{280} ratios close to optimum, while A_{260}/A_{230} ratios indicating comparatively reduced humic content was obtained by method 2.

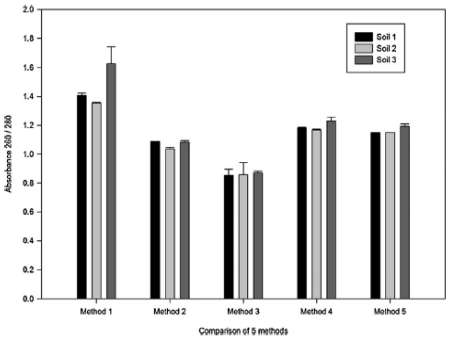

Fig. 2. Purity of DNA (A_{260}/A_{280}) from different Mangrove soils by five methods.

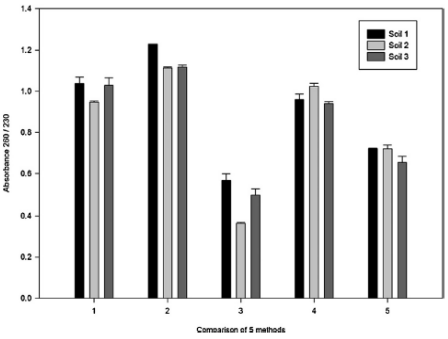

Fig. 3. Purity of DNA (A_{260}/A_{230}) from different Mangrove soils by five methods.

3.3. Visualization of community DNA on agarose gel

Although the quantity of total DNA isolated by the different methods varied considerably, the extracted DNA were of high molecular weight, which was also a DNA quality indicator. The spectophotometric data were supported by the agarose gel analysis. (Fig. 4). Lower DNA concentration obtained by method 2 was clearly visible in the gel picture.

3.4. 16S rRNA gene amplification

PCR amplification of 16S rRNA gene was successful only with DNA obtained by method 2 (Fig. 5), which had comparatively reduced humic acid contaminants.

4. Discussion

To isolate high molecular weight, contaminant free and PCR amplifiable DNA, five different methods of total DNA isolation were utilised. Various environmental DNA isolation protocols have been previously studied [10,11]. Extracting pure DNA from environmental samples is practically as important as yield, however it is also one of the most complex problems associated with the application of molecular techniques on environmental samples. Heterogeneous nature of the environmental samples requires each extraction procedure to be precise and optimised for every soil sample. Most DNA extraction procedures co-extract humic acids, pigments, heavy metals, and other contaminants. Humic

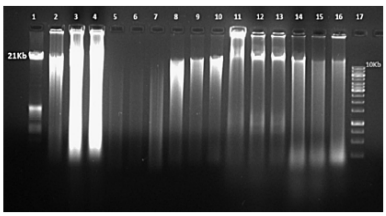

Fig. 4. Agarose gel electrophoresis of metagenomic DNA from three Mangrove soils by five methods. Lane 1–21 Kb ladder (Thermo Scientific, MA, USA) lane 17–1 Kb ladder, lane 2–4: DNA isolated by method 1, lane 5–7: DNA isolated by method 2, lane 8–10: DNA isolated by method 3, lane 11–13: DNA isolated by method 4, lane 14–16: DNA isolated by method 5.

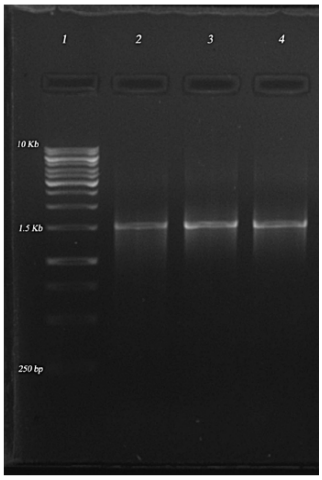

Fig. 5. 16S rDNA amplification profile of DNA by method 2. Lane 1 Kb ladder (Thermo Scientific, MA, USA), lane 2–4 – 16S rDNA amplicon.

