CHARACTERIZATION AND APPLICATIONS OF TWO PROTEASE ENZYMES OBTAINED BY CULTURE DEPENDENT AND INDEPENDENT APPROACHES FROM MANGROVE SEDIMENTS

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

HELVIN VINCENT
Reg. No. 3369

Under the guidance of

Dr. Sarita G. Bhat

MICROBIAL GENETICS LABORATORY
DEPARTMENT OF BIOTECHNOLOGY
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN - 682 022, KERALA, INDIA.

MAY 2014

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
Associate Professor

07. 05. 2014

CERTIFICATE

This is to certify that the research work presented in the thesis entitled "**Characterization and applications of two protease enzymes obtained by culture dependent and independent approaches from mangrove sediments"** is based on the original research work carried out by Ms. Helvin Vincent under my guidance and supervision at the Department of Biotechnology, Cochin University of Science and Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, and that no part thereof has been presented for the award of any degree.

Dr. SARITA G. BHAT
Supervising Guide

DECLARATION

I hereby declare that the thesis entitled "**Characterization and applications of two protease enzymes obtained by culture dependent and independent approaches from mangrove sediments"** is the authentic record of research work carried out by me at the Department of Biotechnology, Cochin University of Science and Technology for my doctoral degree, under the supervision and guidance of Dr. Sarita G. Bhat, Associate 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 or diploma, associateship or other similar titles or recognition.

Cochin-22

07. 05. 14 **Helvin Vincent**

Acknowledgement

This is a note of acknowledgement to all my kith and kin, who have contributed in one way or the other towards the successful completion of my Ph.D thesis.

First and foremost, let me thank my Almighty, for his plans in my life which have enabled me to pursue anything and everything I am till now. His unseen hands carried me all throughout this venture, giving me strength, wisdom, courage and patience, to overcome all the hurdles I had to pass through and to execute all my responsibilities to a good extent.

Words cannot express my gratitude towards my mentor, Dr. Sarita G Bhat, for giving me an opportunity to do research in her lab and for her excellent supervision during the course of this research work. I would not have thought of a doctoral degree if Ma'am had not called me to join her lab. She gave me all freedom to plan my work and her appreciation and encouragement even for my smallest achievement needs special mention. She introduced the concept of metagenomics to me and led me through the experiments of molecular cloning and library construction. She has being very much patient and kind to me in all my flaws and her care and concern in all personal matters is truly appreciable. She has never hesitated the numerous leaves I had asked her at times of my emergency. Thank you ma'am for being so lenient and generous to me and this thesis is my humble dedication to you. Also I apologise you for my various mistakes and the botherations I had caused you.

I am grateful to Dr. Padma Nambisan, Head, Dept. of Biotechnology for her valuable suggestions and advice during the doctoral committee presentations. I express my profound gratitude to Prof. (Dr.) C.S. Paulose for his encouragement and motivation throughout the research period. The valuable suggestions and backup offered by Prof. Dr. M. Chandrasekharan, Dr. Elyas K.K. and Dr. E. Vijayan are gratefully acknowledged.

I am thankful to Dr. P. M. Sherief, College of Fisheries, Panangad and Dr. T. V. Sankar, Central Institute of Fisheries Technology for serving as the external experts of my SRF assessment committee.

I express my heartfelt gratitude to all the teachers of my school days as well as college days, especially Delna miss of St. Antony's H.S.S and Sr. Francis Ann, Gladis miss, Reema miss, Meera miss, Mini miss and Romily miss of St.Teresa's College, for their

prayers and blessings during all stages of my life. Their inspirations and encouragements in my academic career is greatly acknowledged.

I am thankful to Cochin University for providing the financial support in the form of Junior and Senior Research Fellowship. The financial assistance provided by University Grants Commission (UGC) in the form of Maulana Azad National Fellowship (MANF) is also acknowledged.

I thank all the present and past office staff of the Department of Biotechnology for their prompt support and help. My special thanks to the higher authorities and administrative staff of Cochin University of Science & Technology for their help and co-operation. The services rendered by the staff of academic, audit, cash and examination sections is thankfully remembered with special mention to Santhakumari madam for her constant care and concern.

My special thanks to Dr. I.S. Bright Singh, National Centre for Aquatic Animal Health for providing the PCR facility in his lab. I am grateful to Dr. Divya, St.Xavier's College for helping me with bioinformatic tool for phylogenetic diversity analysis.

I feel short of words to express my heartfelt gratitude to my companions at the Microbial Genetics lab. The cordial and stimulating relationship extended by each of them provided a relaxed working environment. My first companion at lab, Dr. Jeena Augustine, motivated me at all times through her scientific discussions and unconditional love. I have never met such an innocent character in my life. Dr. Raghul Subin helped me with various bioinformatic tools and had always extended a brotherly concern.

Ms. Smitha S had been my companion at all stages of the research career, right from the time of Ph.D registration. We shared together some of the wonderful stages of our life and the tensions during the final stages of our research work. I am always indebted to her for sparing me one of her isolates, BTKM4, for my comparative study. I was fortunate to have her as my companion as she very well passified me whenever my energy levels went down. She explained me the basics of protein work and helped me at all stages of my protein work. Thank you Smitha for being a supporting hand at all hard times, your companionship will always be cherished.

Mr. Siju M. Varghese has always inspired me with his perfection in molecular biology experiments as well as bioinformatic tools. My teacher turned colleague, Vijaya miss, had always been a great moral support. She gave me the strength and courage to move on especially doing the stressed stages of thesis writing. Thank you Miss for your

valuable prayers and I am privileged to be your student once. I appreciate Ms. Linda Louis for her valuable practical suggestions and for the mental support through her spiritual views and prayers. Her entertainment by free talk-time of stories and jokes are memorable.

Ms. Harisree P. Nair has been my right hand in all the experiments right from the time she joined the lab as trainee. The relentless and selfless help extended by her needs special appreciation and acknowledgement which helped me to achieve my goals easier and with a good pace. Her helping hands enabled me to keep my work on track even in my absence. She helped me with computer related tasks like sequence submission, tree construction, reference formatting and so on. Her willingness to help even during the late hours of night needs special mention. Her moral support and practical suggestions helped me to tide over many hard times. Thank you dear for all your help and your sisterly affection will always be cherished.

I specially thank Ms. Mridula V.G. for her loving support and affection, Ms. Laxmi M, Mr. Noble Kurian and Ms. Nandita for their care and concern at all stages of thesis writing, Ms. Anu and Ms. Lakshmy Ramakrishnan for their friendship and cooperation. Their effort towards proof reading my thesis chapters as well as formatting the references is greatly acknowledged.

I also express my gratitude to Ms. Emil, Ms. Nesgin Dias, Mr. Jayanath, Mr. Satheesh Kumar MK, Mr. Dominic and Mr. Jijith for their support during different stages of my work. I sincerely acknowledge my senior colleagues Dr. Bernard Rajeev, Dr. Manjusha S. and Dr. Archana Kishore for their support and valuable suggestions in my research work.

I am thankful to Dr. Manzur Ali PP and Dr. Sapna K of Immuno Technology lab for their valuable practical suggestions in the protein work. Their willingness to find solutions to any problem regarding protein work, even over the phone is greatly acknowledged. Special appreciations to Dr. Manzur Ali P.P. for informing me about the UGC-Maulana fellowship. I am thankful to Ms. Rekha mol K.R. for her affection, care, companionship and for her valuable suggestions. Dr. Abraham Mathew always inspired me with his hard work and enthusiasm in the research. I thank Mr. Ramesh Kumar for his care and helpful advices.

My heartfelt thanks to seniors of the Microbial Technology lab, Dr. Soorej M. Basheer, Dr. Sreeja, Dr. Jissa G Krishnan, Dr. Madhu, Dr. Jasmin, Dr. Lailaja and Dr. Beena for their valuable suggestions, encouragement and unconditional help during

various stages of my research work. I thank Mr. Sijo, Mr. Ajith M. Thomas, Dr. Roseline Alex, Dr. Karthikeyan, Ms. Manjula P, Mr. Sajan, Mr. Doles, Mr. Cikesh P.C., Ms. Bindiya and Ms. Tina for their friendship and support.

I am grateful to Dr. Jikku Jos of Plant Biotechnology lab for her co-operation and for sharing the worries being my Co-Maulana fellow. I thank Dr. Jasmine Koshy, Ms. Aneeta, Ms. Thresia Regimol, Ms. Ummu Habeeba, Ms. Elizabeth, Ms. Sudha Hariharan, Ms. Soumya, Ms. Kiran Lakshmi, Ms. Arrinya, Ms. Anala, Ms. Nayana and Ms. Aiswarya for their love, encouragement and support throughout the work.

I wish to thank all my seniors and friends of Neurobiology lab, Dr. Binoy Joseph, Dr. Balarama Kaimal S, Dr. Akash K George, Dr. Gireesh G, Dr. Finla Chathu, Dr. Reas Khan S, Dr. Savitha Balakrishnan, Dr. Remya Robinson, Dr. Amee, Dr. Anu Joseph, Dr. Pretty Mary Abraham, Dr. Jobin Mathew, Dr. Peeyush Kumar, Dr. Sherin Antony, Dr. Anju TR, Dr. Jes Paul, Dr. Nandhu MS , Dr. Korah Kuruvila , Dr. Smijin Soman, Dr. Anita, Dr Chinthu, Dr. Jayanarayan, Dr. Shilpa, Ms. Roshini, Mr. Naigil George and Mr. Ajayan for their cooperation and friendship.

I thankfully remember all the M.Sc. students of the Department especially Rinu, Siktha and Preethi for their friendship and affection.

The presence of our little kids Aardra, Anna, Aami, Diya, Esa, Anthony, Alloos, Appu, Anwi, Adrija, Jiya, Rukku and Vasu is fondly remembered with sepecial mention to Pippi for her small helps like filling up the tip boxes and to her little sister Nunnu for her innocent smile and kiddy talks which were relaxations from the tensions of thesis writing.

My warm gratitude to my dearest friends Anila, Anisha, Ann and Aswathy for their friendship, care, prayers and support.

Words cannot express my gratitude to my dear parents for their love, motivations and prayers from my childhood. I am deeply indebted to them for their sustained encouragement and effort which made me accomplish my goals and for taking care of my kids during my busy schedule. My sister, Heslin, has always shown keen interest in my research, and was always eager to know about my results. Thank you dear for your prayers, affection and concern.

I feel short of words to express my love and gratitude to my husband, Mr. Joby. Achayan's unconditional love, inspirations and patience have been instrumental in all my

accomplishments. I would not have completed even my graduation without his support. I feel privileged to be his better half and I am indebted to him for his cooperations throughout my research career, right from sample collection upto designing of the thesis cover page. I thank him for being with me and for covering up all my tensions, frustrations and emotions.

My little kids, Kunjoos and Appoos, cooperated extremely throughout my entire research period. They adjusted a lot with my research and sacrified their happy moments during my hectic times of thesis work. My Kunjoos always showed her concern about my thesis completion. I can't express how much I owe them….I truly treasure them as my Lord's greatest gifts to me…..

I express my sincere gratitude to my kids' caretakers, Alice aunty, Gracy aunty, Mary chechi, Sindhu chechi, Binni chechi and Elizabeth chechi for taking care of my kids in my absence. They helped me a lot to relieve my tensions about my Kunjoos and Appoos.

I am thankful to my in-laws, Ammachi, Biju, Riju, Deepa and Ofelia and the little ones, Eleesha and Evaan, for their affection and moral support. I specially thank Ammachi for her prayers and for the scooter she bought me, which made my conveyance easier.

I take this opportunity to thank all my relatives- uncles, aunts and cousins for their love, prayers and encouragement especially to my Densy aunty for her special affection and for laying strong foundations for my academic career. I extend my prayers to my grandmother for her soul strenghthens me in every walk of my life.

I thank all my well wishers who I may have unknowingly not mentioned, for their generous help and prayers for the successful completion of my research work.

Helvin Vincent

CONTENTS

LIST OF TABLES AND FIGURES

ABBREVIATIONS

%	-	Percentage
~	-	Approximately
<	-	less than
>	-	greater than
°C	-	Degree Celsius
APS	-	Ammonium persulfate
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
DEAE	-	Diethylaminoethyl
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
GRAS	-	Generally Recognized As Safe
h	-	Hours
HCl	-	Hydrochloric acid
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
NA	-	Nutrient agar
NaCl	-	Sodium chloride
NaOH	-	Sodium hydroxide
NB	-	Nutrient broth
NCBI	-	National Center for Biotechnology Information
ng	-	Nanogram
nm	-	Nanometer
No.	-	Number
OD	-	Optical density
OD_{230}	-	Optical density at 230 nm
OD_{260}	-	Optical density at 260 nm
OD_{280}	-	Optical density at 280 nm
ORF	-	Open reading frame
OTU	-	Operational taxonomic unit
PAGE	-	Polyacrylamide gel electrophoresis
PCMB	-	p-Chloromercuribenzoic acid
PCR	-	Polymerase chain reaction

pH	-	Power of Hydrogen
pI	-	Isoelectric point
rpm	-	Revolutions per minute
rRNA	-	ribosomal Ribonucleic acid
rRNA	-	Ribosomal RNA
s	-	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
TLCK	-	Tosyl lysine Chloromethyl Ketone
TPCK	-	Tosyl phenylalanyl chloromethyl ketone
UF	-	Ultra filtration
UV	-	Ultraviolet
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
ZB	-	Zobell marine broth
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
Al	-	Aluminium
Ba	-	Barium
Ca	-	Calcium
Cd	-	Cadmium
Co	-	Cobalt
Cu	-	Copper
Fe	-	Iron
Mg	-	Magnesium
Mn	-	Manganese
Na	-	Sodium
Ni	-	Nickel
Zn	-	Zinc

Chapter - **1**

INTRODUCTION

Enzymes as biocatalysts carry out large number of chemical reactions and are commercially exploited in various industries (Kumar and Takagi, 1999). The estimated value of the worldwide sales of industrial enzymes was US $2.5 billion in 2009 (Rajan, 2004). Among industrial enzymes, proteases represent one of the three largest groups and account for 60% of the overall worldwide sale of enzymes (Rao *et al*., 1998).

Proteases are hydrolytic enzymes which catalyse the cleavage of specific peptide bonds in their target proteins. They are essential for various cellular and metabolic processes, such as sporulation and differentiation, cell migration and invasion, protein turnover, maturation of enzymes and hormones and maintenance of the cellular protein pool. Proteases execute a large variety of functions from the cellular level to the organ and organism level, to produce cascade systems like homeostasis and inflammation; and complex processes involved in the normal physiology as well as in abnormal pathophysiological conditions. They have also gained considerable attention in the industrial community mainly in detergent, pharmaceutical, food, diagnostics, leather, waste management and silver recovery sectors (Gupta *et al*., 2002).

According to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, proteases are classified in subgroup 4 of group 3 (hydrolases) (IUBMB, 1992). Proteases are broadly divided into two major groups, exopeptidases and endopeptidases, depending on their site of action. Exopeptidases cleave the peptide bond at the amino or carboxy termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate. Proteases are further classified into seven prominent groups, aspartic

proteases, cysteine proteases, glutamic proteases, metalloproteases, serine proteases, threonine proteases and asparagine peptide lyases based on the functional group present at their active site (Hartley, 1960; Fujinaga *et al.*, 2004; Rawlings *et al.*, 2011).

Proteases are also classified based on their structure and amino acid sequence similarity, those with significant similarities in amino acid sequences are grouped into families and families with related structures are grouped into clans (Argos, 1987) and are included in the MEROPS database, which is a database specifically dedicated for peptidases. Each family of peptidases is assigned a code letter denoting the type of catalysis, *i.e*., A, C, G, M, N, S, T or U for Aspartic, Cysteine, Glutamic, Metallo, Asparagine, Serine, Threonine or unknown type, respectively, along with a unique number representing the clan to which they belong. The letter "P" is used for families of proteases with more than one of the catalytic types, serine, threonine and cysteine (Rawlings and Barrett, 2012).

Being physiologically essential for all living organisms, proteases being ubiquitous are found in a wide diversity of sources such as plants, animals, and microorganisms. Microbial proteases account for approximately 40% of the total worldwide enzyme sales (Godfrey and West, 1996). Microorganisms represent an excellent source of enzymes owing to their broad biochemical diversity and their susceptibility to genetic manipulation. The inability of the plant and animal proteases to meet current world demands has led to an increased interest in microbial proteases (Rao *et al*., 1998).

Microbes serve as an ideal source of enzymes because of their fast growth, the limited space needed for their cultivation, and the easiness with which they can be genetically manipulated to produce new enzymes with altered features desirable for diverse applications. Microbial proteases are generally extracellular in nature and are directly secreted into the fermentation broth, which simplifies the

downstream processing of the enzyme as compared to proteases obtained from plants and animals (Rao *et al*., 1998).

Microbial proteases are classified into acidic, neutral and alkaline groups, depending on the conditions in which they are optimally active. Most commercial proteases, mainly neutral and alkaline, are produced by organisms belonging to the genus *Bacillus* including *Bacillus alcalophilus*, *B. amyloliquefaciens*, *B. licheniformis*, *B. circulans*, *B. coagulans*, *B. firmus*, *B. firmus*, *B. thuringiensis* and *B. subtilis* (Kumar and Takagi, 1999; Rao *et al.*, 1998; Gupta *et al.*, 2002). Some of the Gram-negative bacteria producing alkaline proteases are *Pseudomonas aeruginosa* (Morihara *et al.*, 1963), *Pseudomonas maltophila* (Kobayashi *et al.*, 1985), *Pseudomonas* sp. strain B45 (Chakraborty and Srinivasan, 1993), *Xanthomonas maltophila* (Debette, 1991), *Vibrio alginolyticus* (Deane *et al.*, 1987) and *Vibrio metschnikovii* strain RH530 (Kwon *et al.*, 1994).

The microbial extracellular proteases are of commercial importance and find multiple applications in various industrial sectors like detergent, food, photographic, leather and pharmaceutical industries. Although there are many microbial sources available for producing proteases, only a few are recognized as commercial producers. Bacterial alkaline proteases are characterized by their high activity at alkaline pH and their optimal temperature is around 60°C. These properties make them suitable for use in the detergent industry. A good number of bacterial alkaline proteases are commercially available like subtilisin Carlsberg, subtilisin BPN′ and Savinase, with their major application as detergent enzymes (Gupta *et al*., 2002).

Microbial proteases have been exploited in food industries in many ways, especially alkaline proteases are used in the preparation of protein hydrolysates used in infant food formulations, specific therapeutic dietary products and in the fortification of fruit juices and soft drinks. The protein hydrolysates are of high

nutritional value and play important role in blood pressure regulation (Neklyudov *et al.,* 2000; Ward, 1985) and are derived from casein (Miprodan; MD Foods, Viby, Germany), whey (Lacprodan; MD Foods) and soy protein (Proup; Novo Nordisk, Bagsvaerd, Denmark). Alkaline proteases are also employed in the preparation of proteinaceous fodder from waste feathers by virtue of their keratinolytic activity (Dalev, 1990, Cheng *et al.,* 1995). They are also used for meat tenderization (Takagi *et al.,* 1992, Wilson *et al.,* 1992).

Alkaline proteases with elastolytic and keratinolytic activity can be used in leather-processing industries. Proteases find their use in soaking, dehairing and bating stages of preparing skins and hides. The enzymatic treatment destroys undesirable pigments, increases the skin area and thereby clean hide is produced. Alkaline proteases from *B. subtilis* IIQDB32 (Varela *et al.,*1997), *B. amyloliquefaciens* (George *et al.,* 1995), *B. subtilis* K2 (Hameed *et al.,* 1999), *Aspergillus flavus* (Malathi and Chakraborty, 1991) and *Streptomyces* sp. (Mukhopadhyay and Chandra, 1993) have been successfully used in dehairing hides and skin and in leather processing (Nilegaonkar *et al*., 2007, Jian *et al*., 2011).

Alkaline proteases play a crucial role in the bioprocessing of used X-ray or photographic films for silver recovery in the photographic industry. These waste films contain 1.5–2.0% silver by weight in their gelatin layer, which can be used as a good source of silver for a variety of purposes. Conventionally, this silver is recovered by burning the films causing undesirable environmental pollution. Enzymatic hydrolysis of gelatin not only helps in extracting silver, but also recycling the polyester film base. Alkaline protease from *B. subtilis* (Fujiwara *et al.,* 1989), *Bacillus* sp. B21-2 (Ishikawa *et al.,* 1993), *Bacillus* sp. B18 (Fujiwara *et al.,* 1991) and *B. coagulans* PB-77 (Gajju *et al.,*1996) proved efficient in decomposing the gelatinous coating on used X-ray films for recovery of silver (Najafi *et al.,* 2005, Shankar *et al*., 2010).

Although there exists an abundance of microbial proteases, majority are isolated from mesophilic organisms. These enzymes mainly function in a narrow range of pH, temperature, and ionic strength. Moreover, the currently available enzymes have become unrecommendable under demanding industrial needs. Hence, in the search for new microbial sources, microorganisms from diverse and exotic environments are screened for specific properties expecting to result in novel process applications (Kumar and Takagi, 1999).

Mangroves are boundary landform 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 (Andrade *et al*., 2012; Lyimo *et al*., 2009).

Forty two percent of the world's mangrove regions are in Asia, located along the south coast and especially throughout the Indian subcontinent. In tropical mangroves, bacteria and fungi constitute 91% of the total microbial biomass, whereas algae and protozoa represent only 7% and 2%, respectively (Alongi, 1988). Microbial activity is responsible for major nutrient transformations within a mangrove ecosystem. Studies of microbes and their interactions with other ecosystem components are critical for understanding the mangrove ecosystem, however very little is known about the microbial communities in mangrove sediment (Gray and Herwig, 1996; Ghosh *et al*., 2010).

It has to be noted that the vast biodiversity of the microbial world is still unknown as barely less than 1% of the total microbiota has been isolated and

characterized by standard culturing conditions. As indicated by the 'great plate count anomaly' (Staley and Konopka, 1985) most of the environmental microbes observed under microscope cannot be cultured under standard laboratory conditions, as some of them may be non viable while others are viable but nonculturable (VBNC). More than 99% of the microbial biodiversity thus remains unexploited and underutilized mainly because of the unavailability of suitable cutlturing conditions and hence their bioresources remain inaccessible (Amann *et al.,* 1995).

In this context culture independent methods are to be attempted to explore the vast biodiversity of the unculturable world. In 1985, Pace and colleagues introduced a cultivation-independent method involving direct analysis of 5S and 16S rRNA gene sequences in environmental samples to describe the diversity of microorganisms (Lane *et al.,* 1985; Stahl *et al.,* 1985). With the development of PCR based techniques it was understood that the uncultured world is much more diverse and also a reservoir of many potent biomolecules.

Woese (1987) pioneered the use of 16S rRNA for phylogenetic studies as it is highly conserved between different species of bacteria and archaea. When Woese originally proposed a 16S rRNA-based phylogeny, 12 bacterial phyla were recognized, each with cultured representatives. However the number of recognizable bacterial phyla continues to increase due to culture based activities and environmental rRNA gene surveys. Recently, public databases collectively identified more than 70 different phyla with only half of them consisting of cultured representatives (Pace, 2009). Consequently, alternative methods are to be envisaged to explore this uncultured world which is a hidden bounty of genetic diversity.

Metagenomics is a culture independent method which involves directly accessing and analyzing the genomes of microorganisms in an environment,

sidestepping the need to culture them. The term "metagenomics" was first used by Jo Handelsman and others of the Department of Plant Pathology, University of Wisconsin, Madison, in their study of natural products from soil microbes (Handelsman *et al.*, 1998). The technique involves isolating the total DNA of organisms in an environmental sample, cloning it into suitable cultivable host organisms and screening the resultant clones for new genes and gene products. The approach focuses on the entire population in an environmental sample and hence can also be referred to as environmental genomics, ecogenomics and community genomics (Handelsman, 2004).

The metagenomic clones can be screened for phylogenetic markers or "anchors," such as 16S rRNA and recA, or for other conserved genes by PCR or for expression of specific traits, such as enzyme activity or antibiotic production, or they can be sequenced randomly (Handelsman, 2004). Accordingly there are two approaches to metagenomic analysis, sequence based approach and function based approach.

Sequenced-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. This was a powerful approach, first proposed by DeLong and coworkers, as they identified an insert from seawater bacteria containing a 16S rRNA gene that affiliated with the *Gammaproteobacteria*. The sequence of flanking DNA showed similarity to a bacteriorhodopsin-like gene (Stein *et al.*, 1996). This result provided the first indication that rhodopsins are not limited to the Archaea, but is in fact abundant among the *Proteobacteria* of the ocean (Beja *et al.*, 2000).

In function based approach, the metagenomic clones with a desirable character such as enzyme production or antibacterial property were identified followed by characterization of the active clones by sequence and biochemical analysis. Heterologous expression of the gene or genes of interest and secretion of the gene product determines the success of the approach. Heterologous gene expression can be improved with shuttle vectors that facilitate screening of the metagenomic DNA in diverse host species.

A wide range of activities have been studied using this approach, such as proteases, lipases/esterases, β-lactamases, nitrilases, oxidoreductases, polysaccharide-modifying enzymes, dehydrogenases, and so on (Ferrer *et al*., 2005; Hu *et al*., 2008; Jeon *et al*., 2009). Also entire operons implicated in the degradation of aromatic compounds (Suenaga *et al*., 2007; Uchiyama *et al*., 2005) or in the biosynthesis of antibiotic/antifungal have been described using this approach (Chung *et al*., 2008; Lim *et al*., 2005) with the use of large DNA cloned insert libraries. Although there is an abundance of novel enzymes isolated by metagenomic approaches there are relatively little information regarding metagenome-derived proteases, either due to unsuccessful screening for proteolytic activities (Laemmli, 1970; Rondon *et al*., 2000), or resultant false-positive clones (Jones *et al*., 2007).

OBJECTIVES

Our planet's greatest but least developed resource for biotechnological innovation is its innate natural and in particular microbial biodiversity. Looking into the depth of microbial diversity and to the exotic environment to which they belong, there is always a likelihood of finding microorganisms harboring novel enzymes with better properties, and capable for commercial exploitation (Gupta *et al*., 2002).

. In this context an attempt was made to characterize two protease enzymes obtained from mangrove sediments by metagenomic functional screening which is an innovative culture independent approach and by a traditional culture dependent method. The microbial diversity of the mangrove sediment was also analyzed based on 16S rDNA sequences. The potential of the protease enzymes for application in laundry as well as in treatment of photographic waste was also established.

Therefore the specific objectives of the present work include

1. **Extraction of metagenomic DNA of microbial communities in mangrove sediments.**
2. **Analysis of the phylogenetic diversity of the mangrove metagenome based on 16S rDNA sequences.**
3. **Screening for protease production by culture independent and culture dependent methods.**
4. **Characterization of the partial protease gene obtained by culture independent and culture dependent methods.**
5. **Characterization of the protease enzyme obtained by culture independent and culture dependent methods.**
6. **Application studies of proteases.**

Chapter - 2

REVIEW OF LITERATURE

Microorganisms are ubiquitous in nature, contributing to the vast majority of species on earth, and have an important effect on people's lives, as they play a key role in the regulation of ecosystem processes. The first step in investigating microbial processes is to obtain them in pure culture. Bacteria were initially isolated from habitats commonly associated with humans, at near neutral pH and ambient temperature. Later these were isolated from even the most extreme environments, like thermal vents, acidic ponds, saturated brine and glaciers (Madigan and Marrs, 1997). Successful isolation of bacteria from earth's crust (Kerr, 1997) and polar ice (Rothschild and Mancinelli, 2001) has led to the belief that bacterial life may exist even on other planets.

2.1 The uncultivable majority

Until 1980's it was not understood that most microorganisms in the environment resist cultivation. The "great plate count anomaly" (Staley and Konopka, 1985) indicated that all environmental microbes observed under microscope cannot be cultured under standard laboratory conditions, as some of them may be non viable while others are viable but nonculturable (VBNC). However, standard culturing techniques account for 1% or less of the bacterial diversity in most environmental samples and therefore 99% microorganisms in nature typically remain uncultivated and consequently unexploited for their ecological functions (Amann *et al.*,1995). This vast majority of microbes remain uncultured mainly because of the inability to provide appropriate culture conditions which may be impossible or even hard to imitate in laboratory conditions. Also, some organisms require interdependence with other organisms in the nature and hence they cannot be grown as pure culture in lab condition.

Some significant breakthroughs have resulted from recent attempts to culture the as-yet-unculturable bacteria by variations in their culture conditions. In the earliest cultivation attempts, media with very low nutrient were used and successful isolation of many new genera including members of the candidatus clade like SAR11 was made possible (Rappe *et al.,* 2002). Recently, using simple media and physiological conditions, many previously unknown bacteria have been isolated which belong to diverse families and phyla (Joseph *et al.,* 2003, Stevenson *et al.,* 2004). Extinction culturing technique with low nutrient media was used to culture previously uncultured marine bacteria belonging to SAR11, OM43, SAR92 and OM60/OM241 clades of *Proteobacteria* (Connon and Giovannoni, 2002). A collection of 350 isolates from soil was obtained using a variety of simple solid media in petri dishes. Twenty seven per cent of these isolates belonged to 20 as yet unnamed family level groupings. Many of these isolates were members of poorly studied subdivisions of phyla *Acidobacteria*, *Verrucomicrobia, Gemmatimonadetes* and *Actinobacteria* (Joseph *et al.,* 2003). Recently, bacteria from soil and termite gut were isolated by an integrated approach using various growth parameters like low nutrient media, varying oxygen and carbon dioxide concentrations, long incubation period and additives like humic acids and quorum signaling molecules. It was observed that more isolates belonging to *Acidobacteria* were obtained in presence of higher concentration of carbon dioxide (Stevenson *et al.,* 2004).

Other innovative and successful approaches include simulation of natural environment, community interactions and cell–cell communication important for the cultivation. Kaeberlein *et al.,* (2002) designed a diffusion chamber that allowed cultivation of previously uncultivated bacteria in a simulated natural environment. The isolates did not grow on artificial media alone but formed colonies in presence of other microorganisms indicating that they required specific signals originating from their neighbours that point to the presence of familiar

environment. In another method, single cells were encapsulated in gel microdroplets and allowed to grow with nutrients that were present at environmental concentrations in a single vessel. The pore size in gel micro droplets allowed exchange of metabolites and signaling molecules, which might be the reason for the enhanced culturability (Zengler *et al.*, 2002)

These approaches helped the culturing of many 'unculturables' to some extent. Nevertheless the vast majority remained uncultured. Hence culture-independent techniques were tried to complement efforts to culture the thousands or millions of unknown species in the environment. In 1985, Pace and colleagues introduced a cultivation-independent method involving direct analysis of 5S and 16S rRNA gene sequences in environmental samples to describe the diversity of microorganisms (Lane *et al.,* 1985a; *et al.*, 1985b; Stahl *et al.*,1985). With the development of PCR based techniques it was understood that the uncultured world is much more diverse and also a reservoir of many potent biomolecules.

Most of the PCR based culture independent approaches have relied on sequences of the small subunit ribosomal RNA (SSU rRNA/16S rRNA) gene. 16S ribosomal RNA (or 16S rRNA) is a component of the 30S small subunit of prokaryotic ribosomes and is 1.542 kb in length. Woese (1987) pioneered the use of 16S rRNA for phylogenetic studies as it is highly conserved between different species of bacteria and archaea (Weisburg *et al.,* 1991; Coenye and Vandamme, 2003). Bacterial 16S ribosomal RNA (rRNA) genes contain nine "hypervariable regions" (V1 – V9) that demonstrate considerable sequence diversity among different bacteria (Peer *et al.,* 1996). Hypervariable regions are flanked by conserved stretches in most bacteria, enabling PCR amplification of target sequences using universal primers. Primers were designed to bind to conserved regions and amplify variable regions (Fig. 2.1) (Baker *et al.,* 2003; Lu *et al.,* 2000; McCabe *et al.,* 1999; Munson *et al.,* 2004)

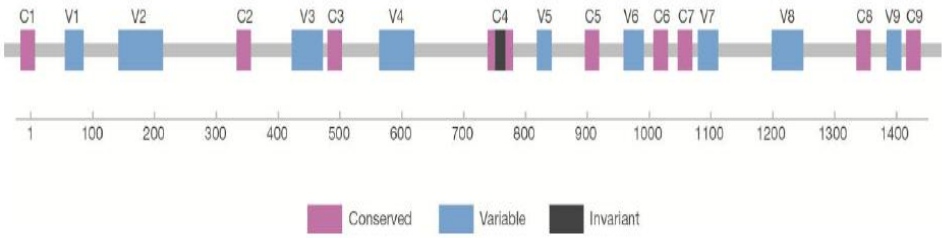

Fig 2.1 Schematic representation of the 16S rRNA gene.

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).

When Woese (1987) originally proposed a 16S rRNA-based phylogeny, 12 bacterial phyla were recognized, each with cultured representatives. Later Rappe and Giovannoni (2003) reported 14 additional phyla with cultured representatives. In addition, 16S rRNA gene sequence analysis suggested 26 candidate phyla that have no known cultured representatives. Recently, public databases collectively identified more than 70 different phyla with only half of them consisting of cultured representatives (Pace, 2009). However the number of recognizable bacterial phyla continues to increase due to culture activities and, particularly, environmental rRNA gene surveys. Hence, alternative methods are to be envisaged to explore this expanding uncultured world which is expected to be a huge resource of genetic diversity.

2.2 Metagenomics

Metagenomics is a culture independent method which involves directly accessing the genomes of microorganisms in an environment that cannot be, or have not been cultured, by isolating their DNA, cloning it into culturable organisms and screening the resultant clones for the production of new chemicals. The technique is partly analogous to genome library construction and screening, with the difference that the cloned DNA does not originate from a single known

microorganism, but rather from the entire population in an environmental sample and hence can also be referred to as environmental genomics, ecogenomics and community genomics. The term "metagenomics" was first used by Jo Handelsman and others of the Department of Plant Pathology, University of Wisconsin, Madison, in their study of natural products from soil microbes (Handelsman *et al.*, 1998).

2.2.1 History of metagenomics

Early molecular work in the field was conducted by Norman R. Pace and colleagues in the Department of Biology at Indiana University, who used PCR to explore the diversity of ribosomal RNA sequences (Lane *et al.*, 1985a). The idea of cloning DNA directly from environmental samples was first proposed by Pace (1985), and the first report of isolating and cloning bulk DNA from an environmental sample in a phage vector was published by Pace and colleagues in 1991 (Schmidt *et al.*, 1991). The next advance was the construction of a metagenomic library with DNA derived from a mixture of organisms enriched on dried grasses in the laboratory (Healy *et al.*,1995). Clones expressing cellulolytic activity were found in these libraries, which were referred to as zoolibraries. After leaving the Pace laboratory, Edward DeLong continued in the field and reported libraries constructed from prokaryotes in seawater (Stein *et al.*,1996) which laid the groundwork for environmental phylogenies based on signature 16S sequences. They identified a 40 kb insert clone that carried a 16S rRNA gene indicating that the clone was derived from an archaeon that had never been cultured.

Construction of libraries with DNA extracted from soil lagged due to difficulties associated with maintaining the integrity of DNA during its extraction and purification from soil matrix. Rondon *et al.* (2000) reported cloning of soil metagenome in BAC vector and identified clones with antibacterial, lipase, amylase, nuclease, and hemolytic activities. In 2002, Breitbart, Rohwer, and

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

The application of whole genome shotgun (WGS) sequencing technology revolutionized the field of metagenomics. In 2004, Tyson, Banfield, and colleagues at the University of California, Berkeley and the Joint Genome Institute sequenced DNA extracted from a natural acidophilic biofilm of an acid mine drainage system using random shotgun sequencing and reported reconstruction of near-complete genomes of *Leptospirillum* group II and *Ferroplasma* type II, and partial recovery of three other genomes (Tyson *et al.,* 2004).

Craig Venter, as part of Global Ocean Sampling (GOS) expedition, collected seawater samples from the Sargasso Sea near Bermuda and whole genome shotgun sequencing of the microbial populations estimated at least 1800 genomic species based on sequence relatedness, including 148 previously unknown bacterial phylotypes (Venter *et al.*, 2004).

2.2.2 Approaches to metagenomic analysis

Metagenomic analysis involves isolating DNA from an environmental sample, cloning the DNA into a suitable vector, transforming the clones into a host bacterium, and screening the resulting transformants. The clones can be screened for phylogenetic markers or "anchors," such as 16S rRNA and recA, or for other conserved genes by PCR or for expression of specific traits, such as enzyme activity or antibiotic production, or they can be sequenced randomly (Handelsman, 2004). Accordingly there are two approaches to metagenomic analysis, sequence based approach and function based approach (Fig. 2.2).

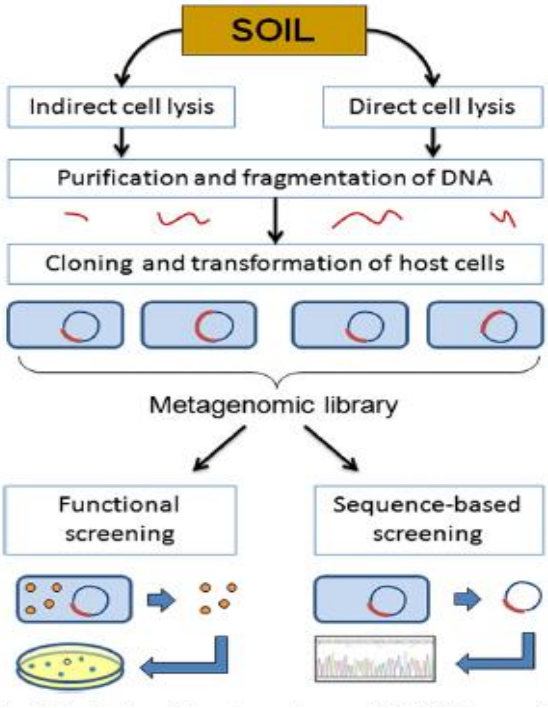

Fig. 2.2 Schematic representation metagenomic approaches
(Adapted from Mocali and Benedetti, 2010)

2.2.2.1 Sequence based approach

Sequence-based analysis involves sequencing of clones containing 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. This was a powerful approach, first proposed by the DeLong and coworkers, as they identified an insert from seawater bacteria containing a 16S rRNA gene that affiliated with the γ-*Proteobacteria*. The sequence of flanking DNA showed similarity to a bacteriorhodopsin-like gene (Stein *et al.*, 1996). This result provided the first indication that rhodopsins are not limited to the Archaea, but is in fact abundant among the *Proteobacteria* of the ocean (Beja *et al.*, 2000). Subsequent

heterologous expression of the bacteriorhodopsin gene in *E. coli* produced a functional biochemical characterization of the protein, thereby enabling phylogeny to be linked to function.

The recent monumental sequencing efforts, which included reconstruction of the genomes of uncultured organisms in a community in acid mine drainage (Tyson *et al.,* 2004) and the Sargasso Sea (Venter *et al.,* 2004), illustrate the power of large-scale sequencing efforts to enrich our understanding of uncultured communities.

The advantage of this approach is that it can be performed using either PCR-based or hybridization-based procedures, but the genes obtained through this approach are limited to those having homologies to the probe sequence, or primer sequence and may not allow us to obtain novel genes. Since the primers are designed based on conserved sequences of gene and gene products, only internal gene sequence is obtained, whereas the flanking gene sequence also is required to obtain the whole gene.

Primers designed based on flanking conserved gene sequences may therefore be used to obtain complete gene. Integrons carries gene cassettes that are commonly flanked by *attC* sequences. By designing primers based on *attC* sequences, complete gene cassettes in the integrons may be obtained (Elsied *et al.*, 2009, Koenig *et al.*, 2009).

Another approach based on sequence based homology involves retrieving similar gene sequences from public databases to synthesize genes and its expression in appropriate host as reported by Bayer and colleagues (2009). They screened for methyl halide transferases in the public NCBI sequence database, synthesized all of the candidate genes with codons optimal for *E. coli* and *Saccharomyces cerevisiae* and expressed them in the hosts. Surprisingly, 94% of

the proteins showed methyl halide transferase activities with various halide (Cl, Br, I) specificities.

The limitation of this approach is the small number of available markers that provide reliable placement in the Tree of Life. The taxonomic position of a fragment of DNA with a desired trait can be determined only if it carries a dependable marker. 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).

2.2.2.2 Function based approach

In function based approach, clones carrying a desired trait were identified followed by characterization of the active clones by sequence and biochemical analysis. Heterologous expression of the gene or genes of interest and secretion of the gene product determines the success of the approach. It also depends on the availability of an assay for the function of interest that can be performed efficiently on vast libraries, because the frequency of active clones is quite low. Heterologous gene expression can be improved with shuttle vectors that facilitate screening of the metagenomic DNA in diverse host species.

2.2.2.2.1 Selection of host and vector

In most cloning experiments, *E.coli* is used as a surrogate host. They lack genes for homologous recombination (*recA*, *recBC*) and restriction (*mcrA*, *mcr BC*) which are useful for cloning variously modified foreign DNA into *E. coli* (Uchiyama and Miyazaki, 2009).

The choice of a vector depends largely on the length of the insert. Plasmids are suitable for cloning smaller than 10 kb DNA fragments, and cosmids (25–35 kb), fosmids (25–40 kb), or BACs (100–200 kb) can be used to clone larger fragments (Angelov *et al.*, 2009; Kakirde *et al.*, 2011; Uchiyama and

Miyazaki, 2009; Van Elsas *et al.,* 2008). Among these vectors, plasmids have high copy numbers and strong vector-borne promoters.

In order to eliminate the limitations generated by using *E.coli* as a single host, shuttle vectors and non-*E. coli* host systems have been developed. Bacterial strains from genera like *Burkholderia, Bacillus, Sphingomonas, Streptomyces*, and *Pseudomonas* have thus been reported as alternative hosts (Courtois *et al.,* 2003; Eyers *et al.,* 2004; Martinez *et al.,* 2004; Van Elsas *et al.,* 2008a)

Some researchers use *E. coli* for library construction and a second host for functional screening. Shuttle vectors may be used in this context and the low-level gene expression of the metagenome can be enhanced. Wang *et al.* (2006) used an *E. coli–Sinorhizobium* shuttle vector and screened for clones which utilize D-3-hydroxybutyrate in a *bdhA*$^{-}$ mutant of *Sinorhizobium meliloti* and *atoC*$^{-}$ mutant of *E. coli*, both of which lacked the ability to utilize D-3-hydroxybutyrate. Li *et al.* (2005) used an *E. coli–Rhizobium* shuttle vector and screened for tryptophan synthetic pathway genes using two different trp mutants of *E. coli* and *Rhizobium leguminosarum*. In both studies, most of the clones selected using non-*E. coli* hosts failed to complement the deficiency of *E. coli* (and vice versa). This indicates that the expression profiles differ depending on the hosts and it demonstrates the advantage of using a second host for functional screening.

The development of novel shuttle vectors that have extended host ranges is important in this context. Ono *et al.* (2007) used a cosmid, which contained both colE1 and RK2 replicons so that it can replicate in a range of Gram positive and Gram negative bacteria. Aakvik *et al.* (2009) also used the RK2 replicon to modify fosmid and BAC vectors so that they were stably maintained in alternate hosts, *Pseudomonas fluorescens* and *Xanthomonas campestris*. The main features of this vector are (1) inducible copy number for controlled gene expression, which minimizes possible gene product toxicity but allows high-level gene expression

for effective detection in screenings, (2) the ability to stably hold inserts of up to 200 kb, and (3) a high capacity to be efficiently transferred to a wide range of hosts.

Craig *et al. (*2010) reported the construction of soil metagenomic libraries with a broad-host-range cosmid vector, IncP1-α, using six selected proteobacterial host strains, *i.e*., *Agrobacterium tumefaciens, Burkholderia graminis, Caulobacter vibrioides, E. coli, Pseudomonas putida*, and *Ralstonia metallidurans* and illustrated totally different expression profile of the same metagenomic library purely based on the expression host. Martinez *et al.* (2004) reported diverse expression profiles of antibiotic-producing genes upon parallel screening of metagenomic libraries in multiple hosts.

The development of host–vector systems with environmentally prevalent strains from phyla that are relatively incompatible with the *E. coli* expression machinery (like *Acidobacteria* and *Verrucomicrobia*) holds great potential to increase the rates of expression of genes from metagenomes (Ekkers *et al.,* 2012).

The use of dual-orientation promoters on the vector also effectively increases the rate of successful gene expression (Lämmle *et al.,* 2007). However, such promoters are probably most useful in small-fragment libraries where native promoters may not be present in the insert.

Codon usage bias is another important factor in the successful expression of foreign genes in a metagenomic host. As the preference for codon usage for generating proteins differs in different organisms, the gene expression levels also vary. There was 250 fold variation among the different variants of Green Flourescent protein (GFP) which were generated by synthesizing and expressing the 154 genes for GFP with random silent mutations in the third base position (Kudla *et al.,* 2009). Besides overall codon usage, the preference for start codons can also vary greatly across bacterial species (Villegas and Kropinski, 2008).

Furthermore, codon usage bias has been shown to be important in translation (Sørensen *et al.,* 1989), protein folding (Zalucki *et al.,* 2009), and secretion (Power *et al.,* 2004; Zalucki and Jennings, 2007).

2.2.2.2.2 Screening methods

Another crucial step involved in function based metagenomic studies is the screening methods adopted. The gene expression may not be easily detectable if the screening method is not highly sensitive (Gabor *et al.,* 2004a). A broad range of screening methods can be used with three general detection strategies (Simon and Daniel, 2009).

2.2.2.2.2.1 Phenotypic insert detection (PID)

This is the most commonly used approach for the functional screening of metagenomic libraries, a strategy wherein the expression of a particular trait is used to identify positive clones. The positive clones express some positive traits which are directly observable like colony pigmentation, irregular colony morphology, or halo formation on plate overlays (Brady, 2007; Craig *et al.,* 2010; LeCleir *et al.,* 2007). In another approach, indicator media are used to detect particular small molecules, chemical reactions, or metabolic, catabolic, or antibiotic capabilities of a clone. Isolation of novel metallo-proteases from metagenomic libraries using milk-infused plates is a nice illustration of this strategy (Waschkowitz *et al.,* 2009; Vasconcellos *et al.,* 2010; Fan *et al.,* 2011; Morohoshi *et al.,* 2011; Tirawongsaroj *et al.,* 2008). The screening can be based on coexpression of a reporter gene like beta-galactosidase, which is linked to the target gene in the library (Li *et al.,* 2007). LeCleir *et al.* (2007) showed the presence of chitinolytic enzymes in an estuarine metagenomic library by cleavage of fluorogenic analogs of chitin.

2.2.2.2.2.2 Modulated detection (MD)

This is a strategy that does not rely on the direct detection of an expressed gene, but it uses a predesigned expression route. By modulating the expression host and/or vector systems, selection and detection of inserted genes can be manipulated by the coexpression of reporter genes or heterologous complementation. The *lacZ* gene encoding beta-galactosidase (resulting in colony coloring upon growth on X-Gal-containing medium) is frequently used as a reporter gene. Schipper *et al.* (2009) used *A. tumefaciens* strain containing a *traI-lacZ* gene fusion to screen for metagenomic clones containing genes that interfere with quorum sensing (QS). By inducing the traI gene with the QS signal molecule homoserine lactone 3-oxo-C8-HSL, *lacZ* is activated. The quorum sensing-inhibitory or degradation activity of the clone degrades 3-oxo-C8-HSL leading to inhibition of *lacZ* induction. The experiment yielded 438 positive clones showing QS inhibition.

2.2.2.2.2.3 Substrate induction

This is a strategy that is based on the induced expression of cloned genes via a specific substrate and is mainly applied for the detection of catabolic genes. The regulatory elements of these catabolic genes are generally situated close to the genes themselves. These elements have been shown to work in host organisms like *E. coli*.

Uchiyama *et al.* (2005) developed the so-called substrate-induced gene expression (SIGEX) system for use as a screening method for particular catabolic genes. To make this method high-throughput compatible, an operon-trap GFP expression vector was used, resulting in co-expressed GFP upon substrate-induced expression of any responsive inserted gene. This GFP expression subsequently enabled the separation of positive clones by fluorescence-assisted cell sorting (FACS). They studied metagenomic genes from groundwater that could be

induced by benzoate and naphthalene. This yielded 58 benzoate- positive and four naphthalene-positive clones.

2.3 Community metagenomics

Metagenomic libraries have already been constructed from a broad range of environments, from oceans to humans, to access the genetic potential as well as the phylogenetic diversity of the microbial communities present.

2.3.1 Soil

Soil is one of the most challenging environmental sources to analyze microbial diversity. Several parameters of soil, such as particle size, permeability, porosity, water content, mineral composition, and plant cover, can influence microbial composition. In addition, other factors such as collection and storage of soil sample, DNA extraction methods, host-vector systems used for DNA cloning, and representative soil sampling, can also influence the results of microbial content (Ishoey *et al.,* 2008; Kowalchuk *et al.,* 2002; Henne *et al.,* 2000). It is estimated that, depending on the sample and methods used, the number of bacteria in soil may vary from 6,400 to 38,000 per g while that in oceans is 160 per ml (Curtis *et al.,* 2002) which clearly suggest that microbial content is several orders of magnitude less in the sea in comparison to soil environments (Neelakanta and Sultana, 2013).

By direct cloning into plasmid, cosmid, or BAC vectors, novel genes from soil microbes that encode enzymes and antibiotics have been discovered. These genes share little homology with known genes, thus illustrating the enormous potential of soil metagenomics in isolating novel classes of genes. Some of the genes that were isolated from soil microorganisms include lipases/esterases (Lee *et al.,* 2004; Henne *et al.,* 2000; La¨mmle *et al.,* 2007; Couto *et al.,* 2010; Litthauer *et al.,* 2010; Rhee *et al.,* 2005; Rondon *et al.,* 2000; Voget *et al.,* 2003;

Jeon *et al.,* 2009), proteases (Lee *et al.,* 2007; Pushpam *et al.,* 2011; Waschkowitz *et al.,* 2009; Neveu *et al.,* 2011; Gupta *et al.,* 2002), amylases (Yun *et al.,* 2004; La¨mmle *et al.,* 2007; Rondon *et al.,* 2000; Richardson *et al.,* 2002; Gabor *et al.,* 2003), phosphatase (La¨mmle *et al.,* 2007), β-glucosidases and β-lactamases (Gabor *et al.,* 2003; Song *et al.,* 2005), chitinases (Cottrell *et al.,* 1999), amidases (Gabor *et al.,* 2004b), nitrilases (Robertson *et al.,* 2004; DeSantis *et al.,* 2002), xylanase (Hu *et al.,* 2008), pectate lyases (Solbak *et al.,* 2005), antimicrobials (MacNeil *et al.,* 2001; Gillespie *et al.,* 2002), 4-Hydroxybutyrate dehydrogenase (Henne *et al.,* 1999), carbonyls (Knietsch *et al.,* 2003a), monooxygenase (Hellemond *et al.,* 2007), nickel resistance gene (Mirete *et al.,* 2007) and oxidoreductases (Knietsch *et al.,* 2003b).

Moreover, with the use of large DNA cloned insert libraries, the screening is not limited to a single gene encoding activity, and entire operons involved in the degradation of aromatic compounds (Suenaga *et al.,* 2007; Uchiyama *et al.,* 2005) or in antibiotic/antifungal biosynthesis have been described using this approach (Chung *et al.,* 2008; Lim *et al.,* 2005). Microbial diversity of extreme environmental conditions like hot spring, polar ice, acid mine drainage and complex environments like human gut and skin is revealed using metagenomics approach.

2.3.2 Hot spring

Studies from Barns and colleagues (1994) have provided information on microbial diversity in hot spring environments. Archaea similar to Crenarchaeota phylotype are found to be the abundant species in the sediments of Yellowstone National Park hot springs near Cauldron, USA, which has several boiling areas (93ºC). Analysis in the same hot spring revealed more bacterial numbers distributed in twelve new division-level lineages including *Desulfurococcus/ Pyrodictium* clade, *Pyrobaculum* sp. *Thermofilum pendens* and *Archaeoglobus*

fulgidus. Furthermore, Blank and colleagues (2002) showed differences in microbial content in the samples collected from seven different hot springs of Yellowstone National park at close proximity with similar temperatures and comparable pH values. It was also found that *Thermocrinis ruber*, a member of the order *Aquificales* is ubiquitous, in all the springs and indicates that primary production in these springs is driven by chemoauxotrophic hydrogen oxidation. Microbial community analysis of one of the Gedongsongo hot spring, GS-2, revealed the presence of unique strains belonging to *β-Proteobacteria and Firmicutes* (Aminin *et al.,* 2008).

2.3.3 Polar ice

Sequencing of polar ice caps has revealed the presence of algal population and several heterotrophic bacteria in ice matrix at low temperatures and low levels of light. Bowman and colleagues (1997) investigated the bacterial populations associated with sea ice sampled from Antarctic coastal areas and several psychrophilic as well as psychrotrophic isolates belonging to four phylogenetic groups, the alpha and gamma subdivisions of the *Proteobacteria*, the gram-positive branch, and the *Flexibacter-Bacteroides-Cytophaga* phylum were obtained. 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. Psychrotrophic strains were found to be members of the genera *Pseudoalteromonas, Psychrobacter, Halomonas, Pseudomonas, Hyphomonas, Sphingomonas, Arthrobacter, Planococcus,* and *Halobacillus*. Later Brown and Bowman (2001) studied the bacterial biodiversity of seven Antarctic sea-ice samples and one Arctic sea-ice sample. Bacterial phylotypes grouped within α and γ *Proteobacteria*, the *Cytophaga-Flavobacterium-Bacteroides* division, the Gram-Positive bacteria and the orders *Chlamydiales* and *Verrucomicrobiales*. Recent

studies have also found dominance of archaea *Salinibacter ruber* in hyper-saline environments (Benlloch *et al.*, 2002).

2.3.4 Acid mine drainage

The microbial diversity of non-thermal environments with extreme acidic conditions like acid mine drainage, have been investigated and shown to contain archaea of *Ferroplasma* and *Thermoplasma* groups. Several bacterial species that include *Acidiphilium, Acidithiobacillus, Leptospirillum* and *Sulfobacillus* have also been found to be abundant in extreme acidic environments (Baker and Banfield, 2003). The bacterial lines of descent are divisions within the *Proteobacteria* (α, β, γ and δ), *Nitrospira, Firmicutes*, and *Acidobacteria*. By whole genome shotgun (WGS) sequencing Tyson and colleagues (2004) sequenced 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.

2.3.5 Human gut

Metagenomics have provided great insights to the human gut microbiome over the past decade. Several factors such as specific microbial species colonizing the gut, niches they occupy, time, space, factors unique to the environment of each human being such as different dietary needs, and interactions with host cells can all influence taxonomic composition of the human microbiome. Metagenomics have uncovered nearly 1000 human-associated microorganisms' draft genome sequences, along with 3.3 million unique microbial genes derived from the intestinal tract of over 100 European adults (Claesson *et al.*, 2001; Dutton and Turnbaugh, 2012). Metagenomics of the human gut microbiome also revealed interesting functions carried by microorganisms within the gut, ranging from its role in newly discovered signaling mechanisms, vitamin production, glycan production, amino-acid, and xenobiotic metabolism. Several studies have also

reported that microbial composition of the human gut is greatly affected by genetic background, age, diet, and health status of the host (Maccaferri *et al.*, 2001). Differences in microbial content were seen in all age groups of human beings. Babies (breast milk fed and formula fed), healthy and malnourished infants, youngsters, the elderly, humans that were either lean or obese, and humans with inflammatory bowel diseases (IBD) showed differences in microbial composition. A metagenomic study from De Filippo *et al.* (2010) showed that European children who consumed a diet high in animal protein, sugar, starch, and fat, and low in fiber showed differences in gut microbial content in comparison to children fed on vegetarian diet consisting of carbohydrates, fiber, and non-animal protein. Interestingly, the microbiome of European children was enriched with *Firmicutes* and *Proteobacteria*, whereas the African microbiome was enriched with *Actinobacteria* and *Bacteroidetes*. Members of *Xylanibacter* and *Prevotella* were only present in Children from Europe. These results clearly suggest that host dietary habits influence gut microbial content. Metagenomics have also revealed an interesting link with microbial content in the gut to host metabolism and disease development. The causes of intestinal diseases such as IBD, Crohn's disease (CD), and ulcerative colitis (UC) have all been linked with both human gene- associated and microbiome-associated factors (Manichanh *et al.*, 2006).

2.3.6 Skin microbiome

Recent studies have employed metagenomic approaches in determining the microbial diversity of the human skin. Skin serves as a good host of microbes that include both commensal and pathogenic bacteria. Bacteria belonging to *Proteobacteria*, such as *Pseudomonas* sp. and *Janthinobacterium* sp., were found to be abundant in both human and mice skin biopsies. The presence of other bacteria belonging to *Alphaproteobacteria, Gammaproteobacteria, Betaproteobacteria, Actinobacteria* sp., such as *Kocuria* sp., *Pripionibacteria* sp., *Firmicutes*, and *Bacteroidetes* were all evident in human skin biopsies (Grice *et*

al., 2008; Kong *et al.,* 2012; Courtois *et al.,* 2003). There is substantial evidence that viruses also represent a significant part of the skin microbiome. The presence of beta and gamma human papillomaviruses, polyomaviruses, and circoviruses on normal appearing skin has been reported (Grice *et al.*, 2009).

2.4 Metagenomic DNA isolation

The first step in any metagenomic study is the isolation of community DNA. This is also a critical step as it carries all the information for further studies. The isolated DNA should be representative of all cells present in the sample and should be ideal for generation of genomic libraries (Neelakanta and Sultana, 2013). A major difficulty associated with the metagenome approach is the contamination of isolated DNA with polyphenolic compounds, which are copurified with the DNA. These compounds are difficult to remove, as the polyphenols also interfere with enzymatic modifications of the isolated DNA (Tsai and Olson, 1992)

Extraction of microbial community DNA for use in metagenomic library construction can be roughly divided into two strategies: The first approach consists of the direct extraction of nucleic acids from soil through *in situ* cell lysis followed by DNA purification. The alternative approach is based on the separation of bacteria from the soil particles followed by cell lysis and then DNA purification (Robe *et al.*, 2003; Van Elsas *et al.*, 2008). Both methods have their own specific advantages and biases. Four key parameters that define the suitability of the DNA extracted by each method for subsequent metagenomics analysis have been identified as yield, purity, fragment size, and representativeness (Ekkers *et al.*, 2012). However, these factors often stand in negative relation to one another, enhancing one will often have a negative effect on other factors. The extractions often result in either low-yield extracts containing large DNA fragment sizes or high-yield with small-fragment DNA. A low average fragment size mostly

impedes the subsequent analysis of larger operons, for which larger insert libraries are needed (Williamson *et al.*, 2011).

2.4.1 Direct lysis extraction method

This is the widely used method and it provides the highest DNA yields with acceptable processing time. The first step is the lysis of microbial cell wall to liberate the nucleic acid from all bacteria, which is followed by the separation of nucleic acids and its purification.

2.4.1.1 Cell lysis

The methods adopted for cell lysis or membrane disruption can be broadly divided into three, physical, chemical or enzymatic, which is used either alone or in combination.

Physical treatments have the greatest access to the whole bacterial community, including bacteria deep within soil microaggregates (Robe *et al.*, 2003). The most commonly used physical disruption methods include freezing–thawing or freezing–boiling (Degrange and Bardin, 1995; More *et al.*, 1994; Tsai *et al.*, 1991; Kauffmann *et al.*, 2004) and bead-mill homogenization (Maarit Niemi *et al.*, 2001; Miller, 2001; Miller *et al.*, 1999; Steffan *et al.*, 1988).

An easier method of cell lysis is bead-beating, which can be used for obtaining inserts 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) and ZR Soil Microbe DNA Kit™ (Zymo Research). Additionally, certain devices available, such as Beadbater™, allow very efficient disintegration with the use of beads of various sizes, made of glass, steel, zircon or silica. This type of DNA isolation kit, manufactured by Qbiogene, Inc. (FastDNA® spin kit) has been used for creating a metagenomic library from a

sample of geothermal sediment, with the use of pCR-XLTOPO vector (Wilkinson *et al.,* 2002).

Other physical methods for cell lysis are mortar mill grinding (Tebbe and Vahjen, 1993), grinding under liquid nitrogen (Volossiouk *et al.,* 1995), ultrasonication (Picard *et al.,* 1992*;* Porteous *et al.,* 1997) and microwave thermal shock (Orsini and Romano-Spica, 2001; Lakay *et al.,* 2007).

Bead beating method has the advantage of increased DNA yields with longer beating times, higher speed and reduced extraction buffer volumes but at the cost of DNA shearing (Bürgmann *et al.,* 2001). Also, drying the soil sample before grinding improved the lysis efficiency (Frostegard *et al.,* 1999). Physical methods also have efficiencies for disruption of vegetative forms, small cells and spores (More *et al.,* 1994). The limitation of the method is that they often result in significant DNA shearing (Liesack *et al.,* 1991; Simonet *et al.,* 1991).

Chemical methods of cell lysis involves the use of different chemical agents and the most commonly used chemical is the detergent, sodium dodecyl/lauryl sulphate (SDS), which dissolves the hydrophobic material of cell membranes. Detergents are often used in various concentrations (0.1% – 20%), at elevated temperature (65°C) and with chelating agents such as EDTA, Chelex 100 (Herron and Wellington, 1990; Jacobsen and Rasmussen, 1992) and diverse Tris buffer or sodium phosphate buffers. Increasing the EDTA concentration increased the strength of extraction resulting in higher yields, but lower purity of isolated nucleic acids (Krsek and Wellington, 1999).

Other chemicals used for cell lysis are cetyl/hexadecyl trimethyl-ammonium bromide (CTAB) and polyvinylpolypyrrolidone (PVPP) (Krsek and Wellington, 1999; Zhou *et al.,* 1996; Nalin *et al.,* 1999). Both CTAB and PVPP can partially remove humic compounds, but (contrary to CTAB) PVPP resulted in DNA loss (Zhou *et al.,* 1996; Roose-Amsaleg *et al.,* 2001). CTAB forms insoluble

complexes with denatured proteins, polysaccharides and cell debris (Saano *et al.*, 1995). PVPP was shown to be ineffective during cell lysis, but efficient when used as a spin column during the nucleic acids purification step (Krsek and Wellington, 1999).

In enzymatic lysis, different enzymes like lysozyme, proteinase K, pronase and achromopeptidase are used. Lysozyme treatment is most commonly used and its hydrolytic action on the glycisidic bonds of humic components enhance DNA purity (Bruce *et al.*, 1992; Maarit Niemi *et al.*, 2001; Rochelle *et al.*,1992; Tebbe and Vahjen, 1993; Tsai and Olson, 1991). Proteinase K was used to digest contaminating proteins (Maarit Niemi *et al.*, 2001; Porteous and Armstrong, 1991; Zhou *et al.*, 1996), whereas another enzyme achromopeptidase was used to improve the lysis of the recalcitrant Gram positive *Frankia* (Simonet *et al.*, 1984).

2.4.1.2 Extraction of nucleic acids

Extraction of nucleic acids from soil components is crucial, as a major soil component, humic acid inhibits downstream processes like restriction enzyme digestion of DNA and polymerase chain reaction (Tebbe and Vahjen, 1993). 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). The first step following cell lysis is organic solvent extraction by either phenol alone or in combination like phenol:chloroform (50:50, v:v) (Ranjan *et al.*, 2005), phenol:chloroform:isoamyl alcohol (25:24:1, v:v:v) (Zhou *et al.*,1996). The application of chloroform brings very good results, while phenol is rather avoided due to its toxicity and the ability to separate DNA together with other compounds (Porteous and Armstrong, 1991). Isolated nucleic acids are condensed and purified by precipitation with ethanol, isopropanol, sodium acetate or polyethylene glycol (PEG) (Porteous & Armstrong, 1991; Roose-Amsaleg *et*

al., 2001). PEG frequently replaces isopropanol, since alcohol has the ability to precipitate DNA with humus substances of soil (Porteous & Armstrong, 1991). The addition of sodium acetate during precipitation improves the purity of the DNA.

Hydroxyapatite column was successful for extracting DNA (Ogram *et al.,* 1987; Steffan *et al.,* 1988) and rRNA (Purdy *et al.,* 1996) from soil and sediment samples. Cesium chloride (CsCl) density gradient centrifugation was also used for purification of nucleic acids (Ogram *et al.,* 1987; Porteous and Armstrong, 1991). However both these methods did not completely remove humic acid from DNA preparations. Agarose gel electrophoresis assisted by polyvinylpyrrolidone (PVP) facilitated separation of DNA from humic acids (Hugenholtz *et al.,* 1998; LaMontagne *et al.,* 2002), since PVP retarded the phenolic compounds that usually co-migrate with nucleic acids (Lee *et al.,* 1996; Young *et al.,* 1993).

Various gel filtration resins including sephadex G200 (Kuske *et al.,* 1998) and G150, Sepharose 2B, 4B and 6B, Biogel P100 and P200 (Jackson *et al.,* 1997; Maarit Niemi *et al.,* 2001; Miller, 2001) Microspin Sephacryl S-300 and S-400 columns (Pharmacia Biotech) were also used to purify crude DNA extracts contaminated with humic acids (Edgcomb *et al.,* 1999; Frostegard *et al.,* 1999). Other commercial purification products investigated 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 & 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.,* 2011). Jacobsen (1995) successfully removed the PCR-inhibitory effect of humic acids through alternative magnetic capture hybridization (MCH) approach.

Although recovery of large DNA fragments (40–90 kb) using gentle lysozyme–SDS-based methods was reported (Krsek and Wellington, 1999), in most studies, direct extraction did not lead to the recovery of DNA fragments larger than 20 kb.

2.4.2 Indirect lysis extraction method

Faegri *et al.* (1977) proposed the method of genomic DNA isolation from environmental samples after a preliminary separation of microbial cells. The separation of cells from the soil sample is aided by centrifugation according to sedimentation velocities, buoyant density or both. An indirect method of metagenomic DNA isolation consists of the following steps: dispersion of the environmental sample, cell separation, cell lysis, DNA isolation and purification (Robe *et al.*, 2003)

Dispersion of the soil sample is done to release the microbial cells from within the soil aggregates and to reduce the adhesion of cells to the soil particles. This is done by using blenders (Bakken and Lindahl, 1995; Faegri *et al.*, 1977; Hardeman and Sjoling, 2007), sonication (Ramsay, 1984), mild dispersal by shaking (Turpin *et al.*, 1993), and rotating pestle (Lindahl and Bakken, 1995). Care has to be taken not to exceed the time of mechanical impact on cells over 18 min, as some of the microorganisms are subject to disintegration beyond that time (Lindahl and Bakken, 1995)

Separation of bacteria from soil particles according to sedimentation velocities is based on two successive centrifugations, first is a low speed centrifugation ranging from 500 × g to 1000 × g lasting 2–15 min, respectively, in order to remove the soil debris, fungal mycelia and heavy soil particles (Faegri *et al.*, 1977). A subsequent high speed centrifugation of the cell-containing supernatant produces the bacterial fraction. Supernatants resulting from the low speed centrifugation might still contain non-cellular material and soil

contaminants such as humic acids. Flocculation of the cell debris and clay particles using $CaCl_2$ resulted in a considerable decrease in cell recovery (Jacobsen and Rasmussen, 1992).

An alternative is a high-speed centrifugation method based on density gradient centrifugation (Bakken, 1985) in order to separate bacteria according to their buoyant density. Several multi-gradient media tested include Percoll, metrizamide and Nycodenz (Bakken and Lindahl, 1995; Delmont *et al.,* 2011) with Nycodenz providing the best results. Bacteria will settle on top of the Nycodenz gradient, and organic and mineral particles of greater density will sediment to the bottom of the tube.

After separation of microbial cells, cell lysis resulted in release of DNA which is then 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 before performing a gentle bacterial lysis recovered DNA fragments of a few hundred kilobases with limited mechanical shearing and BAC libraries were prepared from several microorganisms including *Methanosarcina thermophila* (Diaz-Perez *et al.,* 1997), *Bacillus cereus* (Rondon *et al.,* 1999), *Mycobacterium tuberculosis* (Brosch *et al.,* 1998) and *Pseudomonas aeruginosa* (Dewar *et al.,* 1998).

An integrated approach combining centrifugation-based cell separation from soil particles, in 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). DNA fragments recovered using this method is more than 300 kb in size and of adequate purity for further molecular cloning procedures (Nalin *et al.,* 2001).

The inevitable biases inherent to any DNA extraction method can lead to unrepresentative (biased) microbial community DNA. An interesting new method for separating DNA from highly contaminated samples, called synchronous coefficient of drag alteration, 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 nanogram to microgram amounts of DNA (Delmont *et al.,* 2011; Ishoey *et al.,* 2008). In the case of samples that yield less DNA, pre-amplification methods for the DNA is recommended. Multiple displacement amplification using random hexamers and phage phi29 polymerase has been reported to successfully amplify femtograms of DNA in order to produce micrograms of product (Lasken, 2009; Eisen, 2007).

Direct extraction methods produced DNA with high yields, but with less purity and increased DNA shearing when compared to indirect extraction methods. The recovered bacterial fraction obtained by indirect methods represents only 25–50% of the total endogenous bacterial community (Bakken and Lindahl, 1995) whereas direct extraction lysis recovered more than 60% of the total theoretical bacterial DNA (More *et al.,* 1994). Direct lysis procedures are preferred when large quantities of nucleic acids are required for DNA consuming methods, statistically significant detection of non-abundant microorganisms, and when the entire diversity of an environmental sample must be investigated with minimum bias. Indirect extraction methods are preferred for targeting prokaryotic communities, when high DNA purity is required for inhibitor-sensitive methods, and when recovery of high molecular weight DNA is necessary (Robe *et al.,* 2003).

2.5 Bioinformatic tools

Several bioinformatic tools and databases, either web based or downloaded versions, like BLAST, ClustalX/ClustalW, BioEdit, MEGA and RDP help in the analysis of nucleic acid and protein sequences.

2.5.1 BLAST

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between amino-acid sequences of different proteins or the nucleotides of DNA sequences (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi). The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. Input sequences are in FASTA or Genbank format and weight matrix and output can be delivered in a variety of formats including HTML, plain text, and XML formatting. Different types of BLASTs are available according to the query sequences. BLAST is actually a family of programs (all included in the blastall executable) (Altschul *et al*., 1990).

Nucleotide BLAST (blastn): A nucleotide database is searched using a nucleotide query and returns the most similar DNA sequences.

Protein BLAST (blastp): A protein database is searched using a protein query and the most similar protein sequences are returned.

Nucleotide 6-frame translation-protein (blastx): This program compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

Protein-nucleotide 6-frame translation (tblastn): This program compares a protein query against the all six reading frames of a nucleotide sequence database.

Nucleotide 6-frame translation-nucleotide 6-frame translation (tblastx): This program is the slowest of the BLAST family. It translates the query nucleotide sequence in all six possible frames and compares it against the six-frame translations of a nucleotide sequence database. The purpose of tblastx is to find very distant relationships between nucleotide sequences.