contaminants due to their three dimensional structure and functional reactive groups bind with organic compounds [12] and are therefore one of the major problems associated with any soil community DNA isolation. Depending on soil types, crude DNA extracts can be contaminated by approximately 0.7–3.3 μg/μL of humic acid [13]. In addition, due to similar physicochemical properties with nucleic acid they easily co-precipitate with nucleic acid. These contaminants may not only hinder PCR reactions acting as inhibitor, but also can degrade the DNA during storage. Humic acid may through specific binding to DNA inhibit amplification in PCR reactions by limiting the amount of available template [14]. Purification of DNA employing polyvinylpolypyrrolidone, embedding DNA in agarose blocks followed by successive washing steps or by using sephadex columns can help improve quality of soil DNA and subsequent PCR amplification [15–17]. The aim of any extraction protocol is to succeed in obtaining genomic DNA which is a representative of the microbial diversity present within a soil. However different extraction procedures targets only specified group of microbiota present within a soil which results in biased estimates of DNA quantity, evidently due to differences in individual component steps, sorption of DNA to soil particles, DNA degradation or co-extraction of inhibitors [18–20] suggesting that additional measures should be considered when divergent soil types are compared or when comprehensive community analysis is required. SDS based cell lysis is the most widely used DNA extraction method, whereby DNA yield is more compared to freeze thawing and use of other detergents [21]. Physical treatments such as grinding, sonication and bead beating homogenises soil particles and can access individual microbial cells within a sample but with greater possibility of DNA shearing. Previous studies revealed that a combination of chemical and mechanical lysis can yield twice the amount of DNA than by any single method alone [20]. In the present study mechanical disruption of cell wall by grinding with liquid nitrogen and bead beating (method 2 and 3) resulted in increased DNA shearing, when compared to the gentle freeze-thawing method 4. Although the liquid nitrogen method yielded the shortest DNA fragments, it also has reduced amounts of contaminants. Consequently a combination of chemical lysis along with mild physical methods can greatly influence the total DNA content in terms of quantity and quality. Despite the shearing of DNA in all 3 soil samples employing liquid nitrogen extraction technique, they yielded 16S rRNA gene amplification using a single set of primer without the addition of any PCR enhancers or additives, thereby suggesting the suitability of the method in diverse soils and also in diversity studies. Commercial DNA extraction kits are now commonly used for extraction of high molecular weight DNA from complex habitats. Studies evaluating various commercial kits to other methods have shown that DNA yield and purity vary based on methodology and soil type. The mechanism of purification of these kits is based on the adsorption and desorption of the nucleic acids in presence of chaotropic salts [22] which results in contaminants free DNA but the quantity of DNA obtained will be less compared to classical method of DNA extraction. Previous studies recommended that slight modification of protocols employing commercial kits or a combination of classical isolation methods followed by purification of DNA using commercial kits can greatly affect the quantity and quality of the isolated DNA [23,24,25]. In the present study maximum DNA yield was obtained in lysozyme-freeze-thawing protocol (method 4), although the presence of residual amounts of humic and protein contaminants hindered PCR reaction.

In conclusion all methods yielded an acceptable amount of DNA, but were not suitable for further downstream processing, except that obtained by method 2. Previous studies suggested that addition of carriers and polyvalent polymers helps to reduce DNA loss due to adsorption and degradation [26]. Similarly in method 2, addition of skim milk prior to addition of extraction buffer may have helped to retain high quality DNA. Our results suggested that addition of skim milk helped to extract DNA amenable to PCR with the three soil samples tested which is in agreement with previous reports [5,27–29] as skim milk by acting as a carrier can reduce the adsorption and degradation of nucleic acids. On the other hand precipitating DNA with isopropanol improved DNA yield compared to the original study which used absolute alcohol instead [5]. Observations from the present study suggest that starting with a low gram weight of soil for DNA isolation as seen in method 2 and addition of skim milk during extraction can possibly help to reduce the humic contaminants, which would otherwise interfere with all other downstream processing of DNA, like amplification and cloning to name a few.