2.5.2 Clustal

Clustal is a widely used computer program for multiple sequence alignment of nucleic acid and protein sequences. There are three main variations, 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 addition to the Clustal family. It offers a significant increase in scalability over previous versions, allowing hundreds of thousands of sequences to be aligned in only a few hours. It will also make use of multiple processors, where present. In addition, the quality of alignments is superior to previous versions, as measured by a range of popular benchmarks (Sievers *et al*., 2011). This program accepts a wide range of input formats, including NBRF/PIR, FASTA, EMBL/Swiss-Prot, Clustal, GCC/MSF, GCG9 RSF, and GDE. The output format can be one or many of the following: Clustal, NBRF/PIR, GCG/MSF, PHYLIP, GDE, or NEXUS (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

2.5.3 BioEdit Sequence Alignment Editor

BioEdit Sequence Alignment Editor (version 7.1.3.0) is a biological sequence editor that runs in Windows 95/98/NT/2000/XP and is intended to provide 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. Sequence lengths and numbers are limited only by

available system memory. Alignments >100 Mb have been edited on an average desktop with reasonable efficiency (Hall, 1999).

2.5.4 MEGA

The MEGA (Molecular Evolutionary Genetics Analysis) software was developed with an aim of providing a biologist-centric, integrated suite of tools for statistical analyses of DNA and protein sequence data from an evolutionary perspective. Different versions have been released so far and it included tools for sequence alignment, phylogenetic reconstruction, phylogeny visualization, estimating sequence divergences, 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). With the fifth major release (MEGA5), the collection of analysis tools in MEGA has now broadened to include the Maximum Likelihood (ML) methods for molecular evolutionary analysis.

MEGA5 automatically infers the evolutionary tree by the Neighbor-Joining (NJ) algorithm that uses a matrix of pairwise distances estimated under the Jones-Thornton-Taylor (JTT) model for amino acid sequences or the Tamura and Nei (1993) model for nucleotide sequences. The principle of this method is to find pairs of operational taxonomic units (OTUs [= neighbors]) that minimize the total branch length at each stage of clustering of OTUs starting with a star like tree. (Saitou and Nei, 1987)

2.5.5 Ribosomal Database Project

The Ribosomal Database Project (RDP) is a curated database which offers ribosome-related data, analysis services and associated computer programs (http://rdp.cme.msu.edu/classifier/classifier.jsp). The services include

phylogenetically ordered alignments of ribosomal RNA (rRNA) sequences, derived phylogenetic trees, rRNA secondary structure diagrams and various softwares for handling, analyzing and displaying alignments and trees. The ribosomal RNA sequences in the RDP alignments are drawn from major sequence repositories (GenBank and EBI) and direct submissions to the RDP. They are organized and presented in an aligned and phylogenetically ordered form. Each sequence is annotated with its organismal source (for cultured organisms: the genus, species, culture collection numbers, *etc*.), cellular compartment, origin of sequence data (usually a literature citation) and other relevant information (Maidak *et al*., 1996).

The RDP Naïve Bayesian Classifier offers two hierarchy models for 16S rRNA and Fungal LSU (large subunit rRNA) genes. The existing hierarchy model used by the 16S rRNA Classifier comes from the new phylogenetically consistent higher-order bacterial taxonomy with some minor changes for lineage with few cultivated members. From the highest to the lowest, the major recognized taxonomic ranks are: domain, phylum, class, order, family and genus. There are occasional intermediate ranks such as "subclass" and "suborder". The classifier requires a sequence with at least 50 bases to get a good classification result. The number of query sequences is limited to 100000. Currently FASTA, GenBank and EMBL formats are allowed as input format in either uppercase or lowercase (Wang *et al*., 2007).

RDP's Pipeline aims to simplify the processing of large rRNA sequence libraries (including single-strand and paired-end reads) obtained through high-throughput sequencing technology. This site offers 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. For taxonomy-based analysis, the RDP Classifier provides fast and reliable classification of short sequence reads (Liu *et al*., 2008; Wang *et al*., 2007).

For taxonomy-independent alignment, the trimmed reads are aligned using the fast Infernal aligner. Reads are then clustered into Operational Taxonomic Units at multiple pairwise distances using custom code implementing the complete-linkage clustering algorithm (Cole *et al*., 2011).

Rarefaction is a technique to assess species richness from the results of sampling and allows the calculation of species richness based on the construction of so-called rarefaction curves. This curve is a plot of the number of species as a function of the number of samples. On the left, the steep slope indicates that a large fraction of the species diversity remains to be discovered. If the curve becomes flatter to the right, a reasonable number of individual samples have been taken and more intensive sampling is likely to yield only few additional species.

In the context of analyzing the microbial diversity based on 16S rRNA sequences, the rarefaction curve represents a plot of the number of Operational Taxonomic Units (OTUs) as a function of the number of 16S rRNA sequences from an individual sample. The OTUs are defined as clusters of sequences with a similarity of 97%, 95%, or 90% (*i.e.*, a distance of 0.03, 0.05, or 0.10)

Several other computational tools 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) which can be used to analyse microbial ecology data.

2.5.6 Phyre^2

Phyre and Phyre^2 (**P**rotein **H**omology/Analog**Y** **R**ecognition **E**ngine) are freely available web-based services used for protein structure prediction (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). 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*., 2000). 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.

A user-submitted sequence, hereafter known as the 'query', is also 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 detailed in 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. Where 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.

The Phyre system is typical of many of the freely available structure prediction systems on the Web, and to use the Phyre system, a user simply pastes the amino-acid sequence in FASTA format into a Web page together with their email address. Approximately 30 min later, the user will receive an email containing, among other things, a link to a Web page of results, including full

downloadable 3D models of their protein and associated confidence estimates. Recent major developments to the core algorithms of Phyre have placed it among the best servers in the most recent CASP8 (Critical Assessment of Structure Prediction) (2009) preliminary assessment. 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.

2.6 Mangrove ecosystem

Mangroves are boundary landform ecosystems present in tropical and subtropical regions, located in the intersection between the land and the sea. The majority (60-70%) of the world tropical and subtropical coastlines are covered with mangrove ecosystems (Aksornkoae *et al.,* 1984). Mangroves are known to be highly productive ecosystems and have immense ecological values as they protect and stabilize the costal zones, nourish and nurture the coastal water with nutrients. They play important role as the feeding and breeding areas of many organisms including plants, animals and micro-organisms. Mangrove sediments form a unique environment, with varying salinity and nutrient availability and are predominantly anaerobic. The microbial community in the mangrove sediment is strongly influenced by bio-geographical, anthropological and ecological properties. These properties include food web in the ecosystem, nutrient cycling and the presence of organic and inorganic matters (Ghosh *et al.,* 2010).

Mangrove ecosystems are generally nutrient-deficient, especially of nitrogen and phosphorus. Dissolved inorganic nutrients were dominated by high (~200–500 μM) concentrations of silicates, but porewater phosphate levels were usually below detection limits (<0.02 μM). Measured rates of nutrient regeneration were either slow into the sediment, or undetectable, but for some solutes such as silicate there was a high concentration gradient. Rates of bacterial DNA and protein synthesis and patterns of benthic primary production, indicate uptake of nutrients at the sediment-water interface by epibenthic microalgae and

sequestering of porewater solutes by very active, subsurface bacterial communities. Rapid growth of these bacteria may be partially maintained by the decomposition and release of nutrients of mangrove roots and rhizomes, as suggested by the dominance of silicate in the porewater. Correlation analysis supports the notion of nutrient limitation (mainly phosphorous) of bacteria and microalgae in mangrove muds. It appears that a close microbe-nutrient-plant connection serves as a mechanism to conserve scarce nutrients necessary for the existence of these tropical tidal forests (Alongi *et al.*, 1993).

In spite of this, mangroves are highly productive. Microbial activity is responsible for major nutrient transformations within a mangrove ecosystem. The highly productive and diverse microbial community living in tropical and subtropical mangrove ecosystems continuously transforms nutrients from dead mangrove vegetation into sources of nitrogen, phosphorus, and other nutrients that can be used by the plants. In turn, plant-root exudates serve as a food source for the microorganisms living in the ecosystem with other plant material serving similarly for larger organisms like crabs (Holguin *et al.,* 2001).

In tropical mangroves, bacteria and fungi constitute 91% of the total microbial biomass, whereas algae and protozoa represent only 7% and 2%, respectively (Alongi, 1988). Bacteria are responsible for most of the carbon flux in tropical mangrove sediments. They process most of the energy flow and nutrients, and act as a carbon sink. For example, in semiarid mangrove ecosystems on the Indus river in Pakistan, bacteria were attached to the sediment particles and processed most of the ecosystem nutrients (Bano *et al.,* 1997). Several studies have shown the uniqueness of mangrove sediments with respect to their microbial composition (Polymenakou *et al.,* 2005; Urakawa *et al.,* 1999). Hence studies on microbial diversity in the mangrove sediments are important to understand the process of biogeochemical cycling and pollutants removal (Roy *et al.,* 2002).

2.7 Proteases

Proteases are hydrolytic enzymes that catalyze the cleavage of peptide bonds in proteins. They are ubiquitous in occurrence, being found in all living organisms, including prokaryotes, fungi, plants and animals. Proteases are essential for cell growth and differentiation and have gained considerable attention in the industry community. They conduct highly specific and selective modifications of proteins such as activation of zymogenic forms of enzymes by limited proteolysis, blood clotting and lysis of fibrin clots, and processing and transport of secretory proteins across the membranes. Proteases represent one of the three largest groups of industrial enzymes, and most of these are of bacterial origin (Orhan *et al.,* 2005). Microorganisms account for 60% share of commercial protease production in the world (Kumar and Takagi, 1999; Gupta *et al.,* 2002; Rao *et al.,* 1998).

2.7.1 Types of proteases

Microbial proteases are of two types, intracellular and extracellular. Intracellular proteases are important for various cellular and metabolic processes, such as sporulation and differentiation, protein turnover, maturation of enzymes and hormones and maintenance of the cellular protein pool. Extracellular proteases are important for the hydrolysis of proteins in cell-free environments, whereafter the resulting hydrolytic products (small peptides and amino acids) can be transported into the cells and utilized as carbon or nitrogen sources (Rao *et al.,* 1998; Gupta *et al.,* 2002; Kalisz, 1988).

Especially, extracellular proteases are of industrial importance and are used in various industries like food, leather, detergent, pharmaceutical, waste management and silver recovery and for diagnosis of illness. Proteases constitute two thirds of the total number of enzymes used in industry (Gupta *et al.,* 2002; Godfrey and West, 1996). They have been widely used in the detergent industry,

since their introduction in 1914 as detergent additives (Kalisz, 1988). However the largest share of the enzyme market has been held by detergent alkaline proteases active and stable in the alkaline pH range.

Proteases are classified in subgroup 4 of group 3 (hydrolases, EC 3.4) by Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB, 1992). Proteases are broadly divided into two major groups, exopeptidases and endopeptidases, depending on their site of action. Exopeptidases cleave the peptide bond proximal to the amino or carboxy termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate. Proteases are further classified into seven prominent groups, serine proteases, aspartic proteases, cysteine proteases, glutamic proteases, asparagine peptide lyses, threonine proteases and metalloproteases, based on the functional group present at the active site (Hartley, 1960). However, certain proteases do not precisely fit into the standard classification, *e.g*., ATP-dependent proteases which require ATP for activity (Menon and Goldberg, 1987). Microbial proteases are also classified into acidic, neutral, or alkaline proteases, dependent on conditions for their optimum activity.

Proteases are also classified based on structure and sequence similarity, those with statistically significant similarities in amino acid sequences, are grouped into families, and families with related structures are grouped into clans (Argos, 1987) and are included in the MEROPS database. Each family of peptidases has been assigned a code letter denoting the type of catalysis, *i.e*., A, C, G, M, N, S, T or U for Aspartic, Cysteine, Glutamic, Metallo, Asparagine, Serine, Threonine or unknown type, respectively, together with a unique number. The letter "P" being used for families of proteases with more than one of the catalytic types serine, threonine and cysteine. Some families are divided into subfamilies because there is evidence of a very ancient divergence within the family, *e.g*. S1A,

S1B. Each clan is identified with two letters, the first letter representing the catalytic type of the families included in the clan (Rawlings and Barrett, 2012).

2.7.1.1 Exopeptidases

The exopeptidases act only near the termini of polypeptide chains. Based on their site of action at the N or C terminus, they are classified as amino and carboxypeptidases, respectively.

2.7.1.1.1 Aminopeptidases

Aminopeptidases (EC 3.4.11) act at a free N terminus of the polypeptide chain and liberate a single amino acid residue, a dipeptide, or a tripeptide. They are known to remove the N-terminal Methionine that may be found in heterologously expressed proteins but not in many naturally occurring mature proteins. Aminopeptidases occur in a wide variety of microbial species including bacteria and fungi (Watson, 1976). In general, aminopeptidases are intracellular enzymes, but extracellular aminopeptidase is also reported from *A. oryzae* (Labbe *et al.,* 1974). Aminopeptidase I from *Escherichia coli*, a large protease of 400,000 Da, has a broad pH optimum of 7.5 to 10.5 and requires Mg^{2+} or Mn^{2+} for optimal activity (De Marco and Dick, 1978)

2.7.1.1.2 Carboxypeptidases

The carboxypeptidases act at C terminals of the polypeptide chain and liberate a single amino acid or a dipeptide. Carboxypeptidases can be divided into three major groups, serine carboxypeptidases (EC 3.4.16), metallo carboxypeptidases (EC 3.4.17), and cysteine carboxypeptidases (EC 3.4.18), based on the nature of the amino acid residues at the active site of the enzymes. The serine carboxypeptidases isolated from *Penicillium* sp., *Saccharomyces* sp., and *Aspergillus* sp. are similar in their substrate specificities but differ slightly in other properties such as pH optimum, stability, molecular weight, and effect of inhibitors. Metallocarboxypeptidases from *Saccharomyces* sp. (Felix and

Brouillet, 1966) and *Pseudomonas* sp. (Lu *et al.*, 1969) require Zn^{2+} or Co^{2+} for their activity.

2.7.1.2 Endopeptidases

Endopeptidases are peptidases that hydrolyse the peptide bonds in the inner regions of the polypeptide chain away from the N and C termini. The presence of the free amino or carboxyl group has a negative influence on enzyme activity. The endopeptidases are divided into seven subgroups based on their catalytic mechanism, serine proteases, aspartic proteases, cysteine proteases, metalloproteases, threonine proteases, glutamate proteases and asparagine peptide lyases.

2.7.1.3 Serine proteases (EC 3.4.21)

The presence of a serine group in the active site is the characteristic of serine proteases. They are numerous and widespread among viruses, bacteria, and eukaryotes, as they are vital to the organisms. Serine proteases are found in the exopeptidase, endopeptidase, oligopeptidase, and omega peptidase groups. Based on their structural similarities, serine proteases have been grouped into 36 families, which have been further, subdivided into about 12 clans with common ancestors (Rawlings and Barrett, 2012). The primary structures of the members of four clans, chymotrypsin (SA), subtilisin (SB), carboxypeptidase C (SC), and *Escherichia* D-Ala–D-Ala peptidase A (SE) are totally unrelated, indicating that there are at least four separate evolutionary origins for serine proteases. Clans SA, SB, and SC have a common reaction mechanism consisting of a common catalytic triad of the three amino acids, serine, histidine and aspartate (Ser-His-Asp triad) (Rao *et al.*, 1998). The nucleophile is the hydroxyl on the side chain of the active site serine (Rawlings *et al.,* 2011). Although the geometric orientations of these residues are similar, the protein folds are quite different, forming a typical example of a convergent evolution. The catalytic mechanisms of clans SE and SF (repressor LexA) are distinctly different from those of clans SA, SB, and SE, as

they lack the classical Ser-His-Asp triad. Another interesting feature of the serine proteases is the presence of a conservation of glycine residues in the vicinity of the catalytic serine residue to form the motif Gly-Xaa-Ser-Yaa-Gly (Brenner, 1988).

Serine proteases are recognized by their irreversible inhibition by 3,4-dichloroisocoumarin (3,4-DCI), L-3-carboxytrans 2,3-epoxypropyl-leucylamido (4-guanidine) butane (E.64), diisopropylfluorophosphate (DFP), phenyl methyl sulfonyl fluoride (PMSF) and tosyl-L-lysine chloromethyl ketone (TLCK). Some of the serine proteases are inhibited by thiol reagents such as p-chloromercuribenzoate (PCMB) due to the presence of a cysteine residue near the active site (Govind et al., 1981)

Serine proteases are generally active at neutral and alkaline pH, with an optimum pH between 7 and 11, nevertheless, proteases with higher pH optima (10.0–12.5) from *Bacillus* sp. YaB have also been reported (Shimogaki *et al.*, 1991). They have broad substrate specificities including esterolytic and amidase activity. Their molecular masses range between 18 and 35 kDa, but with exemption as serine protease of high molecular mass (90 kDa) from *B. subtilis* (natto) (Kato *et al.*, 1992; Yamagata *et al.*,1995) and another serine protease from *Blakeslea trispora* with molecular mass 126 kDa (Govind *et al.*,1981) are also reported. The isoelectric points of serine proteases are generally between pH 4 and 6 (Rao *et al.*, 1998). There are various types of serine proteases, of which, chymotrypsin like proteases and subtilisin type proteases form a major group (Gupta *et al.*, 2002).

2.7.1.3.1 Chymotrypsin-like proteases

Chymotrypsin, a mammalian digestive protease, has structural homology with trypsin, elastase and thrombin (Graycar, 1999). It is specific for basic amino acids and is most active at pH 8. The most important inhibitors of this class are TLCK, TPCK, DFP and soyabean trypsin inhibitor. The molecular weight is

generally around 20 kDa. This group is well represented by proteases of animal origin and those belonging to various species of *Streptomyces,* like *S. erythreus, S. fradiae* and *S. griseus* (Ward, 1985; Kalisz, 1988).

2.7.1.3.2 Subtilisin-like proteases or subtilases

Subtilisin-like serine proteases are generally bacterial in origin, although there are reports in favor of other organisms. They are generally secreted extracellularly for the purpose of scavenging nutrients (Graycar, 1999). This class of proteases is specific for aromatic or hydrophobic residues (at position P1), such as tyrosine, phenylalanine and leucine. They are highly sensitive towards PMSF, DFP and potato inhibitor. They are most active around pH 10, with a molecular weight range of 15–30 kDa and an isoelectric point near pI 9. This class of protease is well represented by various species of *Bacillus*, like *B. amyloliquifaciens, B. licheniformis* and *B. subtilis* (Rao *et al.,* 1998). It is also produced by *Flavobacterium* (Morita *et al.,* 1998).

Two different types of subtilisins have been identified, subtilisin Carlsberg and subtilisin Novo or bacterial protease Nagase (BPN'). Subtilisin Carlsberg was produced by *Bacillus licheniformis* and was discovered in 1947 by Linderstrom and Ottesen at the Carlsberg laboratory. Subtilisin Novo or BPN' is produced by *Bacillus amyloliquefaciens*. Subtilisin Carlsberg is widely used in detergents. Both subtilisins have a molecular mass of 27.5 kDa, but differ from each other by 58 amino acids. They have similar properties such as optimal temperature of 60°C, an optimal pH of 10, exhibit broad substrate specificity and have an active-site triad made up of Ser221, His64 and Asp32. The serine alkaline protease from the fungus *Conidiobolus coronatus* was shown to possess a distinctly different structure from subtilisin Carlsberg in spite of their functional similarities (Phadatare *et al.,* 1993).

2.7.1.4 Aspartic proteases (EC 3.4.23)

Aspartic acid proteases are the endopeptidases that depend on aspartic acid residues for their catalytic activity and hence commonly called acidic proteases. Acidic proteases have been grouped into three families, namely, pepsin (A1), retropepsin (A2), and enzymes from pararetroviruses (A3) (Barett, 1994), and have been placed in clan AA of MEROPS database. There are 16 families belonging to 5 clans. The members of families A1 and A2 are known to be related to each other, while those of family A3 show some relatedness to A1 and A2. Most aspartic proteases show maximal activity at low pH (pH 3 to 4) and have isoelectric points in the range of pH 3 to 4.5. Their molecular masses are in the range of 30 to 45 kDa. The members of the pepsin family have a bilobal structure with the active-site cleft located between the lobes (Sielecki *et al.*, 1991). In aspartic peptidases, the nucleophile is a water molecule, which is activated by two aspartates. The active-site aspartic acid residue is situated within the motif Asp-Xaa-Gly, in which Xaa can be Ser or Thr. The aspartic proteases are inhibited by pepstatin. They are also sensitive to diazoketone compounds such as diazoacetyl-DL-norleucine methyl ester (DAN) and 1, 2-epoxy-3-(p-nitrophenoxy) propane (EPNP) in the presence of copper ions. Microbial aspartic proteases can be broadly divided into two groups, (i) pepsin-like enzymes produced by *Aspergillus, Penicillium, Rhizopus*, and *Neurospora* and (ii) rennin-like enzymes produced by *Endothia* and *Mucor* sp.

2.7.1.5 Cysteine/thiol proteases (EC 3.4.22)

Cysteine proteases occur in both prokaryotes and eukaryotes. About 20 families of cysteine proteases have been recognized. The nucleophile is the thiol on the side chain of the active site cysteine. The activity of all cysteine proteases depends on a catalytic dyad consisting of cysteine and histidine. The order of Cys and His (Cys-His or His-Cys) residues differ among the families (Barett, 1994).

Generally, cysteine proteases are active only in the presence of reducing agents such as hydrogen cyanide (HCN) or cysteine. Based on their side chain specificity, they are broadly divided into four groups: (i) papain-like, (ii) trypsin- like with preference for cleavage at the arginine residue, (iii) specific to glutamic acid, and (iv) others. Papain is the best-known cysteine protease. Cysteine proteases have neutral pH optima, although a few of them, *e.g*., lysosomal proteases, are maximally active at acidic pH. They are susceptible to sulfhydryl agents such as PCMB but are unaffected by DFP and metal-chelating agents. Clostripain, produced by the anaerobic bacterium *Clostridium histolyticum*, exhibits a stringent specificity for arginyl residues at the carboxyl side of the splitting bond and differs from papain in its obligate requirement for calcium. Streptopain, the cysteine protease produced by *Streptococcus* sp., shows a broader specificity, including oxidized insulin B chain and other synthetic substrates. Clostripain has an isoelectric point of pH 4.9 and a molecular mass of 50 kDa, whereas the isoelectric point and molecular mass of streptopain are pH 8.4 and 32 kDa, respectively (Gilles et al., 1979).

2.7.1.6 Metalloproteases (EC 3.4.24)

Metalloproteases are the most diverse of the catalytic types of proteases and are characterized by the requirement for a divalent metal ion, zinc or cobalt for their activity. The nucleophile involved in the catalysis is a water molecule which is activated by one or two metal ions (usually zinc, but also cobalt, manganese, nickel, copper, and iron). They include enzymes from a variety of origins such as collagenases from higher organisms, hemorrhagic toxins from snake venoms, and thermolysin from bacteria. Zinc-containing metalloproteases have the consensus amino acid sequence HEXXH (H-Histidine, E-Glutamate and X-any non conserved amino acid) as the zinc binding motif. In this motif, two histidine residues function as the first and second zinc ligands.

About 70 families of metalloproteases have been recognized, belonging to 15 different clans, of which 17 contain only endopeptidases, 12 contain only exopeptidases and one family, M13 contains both endo and exopeptidases. Families of metalloproteases have been grouped into different clans based on the nature of the amino acid that completes the metal-binding site; *e.g*., clan MA has the sequence HEXXH-E and clan MB corresponds to the motif HEXXH-H (Rao *et al.,* 1998).

Zinc-containing metalloproteases are widely distributed from prokaryotes to eukaryotes and are classified into four groups. The group possessing the HEXXH zinc binding motif is called the 'zincins superfamily'. Although this superfamily is further subdivided into at least ten families on the basis of location of the third zinc ligand, bacterial metalloproteases may fall into three families: thermolysin, serralysin and neurotoxin families of which prototype enzymes are produced by *Bacillus thermoproteolytics*, *Serratia marcescens* and *Clostridium botulinum* or *C. tetani*, respectively (Hooper *et al*., 1994)

2.7.1.6.1 Thermolysin

Analysis of the three-dimensional structure of thermolysin produced by *B. thermoproteolytics* clearly indicated the existence of three amino acid residues and one water molecule binding to a zinc(II) ion. In addition to two histidine residues in the HEXXH motif, glutamic acid, which is 25 residues towards the C-terminal from the motif (GXXNEXXSD), functions as the third zinc ligand. *P. aeruginosa* elastase and metalloproteases from *Legionella pneumophila, V. cholerae* and *V. vulnificus* were members of this family (Jiang and Bond, 1992). The enzymes in the thermolysin family hydrolyze the peptide bond at the amino group side of the P1 amino acid residue, usually a hydrophobic amino acid residue. Synthetic peptides, such as carbobenzoxy-glycyl-L-phenylalanine amide or carbobenzoxy-glycyl-L-leucine amide, are thus commonly used as the specific substrate for this

family. Phosphoramidon [N-(a-rhamnopyranosyloxyhydroxyphosphanyl) Leu-Trp] and zincov [2-(N-hydroxycarboxamido)-4-methylpentanoyl-Ala-Gly amide] are well-known competitive inhibitors for this family's proteases (Matthews, 1988).

2.7.1.6.2 Serralysins

The serralysin family includes *S. marcescens* metalloprotease (serralysin) and *P. aeruginosa* alkaline protease. Both metalloproteases have the extended zinc-binding motif HEXXHXXGXXH, and the third histidine of the motif and a water molecule act as the third and fourth zinc ligands, respectively. There is also a potential fifth ligand, tyrosine, at position 41. Both enzymes hydrolyze the peptide bond at the carboxy group side of the P1 amino acid residue (Matsumoto *et al.,* 1984; Shibuya *et al.,* 1991). They therefore possess the amidolytic activity toward peptide-4-methyl-coumaryl-7-amide substrates which were originally developed for serine- proteases and/or cysteine-proteases. Moncrief *et al.* (1995) documented that *B. fragilis* enterotoxin was a zinc-containing metalloprotease possessing the HEXXHXXGXXH motif. However, the putative fifth zinc ligand is not found in the amino acid sequence reported.

2.7.1.6.3 Neurotoxins

The neurotoxins produced by *C. botulinum* and *C. tetani* are known to inhibit the release of the neurotransmitter, acetylcholine. Genetic investigations clarified that these neurotoxins are a new family of zinc metalloproteases. Like other bacterial metalloproteases, a zinc(II) ion may be coordinated by two histidine residues in the HEXXH motif (Oguma *et al.,* 1995). To date, there is no evidence of the third zinc ligand; however, the most likely residue is a conserved glutamic acid, which is 40 residues towards the C-terminal from the motif.

2.7.1.7 Threonine proteases (EC 3.4.25)

Threnonine proteases are a family of proteolytic enzymes harbouring a threonine (Thr) residue within the active site. Threonine proteases use the secondary alcohol of their N-terminal threonine as a nucleophile to perform catalysis (Seemu¨ller *et al.,* 1995; Brannigan *et al.,* 1995; Cheng and Grishin, 2005; Rawlings and Barrett, 2012). The threonine must be N-terminal since the terminal amide of the same residue acts as a general base by polarising an ordered water which deprotonates the alcohol to increase its reactivity as a nucleophile (Dodson and Wlodawer, 1998; Ekici *et al.,* 2008). Catalysis takes place in two steps, initially the nucleophile attacks the substrate to form a covalent acyl-enzyme intermediate, releasing the first product. Then the intermediate is hydrolysed by water to regenerate the free enzyme and release the second product. The prototype members of this class of enzymes are the catalytic subunits of the proteasome, however the acyltransferases.

2.7.1.8 Glutamic proteases

The sixth catalytic type of proteases was identified in 2004 when certain fungal endopeptidases now known as eqolysins were discovered to be glutamate peptidases (Fujinaga *et al.,* 2004). They have very limited distribution, with most of them found in human or plant pathogenic fungi (Oda, 2012). These types of proteases are represented by 2 families, G1 and G2, in the MEROPS database belonging to 2 different clans, GA and GB. They are endopeptidases and the nuclephile involved in catalytic reaction is a water molecule bound by a glutamate residue.

2.7.1.8.1 Family G1 Glutamic peptidases

This family contains fungal pepstatin-insensitive endopeptidases such as aspergilloglutamic peptidase and scytalidoglutamic peptidase or eqolisin. They are

active at very low pH and inhibited by EPNP. The nucleophile involved in catalysis is a water molecule bound by Glu and Gln, hence the name eqolisins, after the single letter abbreviations of these amino acids. EPNP binds covalently with Glu53 in scytalidoglutamic peptidase and the equivalent residue in aspergilloglutamic peptidase is Gln (Fujinaga *et al.*, 2004).

2.7.1.8.2 Family G2 Glutamic proteases

They are self cleaving endopeptidases. Tail spike protein of bacteriophage phi29 belonged to family G2 and Glu695 has been identified as the active site nucleophile. In the self-processing of the tail spike protein of bacteriophage phi29, Glu695 from one chain is thought to activate a water molecule, which attacks the peptide bond Trp690+Ser on another chain (Schulz & Ficner, 2011; Xiang *et al.*, 2009).

2.7.1.9 Asparagine peptide lyases

Asparagine peptide lyases are considered as the seventh catalytic type of proteolytic enzymes and are not hydrolases but lyases utilizing asparagine as a nucleophile. They are self-cleaving proteins which include the Tsh protein precursor of *Escherichia coli*, in which the large C-terminal propeptide acts as an autotransporter; certain viral coat proteins; and proteins containing inteins. Proteolysis is the action of an amidine lyase (Rawlings *et al.*, 2011). These proteases utilize the peculiar characteristic of asparagine which can be induced to attack its own carbonyl bond leading to the formation of a succinimide ring and peptide bond cleavage and it happens when a second active site residue, an aspartate or glutamate, is brought into close proximity. Cleavage is intramolecular in *cis*. No examples of asparagine peptide lyases have been identified in plants and animals, and homologues are only known from viruses, archaea, bacteria, single-celled eukaryotes and some algae. Ten families of peptide lyases have been included in the MEROPS database, two families of autotransporter proteins, five

families of viral coat proteins, and three families of intein containing proteins (Rawlings and Barrett, 2012).

2.7.2 Sources of proteases

Proteolytic enzymes are ubiquitous in occurrance, being found in all plants, animals and microorganisms. The inability of the plant and animal proteases to meet current world demands has led to an increased interest in microbial proteases. Microorganisms represent an excellent source of enzymes owing to their broad biochemical diversity and their susceptibility to genetic manipulation. They represent an attractive source of proteases as they can be cultured in large quantities in a relatively short time by established fermentation methods, and they produce an abundant, regular supply of the desired product. Generally microbial proteases are extracellular in nature and are directly secreted into the fermentation broth by the producer, thus simplifying downstream processing of the enzyme as compared to proteases obtained from plants and animals. Proteases from microbial sources are preferred to the enzymes from plant and animal sources since they possess almost all the characteristics desired for their biotechnological applications.

Microorganisms account for a two-third share of commercial protease production in the world (Kumar and Takagi, 1999) and approximately 40% of the total worldwide enzyme sales (Godfrey and West, 1996). Despite the long list of protease producing microorganisms, only a few are considered as appropriate for commercial exploitation, being 'Generally recognized as safe' (GRAS), non-toxic and non-pathogenic. A large number of microbes belonging to bacteria, fungi, yeast and actinomycetes are known to be producers of alkaline protease.

2.7.2.1 Bacteria

Most commercial proteases, mainly neutral and alkaline, are produced by organisms belonging to the genus *Bacillus*. Most potential alkaline protease

producing bacilli are strains of *Bacillus alcalophilus*, *B. amyloliquefaciens*, *B. licheniformis* and *B. subtilis* (Kumar and Takagi, 1999; Rao *et al.*, 1998; Gupta *et al.*, 2002). Some of the Gram-negative bacteria producing alkaline proteases were identified as *Pseudomonas aeruginosa* (Morihara *et al.*, 1963); *Pseudomonas maltophila* (Kobayashi *et al.*,1985); *Pseudomonas* sp. strain B45 (Chakraborty and Srinivasan,1993); *Xanthomonas maltophila* (Debette, 1991); *Vibrio alginolyticus* (Deane *et al.*,1987); and *Vibrio metschnikovii* strain RH530 (Kwon *et al.*, 1994). Alkaline proteases are also produced by some rare microorganisms. *Kurthia spiroforme*, a spiral shaped Gram-positive bacterium possessing a distant relationship to genus *Bacillus*, was reported to produce alkaline proteases (Steele *et al.*, 1992). Further, a bacterial isolate capable of producing alkaline protease and showing a symbiotic relationship with a marine shipworm, *Psiloteredo healdi*, was also reported by Greene *et al.* (1989).

Bacterial neutral proteases are active in a narrow pH range (pH 5 to 8) and have relatively low thermotolerance. The bacterial neutral proteases are characterized by their high affinity for hydrophobic amino acid pairs. Their low thermotolerance is advantageous for controlling their reactivity during the production of food hydrolysates with a low degree of hydrolysis. Some of the neutral proteases belong to the metalloprotease type and require divalent metal ions for their activity, while others are serine proteinases, which are not affected by chelating agents.

Bacterial alkaline proteases are characterized by their high activity at alkaline pH, *e.g.*, pH 10, and their broad substrate specificity. Their optimal temperature is around 60°C. These properties of bacterial alkaline proteases make them suitable for use in the detergent industry (Rao *et al.*, 1998).

2.7.2.2 Halophiles

Halophiles that produce alkaline proteases included *Halobacterium* sp. (Ahan *et al.,* 1990, *Halobacterium halobium* ATCC 43214 (Ryu *et al.,* 1994) and *Halomonas* sp. ES-10 (Kim *et al.,* 1991; Kim *et al.,* 1992). The alkalopsychrotrophic and alkalopsychrophilic bacteria represent a new potential source for alkaline proteases (Margesin and Schinner, 1994). These organisms are characterized by their adaptation to both cold temperatures and alkaline conditions. An alkalo-psychrotrophic *Bacillus* sp. capable of producing alkaline proteases of high activity at low temperatures was isolated by Margesin *et al.* (1992). Despite the many published reports on alkaline proteases from alkalophilic *Bacillus* sp., very few reports exist on thermostable alkaline proteases from alkalophiles. Many of the thermophilic alkalophiles have growth temperatures of >60ºC (Fujiwara *et al.,* 1991; Rahman *et al.,* 1994), with a few exceptions of <60ºC (Durham *et al.,* 1987; Takami *et al.,* 1989).

2.7.2.3 Fungi

Fungi elaborate a wider variety of enzymes than do bacteria. For example, *Aspergillus oryzae* produces acid, neutral, and alkaline proteases. The fungal proteases are active over a wide pH range (pH 4 to 11) and exhibit broad substrate specificity. However, they have a lower reaction rate and low thermo tolerance than bacterial proteases. Fungal enzymes can be conveniently produced in a solid-state fermentation process. Fungal acid proteases have an optimal pH between 4 and 4.5 and are stable between pH 2.5 and 6.0. They are particularly useful in the cheese making industry due to their narrow pH and temperature specificities. Fungal neutral proteases are metalloproteases that are active at pH 7.0 and are inhibited by chelating agents. Fungal alkaline proteases are also used in food protein modification.