Acknowledgment

This work is supported by grants from University Grants Commission (major project) vide F.No. 41/527/2012 (SR). A portion of the research was also supported by Cochin University of Science and Technology.

Reference

[1] J. Handelsman, Metagenomics: application of genomics to uncultured microorganisms, Microbiol. Mol. Biol. Rev. 68 (2004) 669–685.
[2] A. Ogram, G.S. Sayler, T. Barkay, The extraction and purification of microbial DNA from sediments, J. Microbiol. Methods 7 (1987) 57–66.
[3] W.E. Holben, J.K. Jansson, B.K. Chelm, J.M. Tiedje, DNA probe method for the detection of specific microorganisms in the soil bacterial community, Appl. Environ. Microbiol. 54 (1988) 703–711.
[4] J. Zhou, M.A. Bruns, J.M. Tiedje, DNA recovery from soils of diverse composition, Appl. Environ. Microbiol. 62 (1996) 316–322.
[5] T. Volossiouk, E.J. Robb, R.N. Nazar, Direct DNA extraction for PCR-mediated assays of soil organisms, Appl. Environ. Microbiol. 61 (1995) 3972–3976.
[6] D. Dong, A. Yan, H. Liu, X. Zhang, Y. Xu, Removal of humic substances from soil DNA using aluminium sulfate, J. Microbiol. Methods 66 (2006) 217–222.
[7] Y.L. Tsai, B.H. Olson, Rapid method for direct extraction of DNA from soil and sediments, Appl. Environ. Microbiol. 57 (1991) 1070–1074.