Many of the fungi have been reported to produce extracellular alkaline proteases (Matsubara and Feder, 1971), with *Aspergillus* sp., being the most exploited group (Chakrabarti *et al.,* 2000; Rajamani and Hilda, 1987). Other fungal producers include *Conidiobolus* sp. (Bhosale *et al.,* 1995), *Penicillium* sp. (Germano *et al.,* 2003) and *Rhizopus* sp. (Banerjee *et al*., 1993). Some alkaline protease producing strains of imperfect fungi, such as *Dendryphiella* sp. and *Scolebasidium* sp., have found application in detergents (Pedersen *et al.,* 1992). Yeasts reported to produce alkaline proteases include *Candida lipolytica* (Tobe *et al.,* 1976), *Yarrowia lipolytica* (Ogrydziak, 1993) and *Aureobasidium pullulans* (Donaghy and McKay, 1993).

2.7.2.4 Actinomycetes

Very few studies exist on the alkaline protease producing alkalophilic actinomycetes (Mikami *et al.,* 1986). The different species of *Streptomyces* reported to produce alkaline proteases include *Streptomyces rectus* var. *proteolyticus* (Mizusawa *et al.,* 1964; 1966), *Streptomyces griseus* (Su and Pan, 1972), *Streptomyces* sp. (Nakanishi *et al.,*1974; Kumar and Tiwari, 1994), *Streptomyces moderatus* NRRL 3150 (Chandrasekaran and Dhar, 1983), *Streptomyces* sp. YSA 2 130 (Yum *et al.,* 1994), *S. diastaticus* SS1 (Chaphalkar and Dey, 1994; 1998), *S. corchorusii* ST36 9 (El-Shanshoury *et al.,* 1995) and *S. pactum* DSM 40530 (Böckle *et al.,*1995). Other types of alkalophilic actinomycetes include *Nocardiopsis dassonvillei* (Liu *et al.,* 1988; Tsujibo *et al.,* 1990; Kim *et al.,* 1993) and *Oerskovia xanthineolytica* TK-1 (Tamai *et al.,* 1994).

2.7.2.5 Viruses

Viral proteases are involved in the processing of proteins of viruses that cause certain fatal diseases such as AIDS and cancer. Viral proteases are mainly serine, aspartic and cysteine peptidases (Rawlings and Barrett, 1993). All of the virus-encoded peptidases are endopeptidases; there are no metallopeptidases.

Retroviral aspartyl proteases that are required for viral assembly and replication are homodimers and are expressed as a part of the polyprotein precursor. The mature protease is released by autolysis of the precursor. Kuo and Shafer (1994) have extensively studied the expression, purification, and enzymatic analysis of retroviral aspartic protease and its mutants. Extensive research has focused on the three-dimensional structure of viral proteases and their interaction with synthetic inhibitors with a view to designing potent inhibitors that can combat the relentlessly spreading and devastating epidemic of AIDS (Rao *et al.,* 1998).

Thus, although proteases are widespread in nature, microbes serve as a preferred source of these enzymes because of their rapid growth, the limited space required for their cultivation, and the ease with which they can be genetically manipulated to generate new enzymes with altered properties that are desirable for their various applications.

2.7.3 Protease gene

Proteases from *Bacillus* sp. have been extensively studied and are found to be highly diverse with respect to their function and characteristics. Even a single organism harboured different types of proteases (Chen *et al.,* 2004). By elucidating the entire genome sequence of *Bacillus subtilis* strain 168, genes responsible for different types of proteases like extracellular alkaline serine protease (*aprE*), intracellular alkaline serine protease (*aprX*), minor extracellular serine protease (*epr*), major intracellular serine protease (*ispA*), extracellular metalloprotease (*mpr*), extracellular neutral protease B (*nprB*), extracellular neutral metalloprotease (*nprE*) were found (Kunst *et al.,* 1997). Despite the great diversity of their nucleotide sequences the structures of active site for proteins with similar function are highly conserved.

Donovan *et al.* (1997) reported cloning of *nprA* gene, encoding *Bacillus thuringiensis* neutral protease A. The size of neutral protease A deduced from the

nprA sequence was 566 amino acids with a molecular weight of 60,982 Da. Sidler *et al.* (1986) reported a 317 amino acid long neutral protease from *B. cereus* with a molecular weight of 35 kDa. Sadeghi and colleagues (2009) cloned 1326 bp long alkaline protease gene of *B. subtilis* 168 encoding a 442 aa protein. The gene encoding subtilisin Carlsberg from *B. licheniformis* have been cloned in pBR322 vector by Jacobs *et al.* (1985) that has a 1137 bp open reading frame (ORF) encoding 379 aa. Peng *et al.* (2004) have cloned the subtilisin DFE gene from *B. amyloliquefaciens* DC-4 in pGEM-T plasmid that has a 1146 bp ORF encoding 382 aa. Also *B. alkalophilus* alkaline protease genes have been cloned and sequenced which encode 380 aa proteins (Laan *et al.*, 1991). Aoyama *et al.* (2000) reported 1149 bp gene encoding an alkaline serine proteinase of *Bacillus pumilus* TYO-67, encoding 383 amino acids. The *espI* gene of 2133 bp, encoding an algicidal extracellular serine protease was isolated from *Pseudoalteromonas* sp. strain A28 which could code for a protein of 711 amino acid residues having a calculated molecular weight of 73.2 kDa (Kohno *et al.*, 2007).

2.7.4 Purification of proteases

Purification of proteases to homogeneity is a prerequisite to study their mechanism of action and behaviour. There are various methods adopted for the purification of proteases like ultrafiltration, chromatography, preparative gel electrophoresis and so on. Various chromatographic techniques like ion exchange chromatography (IEC), affinity chromatography (AC), hydrophobic interaction chromatography (HIC) and gel filtration chromatography are most widely used (Kumar and Takagi, 1999).

2.7.4.1 Precipitation of proteins

The initial step of purification is to separate the extracellular protein fraction from the cellular fraction, which is done by filtration or centrifugation. Precipitation also helps in partial purification as well as concentration of proteins.

The proteins are then precipitated from the crude extract by the addition of reagents such as salt or an organic solvent, which lowers the solubility of the desired proteins in an aqueous solution (Bell *et al.,* 1983).

Ammonium sulphate is the commonly used salt for protein precipitation, though sodium sulphate is also rarely used. Sodium sulphate has better precipitating qualities over ammonium sulphate, but due to poor solubility at low temperatures its use is restricted (Shih *et al.,* 1992). The precipitated protein fraction after dialysis can be concentrated using membrane filter (Amicon-Ultracel-3 kDa) (Ghafoor and Hasnain, 2010).

There are many reports of the use of the organic solvent, acetone at different volume concentration like 5 volumes (Horikoshi, 1971), 3 volumes (Tsujibo *et al.,* 1990; Kim *et al.,* 1996) and 2.5 volumes (Kumar *et al.,* 1997) for the recovery of alkaline proteases. There are also various reports on protein precipitation with acetone at different concentrations like 80% (v/v) (Durham, 1987; Kwon *et al.*,1994), 66% (v/v) (Yamagata *et al.*,1995) or 44%, 66%, and 83% (v/v) (El-Shanshoury *et al.*,1995) followed by centrifugation and/or drying. Precipitation of enzymes is also achieved by the use of water-soluble, neutral polymers such as polyethylene glycol (PEG 35000) (Larcher *et al.,* 1996), activated charcoal, temperature sensitive hydrogel (Han *et al.,* 1995), heat treatment of enzyme (Rahman *et al.,* 1994) and lyophilisation (Manonmani and Joseph, 1993).

2.7.4.2 Ion-exchange chromatography (IEC)

Alkaline proteases are generally positively charged and are not bound to anion exchangers (Tsai *et al.,* 1983; Fujiwara *et al.,* 1993). Hence, cation exchangers are used and the bound molecules are eluted from the column by an increasing salt or pH gradient (Tsuchiya *et al.,* 1992). Commonly used matrix materials include diethyl amino ethyl (DEAE) and carboxy methyl cellulose which

gets adsorbed to the charged protein molecules, thereby adsorbing the protein into the matrices and the adsorbed molecule is eluted by a gradient change in the pH or ionic strength of the buffer or the solution (Kumar and Takagi, 1999; Gupta *et al.*, 2002). Ion exchange chromatography on DEAE-Cellulose was performed for purification of alkaline proteases from *Bacillus altitudans* (Kumar, 2011), *Engyodontium album* (Chellappan *et al.,* 2006), Gammaproteobacterium (Sana *et al.,* 2006), *Bacillus cereus* (Shah *et al.,* 2010), *Bacillus* sp PN51 (Tanskul *et al.,* 2009) and *Streptomyces clavuligerus* (Thumar and Singh, 2007). Ion exchange chromatography using DEAE-Toyopearl column was used for purification of collagenolytic subtilisin like protease from *Alkalimonas collagenimarina* AC40 (Kurata *et al.,* 2010), Mono Q-sepharose column for alkaline protease from *Bacillus licheniformis* MP11 (Jellouli *et al.,* 2011), Q-sepharose HP column for alkaline protease from *Salinivibrio* sp strain AF-2004 (Karbalaei-Heidari *et al.,* 2008), *Bacillus sphaericus* (Singh *et al.,* 2001), DEAE-Sepharose CL-6B and phenyl sepharose for purification of nattokinase from *Bacillus subtilis* TKU007 (Wang *et al.,* 2006), CM-Sephadex C50-F for *Bacillus subtilis*-150 (Normurodova *et al.,* 2010) and CM-cellulose chromatography for alkaline protease from *Bacillus halodurans* JB99 (Shrinivas and Naik, 2011).

2.7.4.3 Affinity chromatography

Reports on the purification of alkaline proteases by different affinity chromatographic methods showed that an affinity adsorbent hydroxyapatite was used to separate the neutral protease (Keay and Wildi, 1970) as well as purify the alkaline protease from a *Bacillus* sp. (Kobayashi *et al.,* 1996). Other affinity matrices used were Sephadex-4-phenylbutylamine (Ong and Gaucher, 1976), immobilized casein glutamic acid (Manonmani and Joseph, 1993), aprotinin agarose (Petinate *et al.,* 1999), casein agarose (Böckle *et al.,* 1995; Manachini *et al.,* 1988) or N-benzoyloxycarbonyl phenylalanine immobilized on agarose adsorbents (Larcher *et al.,* 1996). However, the major limitations of affinity

chromatography are the high cost of enzyme supports and the labile nature of some affinity ligands.

2.7.4.4 Aqueous two-phase systems

This technique is used for purification of alkaline proteases using mixtures of polyethylene glycol (PEG) and dextran or PEG and salts such as H_3PO_4 and $MgSO_4$ (Lee and Chang, 1990; Sinha *et al.,* 1996; Hotha and Banik, 1997).

2.7.4.5 Gel filtration chromatography

Gel filtration is used for the rapid separation of macromolecules based on size and are used in early to middle stage of purification (Gupta *et al.,* 2002). Sephacryl S-200 gel filtration chromatography was used for the purification of alkaline protease from *Bacillus cereus* BG1 (Ghorbel-Frikha *et al.,* 2005) and keratinase from *Streptomyces* sp. strain 16 (Xie *et al.,* 2010), Sephadex G-75 for alkaline serine protease from *Bacillus mojavensis* A21 (Haddar *et al.,* 2009), BIO-GEL P-100 gel column for protease from *Pseudomonas* sp. CL1457 (Shastry and Prasad, 2002), Sephadex G-100 for alkaline protease from *Bacillus halodurans* JB99 (Shrinivas and Naik, 2011) and Sephadex G-50 column for alkaline protease from *Bacillus cereus* VITSN04 (Sundararajan *et al.,* 2011).

2.7.5 Characteristics of proteases

Proteases have been isolated and studied extensively and based on their properties they are used in various industries. The major characteristics like molecular weight, optimum pH and optimum temperature of some of the most studied proteases are presented in the Table 2.1.

Table 2.1 Characterisitics of some of the bacterial proteases

Microorganism	pH optimum	Temp optimum (°C)	MW (kDa)	Reference
Alcaligenes faecalis	9.0	55	67	Thangam and Rajkumar, 2002
Arthrobacter nicotianae 9458	9.0, 9.5	55-60, 37	55, 70-72	Smacchi *et al.,*1999
Pimelobacter sp.	9.0	50	23	Oyama *et al.,*1997
Pseudomonas aeruginosa PST-01	8.5	55	38	Ogino *et al.,*1999
Serratia marcescens ATCC 25419	9.5	48	38	Romero *et al.,* 2001
Bacillus sp. JB99	11.0	70	29	Johnvesly and Naik, 2001
Bacillus sp. NG-27	9.2	40	n.s	Sumandeep *et al.,*1999
Bacillus sp. KSM-KP43	11.0	70	n.s	Saeki *et al.,* 2002
Bacillus sp. NCDC-180	11.0, 12.0	50, 55	28, 29	Kumar *et al.,*1997
Bacillus sp. PS179	9.0	75	42	Hutadilok-Towatana *et al.,* 1999
Bacillus sp. SSRI	10.0	40	29	Singh *et al.,*2001
Bacillus brevis MTCC B0016	10.5	37	n.s	Banerjee *et al.,*1999
Bacillus mojavensis	10.5	60	30	Beg and Gupta, 1999
Bacillus pumilis MK6-5	11.5	55-60	28	Kumar, 2002
Bacillus pumilis	10.0	55	32	Huang *et al.,* 2003
Bacillus subtilis PE-11	10.0	60	15	Adhinarayana *et al.,* 2003
Bacillus subtilis EAG-2	8.5	65	27	Ghafoor and Hasnain, 2010
B. subtilis Y-108	8.0	50	44	Yang *et al.,* 2000
Bacillus cohnii APT5	11.0	50	n.s	Tekin *et al.,* 2012
Bacillus sp. SBP-29	9.5	60	n.s	Saurabh *et al.,* 2007

n.s – not specified

2.7.5.1 Molecular mass

The molecular masses of alkaline proteases generally ranges from 15 to 30 kDa (Fogarty *et al.*, 1974) with few reports of higher molecular masses of 31.6 kDa (Freeman *et al*, 1993), 33 kDa (Samal *et al*, 1991, Larcher *et al*., 1996), 36 kDa (Tsujibo *et al.*, 1990) and 45 kDa (Kwon *et al.*, 1994). However, an enzyme from *Kurthia spiroforme* had an extremely low molecular weight of 8 kDa (Steele *et al.*, 1992).

Rai and Mukherjee (2009) reported that a bacterium may produce arrays of extracellular protease isoenzymes for its survival and growth in a particular

habitat. There are only a limited number of studies dealing with the isoenzymes of alkaline proteases produced by genus *Bacillus*. Among these studies, Mala and Srividya (2010) reported isolating 2 different isoenzymes with molecular weights of 66 kDa and 18 kDa from *Bacillus* sp., which had a close homology with *Bacillus cohnii* YN-2000 protease.

2.7.5.2 Zymography

Zymography is an electrophoretic technique, commonly based on sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE), which contains a substrate copolymerized within the polyacrylamide gel matrix, for the detection of an enzymatic activity. Normally gelatin, either 0.1% (Felicioli *et al*., 1997; Neveu *et al*., 2011; Pushpam *et al*., 2011) or 0.5% (Raut *et al*., 2012), 1% casein (Schmidt *et al*, 1988; Ghafoor and Hasnain, 2010), 1% skim milk (Tekin *et al*., 2012) or fibrin are used for the detection of proteases (Wilkesman and Kurz, 2009). Proteases that have the ability to renature after removal of SDS and to exert proteolytic activity on a co-polymerized substrate can be analyzed with this method. Coomassie Blue staining of the gel reveals sites of proteolysis as translucent bands on a dark blue background (Frederiks and Mook, 2004).

2.7.5.3 Optimum pH and temperature

The optimum pH range of alkaline proteases is generally between 9 and 11, with a few exceptions of higher pH optima of 11.5, 11–12, 12.3 and 12–13 (Rao *et al.,* 1998). The optimum temperature of alkaline proteases ranges from 50 to 70ºC. In addition, the enzyme from an alkalophilic *Bacillus* sp. B189 showed an exceptionally high optimum temperature of 85ºC. Alkaline proteases from *Bacillus* sp., *Streptomyces* sp. and *Thermus* sp. are quite stable at high temperatures, and the addition of Ca^{2+} further enhanced enzyme thermostability. The alkaline and thermotolerant nature of alkaline proteases find their application in detergent industry. Currently, there is a requirement in the detergent industry

for alkaline proteases which are active at low temperatures. Kannase, marketed by Novo Nordisk Bioindustry, Japan is an enzyme active at temperatures as low as 10-20ºC (Gupta *et al.*, 2002).

2.7.5.4 Isoelectric point

The isoelectric point (pI) is the pH at which a particular molecule or surface carries no net electrical charge. Amphoteric molecules such as proteins contain both acidic and basic functional groups. Amino acids that make up proteins may be positive, negative, neutral, or polar in nature, and together give a protein its overall charge. The net charge on the molecule is affected by pH of its surrounding environment and can become more positively or negatively charged due to the gain or loss, respectively, of protons (H^+). At pI the amino acid is neutral, *i.e*. the zwitter ion form is dominant and they do not migrate in an electric field.

At a pH below pI, proteins carry a net positive charge and above their pI they carry a net negative charge. Proteins can, thus, be separated on a polyacrylamide gel according to their isoelectric point (overall charge) using a technique called isoelectric focusing, which uses a pH gradient to separate proteins (Markland and Smith, 1971; Ottesen and Svenden, 1971; Engel *et al*., 1998). Also the isoelectric points of some model proteins and *Bacillus subtilis* neutral protease was determined by the cross partitioning using polyethylene glycol/Dextran aqueous two-phase systems (Han and Lee, 1997). The alkaline proteases have high isoelectric points and are generally stable between pH 6 and 12 (Rao *et al.*, 1998).

2.7.5.5 Metal ion requirement

Alkaline proteases requires a divalent cation like Ca^{2+}, Mg^{2+} and Mn^{2+} or a combination of these cations, for maximum activity. These cations were also

found to enhance the thermal stability of a *Bacillus* alkaline protease (Paliwal *et al.,* 1994). Other metal ions such as Ba^{2+}, Co^{2+}, Fe^{3+} and Zn^{2+} are also used for stabilizing proteases (Johnvesly and Naik, 2001, Ratray *et al.,* 1995). It is believed that these cations protect the enzyme against thermal denaturation and play a vital role in maintaining the active conformation of the enzyme at high temperatures as they impart structural and thermal stability to the protein (Pan and Lin, 1991). In addition, specific calcium binding sites that influence the protein activity and stability apart from the catalytic site were described for Proteinase K (Bajorath *et al.,* 1988).

2.7.5.6 Inhibitors

Inhibition studies give insight into the nature of the enzyme, its cofactor requirements, and the nature of the active site (Sigman and Moser, 1975). In some of the studies, catalytic activity was inhibited by Hg^{2+} ions (Rahman *et al.,* 1994; Shimogaki *et al.,* 1991). Most serine peptidases are inhibited by DFP or PMSF. PMSF sulfonates the essential serine residue in the active site and results in the complete loss of activity (Gold and Fahrney, 1964; Haddar *et al.,* 2009; Kumar, 2002). In addition, some of the alkaline proteases were found to be metal ion dependent in view of their sensitivity to metal chelating agents, such as EDTA, EGTA or 1,10-phenanthroline (Steele *et al.,* 1992; Dhandapani and Vijayaragavan, 1994; Shevchenko *et al.,* 1995). Thiol inhibitors have little effect on alkaline proteases of *Bacillus* sp., although they do affect the alkaline enzymes produced by *Streptomyces* sp. (Yum *et al.,* 1994; El-Shanshoury *et al.,* 1995). Most cysteine peptidases are inhibited by iodoacetate (Barrett, 1994).

2.7.5.7 Substrate specificity

Alkaline proteases are active against many synthetic substrates as well as natural proteins and are more active against casein than haemoglobin or bovine serum albumin. Alkaline proteases are specific against aromatic or hydrophobic

amino acid residues such as tyrosine, phenylalanine, or leucine at the carboxyl side of the cleavage point (Kumar and Takagi, 1999). There are specific types of alkaline proteases like collagenase, elastase, keratinase (Friedrich *et al.*, 1999) and insect cuticle degrading protease (Urtz and Rice, 2000) which are active against specific protein substrates such as collagen, elastin, keratin and cuticle.

Tsai *et al.* (1984) observed that alkaline elastase from *Bacillus* sp. Ya-B hydrolysed elastin and elastase-specific substrates like succinyl-Ala3-p-nitroanilide and succinyl-Ala-Pro-Ala-p-nitroanilide at a faster rate. This enzyme showed a preference for aliphatic amino acid residues, such as alanine, that are present in elastin. Elastolysis was considered to be initiated by the formation of an enzyme–substrate complex through electrostatic interaction between positively-charged residues of the elastase and negatively-charged residues of the elastin in a pH range below 10.6 (Tsai *et al.*, 1986). In keratin, the disulfide bonds form an important structural feature and prevent the proteolytic degradation of the most compact areas of the keratinous substrates. However, clevage of keratin can be accompanied by the use of disulfide-reducing agents like thioglycolic acid or dithiothreitol. A thermostable alkaline protease from an alkalophilic *Bacillus* sp. AH-101 exhibiting keratinolytic activity showed degradation of human hair keratin with 1% thioglycolic acid at pH 12 and 70ºC, and the hair was solubilized within 1 h (Takami *et al.*, 1992). Similarly, enhanced keratin degradation after addition of DTT has also been reported for alkaline proteases of *Streptomyces* sp. (Sinha *et al.*, 1991).

2.8 Metagenomics derived proteases

Despite the abundance of new enzymes isolated by metagenomic approaches and the important industrial potential of proteases, there are relatively few data concerning characterization of metagenome-derived proteases. A fibrinolytic metalloprotease has been characterized from a deep sea sediment

metagenomic library (Lee *et al.,* 2007). Recently, two other metalloproteases with a novel domain structure were isolated from soil metagenomic libraries (Waschkowitz *et al.,* 2009). Pushpam *et al.* (2011) reported an alkaline serine protease from goat skin surface metagenome. Two serine proteases were reported from metagenomic libraries of the Gobi and Death Valley deserts (Neveu *et al.,* 2011). Prabavathi *et al.* (2012) reported a protease enzyme from a metagenomic library constructed from marine soil sediments. In several other functional metagenomic studies, screening for proteolytic activities was unsuccessful (Laemmli *et al.,* 1970; Rondon *et al.,* 2000), or resulted in false-positive clones (Jones *et al.,* 2007).

2.9 Applications of proteases

Bacterial proteases find numerous applications in various industrial sectors with their bulk uses in food and feed industry, leather industry, waste management, photographic industry, medical industry, silk degumming and detergent industry. Alkaline proteases account for a major share of the enzyme market all over the world (Godfrey and West, 1996; Kalisz, 1988). Several products based on bacterial alkaline proteases have been launched successfully in the market in past few years by different companies worldwide (Table 2.2).

Table 2.2 Commercially available microbial proteases

Product trade name	Microbial source	Applications	Supplier
Alcalase	*Bacillus licheniformis*	Detergent, silk degumming	Novo Nordisk, Denmark
Savinase	Alkalophilic *Bacillus* sp	Detergent, textile	Novo Nordisk, Denmark
Esperase	*B. lentus*	Detergent, food, silk degumming	Novo Nordisk, Denmark
Biofeed pro	*B. licheniformis*	Feed	Novo Nordisk, Denmark
Durazym	Protein engineered variant of Savinase®	Detergent	Novo Nordisk, Denmark
Novozyme 471MP	n.s.	Photographic gelatin hydrolysis	Novo Nordisk, Denmark
Novozyme 243	*B. licheniformis*	Denture cleaners	Novo Nordisk, Denmark
Nue	*Bacillus* sp.	Leather	Novo Nordisk, Denmark
Maxacal	*B. alcalophilus*	Detergent	Gist-brocades, The Netherlands
Maxatase	*B. alcalophilus*	Detergent	Gist-brocades, The Netherlands
Subtilisin	*B. alcalophilus*	Detergent	Gist-brocades, The Netherlands
Opticlean	*Bacillus alcalophilus*	Detergent	Solvay Enzymes GmbH, Germany
Optimase	*B. licheniformis*	Detergent	Solvay Enzymes GmbH, Germany
Maxapem	Protein engineered variant of alkalophilic *Bacillus* sp.	Detergent	Solvay Enzymes GmbH, Germany
HT-proteolytic	*B. subtilis*	Alcohol, baking, brewing, feed, food, leather, photographic waste	Solvay Enzymes GmbH, Germany
Protease	*B. licheniformis*	Food, waste	Solvay Enzymes GmbH, Germany
Proleather	Alkalophilic *Bacillus* sp.	Food	Amano Pharmaceuticals Ltd., Japan
Protease P	*Aspergillus* sp.	n.s	Amano Pharmaceuticals Ltd.,

			Japan
Amano protease S	*Bacillus* sp.	Food	Amano Pharmaceuticals Ltd., Japan
Collagenase	*Clostridium* sp.	Technical	Amano Pharmaceuticals Ltd., Japan
Enzeco alkaline protease	*B. licheniformis*	Industrial	Enzyme Development, USA
Enzeco alkaline protease-L FG	*B. licheniformis*	Food	Enzyme Development, USA
Enzeco high alkaline protease	*Bacillus* sp.	Industrial	Enzyme Development, USA
Bioprase concentrate	*B. subtilis*	Cosmetic, pharmaceuticals	Nagase Biochemicals, Japan
Ps. protease	*Pseudomonas aeruginosa*	Research	Nagase Biochemicals, Japan
Ps. elastase	*Pseudomonas aeruginosa*	Research	Nagase Biochemicals, Japan
Cryst. Protease	*B. subtilis* (K2)	Research	Nagase Biochemicals, Japan
Cryst. protease	*B. subtilis* (*bioteus*)	Research	Nagase Biochemicals, Japan
Bioprase	*B. subtilis*	Detergent, cleaning	Nagase Biochemicals, Japan
Bioprase SP-10	*B. subtilis*	Food	Nagase Biochemicals, Japan
Purafect	*B. lentus*	Detergent	Genencor International, Inc., USA
Primatan	Bacterial source	Leather	Genencor International, Inc., USA
Godo-Bap	*B. licheniformis*	Detergent, food	Godo Shusei, Japan
Corolase 7089	*B. subtilis*	Food	Rohm, Germany
Wuxi	*Bacillus* sp.	Detergent	Wuxi Synder Bioproducts, China
Protosol	*Bacillus* sp.	Detergent	Advance Biochemicals, India

n.s – not specified

2.9.1 Food and feed industry

Microbial proteases have been exploited in the food industries in many ways especially alkaline proteases in the preparation of protein hydrolysates of high nutritional value. The protein hydrolysates play an important role in blood pressure regulation and are used in infant food formulations, specific therapeutic

dietary products and the fortification of fruit juices and soft drinks (Neklyudov *et al.,* 2000; Ward, 1985). The commercial protein hydrolysates are derived from casein (Miprodan; MD Foods, Viby, Germany), whey (Lacprodan; MD Foods) and soy protein (Proup; Novo Nordisk, Bagsvaerd, Denmark). Rebeca *et al.* (1991) reported the production of fish hydrolysates of high nutritional value, using *B. subtilis* proteases. Matsui *et al.* (1993) reported on protease hydrolysates having angiotensin-I-converting enzyme-inhibitory activity from sardine muscle obtained after treatment with *B. licheniformis* alkaline protease. Fujimaki *et al.* (1970) used alkaline protease for the production of soy protein hydrolysates. Cheese whey is an abundant liquid by-product of the dairy industry with an estimated world production of 45x106 ton/year. Perea *et al.* (1993) used alkaline protease for the production of whey protein hydrolysate, using cheese whey in an industrial whey bioconversion process. Ohmiya *et al.* (1979) reported the use of immobilized alkaline protease in cheese-making. Tanimoto *et al.* (1991) reported the use of alkaline protease in the enzymatic modification of zein to produce a non-bitter peptide fraction with high Fischer ratio for patients with hepatic encephalopathy. Keratinolytic activity of alkaline protease has also been exploited in the production of proteinaceous fodder from waste feathers or keratin-containing materials. Dalev (1990) and Cheng *et al.* (1995) reported the use of alkaline proteases for the hydrolysis of feather keratin, to obtain a protein concentrate for fodder production.

Further, proteases play a prominent role in meat tenderization, especially of beef. An alkaline elastase (Takagi *et al.,* 1992) and thermophilic alkaline protease (Wilson *et al.,* 1992) have proved to be successful and promising meat tenderizing enzymes, as they possess the ability to hydrolyze connective tissue proteins as well as muscle fibre proteins. The tenderization process can be achieved by sprinkling the powdered enzyme preparation or by immersion in an enzyme solution and/or by injecting the concentrated protease preparation into the

blood stream or meat. A method has been developed in which the enzyme is introduced directly into the circulatory system of the animal shortly before slaughter (Bernholdt, 1975) or after stunning the animal to cause brain death (Warren, 1992). Soluble meat hydrolysates can also be derived from lean meat wastes and from bone residues after mechanical deboning by solubilization with proteolytic enzymes. However, the hydrolysates are usually bitter when the degree of hydrolysis is above 10%, which is needed for sufficient solubilization. Alcalase has been found to be the most appropriate enzyme in terms of cost, solubilization, and other relevant factors. In an optimized process with Alcalase at a pH of 8.5 and temperature of 55–60ºC, a solubilization of 94% was achieved (O'Meara and Munro, 1984a; 1984b). The resulting meat slurry is further pasteurized to inactivate the enzyme, and finds wide application in canned meat products, soups, and seasonings. The cleaned bones may also be used as an excellent raw material for the production of gelatin.

2.9.2 Leather industry

The conventional methods in leather processing involve the use of hydrogen sulfide and other chemicals, which contribute to 100% of sulfide and over 80% of the suspended solids in tannery effluents, creating environmental pollution and safety hazards. Thus, for environmental reasons, the biotreament of leather using an enzymatic approach is preferable as it offers several advantages, *e.g.* easy control, speed and waste reduction, thus being ecofriendly (Andersen, 1998). Alkaline proteases with elastolytic and keratinolytic activity can be used in leather-processing industries. Proteases find their use in the soaking, dehairing and bating stages of preparing skins and hides. The enzymatic treatment destroys undesirable pigments, increases the skin area and thereby clean hide is produced. Bating is traditionally an enzymatic process involving pancreatic proteases. Alkaline proteases speed up the process of dehairing, because the alkaline conditions enable the swelling of hair roots; and the subsequent attack of protease

on the hair follicle protein allows easy removal of the hair. The bating following the dehairing process involves the degradation of elastin and keratin, removal of hair residues, and the deswelling of collagen, which produces good, soft leather mainly used for making leather clothes and goods.

Varela *et al.* (1997) reported the use of *B. subtilis* IIQDB32 alkaline protease for unhairing sheepskin. George *et al.* (1995) used *B. amyloliquefaciens* alkaline protease for unhairing hides and skins. Hameed *et al.* (1999) used *B. subtilis* K2 alkaline protease in bating and leather processing. Successful use of alkaline proteases in leather tanning from *Aspergillus flavus* (Malathi and Chakraborty, 1991) and *Streptomyces* sp. (Mukhopadhyay and Chandra, 1993) are also reported.

2.9.3 Management of industrial and household waste

Proteases solubilize proteinaceous waste and thus help lower the biological oxygen demand of aquatic systems. Recently, the alkaline proteases are used in the management of wastes from various food-processing industries and household activities. Dalev (1990) used alkaline protease from *B. subtilis* for the management of waste feathers from poultry slaughterhouses. Waste feathers make up approximately 5% of the body weight of poultry and are considered to be a high protein source for food and feed, provided their rigid keratin structure is completely destroyed. The use of keratinolytic protease for food and feed industry waste, for degrading waste keratinous material from poultry refuse (Ichida *et al.*, 2001) and as depilatory agent to remove hair from the drains (Takami *et al.*, 1992) has been reported. A formulation containing proteolytic enzymes from *B. subtilis*, *B. amyloliquefaciens* and *Streptomyces* sp. and a disulfide reducing agent (thioglycolate) that enhances hair degradation was helpful in clearing pipes clogged with hair-containing deposits has been commercialised. It was prepared and patented by Genex (Jacobson *et al.*, 1985).

2.9.4 Photographic industry

Alkaline proteases play a crucial role in the bioprocessing of used X-ray or photographic films for silver recovery. These waste films contain 1.5–2.0% silver by weight in their gelatin layer, which can be used as a good source of silver for a variety of purposes. Conventionally, this silver is recovered by burning the films, which causes undesirable environmental pollution. Furthermore, base film made of polyester cannot be recovered using this method. Since the silver is bound to gelatin, it is possible to extract silver from the protein layer by proteolytic treatments. Enzymatic hydrolysis of gelatin not only helps in extracting silver, but also the polyester film base can be recycled. Alkaline protease from *B. subtilis* decomposed the gelatin layer within 30 min at 50–60°C and released the silver (Fujiwara *et al.,* 1989). Ishikawa *et al.* (1993) have reported the use of alkaline protease of *Bacillus* sp. B21-2 for the enzymatic hydrolysis of gelatin layers of X-ray films to release silver particles. The alkaline proteases of *Bacillus* sp. B18 (Fujiwara *et al.,* 1991) and *B. coagulans* PB-77 (Gajju *et al.,* 1996) were also efficient in decomposing the gelatinous coating on used X-ray films from which the silver could be recovered.

2.9.5 Medical usage

Alkaline proteases are also used for developing products of medical importance. Collagenases with alkaline protease activity are increasingly used for therapeutic applications in the preparation of slow-release dosage forms. A new semi-alkaline protease with high collagenolytic activity was produced by *Aspergillus niger* LCF9. The enzyme hydrolyzed various collagen types without amino acid release and liberated low molecular weight peptides of potential therapeutic use (Barthomeuf *et al.,* 1992). Kudrya and Simonenko (1994) exploited the elastolytic activity of *B. subtilis* 316M for the preparation of elastoterase, which was applied for the treatment of burns, purulent wounds, carbuncles, furuncles and deep abscesses. Kim *et al.* (1996) reported the use of

alkaline protease from *Bacillus* sp. strain CK 11-4 as a thrombolytic agent having fibrinolytic activity.

2.9.6 Silk degumming

One of the least explored areas for the use of proteases is the silk industry and only a few patents have been filed describing the use of proteases for the degumming of silk. Sericin, which is about 25% of the total weight of raw silk, covers the periphery of the raw silk fibers, thus providing the rough texture of the silk fibers. This sericin is conventionally removed from the inner core of fibroin by conducting shrink-proofing and twist-setting for the silk yarns, using starch (Kanehisa, 2000). The process is generally expensive and therefore an alternative method suggested is the use of enzyme preparations, such as protease, for degumming the silk prior to dyeing. The silk-degumming efficiency of an alkaline protease from *Bacillus* sp. RGR-14 was studied and results were analyzed gravimetrically (fiber weight reduction) and by scanning electron microscopy (SEM) of treated silk fiber. After 5 h of incubation of silk fiber with protease from *Bacillus* sp., the weight loss of silk fiber was 7.5%. SEM of the fibers revealed that clusters of silk fibers had fallen apart as compared with the smooth and compacted structure of untreated fiber (Gupta *et al*., 2002).

2.9.7 Proteases in the detergent industry

Enzymes are of interest to the detergent industry for their ability to aid in the removal of proteinaceous stains and the largest application of detergent proteases is in laundry detergent formulations. However, they are also popular in the formulation of household dishwashing detergents and both industrial and institutional cleaning detergents (Godfrey and West, 1996; Showell, 1999).

The history of detergent enzymes dates back to 1914, when two German scientists, Rohm and Haas, used pancreatic proteases and sodium carbonate in washing detergents. The product was named Burnus, after the white arab cloak. The first detergent containing the bacterial enzyme was introduced into the market in 1956 under the trade name Bio-40. However, it was only in 1963 when an alkaline protease, alcalase, was effectively incorporated in detergent powder and was marketed by Novo Industry, Denmark under the trade name Biotex. However, detergent proteases faced a setback in the early 1970s, due to unfavorable publicity of allergic reactions which was solved by the introduction of dust-free encapsulated products. Recently, detergent enzymes account for 89% of the total protease sales in the world; and a significant share of the market is captured by subtilisins and/or alkaline proteases from many *Bacillus* species (Gupta *et al.,* 2002).

In the 1980s and early 1990s, the major market share (>55%) of the detergent enzyme was held by Gist-Brocades in The Netherlands, Genencor International in the United States, Solvay in Belgium and Showa-Kenko in Japan. By 2002, Novo Nordisk and Genencor International became the major suppliers of detergent enzymes, supplying up to 95% of the global market of proteases (Gupta *et al.,* 2002). In 1994, the total market for industrial enzymes accounted for approximately $400 million, of which enzymes worth $112 million were used for detergent purposes (Hodgson *et al.,* 1994).

2.9.8 Selection and evaluation of detergent protease performance

There are several parameters involved in the selection of a good detergent protease, such as compatibility with detergent components, *e.g*. surfactants, perfumes and bleaches (Bech *et al.,* 1993; Gupta *et al.,* 1999; Kumar *et al.,* 1998), good activity at relevant washing pH and temperature (Aehle *et al.,* 1993; Oberoi *et al.,* 2001), compatibility with the ionic strength of the detergent solution, stain

degradation and removal potential, stability and shelf life (Showell, 1999). Further, a good detergent enzyme should be effective at low levels (0.4–0.8%) and should also be stable in the presence of oxidizing agents and bleaches. In general, the majority of the commercially available enzymes are not stable in the presence of bleaching/oxidizing agents. Hence, the latest trend in enzyme-based detergents is the use of recombinant DNA technology to produce bioengineered enzymes with better stability. Bleach- stability and oxidation-stability has been introduced through Site Directed Mutagenesis and protein engineering by the replacement of certain amino acid residues (Bech *et al.,* 1993; Estell *et al.,* 1985; Wolff *et al.,* 1996; Yang *et al.,* 2000).

Very few published reports are available on the compatibility of the alkaline proteases with detergents (Durham *et al.,* 1987; Phadatare *et al.,* 1993; Bhosale *et al.,* 1995; Samal *et al.,* 1990) The in-place cleaning of ultrafiltration (UF) and reverse osmosis (RO) membranes forms one of the most important aspects of modern dairy and food industries. The UF and RO membranes are put to a variety of uses, including concentration, fractionation, clarification and/or sterilization of liquid foods such as milk, whey, egg white, fruit juices, wines, and other beverages (Glover, 1985; Cheryan, 1986) The use of thermophilic proteases from *Thermus* sp. strain Rt41A and alkaline proteases from Bacillus sp. strain MK5-6 has also been successful (Kumar, 1997; Coolbear *et al.,* 1992). The use of a cocktail of proteases and lipases to degrade and solubilize protein and fat foulants have also proven beneficial. In addition, contact lens cleaning solutions using an alkaline protease from a marine shipworm bacterium cleaned the contact lens at low temperatures (Greene *et al.,* 1989; Greene *et al.,* 1996). In India, one such enzyme-based optical cleaner in the form of tablets containing Subtilopeptidase A was marketed by M/s Bausch and Lomb (India) Ltd.

Conventionally, detergents have been used at elevated washing temperatures, but at present there is considerable interest in the identification of alkaline proteases which are effective over a wide temperature range (Oberoi *et al.,* 2001). In addition, the current consumer demands and increased use of synthetic fibers, which cannot tolerate high temperatures, have changed washing habits towards the use of low washing temperatures (Kitayama 1992; Nielsen *et al.,* 1981). So enzyme manufacturers, currently look for novel enzymes that can act under low temperatures. Novo Nordisk Bioindustry in Japan has developed a detergent protease called Kannase, which keeps its high efficiency, even at very low temperatures (10–20 °C).

The evaluation of detergent proteases is mainly dependent upon parameters such as the pH and ionic strength of the detergent solution, the washing temperature and pH, mechanical handling, level of soiling and the type of textile. In the case of laundry detergents, protease performance is evaluated by using soiled test fabrics and the efficiency is measured either visually or by measuring the reflectance of light under standard conditions (Durham *et al.,* 1987; Masse and Tilburg, 1983; Nielsen *et al.,* 1981; Wolff *et al.,* 1996). ΔR is the measure of efficiency and is defined as the difference in reflectance between fabric treated with and without enzyme.

The wash performance analysis of a SDS-stable alkaline protease from *Bacillus* sp. RGR-14 was studied using a reflectance meter. Wash performance analysis of RGR-14 protease on grass- stained and blood-stained cotton fabrics showed an increase in reflectance (14% with grass stains, 25% with blood stains) after enzyme treatment. However, enzyme in conjunction with detergent proved best, with a maximum reflectance change of 46% and 34% for grass and blood stain removal, respectively, at 45°C. Stain removal was also effective after protease treatment at 25°C and 60°C (Oberoi *et al.,* 2001).

In addition to these major applications, alkaline proteases are also used to a lesser extent for other applications, such as in molecular biology, for the isolation of nucleic acid (Kwon *et al.,* 1994), pest control (Kim *et al.,* 1999) and selective delignification of hemp (Dorado *et al.,* 2001). Sakiyama *et al.* (1998) reported the use of a protease solution for cleaning the packed columns of stainless steel particles fouled with gelatin and β-lactoglobulin. These applications are however technically interesting, but have not reached commercial success in terms of impressive sales figures.

2.10 National status of metagenomic research

Research work related to metagenomics is going on in a number of research institutions all over India. School of Basic Sciences of Indian Institute of Technology, Mandi is involved in metagenomic sequencing and analysis of highthroughput sequence data obtained from different environments, including human body sites, soil, sediment, etc. Complete genome sequencing and analysis of novel bacteria inhabiting different environments is also carried out with first report of the complete genome sequence of a novel and yet unculturable SFB (Segmented Filamentous Bacteria) (Prakash *et al.,* 2011). They also develop computational tools and algorithms for metagenomic and genomic data analysis and own copyrights for softwares like MetaBin (Sharma *et al.,* 2012), ConjoinG, MetaBioME (Sharma *et al.,* 2009) and CoPS (Prakash *et al.,* 2004).

Energy and Environmental Biotechnology division of IGIB, New Delhi is involved in the preparation of metagenomic libraries of different ecological niches in India and screening for functional clones expressing novel biocatalysts and bioactive compounds. In addition they also carry out phylogenetic analysis of the sequences obtained from metagenomic DNA in order to analyze microbial diversity. Environmental Genomics Division of National Environmental Engineering Research Instituition (NEERI), Nagpur is also engaged in mining the

metagenome for novel antibiotic producing genes and biomolecules for microbial diversity from different ecological niches. Department of Animal Biotechnology of College of Veterinary Science & Animal Husbandry at Anand Agricultural University, Gujarat is engaged in whole genome sequencing and is equipped with Genome Sequencer FLX and Ion Torrent, with capacity to produce 1GB data in a day. Whole genome sequencing of Jaffrabadi buffalo was carried out by this department for the first time in the world.

Metagenomic work is carried out in some of the research instituitions in Kerala also. NIIST, Trivandrum is involved in construction of metagenomic library of Western Ghats to screen for amylases (Vidya *et al.*, 2011). Also, there is an ongoing project at National Centre for Aquatic Animal Health of Cochin University of Science and Technology to evaluate diversity of Nitrifying Bacterial Community in the Sediments of Shrimp Ponds which is financially supported by DST, Govt. of India.

Chapter - 3

MATERIALS AND METHODS

3.1 EXTRACTION OF METAGENOMIC DNA OF MICROBIAL COMMUNITIES IN MANGROVE SEDIMENTS

3.1.1 Collection of mangrove sediment sample

Sediment samples were collected from mangrove and mangrove associated aquafarms, situated in different locations as per the details given in Table 3.1, in sterile polythene bags, tied well and brought to the laboratory in an icebox for further processing.

Table 3.1 Sampling stations

Sl. No.	Code	Sampling location	Sampling site	GPS coordinates
1	MGK	Kannamaly	Mangrove	09°52′43.3″N 76°15′50.6″E
2	AQK	Kannamaly	Mangrove associated aquafarm	09°52′48.3″N 76°15′42.4″E
3	MGM	Mangalavanam	Mangrove	09°59′26.4″N 76°16′27.2″E

3.1.2 Extraction of metagenomic DNA using three different protocols

Total or metagenomic DNA was extracted from the mangrove sediment samples following three different protocols that are frequently used. The yield and purity of DNA obtained by each protocol was compared.

3.1.2.1 Protocol I (Zhou *et al*., 1996)

Weighed 5 g of mangrove sediment sample, mixed with 13.5 mL of DNA extraction buffer (Appendix I), and 100 µL of proteinase K (10 mg/mL) (GeNei, Bengaluru, India) in Oakridge tubes by horizontal shaking (Scigenics Biotech, Chennai, India) at 225 rpm for 30 min at 37°C. After the shaking treatment, 1.5mL

of 20% SDS (SRL, Mumbai, India) was added, and the samples were incubated in a 65°C water bath (MRC, Holon, Israel) for 2 h with gentle end-over-end inversions every 15 to 20 min.

The supernatants were collected after centrifugation at 6,000 x g (Sigma 3K30, Osterode, Germany) for 10 min at room temperature and transferred into 50mL centrifuge tubes. The soil pellets were extracted two more times by adding 4.5 mL of the DNA extraction buffer and 0.5 mL of 20% SDS, vortexed (Remi Cyclomixer, Mumbai, India) for 10 s, incubated at 65°C for 10 min, and centrifuged as before.

Supernatants from the three cycles of extractions were combined and mixed with an equal volume of chloroform: isoamyl alcohol (24:1vol/vol). The aqueous phase was recovered by centrifugation and extraction was repeated two more times. DNA was precipitated with 0.6 volume of isopropanol (SRL) at room temperature for 1 h. The pellet of crude nucleic acids was obtained by centrifugation at 16,000 rpm for 20 min at room temperature, washed with cold 70% ethanol, and resuspended in sterile deionized water, to give a final volume of 200 µL.

3.1.2.2 Protocol II (Dong *et al*., 2006)

About 1 g of sediment sample, 0.35 g of glass beads (diameter 2.0 mm), and 300 µL of 0.1 M phosphate buffer (pH 8.0) (Appendix I) were added to a microcentrifuge tube and mixed well by vortex (Remi). About 250 µL of SDS lysis buffer (Appendix I) was added and vortexed horizontally for 10 min at maximum speed using a Vortex Adapter. After centrifuging (Sigma) at 10,000rpm for 30 s, the supernatant was transferred into another microcentrifuge tube.

Protein was removed by adding 250 µL of chloroform : isoamyl alcohol (24:1 v/v), vortexed for 5 s, incubated at 48°C for 5 min, and centrifuged at

10,000rpm for 1 min. DNA was precipitated by adding 0.5 vol of 7.5 M ammonium acetate (SRL) and 1.0 vol of isopropanol to the supernatant. After incubation at -20°C for 15 min, DNA was pelleted at 12,000 rpm for 10 min and washed three times with 70% ethanol. After being air-dried, pellets were dissolved in 50 µL of 10 mM Tris (SRL), pH 8.0 and flocculated with 10 mM aluminium sulfate (SRL). The brownish precipitate of humic substances was removed by centrifuging at 10,000 rpm for 5 min.

3.1.2.3 Protocol III (Tsai and Olson, 1991)

Sediment samples (1 g) were mixed with 2 mL of 120 mM sodium phosphate buffer (pH 8.0) (Appendix I) by shaking at 150 rpm for 15 min. The slurry was pelleted by centrifugation at 6,000 rpm (Sigma) for 10 min. The pellet was washed again with phosphate buffer, resuspended in 2 mL of lysis solution (Appendix I), and incubated in a 37°C water bath (MRC) for 2 h with agitation at 20 to 30 min intervals, and then 2 mL of 0.1 M NaCl/0.5 M Tris-HCl (pH 8.0)/10% SDS was added.

Three cycles of freezing in liquid nitrogen and thawing in a 65°C water bath were conducted to release DNA from the microbial cells in the soil or sediments. 2 mL of tris saturated phenol (SRL) was added after the freeze-thaw cycles, and the sample was briefly vortexed to obtain an emulsion. The mixture was centrifuged at 6,000 rpm for 10 min. A 3 mL sample of the top aqueous layer was collected and mixed with equal volume of phenol: chloroform: isoamyl alcohol mixture (25:24:1) (HiMedia, Mumbai, India). A 2.5 mL portion of the resulting extract was further extracted with an equal volume of chloroform mixture. Finally, nucleic acids in the extracted aqueous phase (2 mL) were precipitated with 2 mL of cold isopropanol at -20°C for 1 h. The pellet of crude nucleic acids was obtained by centrifugation at 10,000 rpm for 10 min and then air

dried. The nucleic acid pellet was resuspended in 50 µL of TE buffer (pH 8.0) (Appendix I).

3.1.2.4 DNA isolation using kit

UltraClean™ Soil DNA isolation kit (MoBio, CA, USA) was also used for the extraction and purification of DNA from the sediment sample following the manufacturer's instructions. All steps, except beat beating, were performed for the purification of crude DNA extracts.

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

1% agarose (SRL) gel containing ethidium bromide (SRL) (0.5 µg/mL) was 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 marker Lambda DNA *Eco*R I/ *Hin*d III double digest (GeNei).

Electrophoresis was carried out at 80 V for 1 h (GeNei Mini Electrophoresis system, Bengaluru, India). The gel was visualized under ultraviolet illumination and gel pictures were captured using Gel documentation system (Syngene, CA, USA).

3.1.4 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.

The ratio between the readings at 260 nm and 280 nm (OD_{260}/OD_{280}) and 260 nm and 230 nm (OD_{260}/OD_{230}) indicates the purity of the DNA. Pure

preparation of DNA has OD_{260}/OD_{280} ratio in the range of 1.8-2.0, or else it indicates protein contamination. OD_{260}/OD_{230} ratio indicates humic acid contamination, for pure samples the ratio is the range of 1.2-2.0.

3.2 ANALYSIS OF PHYLOGENETIC DIVERSITY OF THE MANGROVE METAGENOME BASED ON 16S rDNA SEQUENCES

3.2.1 PCR amplification of 16S rRNA of metagenomic DNA

The forward and reverse primers (Sigma Aldrich, St Louis, MO, USA) used for the amplification of 16S rDNA sequences of the metagenomic DNA is given in Table 3.2.

Table 3.2 Primers used to amplify 16S rDNA

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

The concentration of different ingredients used for PCR amplification is given in Table 3.3.

Table 3.3 PCR Mix composition (Sambrook *et al*., 2000)

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

PCR amplification was carried out in a Thermal Cycler (BioRad MJ Mini Gradient, CA, USA) using the following program (Table 3.4)

Table 3.4 The program for PCR

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

The amplicon was then electrophoresed and visualized using gel documentation system (Syngene) as described in section 3.1.3. The PCR product was purified using the Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) according to manufacturer's instructions and the purified PCR product was used for cloning.

3.2.2 Construction of 16S rDNA library

The 16S rDNA amplicons were ligated into TA cloning vector pTZ57R/T (Thermo Scientific, MA, USA) and transformed onto competent *E. coli* JM109 host cells using InsTAclone PCR cloning kit (Thermo Scientific) according to manufacturer's instructions. The transformed cells were plated on Luria Bertani agar (HiMedia) plates containing Ampicillin (SRL), X-Gal (Chromous Biotech, Bengaluru, India) and IPTG (GeNei) (Appendix I) and incubated at 37ºC overnight. Appropriate controls were also included. The clones which appeared white on the plates were selected as recombinants and constituted the phylogenetic library.

3.2.3 Isolation of plasmids from phylogenetic clones

Plasmid isolation was done by alkaline lysis method which is a modification of the methods of Birnboin and Doly (1979) and Ish-Horowicz and Burke (1981). The method relies on bacterial lysis by sodium hydroxide (SRL)

and sodium dodecyl sulfate (SDS), followed by neutralization with a high concentration of low-pH potassium acetate (SRL). 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 (HiMedia) agar plate (Appendix II), 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 (Sigma) 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 plasmid 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 plasmid was 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 s and the supernatant was discarded. The pellet was air dried for 10 to 30 minutes. The pellet was resuspended in 50 µL of sterile deionised H_2O and stored at -20ºC.

3.2.4 Agarose gel electrophoresis of isolated plasmids

A sample of 5 µL was loaded on to the wells of agarose gel and electrophoresis was carried out as described in section 3.1.3. Appropriate DNA marker was also included. Plasmids showing band shift when compared with control were selected as recombinants.

3.2.5 Confirmation of recombinants

Confirmation of recombinants was done by reamplification of 16S rDNA inserts from the plasmids. PCR amplification of the 16S rDNA insert was done with 50 ng of the plasmid sample with vector specific forward (5'GTAAAACGACGGCCAGT 3') and reverse (5'CAGGAAACAGCTATGAC 3') primers as described under section 3.2.1

The PCR product was checked for the presence of amplicons by agarose gel electrophoresis as described in section 3.1.3. The presence of amplicons confirmed the presence of 16S rDNA insert within the plasmids of the clones. The clones confirmed to carry recombinant plasmid were segregated as phylogenetic clones and were stored on LB agar plates containing ampicillin and maintained as stock in glycerol.

3.2.6 Glycerol stocking

The clones were inoculated in LB broth containing ampicillin. After 24 h growth at 37°C, 80 µL of culture broth was mixed with 20 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.

3.2.7 Analysis of phylogenetic diversity based on 16S rDNA sequence

The 16S rDNA was sequenced from the recombinant plasmids by Sanger's Dideoxy method using ABI 3730 Excel (Applied Biosystems, CA, USA)

at Scigenom Labs, Kochi, Kerala. The identity of the sequences was determined by comparing with the sequences in the NCBI database using BLAST (Altschul *et al.*, 1990). For this, the sequences in FASTA format was pasted on the NCBI BLAST page (http://blast.ncbi.nlm.nih.gov) and nBlast was carried out. The sequences were compiled and multiple sequence alignment was done using ClustalX program of BioEdit software (Hall, 1999). The phylogenetic tree was constructed using the Neighbor-Joining method (Saitou and Nei, 1987) with 1000 resampling bootstrap using MEGA software version 5.0 (Tamura *et al*., 2011). Taxonomical hierarchy of the sequence was assigned using RDP Naive Bayesian Classifier Version 2.5 (Wang *et al*., 2007) and the distribution of various phyla was represented in pie diagram. The '.fasta' file of the sequences was processed through RDP Pipeline (Cole *et al*., 2014) for aligning and clustering and the '.clust' file was used for plotting rarefaction curve. The sequences were deposited at NCBI database as Sequin file and accession numbers were obtained.

3.3 SCREENING FOR PROTEASE PRODUCTION BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

3.3.1 Culture independent method - Construction of functional metagenomic library

In culture independent approach, the total genomic or metagenomic DNA isolated from mangrove sediment was cloned in *E.coli* DH5α and functional library was constructed.

3.3.1.1 Restriction digestion of metagenomic DNA

The DNA isolated from sediment sample was digested using the restriction enzyme *Sau*3A I (GeNei) at 37ºC for 30 min. The reaction mixture is given in Table 3.5.

Table 3.5 Ingredients of restriction digestion reaction mixture

Ingredient	Quantity
DNA (1 µg/µL)	1 µL
10X assay buffer	2 µL
*Sau*3A I (1U/ µL)	1 µL
Sterile H_20	17 µL
Total volume	20 µL

Prior to further processing, an aliquot of the digestion was analyzed by agarose gel electrophoresis versus non-digested DNA and a size marker to ensure proper digestion. After proper digestion the reaction was stopped by deactivating the enzyme by incubation at 65ºC for 15 min.

3.3.1.2 Restriction digestion of plasmid DNA

The restriction enzyme *Bam*H I (GeNei) was used to digest the plasmid DNA, pUC18, and was carried out in the same manner as explained in section 3.3.1.1. The presence of a single band on agarose gel confirms that all of the plasmid DNA has been converted from circular form to linear molecules. Circular form displays 3 to 5 bands on agarose gel based on the different conformations of plasmid DNA. The enzyme reaction was stopped as explained in section 3.3.1.1.1.

3.3.1.3 Dephosphorylation of digested plasmid DNA (Sambrook *et al*., 2000)

The linearised plasmid DNA was dephosphorylated using the enzyme alkaline phosphatase (GeNei) which removed the phosphate group from the 5' end of the plasmid and thereby prevented recircularisation. The reaction mixture for dephosphorylation of digested plasmid is given in Table 3.6.

Table 3.6 Ingredients of dephosphorylation reaction mixture

Ingredient	Quantity
Linearised plasmid DNA (0.5 µg)	10 µL
10X assay buffer	2 µL
Alkaline phosphatase (1U/µL)	1 µL
Sterile H_20	7 µL
Total volume	20 µL

The reaction mixture was incubated at 37ºC for 30 min. After reaction, the phosphatase activity was inactivated by adding SDS, EDTA (pH 8.0) and proteinase K to final concentrations of 0.5%, 5 mM and 100 µg/mL respectively. Mixed well and incubated for 30 min at 56°C. After cooling the reaction mixture was extracted once with phenol and once with phenol: chloroform. Added 0.1 vol of 3M sodium acetate (pH 7.0) and mixed well. Added 2 vol of ethanol, mixed the solution and stored for 15 min at 0°C. Centrifuged (Sigma) at 12,000 rpm for 10 min at 4°C and the pellet was washed with 70% ethanol at 4°C and recentrifuged. The pellet was air-dried and dissolved in 10 µL TE buffer (pH 8.0). The DNA concentration was quantified spectrophotometrically (Schimadzu) as described under section 3.1.4 and stored at -20°C.

3.3.1.4 DNA ligation (Sambrook *et al*., 2000)

DNA ligations was performed by incubating DNA fragments with linearized plasmid DNA in the presence of buffer and T4 DNA ligase (GeNei) at 4ºC overnight. The insert: vector ratio was 3:1. The reaction mixture for DNA ligation is given in Table 3.7

Table 3.7 Ingredients of ligation reaction mixture

Ingredients	Test	Control
Insert DNA	200-600 ng	-
Vector DNA	200 ng	200 ng
10X ligation buffer	1 μL	1 μL
T4 DNA ligase	0.5 U	0.5 U
Sterile H_2O	upto 10 μL	upto 10 μL

The ligation mixture was then used for transformation of competent *E.coli* DH5α cells.

3.3.1.5 Preparation of competent cells (Sambrook *et al*., 2000)

E.coli DH5α was used as host organism for transformation and competent cell preparation was done using calcium chloride.

A single bacterial colony of *E.coli* DH5α was inoculated into 100 mL Luria Bertani broth and incubated at 37°C in a 1 L flask with vigorous agitation. The bacterial culture was transferred to ice-cold centrifuge tubes and the tubes were kept on ice for 10 min. The culture was centrifuged at 4000 x g for 10 min at 4°C. The supernatant was decanted and the tubes were kept inverted for 1 min.

The pellet was resuspended in 10 mL of ice-cold 0.1 M $CaCl_2$ (SRL) solutions by gentle vortexing and stored on ice for 10 min. The cells were recovered by centrifugation at 4000 x g for 10 min at 4°C. The medium was decanted and the tubes kept inverted for 1 min. The pellet was resuspended by gentle vortexing in 2 mL of ice-cold 0.1 M $CaCl_2$ for each 50 mL of original culture. The cells can be either used directly for transformation or dispensed into aliquots and freezed at -70°C (with 15-30% glycerol).

3.3.1.6 Preparation of competent cells for frozen storage

80 µL of the competent cell suspension was mixed with 20 µL of sterile 100% glycerol, giving a final concentration of 20% glycerol. The competent cells were transferred to -70ºC and can be stored indefinitely. For further use, competent cells can be removed from freezer and allowed to thaw for some time. The ligation mixture was added and incubated for 30 min as in the standard transformation procedure.

3.3.1.7 Transformation of competent *E.coli* cells (Sambrook *et al*., 2000)

200 µL of competent *E.coli* DH5α cells were taken in a sterile, chilled microfuge tube. The ligation mixture (1-2 µL) was added to the competent cells. The contents were mixed by swirling gently and the tube was stored on ice for 30 min. The tubes were placed in a preheated 42°C water bath and stored for exactly 90 s without shaking. The tubes were rapidly transferred to an ice bath and allowed to chill for 2 min. 800 µL of LB broth was added and incubated the cultures for 45 min in a water bath at 37°C. 100 µL of transformed competent cells were transferred onto LB agar plates containing ampicillin, X-gal and IPTG to screen for the recombinants.

Appropriate positive and negative controls were included. The plates were stored at room temperature until the liquid has been absorbed and incubated at 37°C for 12-16 h and observed for blue and white colonies. The white colonies representing the recombinants were picked and stored as functional library.

3.3.1.8 Screening of the library for protease producer by plate assay

For functional screening, the clones were spot inoculated on LB agar plates containing 1% (w/v) skimmed milk and incubated overnight at 37°C (Lee *et al*., 2007). Proteolytic activity was determined by zone of clearance surrounding the growth.

3.3.2 Culture dependent method

Sediment samples collected from different sources were incubated in physiological saline. Samples were then serially diluted and plated onto nutrient agar (HiMedia) (Appendix II) plates by spread plate technique. After incubation, single colonies grown on the plates were picked and transferred to another plate. Quadrant streaking of these isolates was then done to get a monospecific culture based on uniform colony morphology and colour. The isolates were maintained as frozen stock preparations on agar slants and were included in the lab's culture collection.

3.3.2.1 Screening for protease producer

The isolates were screened for protease production by plate assay method as described in section 3.3.1.2. The protease positive clone designated as BTKM4 which showed consistent activity was selected for the study.

3.4 CHARACTERISATION OF THE PARTIAL PROTEASE GENE OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

3.4.1 Amplification of partial protease gene from the recombinant plasmid of clone BTM106

The protease positive clone was inoculated in 5mL LB broth (HiMedia) containing ampicillin and plasmid isolation was carried out as described in section 3.2.3. The recombinant plasmids were subjected to PCR using protease specific degenerate primers (Sigma Aldrich) in order to amplify the partial protease gene (Table 3.8).

Table 3.8 Degenerate primers used to amplify partial protease gene of clone BTM106

Primer	Primer sequence	Reference
npr1F	5'-ScDgcVgtDgaYgcHcaYKM-3'	Chen *et al*., 2004
npr3R	5'-aDYaRRtaMgcBKStttRtt-3'	

The sequences of degenerate primers were labeled by the degenerate DNA genetic code: B = G, T, or C; D= G, A, or T; H= A, T, or C; K = G or T; M = A or C; N = A, T, G, or C; R = A or G; S = G or C; V= G, A, or C; W = A or T; Y = C or T.

PCR amplification was carried out in a Thermal Cycler (BioRad) using the following program (Table 3.9)

Table 3.9 The program for PCR amplification of protease gene

Step	Temperature	Time	
Initial Denaturation	94°C	1.5 min	
Denaturation	94°C	30 s	
Annealing	55°C	30 s	30 cycles
Extension	72°C	2 min	
Final Extension	72°C	10 min	

3.4.2 Amplification of partial protease gene from the genomic DNA of strain BTKM4

The genomic DNA was isolated from the protease positive strain BTKM4 (Ausubel *et al*., 1987) and was subjected to PCR using protease specific primers (Sigma Aldrich) in order to amplify the partial protease gene (Table 3.10).

Table 3.10 Primers used to amplify partial protease gene of strain BTKM4

Primer	Primer sequence	Reference
BclI	5'-GTAACAGGAACGAATAAAGTAGGAACTGGTAAAG-3'	Bach *et al*., 1999
BclII	5'-GTTTACACCAACAGCACTAAATGATTGCTTAAC-3'	

PCR amplification was carried out in a Thermal Cycler (BioRad) using the following program (Table 3.11)

Table 3.11 The program for PCR amplification of protease gene

Step	Temperature	Time	
Initial Denaturation	94°C	2 min	
Denaturation	94°C	30 s	30 cycles
Annealing	54°C	30 s	
Extension	72°C	45 s	
Final Extension	72°C	10 min	

The PCR products were analysed on agarose gel as described under section 3.1.3 for the presence of amplicons.

3.4.3 *In silico* analysis of partial protease gene of clone BTM106 and strain BKM4

The protease gene amplicons were cloned into vector pGEMT (Promega, WI, USA). The clones were confirmed by the reamplification of the gene from the plasmid. The gene was sequenced by Sanger's Dideoxy method using ABI 3730 Excel (Applied Biosystems) at Scigenom Labs and the sequence analysis was done using various bioinformatic tools.

The identity of the sequence was determined by comparing with sequences in the NCBI database using the online tool BLAST (Altschul *et al*.,

1990). The multiple sequence alignment of the sequences was carried out using Clustal W2 (Larkin *et al*., 2007) with the first ten sequences showing maximum identity. The phylogenetic tree was constructed as described under section 3.2.7. The partial protease gene sequences were deposited at NCBI and accession numbers were obtained.

The Open Reading Frame (ORF) in the nucleotide sequences was determined using NCBI ORF finder (Wheeler *et al*., 2003). The deduced amino acid sequence was obtained by translation of the nucleotide sequence using online tool ExPasy (Gasteiger *et al*., 2003) and the identity of the sequence was confirmed by pBLAST. Amino acid sequences were aligned using ClustalW2 (Larkin *et al*., 2007). The phylogenetic tree was constructed as described in section 3.2.7. The active site in the amino acid sequence was predicted using conserved domain database (Bauer *et al*., 2013) of NCBI. Based on the conserved domains present, the proteases were assigned to families in MEROPS database (Rawlings and Barrett, 1993). The multiple sequence alignment of the conserved domain sequence was done using ClustalW2 (Larkin *et al*., 2007). Protein modeling was done using Phyre2 (Kelley, 2009) with the deduced amino acid sequences. The secondary structures encoded by the amino acid sequences were also predicted using the software.

3.5 CHARACTERISATION OF THE PROTEASE ENZYME OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

3.5.1 Extraction of crude protease

3.5.1.1 Extraction and recovery of crude protease from clone BTM106

The protease positive clone BTM106 was inoculated in LB broth containing ampicillin and incubated overnight at 37ºC. The cell free supernatant

was collected after centrifugation (Sigma) at 6000 rpm for 10 min. The protease enzyme obtained from clone BTM106 was designated as P106.

The cell free supernatant of the control (*E.coli* DH5α with pUC18) was also prepared in the same manner and analysed for variation in their protein profile.

3.5.1.2 Extraction and recovery of crude protease from strain BTKM4

The protease positive strain BTKM4 was inoculated in Zobells Marine broth (Appendix II) and incubated overnight at 37ºC. After incubation, the cell free supernatant was prepared as described under section 3.5.1.1. The protease obtained from strain BTKM4 was designated as P4.

The cell free supernatants of clone BTM106 and strain BTKM4 was assayed for proteolytic activity as explained below.

3.5.2 Analytical methods

3.5.2.1 Caseinolytic assay

Protease activity was determined according to the method described by Kunitz with slight modifications (Kunitz, 1947). The TCA (SRL) soluble peptide fractions of casein formed by the action of protease were quantified by comparing with tyrosine (SRL) as standard. Appropriate blanks for the enzyme and the substrate were also included in the assay along with the test.

Four hundred microlitre of 1% (w/v) Hammerstein casein (SRL) prepared in 0.05 M carbonate-bicarbonate buffer (pH 10.0) (Appendix I) was added to 100µL of same buffer. To this, 100 µL of appropriately diluted enzyme solution was added and incubated at 40°C for 30 min. The reaction was terminated by the addition of 500 µL of 0.44 M trichloroacetic acid (TCA) solution. The reaction mixture was centrifuged at 10,000 rpm for 15 min. The absorbance of the clear

supernatant was measured at 280 nm in UV-Visible spectrophotometer (Shimadzu) against appropriate blanks.

One Unit of protease activity is defined as the amount of enzyme required to liberate 1µMol of tyrosine per millilitre of the reaction mixture per minute under assay conditions and was expressed as U/mL.

3.5.2.2 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.

3.5.2.2.1 Bradford reagent

Hundred milligram Coomassie Brilliant Blue G-250 (Sigma Aldrich) was dissolved in 50 mL 95% ethanol, added 100 mL of 85% (w/v) phosphoric acid and diluted to 1 L when the dye has completely dissolved, followed by filtration through Whatman No.1 paper.

3.5.2.2.2 Estimation

Bovine serum albumin (BSA) standard was prepared containing a range of 5 to 25 µg in 100 µL volume. 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 was added and incubated for 5 min at room temperature. The absorbance was measured at 595 nm in a UV-Visible spectrophotometer (Shimadzu). The concentration of the protein sample was determined from the slope of the standard curve.

3.5.2.3 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.

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

3.5.3 PURIFICATION OF PROTEASES

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

3.5.3.1 Ammonium sulphate precipitation of proteases

Ammonium sulphate precipitation of the proteases was done according to the method described by Englard and Siefter (1990). The fractionation using ammonium sulphate precipitation has the advantage of intermediate removal of unwanted proteins with the simultaneous concentration of the protein of interest. Ammonium sulphate (SRL) required to precipitate the proteases was optimized by adding varying concentrations (30%, 60% and 90% saturations) to the crude extract as detailed below.

To precipitate the protein, ammonium sulphate was slowly added, initially at 30% saturation to the crude extract, kept in ice, with gentle stirring. After complete dissolution of ammonium sulphate, the solution was kept for precipitation at 4°C for 4 h. The precipitated protein was collected by centrifugation at 10,000 rpm (Sigma) for 15 min at 4°C. To the supernatant, required ammonium sulphate for next level of saturation was added and the procedure mentioned above was repeated. The precipitation was continued up to 90% of ammonium sulphate saturation.

3.5.3.1.1 Dialysis

The precipitate obtained after ammonium sulphate precipitation was further dialyzed against 0.01 M phosphate buffer (pH 7.5), in order to remove the ammonium sulphate from the precipitate, as detailed below.

3.5.3.1.2 Pretreatment of dialysis tube

Dialysis tube (Sigma-Aldrich) was treated to remove the humectants and protectants like glycerin and sulfur compounds present in it, and to clear the pores of the tube. The treated tube retained most of the proteins of molecular weight 12kDa or greater. The method followed for the treatment of the dialysis tube was as follows.

The tube was washed in running water for 3-4 h and rinsed in 0.3% (w/v) solution of sodium sulfide (SRL), at 80°C for 1 min. The tube was then washed with hot water (60°C) for 2 min and acidified with 0.2% (v /v) sulphuric acid (SRL) rinsed with hot water (60°C).

3.5.3.1.3 Dialysis Procedure

The precipitated protein was resuspended in minimum quantity of 0.01 M phosphate buffer (pH 7.5). The solution was taken in the pretreated dialysis tube (Sigma-Aldrich, cut off value 12 kDa) against 0.01 M solution of phosphate buffer pH 7.5 at 4°C with frequent changes of buffer every 4 h. After 5-6 buffer change, the sample was collected and assayed for enzyme activity and specific activity as described under section 3.5.2.

3.5.3.2 Gel filtration chromatography by Sephadex G-75

Gel filtration chromatography was performed using the concentrated active fraction of ammonium sulphate precipitation.

3.5.3.2.1 Preparation of column

Four grams of sephadex G-75 (Sigma-Aldrich) was suspended in 50mL MilliQ water and allowed to hydrate for 3 h at 90°C in a water bath, and fine particles were removed by decantation. Hydrated gel suspension 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 (0.15 M NaCl) (SRL) passed through the column bed in a descending eluent flow.