[8] P.K. Siddhapura, S. Vanparia, M.K. Purohit, S.P. Singh, Comparative studies on the extraction of metagenomic DNA from the saline habitats of coastal Gujarat and Sambhar Lake, Rajasthan (India) in prospect of molecular diversity and search for novel biocatalysts, Int. J. Biol. Macromol. 47 (2010) 375–379.
[9] S. Shivaji, N.V. Bhanu, R.K. Aggarwal, Identification of Yersinia pestis as the causative organism of plague in India as determined by 16S rDNA sequencing and RAPD-based genomic fingerprinting, FEMS Microbiol. Lett. 189 (2000) 247–252.
[10] T.O. Delmont, P. Robe, I. Clark, P. Simonet, T.M. Vogel, Metagenomic comparison of direct and indirect soil DNA extraction approaches, J. Microbiol. Methods 86 (2011) 397–400.
[11] M.K. Purohit, S.P. Singh, Assessment of various methods for extraction of metagenomic DNA from saline habitats of coastal Gujarat (India) to explore molecular diversity, Lett. Appl. Microbiol. 49 (2009) 338–344.
[12] F.J. Stevenson, Stability constants of Cu^{2+}, Pb_2^{+}, and Cd_2^{+} complexes with humic Acids, Soil Sci. Soc. Am. J. 40 (1976) 665–672.
[13] C.C. Tebbe, W. Vahjen, Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast, Appl. Environ. Microbiol. 59 (1993) 2657–2665.
[14] K.L. Opel, D. Chung, B.R. McCord, A study of PCR inhibition mechanisms using real time PCR, J. Forensic Sci. (2009) .
[15] A. Frostegard, S. Courtois, V. Ramisse, S. Clerc, D. Bernillon, F. LeGall, P. Jeanin, X. Nesme, P. Simonet, Quantification of bias related to the extraction of DNA directly from soils, Appl. Environ. Microbiol. 65 (1999) 5409–5420.
[16] D. Moreira, Efficient removal of PCR inhibitors using agarose-embedded DNA preparations, Nucleic Acids Res. 26 (1998) 3309–3310.
[17] D.N. Miller, J.E. Bryant, E.L. Madsen, W.C. Ghiorse, Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples, Appl. Environ. Microbiol. 65 (1999) 4715–4724.
[18] M.L. Feinstein, W.J. Sul, C.B. Blackwood, Assessment of bias associated with incomplete extraction of microbial DNA from soil, Appl. Environ. Microbiol. 75 (2009) 5428–5433.
[19] E.L. Madsen, A critical analysis of methods for determining the composition and biogeochemical activities of soil microbial communities *in situ*, Soil Biochem. 9 (1996) 287–370.
[20] M.I. More, J.B. Herrick, M.C. Silva, W.C. Ghiorse, E.L. Madsen, Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment, Appl. Environ. Microbiol. 60 (1994) 1572–1580.
[21] J.T. Trevors, H. Lee, S. Cook, Direct extraction of DNA from soil, Microb. Releases 1 (1992) 111–115.
[22] J.P. Gray, R.P. Herwig, Phylogenetic analysis of the bacterial communities in marine sediments, Appl. Environ. Microbiol. 62 (1996) 4049–4059.
[23] L.R. Gutiérrez-Lucas, J.J. Montor-Antonio, N.G. Cortés-López, S. del Moral, Strategies for the extraction: purification and amplification of metagenomic DNA from soil growing sugarcane, Adv. Biol. Chem. 4 (2014) 281–289.
[24] F.M. Lakay, A. Botha, B.A. Prior, Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils, J. Appl. Microbiol. 100 (2006) 1365–1373.
[25] F. Martin-Laurent, L. Philippot, S. Hallet, R. Chaussod, J.C. Germon, G. Soulas, G. Catroux, DNA extraction from soils: old bias for new microbial diversity analysis methods, Appl. Environ. Microbiol. 67 (2001) 2354–2359.
[26] D.T. Denhardt, A membrane-filter technique for the detection of complementary DNA, Biochem. Biophys. Res. Commun. 23 (1966) 641–646.
[27] Y.T. Hoshino, N. Matsumoto, An improved DNA extraction method using skim milk from soils that strongly adsorb DNA, Microbes Environ. 19 (2004) 13–19.
[28] M.D. Garcia-Pedrajas, B.W. Bainbridge, J.B. Heale, E. Perez-Artes, R.M. Jimenez-Diaz, A simple PCR-based method for the detection of the chickpea-wilt pathogen *fusarium axysporum* f.sp. ciceris in artificial and natural soils, Eur. J. Plant Pathol. 105 (1999) 251–259.
[29] S. Ikeda, H. Tsurumaru, S. Wakai, C. Noritake, K. Fujishiro, M. Akasaka, K. Ando, Evaluation of the effects of different additives in improving the DNA extraction yield and quality from andosol, Microbes Environ. 23 (2008) 159–166.

Bio-Genetics Journal
Vol. 1, No. 1 (2013): 1-4
Research Article
Open Access

ISSN: 2347-7407

Culture independent analysis of the soil microbiome to assess microbial diversity of mangrove soil

Harisree P. Nair, Helvin Vincent and Sarita G. Bhat*

Department of Biotechnology, Cochin university of Science and Technology, Kerala, India

* Corresponding author: Sarita G. Bhat; e-mail: saritagbhat@gmail.com

Received: 11 October 2013 Accepted: 24 October 2013 Online: 02 November 2013

ABSTRACT

Mangroves ecosystems are recognized to be highly productive as they harbor vast diversity of marine microorganisms. Analysis of microbial diversity from these ecosystems will facilitate identification of new microbes having specificity for a variety of applications. A culture-independent assessment of microbial diversity of Mangalavanam Mangrove soil was performed by sequence analysis of 16S rDNA clone library. Universal primers were used to amplify 16S rRNA gene from the isolated community DNA. A total of 34 clones were selected from the library and sequence analysis of these clones revealed that bacteria present in the study site belonged to 6 major phyla (Acidobacteria, Firmicutes and Proteobacteria, Actinobacteria, WS3 and Planctomycetes) of bacterial domain with a dominance of deltaproteobacterium class. Further studies are required to understand the diversity in more detail. This study will open new doors in exploring the Viable But Not Culturables (VBNCs) in the Mangalavanam mangrove sediments.

Keywords: metagenome; microbial diversity; VBNCs; 16S rRNA; mangroves

INTRODUCTION

Microbes are the most ubiquitous organisms on earth, present in all habitats, including soil, sediment, water and in extreme environments, playing a key role in the biogeochemical cycles of the biosphere and representing a huge reservoir of novel biomolecules. It is estimated that less than 1% of the total microbial communities from environmental samples are readily cultivable by standard microbiological methods [1]. The uncharacterized ones remain as unexplored reservoir of metabolic and genetic diversity.

Mangrove ecosystem serves as unique ecological niche, with a variety of microbes playing important roles in nutrient recycling and various other ecological processes, thereby requiring a thorough exploration of their microflora. Mangrove soils are generally nutrient rich with high microbial diversity. Similarly Mangalavanam, a conserved mangrove ecosystem located in Kochi, Kerala, India, known as the "Green lungs of Kochi" may turn to be a treasure trove of unexplored microflora.

In this study, the 16S rDNA clone libraries constructed was analyzed to explore the bacterial diversity of Mangalavanam mangrove soil. The partial 16S rRNA gene sequences were compared with public data bases and taxonomic hierarchy was assigned. The present result adds considerably to our understanding of microbial diversity in mangrove sediments.

MATERIALS AND METHODS

Study site and sample collection

Mangrove soil samples were collected from Mangalavanam, Kochi, Kerala, India (9° 54′ 0″ N, 76° 18′ 0″ E). Soil samples were collected by removing the surface leaf litter and collecting the top soil sample. Samples were transferred with a sterilized spatula and the samples were placed in sterile McCartney specimen bottles and were frozen at -20^{0}C until further analysis.

Total community DNA extraction

Community DNA was isolated by the slightly modified method employing liquid nitrogen [2]. Wherein 0.25 g of soil was ground with liquid nitrogen, using sterile mortar and pestle until fine powder remains. The powdered soil was suspended in 0.5 mL of skim milk powder solution (0.1 g skim milk in 25 mL of water), vortexed well and was centrifuged for 10 min at 12,000

x *g* at 4⁰C (Sigma, Germany). To the supernatant 2 mL of SDS extraction buffer (0.3% SDS in 0.14 M NaCl, 50 mM sodium acetate [pH 5.1] was added and vortexed well to mix. An equal volume of Tris-saturated phenol solution was added, and phases were vortexed to mix for 2 min at room temperature. Aqueous phase was separated by centrifugation at 12,000 x *g* for 10 min and the nucleic acids precipitated with 1 volume of ice cold isopropanol at -20⁰C. The DNA was pelleted by centrifugation at 12,000 x *g* for 10 min, the pellet washed twice with cold 70% ethanol and air dried. Pellet was dissolved in sterile deionized water and stored at -20⁰C.

PCR amplification of 16S rRNA gene and library construction

PCR amplification of 16S rRNA gene (1.5 Kb) was performed in a thermal cycler (Biorad, USA) using universal primers [3]. The amplicons were separated electrophoretically in 1% agarose gel [4], visualized using ethidium bromide under ultraviolet illumination and gel pictures were captured using gel documentation system (Syngene, USA). PCR products were purified using PCR clean-up kit (Machery-Nagel, Germany) and were ligated into TA cloning vector pTZ57R/T (Fermentas, India) and transformed onto *E. coli* JM109 host cells. The clones were screened for α-complementation by using X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (isopropyl-β-D thiogalactoside) and positive clones were confirmed by reamplification of 16S rDNA inserts.

DNA sequencing and *in silico* analysis

The PCR amplicons of the 16S r RNA genes were sequenced by Sanger's Dideoxy method using ABI 3730 Excel and compared with public databases using BLAST [5]. The sequences were compiled and aligned using Clustal W program [6] using BioEdit software [7]. The phylogenetic tree was constructed using the Neighbor-Joining method [8] with 1000 resampling bootstrap using MEGA software [9]. Taxonomical hierarchy was assigned to the sequence using RDP Naive Bayesian rRNA Classifier Version 2.5 [10].