3.5.3.2.2 Sample preparation and application

Two milliliters of ammonium sulphate fraction with protein content of 13.5 mg/mL for P106 and 11.5mg/mL for P4 was concentrated using Amicon UF-3 kDa membrane (Millipore, MA, USA). This concentrated protein (0.5 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 0.15 M NaCl with a flow rate of 0.3 mL/min. Five millilitre fractions were collected. Twelve fractions were collected and each fraction from the column was concentrated using Amicon UF-3 kDa membrane and assayed for protease activity. Protein content and specific activity were determined as described under sections 3.5.2.2.2 and 3.5.2.2.3.

3.5.3.3 Calculation of fold of purification

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

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

3.5.4 CHARACTERIZATION OF THE PROTEASE ENZYMES OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS.

The purified proteases were further subjected to characterization for their biophysical and physicochemical properties. The molecular weight was determined by electrophoretic methods and zymogram.

3.5.4.1 Electrophoretic methods

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

The active fractions of protease enzyme after ammonium sulphate 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, CA, USA). Electrophoresis was carried out in a 15% polyacrylamide gel according to the method described by Laemmli (1970). The reagents used in the gel preparation (Appendix I) and their concentration is given in Table 3.12

Table 3.12 Gel preparation for SDS-PAGE

Reagents	Stacking gel (2mL) (5%)	Resolving gel (5mL) (15%)
Stock acrylamide: bis-acrylamide (mL)	0.336	2.5
Stacking gel buffer stock (mL)	0.5	-
Resolving gel buffer stock (mL)	-	1.25
Water (mL)	1.13	1.16
10% SDS (µL)	20	50
Ammonium persulphate (APS) (µL)	40	75
TEMED (µL)	4	5

3.5.4.1.2 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.

3.5.4.1.3 Protein marker for SDS-PAGE

Broad range molecular weight NEB protein marker mix (New England BioLabs, Ipswich, MA) was used for detecting the approximate size. It is a ready-to-load marker. The protein marker was mixed and 7 µL was taken in a tube. It was heated for 5 min at 100ºC. After a quick microcentrifuge spin (Tarsons Spinwin, Kolkota, India) the marker was loaded directly on to the gel (Appendix I).

3.5.4.1.4 Procedure

The gel plates were cleaned and assembled. The resolving gel solution was prepared combining all reagents except APS and TEMED in a beaker. The solution was degassed and APS and TEMED were added. The mixture was immediately poured into the cast and a layer of water was poured over the gel and allowed to polymerize at least for 45 min. The components of stacking gel were added into a beaker except APS and TEMED, mixed gently, and APS and TEMED were finally added. The contents were mixed well and poured into the cast above the resolving gel and immediately inserted the comb between the glass plates. It was allowed to polymerize for at least for 45 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 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 stained by silver staining method.

3.5.4.1.5 Silver staining

Silver staining of the gel after electrophoresis was performed by the method of Blum (1987) with slight modifications. The reagents required for the procedure was prepared (Appendix I).

3.5.4.1.6 Procedure

The SDS-PAGE gel was incubated in the fixer for 30 min. The gel was then washed in wash solution for 15 min, followed by five washes at five minute intervals with Milli-Q water. The gel was incubated in the sensitizer for exactly 60s and washed twice at 60 s intervals with Milli-Q water. The gel was immersed in chilled silver nitrate solution for 25 min and washed twice for 60 s with Milli-Q water. Then, the gel was transferred to developer solution and kept until protein bands were developed. After appropriate staining, the reaction was arrested by adding stop solution.

3.5.4.2 Zymogram

Zymogram was done to determine the approximate size of the protease enzyme. When the enzyme sample was run on a polyacrylamide gel containing gelatin, a clearing was seen at the band responsible for protease activity, due to degradation of gelatin in that region (Felicioli *et al*., 1997).

Polyacrylamide gel was casted with 15% resolving gel containing 0.1% gelatin (HiMedia). Sample was loaded to the gel with standard SDS gel loading buffer as described under section 3.5.4.1.2. The gel was run as described under section 3.5.4.1.4.

After the run, the gel was washed thrice with 0.1% Tween 80 (30 min each), followed by washing with deionized water to remove SDS. The gel was soaked in 100 mM Tris HCl, pH 8.0 with 5 mM $CaC1_2$ for 2 h at 37ºC and then

stained using coommassie staining solution (Appendix I) for 1 h. The gel was then transferred to destaining solution (Appendix I). The protein band responsible for protease activity shows clearing on gelatin-PAGE.

3.5.4.3 Determination of isoelectric point

Isoelectric point (pI) of the purified proteases was determined by isoelectric focusing, performed using 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. The detailed procedure is as given below.

3.5.4.3.1 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 (Bio-Rad) was gently placed gel side down in the equilibration tray and air bubbles were removed if any. 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.

3.5.4.3.2 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 made 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 3.13).

Table 3.13 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.

3.5.4.3.3 Staining of IPG strips after IEF

The gel was subjected to coomassie staining for 1 hour followed by destaining to visualize the isoelectric points of proteases, P106 and P4.

3.5.5 EFFECT OF PHYSICO-CHEMICAL PARAMETERS ON PROTEASE ACTIVITY

The effect of various physicochemical factors like pH, temperature, inhibitors, substrates, metal ions, detergents, oxidizing agent and reducing agent on enzyme activity was studied. The relative activity and residual activity were also determined.

3.5.5.1 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.

$$\textbf{Relative activity} = \frac{\textbf{Activity of the sample}\,(\textbf{U/mL})}{\textbf{Activity of the maximal enzyme activity obtained sample}\,(\textbf{U/mL})} \times \textbf{100}$$

3.5.5.2 Residual activity

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

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

3.5.5.3 Determination of optimum pH for enzyme activity

The optimum pH for maximum activity of the proteases was determined by conducting enzyme assay at various pH levels in the range of 2-13. The enzyme assay was essentially the same as described in section 3.5.2.1 with minor modifications involving the preparation of the substrate in buffer systems of different pH.

The substrate casein was prepared in the respective buffer for each pH. 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) (Appendix I) (Vincent and John, 2009). Preparation of buffers is charted in appendix 1. The enzyme activity and relative activity were calculated as described in sections 3.5.2.1 and 3.5.5.1 respectively.

3.5.5.4 Determination of pH stability of protease enzyme

The stability of the purified proteases over a range of pH was determined by measuring the residual activity at pH10.0 after incubating the enzymes in different buffer systems of pH 2-13 as explained under section 3.5.5.3 for 1 h at 4^{o}C. After incubation the sample was assayed for protease activity as explained in section 3.5.2.1. The enzyme activity and relative activity were calculated as described in sections 3.5.2.1 and 3.5.5.1 respectively.

3.5.5.5 Determination of optimum temperature for protease activity

The optimum temperature for maximum enzyme activity was determined by assaying enzyme activity as described under section 3.5.2.1 at different temperatures ranging from 30°C-100°C. The enzyme activity and relative activity were calculated as described in sections 3.5.2.1 and 3.5.5.1 respectively.

3.5.5.6 Determination of temperature stability of protease enzyme

The thermal stability of the purified proteases was determined by incubating the enzyme samples at different temperatures ranging from 30°C-100°C. After 1h incubation the sample was assayed for protease activity as explained in section 3.5.2.1. The enzyme activity and relative activity were calculated as described in sections 3.5.2.1 and 3.5.5.1 respectively.

3.5.5.7 Effect of inhibitors on protease activity

Effect of different protease inhibitors, Aprotinin (1 mM), EDTA (10 mM), Iodoacetamide (10 mM) and PMSF (10 mM) on protease activity was studied by conducting enzyme assay as explained under section 3.5.2.1 in the presence of inhibitors. The enzyme activity and residual activity were calculated as described in sections 3.5.2.1 and 3.5.5.2 respectively.

3.5.5.8 Determination of substrate specificity of protease enzymes

The substrate specificity of the purified proteases was determined based on their ability to hydrolyze various proteinaceous substrates like bovine serum albumin, casein, gelatin and haemoglobin by conducting enzyme assay as explained under section 3.5.2.1, with 1% of the respective substrates prepared in carbonate-bicarbonate buffer (pH 10.0).

3.5.5.9 Determination of kinetic parameters - Km and Vmax

The purified enzyme was subjected to kinetic studies 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 3.5.2.1 at different substrate [S] concentrations (0.1 mg – 10 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.

3.5.5.10 Effect of various metal ions on protease activity

Effect of various metal ions on activity of proteases was evaluated by incubating the proteases along with 1 mM concentrations of various metals ions for 30 min followed by measuring the protease activity as described under section 3.5.2.1. The metal salts studied included sodium chloride, calcium chloride, magnesium sulphate, ferric chloride, manganese chloride, nickel chloride, barium chloride, cadmium sulphate, zinc sulphate, copper sulphate, cobalt chloride and aluminum sulphate which contributes the metal ions, Na^{+}, Ca^{2+}, Mg^{2+}, Fe^{3+}, Mn^{2+}, Ni^{2+}, Ba^{2+}, Cd^{2+}, Zn^{2+}, Cu^{2+}, Co^{2+} and Al^{3+} respectively. The enzyme activity and residual activity were calculated as described in sections 3.5.2.1 and 3.5.5.2 respectively.

3.5.5.11 Effect of various detergents on protease 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 protease activity was determined by conducting enzyme assay as explained under section 3.5.2.1 in the presence of each detergent. The enzyme activity and residual activity were calculated as described in sections 3.5.2.1 and 3.5.5.2 respectively.

3.5.5.12 Effect of DMSO as oxidizing agent on enzyme activity

Effect of dimethyl sulfoxide (DMSO) as oxidizing agent on protease activity was studied by conducting enzyme assay as explained under section 3.5.2.1 in the presence of DMSO at different concentrations (1-5%, v/v). The

enzyme activity and residual activity were calculated as described in sections 3.5.2.1 and 3.5.5.2 respectively.

3.5.5.13 Effect of β-mercaptoethanol as reducing agent on enzyme activity

Effect of β-mercaptoethanol as reducing agent on protease activity was studied by conducting enzyme assay as explained under section 3.5.2.1 in the presence of β-mercaptoethanol at different concentrations (0.5 mM, 1 mM, 5 mM, 10 mM and 15 mM). The enzyme activity and residual activity were calculated as described in sections 3.5.2.1 and 3.5.5.2 respectively.

3.6 APPLICATION STUDIES OF PROTEASES

3.6.1 Commercial detergent compatibility of the enzymes

The stability of the enzymes in the presence of commercial detergents was determined using various detergents which included Ariel®, Surf Excel®, Sunlight®, Tide® and Wheel® at a concentration of 7 mg/mL. The enzymes already present in the detergent were first heat inactivated by boiling for 10 minutes. The test enzyme solution was mixed with detergent solution for 1 h at 40ºC and enzyme assay was carried out as explained under section 3.5.2.1.

3.6.2 Wash performance studies

The wash performance analysis of the purified proteases was studied on white cotton cloth piece (5 cm x 5 cm) stained with blood. The cloth piece was stained with human blood and dried for 12 h. The stained cloth pieces were taken in separate flasks and subjected to the following wash treatment studies.

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 enzyme sample (P106)

Flask 4: 100 mL detergent solution + stained cloth piece + 1 mL enzyme sample (P4)

After 30 min incubation at 40ºC in a water bath shaker (MRC), the cloth pieces were taken out, rinsed with tap water, dried and visual examination was done to check the effectiveness of stain removal.

3.6.3 Decomposition of gelatin layer of X-ray film

The ability of the proteases to degrade gelatin layer of X-ray film was studied by incubating 2 g of X-ray film (1 cm x 1 cm) in solution with 1 mL of 1 mg/mL enzyme protein. . The experimental set up was as follows.

Flask 1: 20 mL carbonate-bicarbonate buffer + 2 gm X-ray film

Flask 2: 20 mL carbonate-bicarbonate buffer + 2 gm X-ray film + 1 mL enzyme sample (P106)

Flask 3: 20 mL carbonate-bicarbonate buffer + 2 gm X-ray film +1 mL enzyme sample (P4)

The flasks were kept on a rotary shaker (Scigenics Biotech) with 120 rpm at room temperature for 4 h. The X-ray film strips were recovered at every 1 h interval, rinsed with tap water, dried and visual examination was done. The protein stripped into the solution in the flask by the action of the enzyme was also estimated (Bradford, 1976) at 1 h interval as explained under the section 3.5.2.2.

3.7 Statistical analysis

All experiments were repeated thrice, mean and standard deviation was determined using Microsoft Excel 2007 (Microsoft Corporation, Redmond, USA). The graphs were plotted with the help of Sigma plot for Windows Version 11.0 (Systat Software Inc., Germany).

Chapter - **4**

RESULTS

4.1 EXTRACTION OF METAGENOMIC DNA OF MICROBIAL COMMUNITIES IN MANGROVE SEDIMENTS

Metagenomic DNA was isolated from sediments collected from Kannamaly mangrove (MGK), Kannamaly mangrove associated aquafarm (AQK) and Mangalavanam mangrove (MGM).

4.1.1 Agarose gel electrophoresis of metagenomic DNA

DNA isolated using three different protocols (Protocol I, Protocol II and Protocol III) were visualized on agarose gel along with the DNA marker Lambda DNA *Eco*R I/ *Hin*d III double digest (GeNei, Bengaluru, India) (Fig. 4.1).

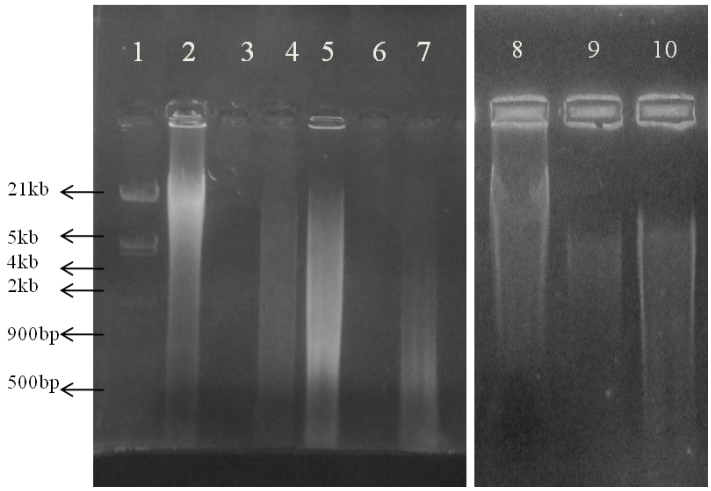

Fig. 4.1 Agarose gel (1%) electrophoresis of metagenomic DNA isolated from three samples using three different protocols.

Lane 1 - DNA marker; Lane 2 - MGK - Protocol I; Lane 3 - MGK - Protocol II; Lane 4 - MGK - Protocol III; Lane 5 - AQK - Protocol I; Lane 6 - AQK - Protocol II; Lane 7 - AQK - Protocol III; Lane 8 - MGM - Protocol I; Lane 9 - MGM - Protocol II; Lane 10- MGM - Protocol III.

The DNA isolated from the three different samples using Protocol I appeared brightly on the gel (Lanes 2, 5 & 8) indicating good DNA yield. Among the three samples, DNA isolated from AQK seemed to be sheared which is observed as a long smear on the gel (Lane 5). The DNA isolated using protocol II appeared as a very faint band on the gel (Lanes 3, 6 & 9) representing poor DNA yield. Aluminium sulphate used in this protocol for the removal of humic contaminants may have degraded DNA and resulted in poor DNA yield. The DNA isolated using protocol III also appeared as a smear on the agarose gel (4, 7 & 10) for all samples analysed, which may be due to repeated freezing and thawing of the sample. Only Protocol I produced significant DNA yield with comparatively less shearing.

4.1.2 Quantification of metagenomic DNA

The concentration of DNA obtained from three samples (sediments collected from Kannamaly mangrove (MGK), Kannamaly mangrove associated aquafarm (AQK) and Mangalavanam mangrove (MGM) using the three different protocols was determined spectrophotometrically and is as represented in Fig. 4.2.

It is evident that protocol I yielded more DNA than the other two methods irrespective of the sample type analysed. However, the DNA yield from Kannamaly mangrove (MGK) sediment was more (57.6 ± 2.26 µg/g sediment) than the other two samples, Kannamaly mangrove associated aquafarm (AQK) and Mangalavanam mangrove (MGM) yielding 45 ± 1.41µg/g and 41.2 ± 1.7µg/g DNA respectively. Protocol II gave poor DNA yield, with less than 13 µg/g for all samples, which is also apparent from the agarose gel. The DNA yield for protocol III was more than that for protocol II, in the range of 12 ± 0.7 µg/g to 26.25 ± 0.35 µg/g sediment, but less when compared to protocol I.

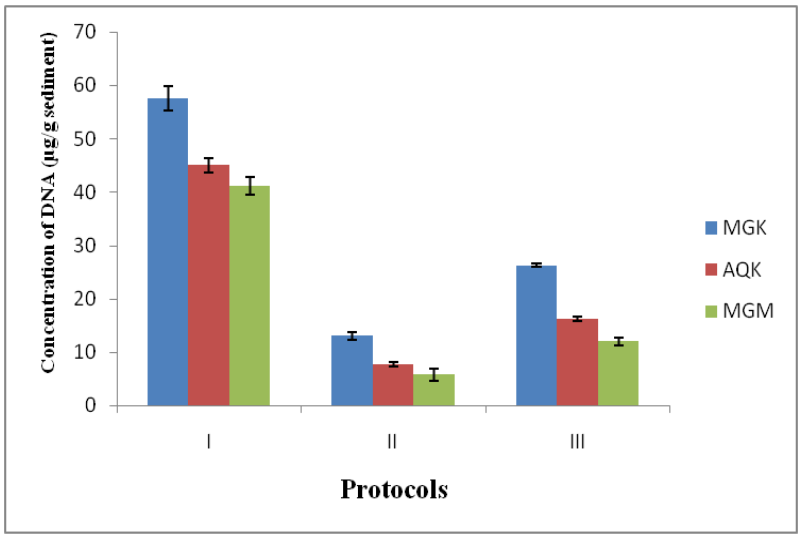

Fig. 4.2 DNA yield from three different samples using three protocols

4.1.3 Purity of metagenomic DNA

The purity of DNA isolated from three different sediment samples using three different protocols was determined by estimating the ratio between the spectrophotometric readings at 260nm and 280nm(OD_{260}/OD_{280}) as well as between 260nm and 230nm (OD_{260}/OD_{230}).

The OD_{260}/OD_{280} ratio of DNA isolated from three different sediment samples using the three different protocols are depicted in Fig. 4.3.

The OD_{260}/OD_{280} ratio gives an indication of protein contamination, with the ratio being in the range of 1.8-2.0 for pure samples. The OD_{260}/OD_{280} ratio of DNA samples isolated utilizing Protocol II was in/near the range 1.8-2.0, indicating that it is more pure compared to the DNA obtained using other two methods. For DNA isolated using protocols I and III the OD_{260}/OD_{280} ratio was above 2.25, indicating protein contamination in both these samples.

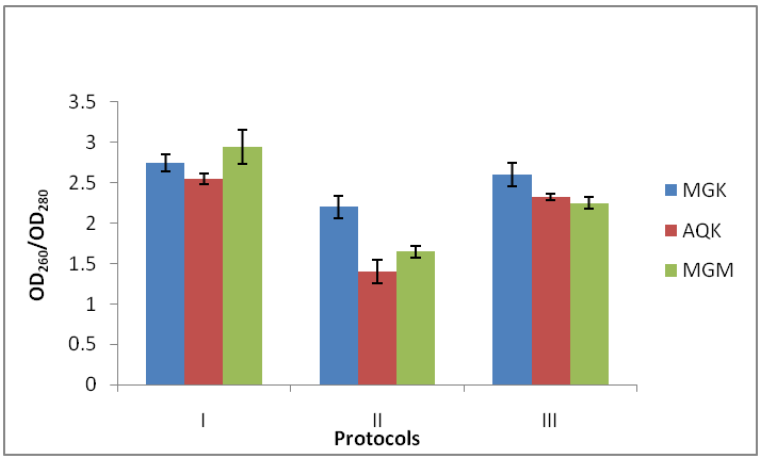

Fig. 4.3 Quality of DNA based on OD_{260}/OD_{280} ratio of different samples

The OD_{260}/OD_{230} ratio of DNA isolated from three different sediment samples using three different protocols are depicted in Fig. 4.4.

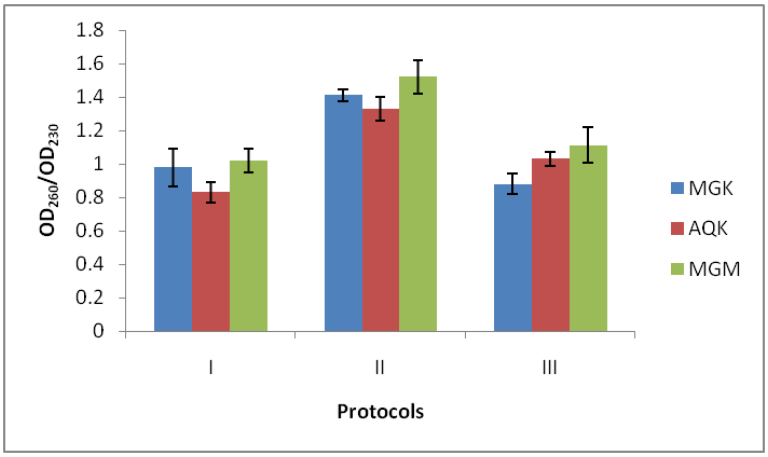

Fig. 4.4 Quality of DNA based on OD_{260}/OD_{230} ratio of different DNA samples

The OD_{260}/OD_{230} ratio gives an indication of the associated humic acid contamination, with the ratio being in the range of 1.2-2.0 for pure samples. The OD_{260}/OD_{230} ratio of DNA sample isolated according to Protocol II was in the

range1.33-1.52, indicating that the sample was devoid of humic contaminants. However for DNA isolated using protocols I and III, the OD_{260}/OD_{230} ratio was below 1.1, indicating humic contamination. Aluminium sulphate used in the protocol II for removal of humic contaminants from the samples is absent in protocol I and protocol III.

4.1.4 Metagenomic DNA isolation using kit

The DNA yield and purity obtained by the three different methods varied, with protocol I yielding more DNA, albeit with less purity. The DNA isolated by Protocol II was more pure, but yield was less than that obtained by Protocol I.

Since the yield and purity of DNA isolated by three different methods was not satisfactory for downstream processes, DNA isolation of Kannamaly mangrove (MGK) sediment sample was also done using the commercially available kit (MoBio UltraClean™ soil DNA isolation kit, CA, USA) and was analysed on agarose gel (Fig. 4.5).

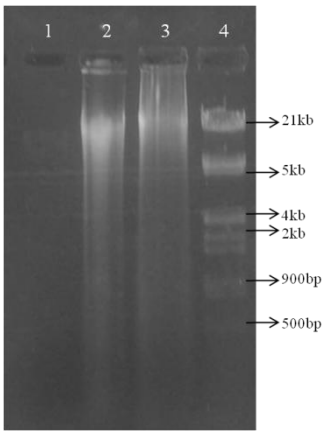

Fig. 4.5 Agarose gel electrophoresis of DNA isolated using kit.

Lane 1 – MGK DNA isolated by kit, Lane 2 – MGK DNA isolated by Protocol I, Lane 3 – MGK DNA isolated by Protocol I and purified using kit, Lane 4 – Lambda DNA *Eco*R I/ *Hin*d III double digest ladder (GeNei)

It was observed that the DNA yield was very low with the kit, which is seen very faintly on the gel (Lane 1). Consequently, the DNA isolated using Protocol I (Lane 2) was purified using the kit and seemed, brightly on the gel (Lane 3) indicating good yield.

The yield and purity of DNA obtained was determined spectrophotometrically and is depicted in Fig. 4.6.

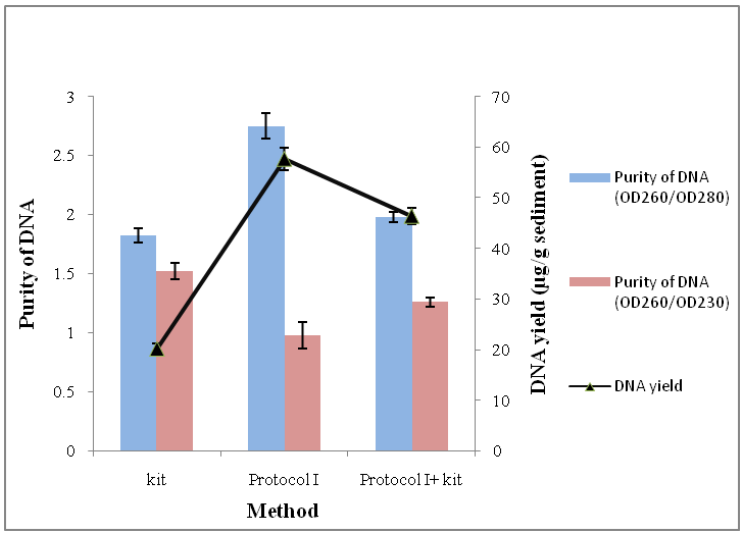

Fig. 4.6 Yield and Purity of DNA isolated by kit

The concentration of DNA obtained using kit was 20.12 ± 1.16 µg/g sediment as compared to that obtained by protocol I (57.6 ± 2.26 µg/g sediment). But the DNA sample was of good quality, as the OD_{260}/OD_{280} and OD_{260}/OD_{230} (1.82 and 1.52 respectively) ratios were within the limits, indicating no protein or humic acid contamination. When DNA isolated using protocol I was purified using the kit, there was significant DNA yield (46.34 ± 1.62 µg/g sediment). Also, the OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios (1.98 and 1.26 respectively) indicated purity of the isolated DNA. Hence the DNA sample isolated from Kannamaly

mangrove (MGK) using protocol I and purified using MoBio kit was used for all downstream processes.

4.2 ANALYSIS OF THE PHYLOGENETIC DIVERSITY OF THE MANGROVE METAGENOME BASED ON 16S rDNA SEQUENCES

In order to study the phylogenetic diversity of the mangrove samples, the 16S rRNA genes in the metagenome was amplified, followed by sequencing and thereafter by sequence analysis.

4.2.1 Amplification of 16S rDNA and construction of phylogenetic library

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

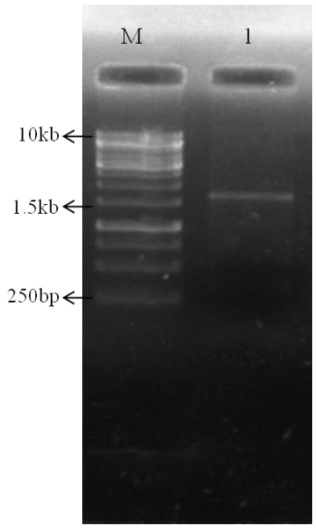

Fig. 4.7 Agarose gel electrophoresis of amplified 16S rDNA

Lane M – 1kb DNA marker (Fermentas), Lane 1 – 16S rDNA amplicon

The 1.5 kb 16S rDNA amplicons were ligated into TA cloning vector pTZ57R/T 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 metagenomic library.

4.2.2 Plasmid isolation from phylogenetic clones

The plasmids were isolated from the segregated white clones and were visualized on agarose gel after electrophoresis. The banding pattern of the recombinant plasmid was compared with the control plasmid, pTZ57R/T (without insert) as depicted in Fig. 4.8.

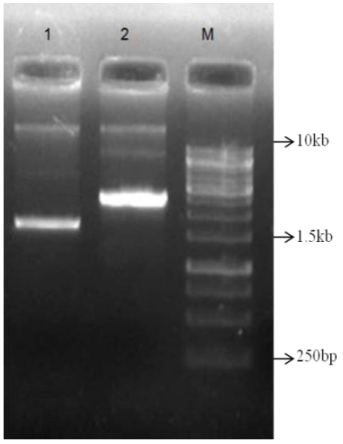

Fig. 4.8 Agarose gel showing band shift of recombinant plasmids

Lane 1 - plasmid pTZ57R/T without insert, Lane 2 - plasmid pTZ57R/T with insert, Lane M - 1kb DNA marker (Fermentas)

The recombinant plasmids, *i.e* plasmid pTZ57R/T with the insert, showed band shift on the agarose gel when compared with the control plasmids without insert. The plasmids showing band-shift on the gel were segregated and the presence of 16S rDNA insert was confirmed by reamplification of the gene from the isolated plasmids.

4.2.3 Reamplification of 16S rDNA from recombinant plasmids

The presence of the 16S rDNA in the segregated recombinant plasmids was confirmed by reamplification of 16S rDNA insert, using vector specific primers and the amplicons were visualized on agarose gel as in Fig. 4.9.

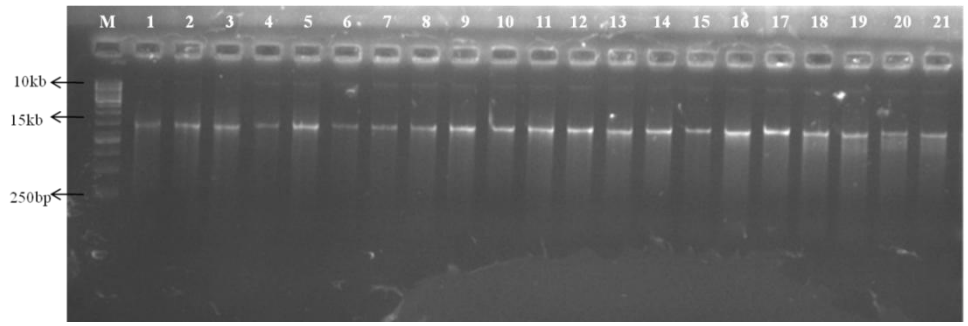

Fig. 4.9 Agarose gel of 16S rDNA reamplification from recombinant plasmids

Lane M - 1kb DNA marker (Fermentas), Lanes 1-20 – 16S rDNA amplicons

A total of 126 clones were confirmed to be recombinants after reamplification of the 16S rDNA insert and constituted the phylogenetic library and were maintained as stock in 20% glycerol for further analysis.

4.2.4 Sequence analysis of 16S rDNA insert of the phylogenetic clones

The 16S rDNA insert of the phylogenetic clones were sequenced. The identity of the sequences was determined using nBLAST of the NCBI database. The sequences were submitted to GenBank and accession numbers were obtained. The 126 accession numbers obtained were JQ805720 - JQ805723 (4 nos), JQ898300 - JQ898305 (6 nos), JQ898307, JQ898308, JQ868598, JQ868599, JQ868601, KF483718 - KF483791 (74 nos) and KF638653 - KF638689 (37 nos).

The clones were classified based on their taxonomic hierarchy by RDP Naïve Bayesian Classifier into 8 phyla of bacterial domain namely, *Proteobacteria*, *Bacteroidetes, Firmicutes, Chloroflexi*, candidate phylum *WS3*,

Planctomycetes, *Spirochaetes* and *Chlorobi*, based on sequence similarity. The distribution of the clones among different bacterial phyla is depicted in Fig. 4.10.

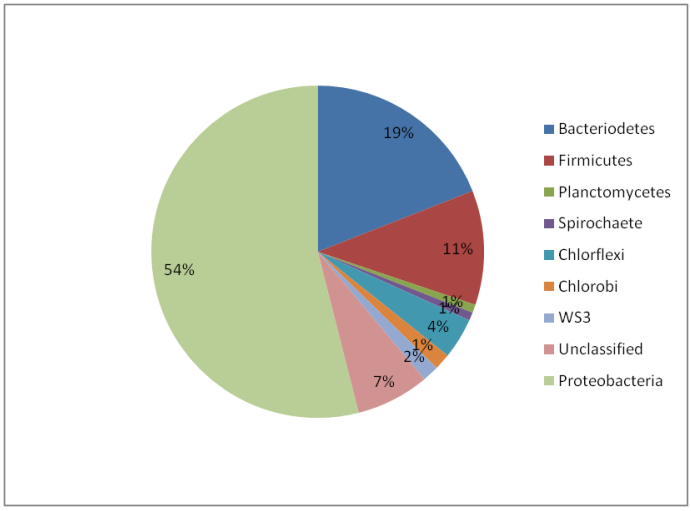

Fig. 4.10 The phylogenetic diversity of the mangrove metagenome based on 16S rDNA sequence analysis

Majority of the clones belonged to phylum *Proteobacteria* (54%), followed by *Bacteroidetes* (19%) and *Firmicutes* (11%). Phylum *Chloroflexi* and WS3 phylum accounted for 4% and 2% respectively. *Planctomycetes*, *Spirochaetes* and *Chlorobi* represented 1% each. Unclassified group accounted for 7% of the total number of clones.

The dominant phylum *Proteobacteria* was represented by four classes, namely, *Gammaproteobacteria*, *Deltaproteobacteria*, *Alphaproteobacteria* and *Epsilonproteobacteria* and their distribution is represented in Fig. 4.11.

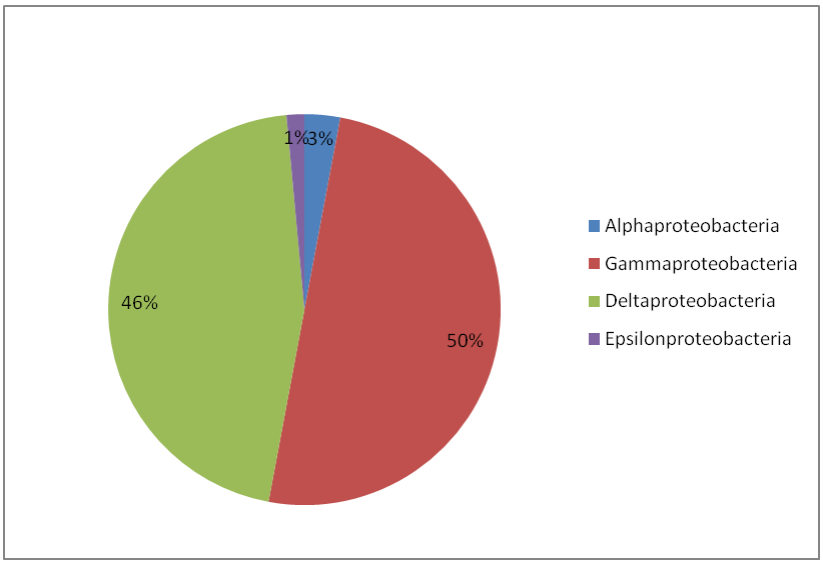

Fig. 4.11 Distribution of phylum *Proteobacteria*

The class *Gammaproteobacteria* represented the dominant group comprising of 50% of the clones, followed by *Deltaproteobacteria* (46%), *Alphaproteobacteria* (3%) and *Epsilonproteobacteria* (1%).

The taxonomic position of the clones as identified by RDP Naïve Bayesian classifier is tabulated in Table 4.1.