RESULTS AND DISCUSSION

DNA extraction and library construction

Soil Metagenomic DNA isolated from Mangalavanam mangrove was used as template for PCR amplification of 16S rRNA gene. The purified 16S rRNA gene was used to construct a clone library and 34 clones from this library were selected for further analysis.

Sequencing and phylogenetic analysis

Sequences of 16S rRNA gene from the selected clones were compared with those in the GenBank data base. Taxonomic positions were assigned to the sequence data using RDP classifier. The phylogenetic tree shown in Fig: 1 is based on the 16S rRNA sequences of the 34 clones. The number at the nodes of the phylogenetic tree are percentages indicating the levels of bootstrap support based on a neighbour-joining analysis of 1,000 resampled data sets. The 16S rRNA gene sequences reported in this study were submitted to the GenBank database with accession numbers JX465646 to JX465653, JX852421 to JX852429 and KC143083-KC143099.

Among the 34 clones analyzed in this study, 15 clones showed maximum similarity to Phylum Proteobacteria which are taxonomically identified as class alpha, beta, delta, Gammaproteobacteria and eight clones were similar to Firmicutes, the library also contained minor representatives from phylum Acidobacterium, Actinobacteria, WS3 and Planctomycetes. The rest of the clones were found similar to uncultured bacterium, i.e. which doesn't fall into any taxonomic hierarchy. In the phylogenetic tree similar phyla claded together and all were distant from the outgroup. This in turn suggests that soil bacterial diversity is immense and much is still to be explored.

In silico analysis of the 34 clones from the16S rDNA library revealed that they belonged to 6 major phyla of bacterial domain, namely Acidobacteria, Firmicutes and Proteobacteria, Actinobacteria, WS3 and Planctomycetes. The dominant phylum in the present study was proteobacteria. Within this large phylum, four proteobacterial subdivisions that were present included alpha, beta, delta and gamma proteobacteria. Previous studies had reported that Proteobacteria represent the largest and metabolically diverse group of soil microbes [11]. Occurrence of diverse phylotypes from various environments like Sundarbans mangrove sediment [12] marine sediments [13, 14, and 15] have already been reported.

Sequencing analysis of the clones revealed the dominance of delta proteobacterial sequences in the library. Sulfur cycle has proved to be the important biogeochemical factor that regulates the flow of electrons along the microbial systems in marine ecosystems. Most of the identified deltaproteobacterial clones from the library showed similarity to the sulfur and sulphate reducing bacteria obtained from different marine sediments [12]. All this facts further supports our findings. In short, Mangalavanam mangrove sediment houses phylogenetically diverse population of bacterial domain.

CONCLUSION

In this study, for the first time culture independent analysis was performed on the sediment from the conserved Mangalavanam mangrove ecosystem. The present analysis revealed the diverse bacterial population in these soils with 6 major phyla of bacterial domain. The study revealed the likelihood of other unknown bacteria in mangrove sediments, the role they play in ecosystem maintenance are also unknown, implying that many more prokaryotic lineages await discovery with the advent of modern molecular methods. Further studies are essential to elucidate the bacterial diversity of the soil in detail, to throw light on this hitherto unexplored hidden treasure of novel yet to be cultured bacteria.