Table 4.1. Taxonomic classification of the 16SrDNA clones by RDP Naïve Bayesian rRNA Classifier

Sl. No.	Clone No	Acc. No.	Phylum	Class	Order	Family	Genus/RDP identity
1	BTM17	KF483719	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		Unclassified *Gammaproteobacteria*
2	BTM9	KF483725	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Marinobacterium*
3	BTM13	KF483727	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*		Unclassified *Alteromonadales*
4	BTM29	KF483731	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
5	BTM35	KF483735	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
6	BTM42	KF483741	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		*Thiohalomonas*
7	BTM52	KF483750	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
8	BTM81	KF483752	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		*Thiohalophilus*
9	BTM86	KF483753	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		*Thiohalophilus*
10	BTM87	KF483754	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
11	BTM92	KF483758	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
12	BTM95	KF483759	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*	Unclassified *Xanthomonadaceae*
13	BTM97	KF483760	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
14	BTM94	KF483764	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
15	BTM98	KF483770	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
16	BTM107	KF483776	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		*Thiohalophilus*
17	BTM108	KF483777	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
18	BTM109	KF483778	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
19	BTM117	KF483783	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
20	BTM118	KF483784	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
21	BTM201	KF638653	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
22	BTM210	KF638658	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		Unclassified *Gammaproteobacteria*
23	BTM215	KF638663	*Proteobacteria*	*Gammaproteobacteria*	*Methylococcales*	*Methylococcaceae*	*Methylocaldum*
24	BTM219	KF638665	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
25	BTM220	KF638666	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Haliea*
26	BTM229	KF638672	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		Unclassified *Gammaproteobacteria*
27	BTM231	KF638673	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		Unclassified *Gammaproteobacteria*
28	BTM233	KF638675	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		Unclassified *Gammaproteobacteria*
29	BTM235	KF638676	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		*Methylohalomonas*
30	BTM238	KF638678	*Proteobacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Oceanospirillaceae*	*Oceaniserpentilla*
31	BTM240	KF638680	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria incertae sedis*		*Thioprofundum*
32	BTM248	KF638683	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
33	BTM257	KF638686	*Proteobacteria*	*Gammaproteobacteria*			Unclassified *Gammaproteobacteria*
34	BT10	JQ805723	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	*Escherichia/Shigella*

35	BTM3	KF483722	*Proteobacteria*	*Deltaproteobacteria*	*Desulfuromonadales*	*Desulfuromonadaceae*	*Pelobacter*
36	BTM11	KF483726	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
37	BTM37	KF483737	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	*Desulfosarcina*
38	BTM44	KF483742	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
39	BTM45	KF483743	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
40	BTM47	KF483745	*Proteobacteria*	*Deltaproteobacteria*	*Desulfuromonadales*	*Desulfuromonadaceae*	*Pelobacter*
41	BTM48	KF483746	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
42	BTM50	KF483748	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
43	BTM53	KF483751	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
44	BTM91	KF483757	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobulbaceae*	*Desulfobulbus*
45	BTM76	KF483761	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
46	BTM84	KF483762	*Proteobacteria*	*Deltaproteobacteria*	*Desulfuromonadales*	*Desulfuromonadaceae*	*Pelobacter*
47	BTM14	KF483765	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	*Desulfosarcina*
48	BTM16	KF483766	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
49	BTM103	KF483774	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
50	BTM114	KF483780	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	*Desulfobacterium*
51	BTM120	KF483786	*Proteobacteria*	*Deltaproteobacteria*	*Syntrophobacterales*		Unclassified *Syntrophobacterales*
52	BTM82	KF483789	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
53	BTM202	KF638654	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
54	BTM207	KF638656	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
55	BTM212	KF638660	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
56	BTM214	KF638662	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
57	BTM218	KF638664	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
58	BTM221	KF638667	*Proteobacteria*	*Deltaproteobacteria*	*Desulfuromonadales*	*Desulfuromonadaceae*	Unclassified *Desulfuromonadaceae*
59	BTM223	KF638668	*Proteobacteria*	*Deltaproteobacteria*	*Desulfuromonadales*	*Desulfuromonadaceae*	*Desulfuromonas*
60	BTM232	KF638674	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
61	BTM236	KF638677	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
62	BTM239	KF638679	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
63	BTM244	KF638681	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
64	BTM251	KF638684	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	Unclassified *Desulfobacteraceae*
65	BTM259	KF638687	*Proteobacteria*	*Deltaproteobacteria*			Unclassified *Deltaproteobacteria*
66	BTM24	KF483721	*Proteobacteria*	*Alphaproteobacteria*			Unclassified *Alphaproteobacteria*
67	BTM255	KF638685	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*		Unclassified *Sphingomonadales*
68	BTM36	KF483736	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacterales*	*Helicobacteraceae*	*Sulfurimonas*
69	BTM49	KF483747	*Bacteroidetes*				Unclassified *Bacteroidetes*
70	BTM79	KF483791	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	Unclassified *Flavobacteriaceae*
71	BTM83	KF483790	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*		Unclassified *Sphingobacteriales*

72	BTM122	KF483788	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Cytophagaceae*	*Ekhidna*
73	BTM209	KF638657	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Flammeovirgaceae*	*Fulvivirga*
74	BTM225	KF638670	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	Unclassified *Flavobacteriaceae*
75	BTM51	KF483749	*Bacteroidetes*				Unclassified *Bacteroidetes*
76	BTM101	KF483772	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	Unclassified *Flavobacteriaceae*
77	BTM25	KF483728	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	Unclassified *Flavobacteriaceae*
78	BTM27	KF483729	*Bacteroidetes*				Unclassified *Bacteroidetes*
79	BTM8	KF483724	*Bacteroidetes*				Unclassified *Bacteroidetes*
80	BTM31	KF483732	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Saprospiraceae*	*Haliscomenobacter*
81	BTM32	KF483733	*Bacteroidetes*				Unclassified *Bacteroidetes*
82	BTM38	KF483738	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Saprospiraceae*	Unclassified *Saprospiraceae*
83	BTM39	KF483739	*Bacteroidetes*				Unclassified *Bacteroidetes*
84	BTM18	KF483767	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Saprospiraceae*	*Lewinella*
85	BTM105	KF483775	*Bacteroidetes*				Unclassified *Bacteroidetes*
86	BTM115	KF483781	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Saprospiraceae*	*Haliscomenobacter*
87	BTM116	KF483782	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Saprospiraceae*	Unclassified *Saprospiraceae*
88	BTM119	KF483785	*Bacteroidetes*				Unclassified *Bacteroidetes*
89	BTM203	KF638655	*Bacteroidetes*	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*Zeaxanthinibacter*
90	BTM211	KF638659	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Flammeovirgaceae*	Unclassified *Flammeovirgaceae*
91	BTM213	KF638661	*Bacteroidetes*				Unclassified_*Bacteroidetes*
92	BTM227	KF638671	*Bacteroidetes*				Unclassified *Bacteroidetes*
93	BT1	JQ898300	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	Unclassified *Bacillaceae*
94	BT2	JQ898301	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*
95	BT3	JQ898302	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*
96	BT5	JQ898303	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Bacillus*
97	BT6	JQ805720	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	Unclassified *Bacillaceae*
98	BT7	JQ805721	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	Unclassified *Bacillaceae*
99	BT9	JQ805722	*Firmicutes*	*Bacilli*	*Bacillales*		Unclassified *Bacillales*
100	BT11	JQ898304	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	Unclassified *Bacillaceae*
101	BT15	JQ868598	*Firmicutes*	*Bacilli*	*Bacillales*		Unclassified *Bacillales*
102	BT16	JQ868599	*Firmicutes*	*Bacilli*	*Bacillales*		Unclassified *Bacillales*
103	BT17	JQ898305	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Falsibacillus*
104	BT22	JQ898307	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	Unclassified *Bacillaceae*
105	BT23	JQ868601	*Firmicutes*	*Bacilli*	*Bacillales*		Unclassified *Bacillales*
106	BT24	JQ898308	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	*Falsibacillus*
107	BTM15	KF483718	*Chloroflexi*				Unclassified *Chloroflexi*
108	BTM22	KF483720	*Chloroflexi*	*Dehalococcoidetes*			*Dehalogenimonas*

109	BTM88	KF483755	*Chloroflexi*				Unclassified *Chloroflexi*
110	BTM93	KF483763	*Chloroflexi*				Unclassified *Chloroflexi*
111	BTM111	KF483779	*Chloroflexi*				Unclassified *Chloroflexi*
112	BTM121	KF483787	*Chlorobi*	*Ignavibacteria*	*Ignavibacteriales*	*Ignavibacteriaceae*	*Ignavibacterium*
113	BTM46	KF483744	*Chlorobi*	*Ignavibacteria*	*Ignavibacteriales*	*Ignavibacteriaceae*	*Ignavibacterium*
114	BTM23	KF483769	*WS3*				*WS3* genera incertae sedis
115	BTM99	KF483771	*WS3*				*WS3* genera incertae sedis
116	BTM19	KF483768	*Spirochaetes*	*Spirochaetes*	*Spirochaetales*	*Spirochaetaceae*	*Spirochaeta*
117	BTM89	KF483756	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	Unclassified *Planctomycetaceae*
118	BTM6	KF483723					Unclassified Bacteria
119	BTM28	KF483730					Unclassified Bacteria
120	BTM34	KF483734					Unclassified Bacteria
121	BTM40	KF483740					Unclassified Bacteria
122	BTM102	KF483773					Unclassified Bacteria
123	BTM224	KF638669					Unclassified Bacteria
124	BTM245	KF638682					Unclassified Bacteria
125	BTM260	KF638688					Unclassified Bacteria
126	BTM261	KF638689					Unclassified Bacteria

The dominant phylum *Proteobacteria*, comprising 68 clones (54%) was represented by classes *Gammaproteobacteria* (50%), *Deltaproteobacteria* (46%), *Alphaproteobacteria* (3%) and *Epsilonproteobacteria* (1%).

Class *Gammaproteobacteria*, comprising 34 clones, was represented by orders *Alteromonadales*, *Xanthomonadales*, *Methylococcales*, *Oceanospirillales* and *Enterobacteriales* including members of the genus *Marinobacterium*, *Thiohalomonas*, *Thiohalophilus*, *Methylocaldum*, *Haliea*, *Methylohalomonas*, *Oceaniserpentilla*, *Thioprofundum* and *Escherichia/Shigella*.

The class *Deltaproteobacteria*, with 31 clones, was represented by the orders *Desulfuromonadales*, *Desulfobacterales* and *Syntrophobacterales*, including members of the genus *Pelobacter, Desulfosarcina, Desulfobulbus, Desulfobacterium and Desulfuromonas.*

The class *Alphaproteobacteria*, with two clones, was represented by the order *Sphingomonadales* and the class *Epsilonproteobacteria*, with single clone, was represented by the order *Campylobacterales* and genus *Sulphurimonas*.

Among the non-proteobacterium group, bacteroidetes was the prominent phylum, with 24 clones (19%) represented by the orders *Flavobacteriales* and *Sphingobacteriales*, comprising members of the genus *Ekhidna, Fulvivirga, Haliscomenobacter, Lewinella* and *Zeaxanthinibacter.*

Fourteen clones (11%) belonged to the phylum *Firmicutes*, represented by a single order *Bacillales* including members of the genus *Bacillus* and *Falsibacillus.*

Five clones (4%) belonged to the phylum *Chloroflexi*, among which one belonged to class *Dehalococcoidetes* and to the genus *Dehalogenimonas* while the other four clones could not be classified further and are grouped as unclassified *Chloroflexi*.

Two clones belonged to the *WS3* candidate phylum and could not be classified beyond phylum level. 2 clones represented the phylum *Chlorobi*, both of them belonging to order *Ignavibacteriales* and to the genus *Ignavibacterium*.

The phyla *Spirochaetes* and *Planctomycetes* were represented by a single clone each belonging to the orders *Spirochaetales* and *Planctomycetales* respectively. Nine clones (7%) could not be classified into any phylum and hence denoted as unclassified group.

The phylogenetic relationship based on partial 16S rDNA sequences of proteobacterial clones was depicted in Fig. 4.12.

The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches and scale bar shows sequence divergence. The analysis involved 68 nucleotide sequences (N=68) and *Oscillatoria splendida* (Accession number EU621365) was used as outgroup. Accession numbers are given in parentheses.

Out of 68 proteobacterial clones, 33 clones belonging to the *Gammaproteobacteria* the predominant group, clustered together and formed a separate clade. The second largest group *Deltaproteobacteria* represented by 31 clones formed another major clade. *Alphaproteobacteria* (2 clones) and *Epsilonproteobacteria* (1 clone) also stood distinctly based on their phylogenetic relationship. The outgroup *Oscillatoria splendida* (Accession number EU621365) was separated from all these clones.

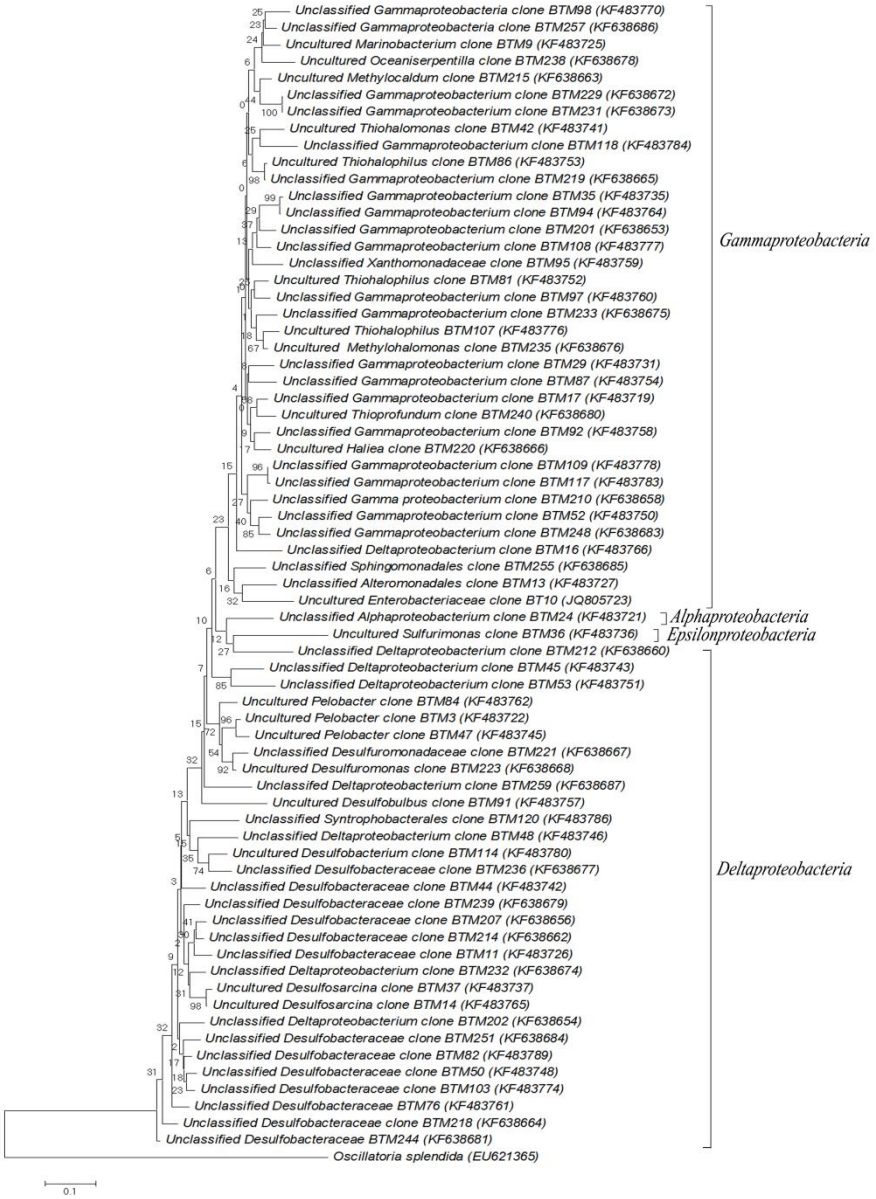

Fig. 4.12 Phylogenetic relationship of proteobacterial clones (N=68)

The phylogenetic relationship based on partial 16S rDNA sequences of 58 non-proteobacterial clones was depicted in Fig. 4.13.

The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches and scale bar shows sequence divergence. The analysis involved 58 nucleotide sequences and *Oscillatoria splendida* (Accession number EU621365) was used as outgroup. Accession numbers are given in parentheses.

From the phylogram, it is clear that the clones belonged to diverse phyla and the members of each phylum grouped together to form separate clades. Out of 58 non-proteobacterial clones, 24 clones belonging to *Bacteriodetes* phylum were grouped together and formed a separate clade. *Firmicutes* was represented by 14 clones and 5 clones belonged to *Chloroflexi* phylum, with each group forming separate clades.

However, 9 clones could not be assigned to any known phyla and formed two separate clusters. Minor representatives like *Chlorobi* (2 clones), *WS3* (2 clones), *Spirochaete* (1 clone) and *Planctomycetes* (1 clone) were also placed distinctly based on their phylogenetic relationship. The outgroup *Oscillatoria splendida* (Accession number EU621365) stood separated from all these clones.

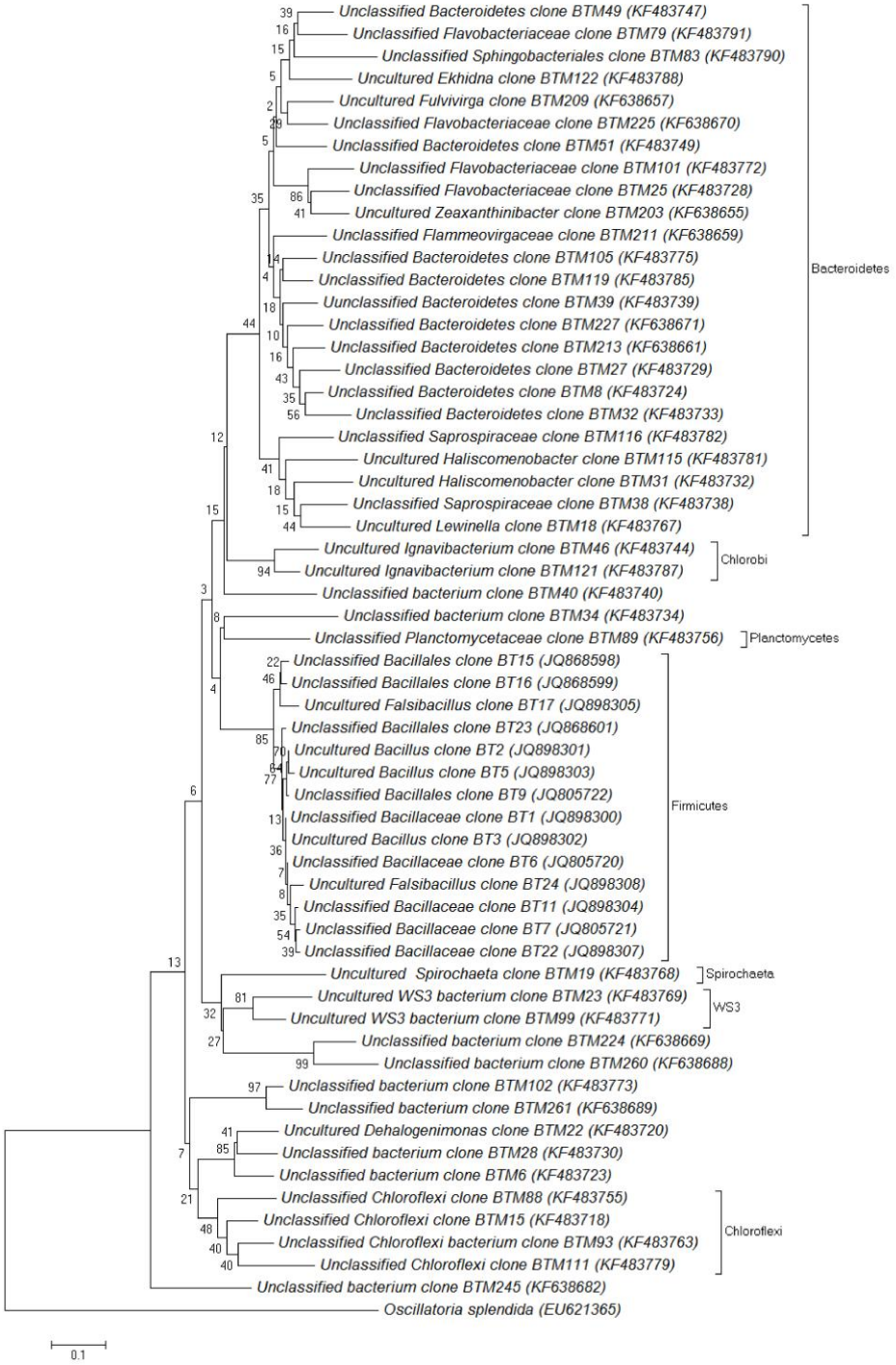

Fig. 4.13 Phylogenetic relationship of non-proteobacterial clones (N=58).

4.2.5 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 126 sequences and rarefaction curve was plotted using the RDP Pipeline (Fig. 4.14). At 3% sequence divergence, 116 different OTUs were observed, with the 126 sequences indicating the species richness of the sample. At 10% and 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 class and phylum level. 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. However, a substantial fraction of the bacterial diversity within the mangrove sediment samples was assessed at class and phylum level by the sampling effort.

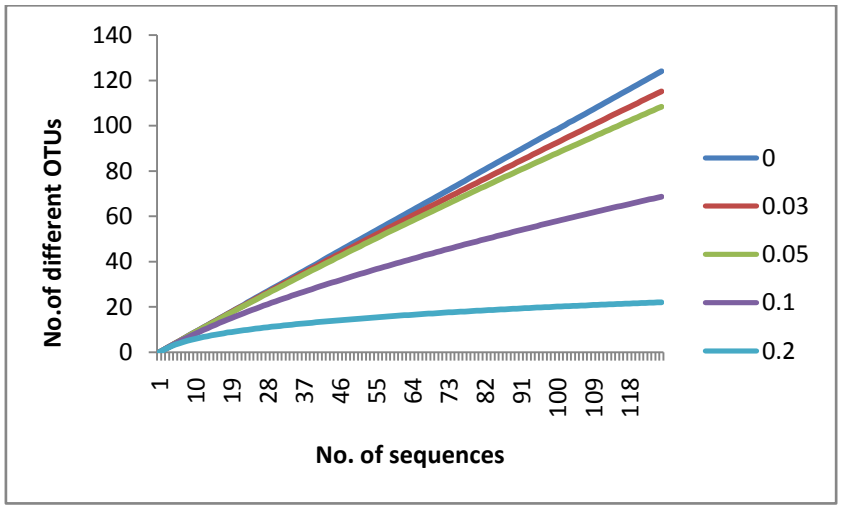

Fig. 4.14 Rarefaction curves indicating the observed number of operational taxonomic units (OTUs) at genetic distances of 3, 5, 10 and 20%, indicated by red, green, violet and blue lines. The reference line represents the OTUs at 0% genetic distance.

4.3 SCREENING FOR PROTEASE PRODUCTION BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

4.3.1 Construction of metagenomic library for screening of protease producer

The functional metagenomic library was constructed for the screening of protease producers by culture independent approach.

4.3.1.1 Restriction digestion of genomic DNA and plasmid DNA

The restriction digestion of the metagenomic DNA (MGK) using *Sau*3A I (GeNei, Bengaluru) generated less than 5kb size DNA fragments. The vector pUC18, upon digestion with enzyme *Bam*H I (GeNei, Bengaluru) was linearised and observed as a single band on agarose gel (Fig. 4.15).

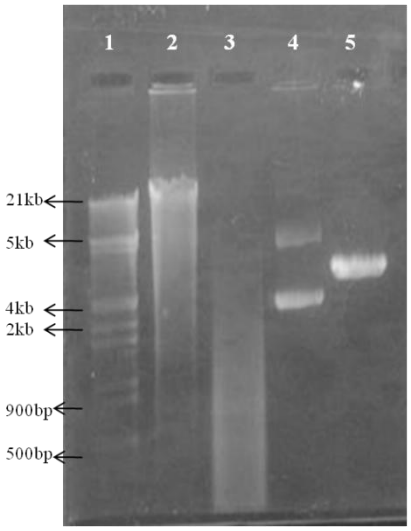

Fig. 4.15 Agarose gel electrophoresis of restriction digestion of metagenomic DNA (using *Sau*3A I) and vector DNA (using *Bam*H I) Lane 1- Lambda DNA *Eco*R 1/ *Hin*d III double digest ladder (GeNei), Lane 2 - Metagenomic DNA, Lane 3 - Digested metagenomic DNA, Lane 4 – pUC18, Lane 5 – digested pUC18

The digested metagenomic DNA fragments of <5kb were ligated into the vector and transformed into competent *E.coli* DH5α. The functional metagenomic library was constructed with 210 clones.

4.3.1.2 Screening of metagenomic clones for protease production

The functional library comprising 210 clones was screened for protease production by plate assay on skimmed milk agar plates (Fig. 4.16). Out of the 210 metagenomic clones screened, clone BTM106 produced extracellular protease enzyme, with an observable zone of clearance on the skimmed milk agar plate. This indicated proteolytic (caseinolytic) activity and therefore this clone was selected as a protease producer.

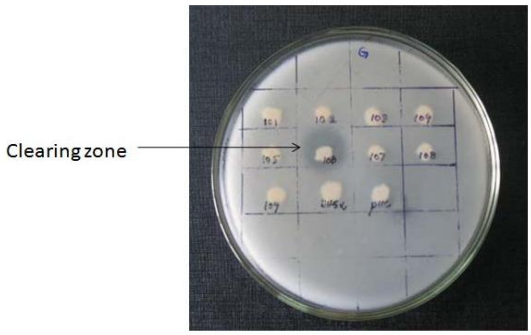

Fig. 4.16 Skimmed milk agar plate showing protease production by clone BTM106. The zone of clearance around the clone indicated proteolytic (caseinolytic) activity.

4.3.2 Screening for protease producer by culture dependent approach

The laboratory isolates obtained by culture dependent approach from the mangrove sediment were also screened for protease production and the strain BTKM4 identified as *Bacillus licheniformis* by 16S ribotyping (GenBank Accession No. HM030820) which produced clearing zone on skim milk agar plate was selected for the study.

The protease gene and the protease enzyme of the clone BTM106, obtained by culture independent approach and the bacterial isolate BTKM4, obtained by culture dependent approach were further characterized.

4.4 CHARACTERISATION OF THE PROTEASE GENE OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

4.4.1 Amplification of protease gene

The protease gene was amplified from the recombinant plasmid of the clone BTM106 using protease specific degenerate primers. The protease gene of strain BTKM4 was also amplified from its genomic DNA using specific primers (Fig. 4.17) and ~ 1kb amplicons were obtained. The amplicons were cloned in the vector pGEMT and sequenced.

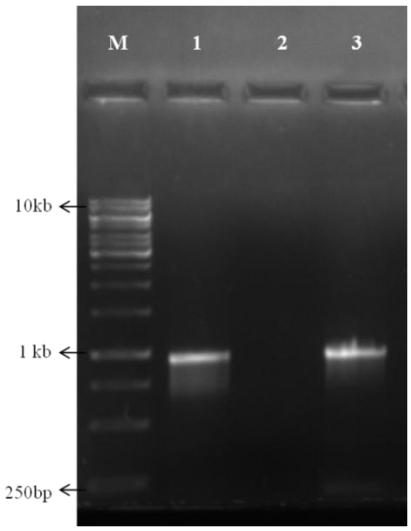

Fig. 4.17 Agarose gel electrophoresis of PCR amplification of protease gene

Arrow indicates amplified protease gene (~1kb) Lane M – 1kb DNA marker (Fermentas), Lane 1 – protease gene amplicon of strain BTKM4, Lane 3 – protease gene amplicon of clone BTM106

4.4.2 Characterisation of protease gene from clone BTM106

The protease gene amplicon from clone BTM106 was sequenced. The identity of the 977 bp long sequence was determined by comparing with sequences in the NCBI database. The GenBank accession numbers and description of the ten hits with maximum identity after nucleotide BLAST of protease gene sequence are as detailed in Table 4.2.

Table 4.2. Nucleotide blast results for the protease gene of clone BTM106 showing the GenBank accession numbers and description of ten hits with maximum identity

Accession No.	Description	Identity
CP001903	*Bacillus thuringiensis* BMB171, bacillolysin, mature peptide	99%
AE016877	*Bacillus cereus* ATCC 14579, bacillolysin	99%
CP005935	*Bacillus thuringiensis* YBT-1518, neutral protease	96%
CP004123	*Bacillus thuringiensis* serovar *thuringiensis* str. IS5056, neutral protease	96%
CP003889	*Bacillus thuringiensis* Bt407, bacillolysin	96%
CP001907	*Bacillus thuringiensis* serovar *chinensis* CT-43, bacillolysin	96%
CP003752	*Bacillus thuringiensis HD-771*, thermolysin-like metalloprotease	96%
DQ196437	*Bacillus cereus* neutral protease (NP) gene, partial cds	94%
CP006863	*Bacillus toyonensis* BCT-7112, neutral protease	94%
CP003687	*Bacillus thuringiensis* MC28, complete genome	94%

The gene sequence from the clone showed 99% identity to the mature peptide of bacillolysin of *Bacillus thuringiensis BMB171* (CP001903) and bacillolysin of *Bacillus cereus* ATCC 14579 (AE016877), which are *Bacillus* metalloendopeptidases (EC 3.4.24.28) included in peptidase M4 family of MEROPS database. The gene sequences were submitted in GenBank under accession number KC912765.