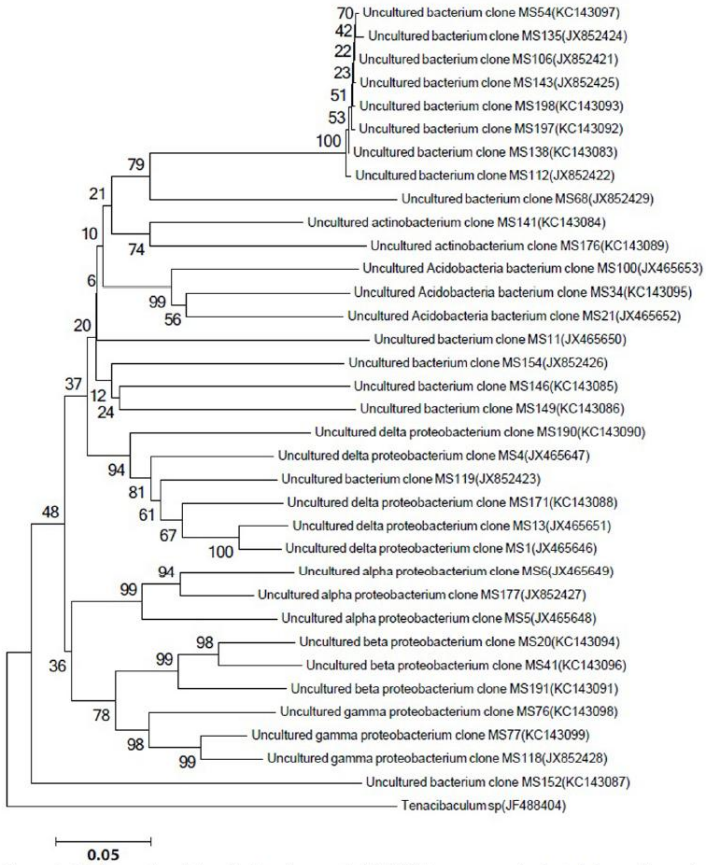

Figure 1. Phylogenetic relationship based on partial 16SrDNA sequences of selected clones. The analysis involved 34 nucleotide sequences. *Tenacibaculum sp.* (accession number JF488404) was used as outgroup. Accession numbers are given in parentheses.

Acknowledgments

First author acknowledges the financial support (Junior Research Fellowship) from Cochin University of Science and Technology. The authors acknowledge the support from University Grants commission (major project) vide F.No. 41/527 /2012 (SR).

REFERENCES

1. Handlesman J (2004) Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev.* 68:669-685
2. Volossiouk T, Robb JE and Nazar RN (1995) Direct DNA Extraction for PCR-Mediated Assays of Soil Organisms. *Appl Environ Microbiol.* 61:3972–3976
3. Shivaji S, Bhanu NV and Aggarwal RK (2000) Identification of *Yersinia pestis* as the causative organism of plague in India as determined by 16S rDNA sequencing and RAPD-based genomic fingerprinting. FEMS Microbiol Lett.189:247-252
4. Sambrook J, Fritsch EF and Maniatis T (2000) Molecular Cloning: a laboratory manual. 3rd ed. Cold Spring Harbor Laboratory Press, New York
5. Altschul SF, Gish W, Miller W et al.(1990) Basic local alignment search tool. *J Mol Biol* . 215: 403-10
6. Thompson JD, Gibson TJ, Plewniak F et al.(1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882
7. Hall TA (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41: 95-8
8. Saitou N and Nei M(1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol and Evol.* 4:406-425
9. Tamura K, Dudley J , Nei M et al.(2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol and Evol.* 24:1596-1599
10. Wang Q, Garrity GM, Tiedje JM et al.(2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol.* 73(16):5261-7

11. Anne M, Lee R and Mostafa S. (2009) Abundance, composition, diversity and novelty of soil *Proteobacteria*. *ISME J*. 3:992-1000.
12. Ghosh A Nirmalya D, Amit B et al.(2010) Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. *Saline Systems*, 6:1: 1-11
13. Gray JP and Herwig RP(1996) Phylogenetic analysis of the bacterial communities in marine sediments. *Appl Environ Microbiol*. 62:4049-4059
14. Urakawa H, Kita-Tsukamoto K and Ohwada K (1999) Microbial diversity in marine sediments from Sagami Bay and Tokyo Bay, Japan, as determined by 16S rRNA gene analysis. *Microbiology*. 145:3305-3315
15. Polumenakous PN, Bertilsson AT and Stephanou EG (2005) Bacterial community composition in different sediments from the Eastern Mediterranean Sea: a comparison of four 16S ribosomal DNA clone libraries. *Microb Ecol*. 50:447-462

*****