4.4.2.1 Multiple sequence alignment of the protease gene of clone BTM106

The genetic variation of the 977 bp protease gene sequence from the clone BTM106 was determined by its multiple sequence alignment with the first ten hits showing maximum identity as obtained by BLAST analysis. The consensus regions in the alignment of the gene sequences are denoted by an asterisk (Fig. 4.18). An ORF was detected by NCBI Orf finder tool having 957 bp. It extended from the start codon ATG at position 21 (indicated in red box) to the stop codon TAA at position 977 (indicated in green box). This ORF was in the reading frame 2 in 5' → 3' direction.

```
BTM106          CGAAAGCTGTTGTAAAGCCTATGGTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGTG 60
CP001903        CGAAAGCAGTAGTAAAACCTATGGTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTG 60
AE016877        CGAAAGCAGTAGTAAAACCTATGGTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTG 60
CP003889        CGAAAGCAGTAGTAAAACCTATGGTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTG 60
CP001907        CGAAAGCAGTAGTAAAACCTATGGTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTG 60
CP004123        CGAAAGCAGTAGTAAAACCTATGGTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTG 60
CP005935        CGAAAGCAGTAGTAAAACCTATGGTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTG 60
CP006863        --AAAGCTGCTGCAAAGCCAATGGTAACAGGAACGAATAAAGTGGGAACTGGTAAAGGTG 58
CP003687        --AAAGCTGCTGCAAAGCCAATGGTAACAGGAACGAATAAAGTGGGAACTGGTAAAGGTG 58
DQ196437        ----------------------GTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGTG 37
CP003752        -----TTAGT--TAATACC------AACAGCACTAAATGATTGCTTAACTGCT---GCTA 44
                                       ***** *   *** *      *****  *   * *

BTM106          TATTAGGAGATACAAAATCACTTAATACAACGTTATCTGGTTCATCTTACTACTTGCAAG 120
CP001903        TATTAGGAGATACAAAATCACTTAATACAACGTTATCTGGTTCATCTTACTACTTGCAAG 120
AE016877        TATTAGGAGATACGAAATCACTTAATACAACGTTATCTGGTTCATCTTACTACTTACAAG 120
CP003889        TATTAGGAGATACGAAATCACTCAATACAACATTATCTGGCTCATCTTACTACTTACAAG 120
CP001907        TATTAGGAGATACGAAATCACTCAATACAACATTATCTGGCTCATCTTACTACTTACAAG 120
CP004123        TATTAGGAGATACGAAATCACTCAATACAACATTATCTGGCTCATCTTACTACTTACAAG 120
CP005935        TATTAGGAGATACGAAATCACTCAATACAACATTATCTGGCTCATCTTACTACTTACAAG 120
CP006863        TATTAGGAGATACGAAATCGCTTAATACAACGTTATCTGGGTCATCTTACTACTTGCAAG 118
CP003687        TATTAGGAGATACGAAATCGCTTAATACAACGTTATCTGGGTCATCTTACTACTTGCAAG 118
DQ196437        TATTAGGAGATACGAAATCGCTTAATACAACGTTATCTGGGTCATCTTACTACTTGCAAG 97
CP003752        CTTCCGCAGA-----AGTCGCACCATATAA-ATCAGCTG---CCGCTTGC-ACT-GCA-- 91
                  *  * ***     * ** *   *** **  * * ***   *  *** * ***  **
```

```
BTM106          ATAATACGCGCGGGGCAACAATTTTCACATATGATGCGAAAAATCGTACTACATTGCCAG 180
CP001903        ATAATACGCGCGGGGCAACAATTTTCACATATGATGCGAAAAATCGTACTACATTGCCAG 180
AE016877        ATAATACGCGCGGGGCAACAATTTTCACATATGATGCGAAAAATCGTACTACATTACCAG 180
CP003889        ATAATACGCGCGGAGCAACGATTTTCACATATGATGCGAAAAATCGTACTACATTGCCAG 180
CP001907        ATAATACGCGCGGAGCAACGATTTTCACATATGATGCGAAAAATCGTACTACATTGCCAG 180
CP004123        ATAATACGCGCGGAGCAACGATTTTCACATATGATGCGAAAAATCGTACTACATTGCCAG 180
CP005935        ATAATACGCGCGGAGCAACGATTTTCACATATGATGCGAAAAATCGTACTACATTGCCAG 180
CP006863        ATAATACGCGCGGGGCAACAATTTTCACATATGATGCGAAAAACCGTTCAACATTACCGG 178
CP003687        ATAATACGCGCGGGGCAACAATTTTCACATATGATGCGAAAAATCGTTCAACATTACCGG 178
DQ196437        ATAATACGCGCGGGGCAACAATTTTCACATATGATGCGAAAAACCGTTCAACATTACCGG 157
CP003752        CCAGCACGAGCTTGGC--TAAATGTAGTAGATTGCGTGAAATATTGT---GTGTTTGCAC 146
                  *  *** **   **    * * *   * **   * **** *  **      **  *

BTM106          GAACTTTATGGGTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCAGCAGT 238
CP001903        GAACTTTATGGGTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCAGCAGT 238
AE016877        GAACTTTATGGGTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCAGCAGT 238
CP003889        GAACTTTATGGTTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCGGCAGT 238
CP001907        GAACTTTATGGTTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCGGCAGT 238
CP004123        GAACTTTATGGTTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCGGCAGT 238
CP005935        GAACTTTATGGTTAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCGGCAGT 238
CP006863        GAACATTATGGGCAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCAGCAGT 236
CP003687        GAACATTATGGGCAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCAGCAGT 236
DQ196437        GAACATTATGGGCAGATGCAGATAA--TGTTTTCAATGCAGCGTATGATGCAGCAGCAGT 215
CP003752        GG---TAATAAAT---TGCACCTAACTTATCTTTAC--CAATGCCCGTTACAGTTACACC 198
                *    * **       ****  ***  * * ** *   **  *   * * ***   **

BTM106          AGA-TGCTCATTACTAT--GCAGGTAGA-ACATATGATTATTA-TAAAGCTACATTTAAC 293
CP001903        AGA-TGCTCATTACTAT--GCAGGTAGA-ACATATGATTATTA-TAAAGCTACATTTAAC 293
AE016877        AGA-TGCTCATTACTAT--GCAGGTAGA-ACATATGATTATTA-TAAAGCTACATTTAAC 293
CP003889        AGA-TGCTCATTACTAT--GCAGGTAAA-ACATATGATTATTA-TAAAGCTACATTTAAT 293
CP001907        AGA-TGCTCATTACTAT--GCAGGTAAA-ACATATGATTATTA-TAAAGCTACATTTAAT 293
CP004123        AGA-TGCTCATTACTAT--GCAGGTAAA-ACATATGATTATTA-TAAAGCTACATTTAAT 293
CP005935        AGA-TGCTCATTACTAT--GCAGGTAAA-ACATATGATTATTA-TAAAGCTACATTTAAT 293
CP006863        AGA-TGCTCATTTCTAT--GCTGGTAGA-ACATATGATTATTA-TAAAGCTACATTTAAT 291
CP003687        AGA-TGCTCATTTCTAT--GCTGGTAGA-ACATATGATTATTA-TAAAGCTACATTCAAT 291
DQ196437        AGA-TGCTCATTTCTAT--GCTGGTAGA-ACATATGATTATTA-TAAAGCTACATTTAAT 270
CP003752        CGAATGCGTACCACCATTTGCTAATAAATAAGCTTGTTTGTTAATGATACCACTGTTTGT 258
                 ** ***  *   * **  **   ** * *    ** ** *** * *  * **  *

BTM106          AGAAACTCTATTAATGATGCAGGAGCTCCGTTAAAATCGACAGTTCATTACGGAAGTAAG 353
CP001903        AGAAACTCTATTAATGATGCAGGAGCTCCGTTAAAATCGACAGTTCATTACGGAAGTAAG 353
AE016877        AGAAACTCTATTAATGATGCAGGAGCTCCGTTAAAATCGACAGTTCATTACGGAAGTAAG 353
CP003889        AGAAACTCTATTAATGATGCAGGGGCACCGTTAAAATCAACAGTTCATTACGGAAGTAAG 353
CP001907        AGAAACTCTATTAATGATGCAGGGGCACCGTTAAAATCAACAGTTCATTACGGAAGTAAG 353
CP004123        AGAAACTCTATTAATGATGCAGGGGCACCGTTAAAATCAACAGTTCATTACGGAAGTAAG 353
CP005935        AGAAACTCTATTAATGATGCAGGGGCACCGTTAAAATCAACAGTTCATTACGGAAGTAAG 353
CP006863        AGAAACTCTATTAATGATGCAGGAGCACCATTAAAATCAACAGTTCATTACGGAAGTAAG 351
CP003687        AGAAACTCTATTAATGATGCAGGAGCACCATTAAAATCAACAGTTCATTACGGAAGTAAG 351
DQ196437        AGAAACTCTATTAATGATGCAGGAGCACCATTAAAATCAACAGTTCATTACGGAAGTAAG 330
CP003752        ATGAACCCCACCGTT-AT-CACTAGTACCAGTGTAAC-----GCTTAGAATAGTGGTCTG 311
                *  *** * *    * ** **   *  **  *  **      * * *  *  *  **  *
```

```
BTM106          TATAATAATGCATTCTGGAATGGCTCACAAATGGTATACGGAGATGGTGATGGTGTAACA 413
CP001903        TATAATAATGCATTCTGGAATGGCTCACAAATGGTATACGGAGATGGTGATGGTGTAACA 413
AE016877        TATAATAATGCATTCTGGAATGGCTCACAAATGGTATACGGAGATGGTGATGGTGTAACA 413
CP003889        TATAATAATGCATTCTGGAACGGCTCACAAATGGTATACGGAGATGGAGATGGTGTAACA 413
CP001907        TATAATAATGCATTCTGGAACGGCTCACAAATGGTATACGGAGATGGAGATGGTGTAACA 413
CP004123        TATAATAATGCATTCTGGAACGGCTCACAAATGGTATACGGAGATGGAGATGGTGTAACA 413
CP005935        TATAATAATGCATTCTGGAACGGCTCACAAATGGTATACGGAGATGGAGATGGTGTAACA 413
CP006863        TATAATAATGCATTCTGGAACGGTTCACAAATGGTATACGGAGATGGTGATGGTGTAACA 411
CP003687        TATAATAATGCATTCTGGAACGGTTCACAAATGGTATACGGAGATGGTGATGGTGTAACA 411
DQ196437        TATAATAATGCATTCTGGAACGGTTCACAAATGGTATACGGAGATGGTGATGGTGTAACA 390
CP003752        GATCACCATATTTCGTTGGATCACTCATAGA------GCGAAGC-------------GCA 352
                 ** *  **   *  * * *    *** * *       ** **              **

BTM106          TTCACTTCATTATCTGGTGGAATTGATGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
CP001903        TTCACTTCATTATCTGGTGGAATTGATGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
AE016877        TTCACTTCATTATCTGGTGGAATTGATGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
CP003889        TTCACTTCATTATCTGGTGGAATTGACGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
CP001907        TTCACTTCATTATCTGGTGGAATTGACGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
CP004123        TTCACTTCATTATCTGGTGGAATTGACGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
CP005935        TTCACTTCATTATCTGGTGGAATTGACGTAATTGGTCACGAATTAACGCATGCTG--TTA 471
CP006863        TTCACTTCATTATCTGGTGGAATTGATGTAATTGGTCACGAATTAACGCATGCTG--TTA 469
CP003687        TTCACTTCATTATCTGGTGGAATTGATGTAATTGGTCACGAATTAACGCATGCTG--TTA 469
DQ196437        TTCACTTCATTATCTGGTGGAATTGATGTAATTGGTCACGAATTAACGCATGCTG--TTA 448
CP003752        TCTCCTGCTTTGCCAGGCGTA--TAA-ATATCT--TCCCCAATTTCCCAAT-CTGGGTTA 406
                *   ** * **  * ** * *  * *  **  *  ** * ****  *  ** ***  ***

BTM106          CTG-----AAAATAGCTCGGACTTAGTTTATCAAAA-----TGAATCAGGAGCGCTAAAT 521
CP001903        CTG-----AAAATAGCTCGGACTTAATTTATCAAAA-----TGAATCAGGAGCGCTAAAT 521
AE016877        CTG-----AAAATAGCTCGGACTTAATTTATCAAAA-----TGAATCAGGAGCGCTAAAT 521
CP003889        CGG-----AAAATAGCTCGGACTTAATTTATCAAAA-----TGAATCAGGAGCGTTAAAT 521
CP001907        CGG-----AAAATAGCTCGGACTTAATTTATCAAAA-----TGAATCAGGAGCGTTAAAT 521
CP004123        CGG-----AAAATAGCTCGGACTTAATTTATCAAAA-----TGAATCAGGAGCGTTAAAT 521
CP005935        CGG-----AAAATAGCTCGGACTTAATTTATCAAAA-----TGAATCAGGAGCGTTAAAT 521
CP006863        CGG-----AAAATAGCTCGGATTTAATTTATCAAAA-----TGAATCAGGAGCGCTAAAT 519
CP003687        CGG-----AAAATAGCTCGGATTTAATTTATCAAAA-----TGAATCAGGAGCGCTAAAT 519
DQ196437        CGG-----AAAATAGCTCGGATTTAATTTATCAAAA-----TGAATCAGGAGCGCTAAAT 498
CP003752        CGGTTATCATAATA-CTCTA--CTAATGTACCAAAAACATCTGAAATAGCTTCATTTAAT 463
                * *     * **** ***     ** * ** *****     ****  **   *  * ***

BTM106          GAAGCAATTTCTGATAT-TTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCG 578
CP001903        GAAGCAATTTCTGATAT-TTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCG 578
AE016877        GAAGCAATTTCTGATAT-TTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCG 578
CP003889        GAAGCGATTTCTGATAT-CTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCA 578
CP001907        GAAGCGATTTCTGATAT-CTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCA 578
CP004123        GAAGCGATTTCTGATAT-CTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCA 578
CP005935        GAAGCGATTTCTGATAT-CTTTGGTACTTTAG--TAGAATTCTATGATAACCGTAACCCA 578
CP006863        GAAGCAATTTCTGATAT-CTTTGGTACTTTAG--TAGAATACTATGATAATCGTAATCCA 576
CP003687        GAAGCAATTTCTGATAT-CTTTGGTACTTTAG--TAGAATACTATGATAATCGTAATCCA 576
DQ196437        GAAGCAATTTCTGATAT-CTTTGGTACTTTAG--TAGAATACTATGATAATCGTAATCCA 555
CP003752        G------CTCCTGATTCATTTTGATAAATTAAATCTGAG--CTATACT--CTGTAA--CA 511
                *       * *****    **** **  ***     **   ****  *    ****  *
```

```
BTM106          GATTGGGA-GATTGGTGA--AG--ATATTTATAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
CP001903        GATTGGGA-GATTGGTGA--AG--ATATTTATAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
AE016877        GATTGGGA-GATTGGTGA--AG--ATATTTATAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
CP003889        GATTGGGA-GATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
CP001907        GATTGGGA-GATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
CP004123        GATTGGGA-GATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
CP005935        GATTGGGA-GATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCGCTT 632
CP006863        GATTGGGA-AATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCACTT 630
CP003687        GATTGGGA-AATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCACTT 630
DQ196437        GATTGGGA-AATTGGTGA--AG--ATATTTACAC-GCCTGGTAAAGCAGGAGATGCACTT 609
CP003752        GCATGCGTTAATTCGTGACCAATTACATCAATACCACCAGATAATGAAG----TGAACGT 567
                *  ** *   *** ****  *   * **  * **  ** * *** * **    **  * *

BTM106          CGCTCTATGAGTGATCCAGCGAAATACG-GTGA-----TCCAGAC--CATTATTCTAAGC 684
CP001903        CGCTCTATGAGTGATCCAGCGAAATACG-GTGA-----TCCAGAC--CATTATTCTAAGC 684
AE016877        CGCTCTATGAGTGATCCAGCGAAATACG-GTGA-----TCCAGAC--CATTATTCTAAGC 684
CP003889        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 684
CP001907        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 684
CP004123        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 684
CP005935        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 684
CP006863        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 682
CP003687        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 682
DQ196437        CGCTCTATGAGTGATCCAGCGAAATATG-GTGA-----CCCAGAC--CATTATTCTAAGC 661
CP003752        TACACCATCACCATCTCCGTATACCATTTGTGAGCCGTTCCAGAATGCATTATTATACTT 627
                  * * ** *      * *   *  *   ****      *****   *******  **

BTM106          GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
CP001903        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
AE016877        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
CP003889        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
CP001907        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
CP004123        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
CP005935        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACA-----AACAGTGGTATCATT 737
CP006863        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACG-----AACAGTGGTATTATT 735
CP003687        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACG-----AACAGTGGTATTATT 735
DQ196437        GTTACACTGGT--ACTAGTGATAACGGTGGGGTTCATACG-----AACAGTGGTATTATT 714
CP003752        ACTTCCGTAATGAACTGTTGATTTTAATGGTGCTCCTGCATCATTAATAGAGTTTCTATT 687
                  * *  *  *  ***  ****     *** * ** * *      ** ** * *   ***

BTM106          AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
CP001903        AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
AE016877        AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
CP003889        AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
CP001907        AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
CP004123        AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
CP005935        AA---CA-------AACAAGCTTATTTATTAGCAA-ATGGTGGTACGCATTCGGGTGTAA 786
CP006863        AA---CA-------AGCAAGCTTATTTATTAGCAA-ATGGCGGTACGCATTCTGGTGTAA 784
CP003687        AA---CA-------AGCAAGCTTATTTATTAGCAA-ATGGCGGTACGCATTCTGGTGTAA 784
DQ196437        AA---CA-------AGCAAGCTTATTTATTAGCAA-ATGGCGGTACGCATTCTGGTGTAA 763
CP003752        AAATGTAGCTTTATAGTAATCATATGTTTTACCTGCATAGTAATGAGCAT-CTACTGCTG 746
                **    *       *  ** * *** * *** *   ** *   *  **** *   **
```

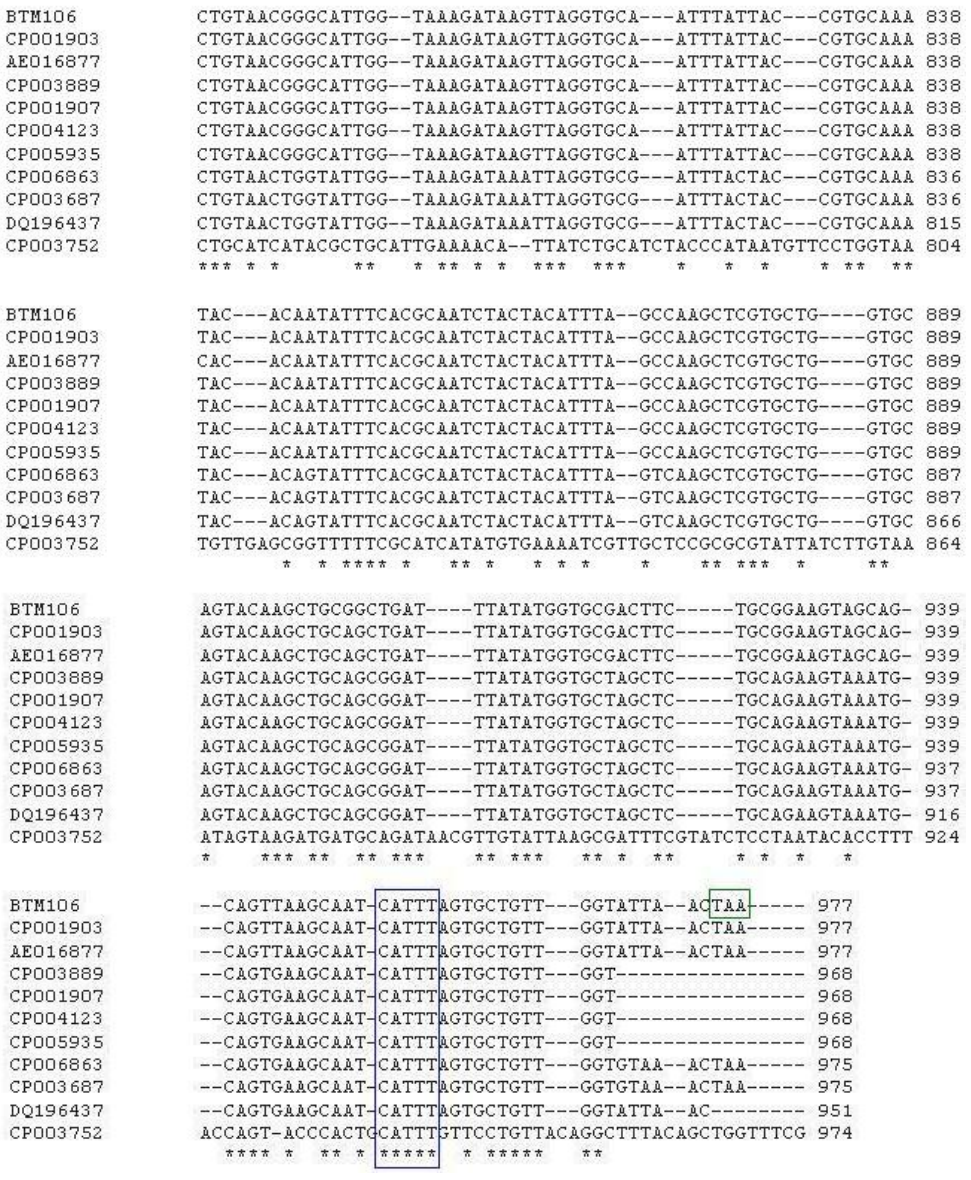

Fig. 4.18 Multiple sequence alignment of the partial protease gene sequence of clone BTM106 with other similar sequences. Some of the consensus sequences are boxed (blue). The start codon and stop codon of the 957 bp ORF are shown in red and green boxes respectively.

The protease partial gene sequence of the clone is represented as BTM106. Other protease partial gene sequences used in the alignment are represented with their GenBank accession numbers CP001903 (*Bacillus thuringiensis* BMB171, bacillolysin, mature peptide), AE016877 (*Bacillus cereus* ATCC 14579, bacillolysin), CP003889 (*Bacillus thuringiensis* Bt407, bacillolysin), CP001907(*Bacillus thuringiensis* serovar *chinensis* CT-43, bacillolysin), CP004123 (*Bacillus thuringiensis* serovar *thuringiensis* str. IS5056, neutral protease), CP005935 (*Bacillus thuringiensis* YBT-1518, neutral protease), CP006863 (*Bacillus toyonensis* BCT-7112, neutral protease), CP003687 (*Bacillus thuringiensis* MC28, complete genome), DQ196437 (*Bacillus cereus* neutral protease (NP) gene, partial cds) and CP003752 (*Bacillus thuringiensis* HD-771, thermolysin-like metalloprotease)

4.4.2.2 Phylogenetic analysis of the protease gene of clone BTM106

Unrooted tree was constructed based on neighbor joining method to determine the phylogenetic interrelationship of partial protease gene of BTM106 and its ten maximum identities (Fig. 4.19).

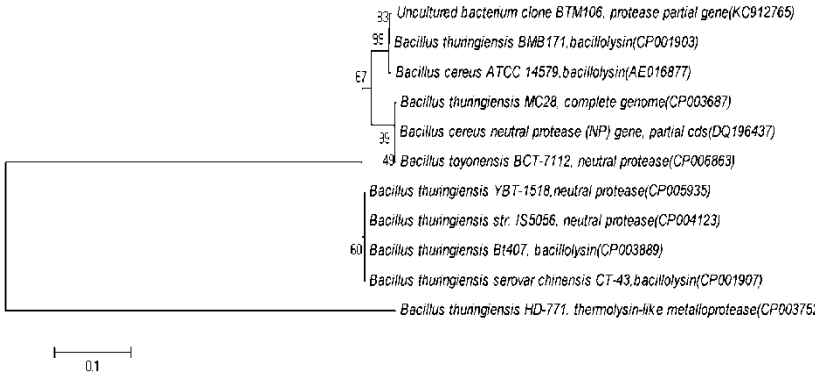

Fig. 4.19 The tree shows the phylogenetic interrelationships of protease gene nucleotide sequences from clone BTM106 with ten other similar sequences. The unrooted tree was constructed by the neighbor-joining method. Bootstrap values are given at the branching points and scale bar shows sequence divergence. Accession numbers are given in parentheses.

From the phylogenetic analysis it is clear that the partial protease gene of clone BTM106 showed similarity towards the bacillolysin of *Bacillus thuringiensis* (CP001903) and *Bacillus cereus* (AE016877), as they all together formed a single cluster.

4.4.3 Characterisation of protease gene sequence of strain BTKM4

The protease gene amplicon of strain BTKM4 was sequenced. The identity of the 954 bp long sequence was determined by comparing with sequences in the NCBI database. The GenBank accession numbers and description of the ten hits with maximum identity after nucleotide BLAST of protease gene sequence of BTKM4 are as detailed in Table 4.3.

Table 4.3. Nucleotide blast results for the protease gene of strain BTKM4 showing the GenBank accession numbers and description of ten hits with maximum identity

Accession No.	Description	Identity
M83910	*Bacillus cereus* neutral protease (NPRC) gene, complete cds	99%
CP001176	*Bacillus cereus* B4264, neutral protease	99%
CP004069	*Bacillus thuringiensis* serovar *kurstaki* str. HD73, bacillolysin	97%
L77763	*Bacillus thuringiensis* neutral protease A (nprA) gene, complete cds	97%
AP007209	*Bacillus cereus* NC7401 genomic DNA, neutral protease	94%
CP001177	*Bacillus cereus* AH187, neutral protease	94%
CP000227	*Bacillus cereus* Q1, bacillolysin	94%
CP000903	*Bacillus weihenstephanensis* KBAB4, thermolysin	94%
CP003763	*Bacillus thuringiensis* HD-789, neutral protease	93%
CP001186	*Bacillus cereus* G9842, neutral protease	93%

The gene sequence from strain BTKM4 showed 99% identity with the neutral protease gene of *Bacillus cereus* (M83910) and neutral protease of *Bacillus cereus* B4264 (CP001176). The BLAST analysis indicated that the gene

sequence obtained was that encoding the neutral protease gene and the sequence was submitted in GenBank under accession number KF769269.

4.4.3.1 Multiple sequence alignment of the protease gene of strain BTKM4

The genetic variation of the 954 bp partial protease gene sequence of the strain BTKM4 was determined by its multiple sequence alignment with the ten hits with maximum identity obtained by BLAST analysis. The consensus regions in the gene sequences are denoted by asterisk (Fig. 4.20). An ORF of 888 bp was detected by the NCBI ORF finder tool. This extended from the start codon ATG at position 67 (indicated in red box) to the stop codon TAA at position 954 (indicated in green box). This ORF was in the reading frame 1 in 5'→ 3' direction.

```
AP007209        GTAACAGGAACAAATAAAGTAGGAACGGGTAAAGGGGTATTAGGAGATACGAAGTCGCTT 60
CP001177        GTAACAGGAACAAATAAAGTAGGAACGGGTAAAGGGGTATTAGGAGATACGAAGTCGCTT 60
CP000227        GTAACAGGAACAAATAAAGTAGGAACGGGTAAAGGGGTATTAGGAGATACGAAGTCGCTT 60
CP004069        GTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGCGTACTAGGAGATACGAAGTCTCTT 60
L77763          GTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGCGTACTAGGAGATACGAAGTCTCTT 60
BTKM4           GTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGTGTATTAGGAGATACGAAGTCGCTT 60
M83910          GTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGTGTATTAGGAGATACGAAGTCGCTT 60
CP001176        GTAACAGGAACGAATAAAGTAGGAACTGGTAAAGGTGTATTAGGAGATACGAAGTCGCTT 60
CP003763        GTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTGTATTAGGAGATACGAAAATCACTT 60
CP001186        GTAACAGGAACAAATAAAGTAGGAACTGGTAAAGGTGTATTAGGAGATACGAAAATCACTT 60
CP000903        GTAACAGGGACGAATAAAGTTGGAACTGGTAAAGGTGTATTAGGAGATACGAAAATCGCTT 60
                ******** ** ******** ***** ******** *** ************** ** ***

AP007209        AATACAACGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGAGCAACAATT 120
CP001177        AATACAACGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGAGCAACAATT 120
CP000227        AATACAACGTTATCTGGATCATCTTACTACTTGCAAGATAATACACGCGGAGCAACAATT 120
CP004069        AATACAATGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGGGCAACGATT 120
L77763          AATACAATGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGGGCAACGATT 120
BTKM4           AATACAATGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGGGCAACGATT 120
M83910          AATACAATGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGGGCAACGATT 120
CP001176        AATACAACGTTATCTGGATCATCTTACTACTTACAAGATAATACACGCGGGGCAACGATT 120
CP003763        AATACAATATTATCTGGTTCATCTTACTACTTACAAGATAATACGCGCGGAGCAACGATT 120
CP001186        AATACAACATTATCTGGTTCATCTTACTACTTACAAGATAATACGCGCGGAGCAACGATT 120
CP000903        AATACAACGTTATCTGGATCATCTTACTATTTACAAGATAATACGCGCGGAGCAACAATT 120
                *******  ******** ************ ** ************ ***** ***** ***
```

```
AP007209        TTCACATATGATGCGAAAAACCGCTCGACATTACCAGGAACATTATGGGCAGATGCAGAT 180
CP001177        TTCACATATGATGCGAAAAACCGCTCGACATTACCAGGAACATTATGGGCAGATGCAGAT 180
CP000227        TTCACATATGATGCGAAAAACCGCTCGACATTACCAGGAACATTATGGGCAGATGCAGAT 180
CP004069        TTCACATATGATGCGAAAAACCGTTCAACATTACCAGGAACATTATGGGCAGATGCAGAT 180
L77763          TTCACATATGATGCGAAAAACCGTTCAACATTACCAGGAACATTATGGGCAGATGCAGAT 180
BTKM4           TTCACATATGATGCGAAAAACCGTTCAACATTACCAGGAACACTATGGGCAGATGCAGAT 180
M83910          TTCACATATGATGCGAAAAACCGTTCAACATTACCAGGAACACTATGGGCAGATGCAGAT 180
CP001176        TTCACATATGATGCGAAAAACCGTTCAACATTACCAGGAACATTATGGGCAGATGCAGAT 180
CP003763        TTCACATACGATGCGAAAAACCGCTCAACATTACCAGGAACATTATGGGCAGATGCAGAT 180
CP001186        TTCACATACGATGCGAAAAACCGCTCAACATTACCAGGAACATTATGGGCAGATGCAGAT 180
CP000903        TTCACATATGATGCGAAAAACCGCTCAACATTACCAGGAACATTATGGGCAGATGCAGAT 180
                ******** ************** ** *************** ******************

AP007209        AATGTTTTCAATGCAGCGTATGATGCAGCGGCAGTAGATGCTCATTACTATGCAGGTAAA 240
CP001177        AATGTTTTCAATGCAGCGTATGATGCAGCGGCAGTAGATGCTCATTACTATGCAGGTAAA 240
CP000227        AATGTTTTCAATGCAGCGTATGATGCAGCGGCAGTAGATGCTCATTACTATGCAGGTAAA 240
CP004069        AATGTTTTCAATGCAGCGTATGATGCAGCAGCGGTAGATGCTCATTACTATGCGGGTATC 240
L77763          AATGTTTTCAATGCAGCGTATGATGCAGCAGCGGTAGATGCTCATTACTATGCGGGTATC 240
BTKM4           AATGTTTTCAATGCAGCGTATGATGCAGCAGCAGTAGATGCTCATTACTATGCAGGTAAA 240
M83910          AATGTTTTCAATGCAGCGTATGATGCAGCAGCAGTAGATGCTCATTACTATGCAGGTAAA 240
CP001176        AATGTTTTCAATGCAGCGTATGATGCAGCAGCAGTAGATGCTCATTACTATGCAGGTAAA 240
CP003763        AATGTTTTCAATGCAGCGTATGATGCAGCGGCAGTAGATGCTCATTACTATGCAGGTAGA 240
CP001186        AATGTTTTCAATGCAGCGTATGATGCAGCGGCAGTAGATGCTCATTACTATGCAGGTAGA 240
CP000903        AATGTTTTCAATGCAGCGTATGATGCAGCGGCAGTAGATGCTCATTACTATGCAGGTAAA 240
                ***************************** ** ******************** ****

AP007209        ACATATGATTACTATAAAGCTACATTTAATAGAAACTCTATTAATGATGCAGGAGCACCG 300
CP001177        ACATATGATTACTATAAAGCTACATTTAATAGAAACTCTATTAATGATGCAGGAGCACCG 300
CP000227        ACATATGATTATTATAAAGCTACATTTAATAGAAACTCTATTAATGATGCAGGAGCACCG 300
CP004069        ACGTATGATTACTATAAGAATACATTTAATCGTAATTCAATTAATGATGCAGGAGCGCCG 300
L77763          ACGTATGATTACTATAAGAATACATTTAATCGTAATTCAATTAATGATGCAGGAGCGCCG 300
BTKM4           ACATATGATTACTATAAAGCTACGTTTAATAGAAATTCTATCAATGATGCAGGAGCGCCG 300
M83910          ACATATGATTACTATAAAGCTACGTTTAATAGAAATTCTATCAATGATGCAGGAGCGCCG 300
CP001176        ACATATGATTACTATAAAGCTACGTTTAATAGAAATTCTATTAATGATGCAGGAGCACCG 300
CP003763        ACATATGATTATTATAAAGCTACATTTAACAGAAACTCTATTAATGATGCAGGAGCACCA 300
CP001186        ACATATGATTATTATAAAGCTACATTTAACAGAAACTCTATTAATGATGCAGGAGCACCA 300
CP000903        ACATATGATTACTATAAAGCTACATTTAACAGAAACTCTATTAATGATGCAGGAGCACCA 300
                ** ******** *****   *** *****  * ** ** ** ************** **

AP007209        TTAAAATCAACAGTCCATTACGGAAGCAATTATAATAATGCATTCTGGAACGGATCACAA 360
CP001177        TTAAAATCAACAGTCCATTACGGAAGCAATTATAATAATGCATTCTGGAACGGATCACAA 360
CP000227        TTAAAATCAACAGTCCATTACGGAAGCAATTATAATAATGCATTCTGGAACGGATCACAA 360
CP004069        TTAAAATCAACAGTTCATTACGGAAGTAATTATAACAATGCATTCTGGAACGGATCACAG 360
L77763          TTAAAATCAACAGTTCATTACGGAAGTAATTATAACAATGCATTCTGGAACGGATCACAG 360
BTKM4           TTAAAATCGACAGTTCATTACGGAAGTAATTATAACAATGCATTCTGGAACGGATCACAG 360
M83910          TTAAAATCGACAGTTCATTACGGAAGTAATTATAACAATGCATTCTGGAACGGATCACAG 360
CP001176        TTAAAATCGACAGTTCATTACGGAAGTAATTATAACAATGCATTCTGGAACGGATCACAG 360
CP003763        TTAAAATCAACAGTTCATTACGGAAGTAAGTATAATAATGCATTCTGGAACGGTTCACAA 360
CP001186        TTAAAATCAACAGTTCATTACGGAAGTAAGTATAATAATGCATTCTGGAACGGTTCACAA 360
CP000903        TTAAAATCAACAGTTCATTACGGAAGTAAGTATAATAATGCATTCTGGAACGGCTCACAA 360
                ******** ***** *********** ** ***** ***************** *****
```

```
AP007209        ATGGTATACGGAGATGGTGATGGTGTAACGTTCACTTCATTATCTGGTGGTATTGATGTT 420
CP001177        ATGGTATACGGAGATGGTGATGGTGTAACGTTCACTTCATTATCTGGTGGTATTGATGTT 420
CP000227        ATGGTATACGGAGATGGTGATGGTGTAACGTTCACTTCATTATCTGGTGGTATTGATGTT 420
CP004069        ATGGTATACGGAGATGGTGATGGTGTAACATTTACTTCATTATCTGGTGGAATTGATGTA 420
L77763          ATGGTATACGGAGATGGTGATGGTGTAACATTTACTTCATTATCTGGTGGAATTGATGTA 420
BTKM4           ATGGTATACGGAGATGGTGATGGTGTAACGTTTACTTCATTATCTGGTGGAATTGATGTA 420
M83910          ATGGTATACGGAGATGGTGATGGTGTAACGTTTACTTCATTATCTGGTGGAATTGATGTA 420
CP001176        ATGGTATACGGAGATGGTGATGGTGTAACGTTTACTTCATTATCTGGTGGAATTGATGTA 420
CP003763        ATGGTATACGGGGATGGTGATGGTGTAACATTCACTTCATTATCTGGTGGAATTGACGTA 420
CP001186        ATGGTATACGGGGATGGTGATGGTGTAACATTCACTTCATTATCTGGTGGAATTGACGTA 420
CP000903        ATGGTATACGGAGATGGTGATGGTGTAACATTCACTTCATTATCTGGTGGCATTGATGTA 420
                *********** ********** ** ********** ***** **

AP007209        ATTGGTCACGAATTAACGCATGCTGTTACGGAAAATAGCTCGAATTTAATTTATCAAAAT 480
CP001177        ATTGGTCACGAATTAACGCATGCTGTTACGGAAAATAGCTCGAATTTAATTTATCAAAAT 480
CP000227        ATTGGTCACGAATTAACGCATGCTGTTACGGAAAATAGCTCGAATTTAATTTATCAAAAT 480
CP004069        ATTGGTCACGAGTTAACGCATGCTGTTACGGAAAATAGTTCAAATCTAATTTATCAAAAT 480
L77763          ATTGGTCACGAGTTAACGCATGCTGTTACGGAAAATAGTTCAAATCTAATTTATCAAAAT 480
BTKM4           ATTGGTCACGAGTTAACGCATGCTGTTACGGAAAATAGTTCAAATCTAATTTATCAAAAT 480
M83910          ATTGGTCACGAGTTAACGCATGCTGTTACGGAAAATAGTTCAAATCTAATTTATCAAAAT 480
CP001176        ATTGGTCACGAGTTAACGCATGCTGTTACGGAAAATAGTTCAAATCTAATTTATCAAAAT 480
CP003763        ATTGGTCACGAATTAACGCATGCTGTTACGGAAAATAGCTCGGACTTAATTTATCAAAAT 480
CP001186        ATTGGTCACGAATTAACGCATGCTGTTACGGAAAATAGCTCGGACTTAATTTATCAAAAT 480
CP000903        ATTGGCCATGAATTAACACATGCTGTTACGGAAAATAGCTCGGATTTAATTTATCAAAAT 480
                ***** ** ** ***** ********************** **  *  ***************

AP007209        GAATCAGGCGCATTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
CP001177        GAATCAGGCGCATTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
CP000227        GAATCAGGCGCATTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
CP004069        GAATCAGGGGCTTTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
L77763          GAATCAGGGGCTTTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
BTKM4           GAATCAGGGGCTTTAAATGAAGCGATTTCTGATATTTTTGGTACTTTAGTAGAATTCTAT 540
M83910          GAATCAGGGGCTTTAAATGAAGCGATTTCTGATATTTTTGGTACTTTAGTAGAATTCTAT 540
CP001176        GAATCTGGGGCTTTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
CP003763        GAATCAGGGGCGTTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
CP001186        GAATCAGGGGCGTTAAATGAAGCGATTTCTGATATCTTTGGTACTTTAGTAGAATTCTAT 540
CP000903        GAATCAGGAGCGTTAAATGAAGCAATTTCTGATATTTTTGGTACTTTAGTAGAATTCTAT 540
                ***** ** ** *********** *********** ************************

AP007209        GATAACCGTAACCCAGATTGGGAGATTGGTGAAGATATTTACACGCCTGGTAAAGCAGGA 600
CP001177        GATAACCGTAACCCAGATTGGGAGATTGGTGAAGATATTTACACGCCTGGTAAAGCAGGA 600
CP000227        GATAACCGTAACCCAGATTGGGAGATTGGTGAAGATATTTACACGCCTGGTAAAGCAGGA 600
CP004069        GATAACCGTAACCCGGATTGGGAGATTGGTGAAGATATTTACACACCTGGTAAAGCAGGA 600
L77763          GATAACCGTAACCCGGATTGGGAGATTGGTGAAGATATTTACACACCTGGTAAAGCAGGA 600
BTKM4           GATAACCGTAACCCGGATTGGGAGATTGGTGAAGATATTTACACACCTGGTAAAGCAGGA 600
M83910          GATAACCGTAACCCGGATTGGGAGATTGGTGAAGATATTTACACACCTGGTAAAGCAGGA 600
CP001176        GATAACCGTAACCCGGATTGGGAGATTGGTGAAGATATTTACACACCTGGTAAAGCAGGA 600
CP003763        GATAACCGTAACCCAGATTGGGAGATTGGTGAAGATATTTACACGCCTGGTAAAGCAGGA 600
CP001186        GATAACCGTAACCCAGATTGGGAGATTGGTGAAGATATTTACACGCCTGGTAAAGCAGGA 600
CP000903        GATAACCGTAACCCAGATTGGGAGATTGGTGAAGATATTTATACACCTAGTAAGGCGGGA 600
                ************** ************************** ** *** **** ** ***
```

```
AP007209        GATGCACTTCGTTCTATGAGTGATCCAGCGAAATATGGTGATCCAGACCATTATTCTAAG 660
CP001177        GATGCACTTCGTTCTATGAGTGATCCAGCGAAATATGGTGATCCAGACCATTATTCTAAG 660
CP000227        GATGCACTTCGTTCTATGAGTGATCCAGCGAAATATGGTGATCCAGACCATTATTCTAAG 660
CP004069        GACGCGCTTCGCTCTATGAGTGATCCTACGAAGTATGGTGATCCAGACCATTATTCTAAG 660
L77763          GACGCGCTTCGCTCTATGAGTGATCCTACGAAGTATGGTGATCCAGACCATTATTCTAAG 660
BTKM4           GACGCACTTCGCTCTATGAGTGATCCTACGAAGTATGGTGATCCAGACCATTATTCTAAG 660
M83910          GACGCACTTCGCTCTATGAGTGATCCTACGAAGTATGGTGATCCAGACCATTATTCTAAG 660
CP001176        GACGCACTTCGCTCTATGAGTGATCCTACGAAGTATGGTGATCCAGACCATTATTCTAAG 660
CP003763        GACGCGCTTCGCTCTATGAGTGATCCAGCGAAATATGGTGACCCAGACCACTATTCTAAG 660
CP001186        GACGCGCTTCGCTCTATGAGTGATCCAGCGAAATATGGTGACCCAGACCACTATTCTAAG 660
CP000903        GATGCACTTCGTTCTATGAGTGATCCAGCGAAATATGGTGACCCAGACCATTATTCTAAG 660
                ** ** ***** **************  **** ******** ******** *********

AP007209        CGTTACACTGGTTCAAGTGATAACGGTGGGGTTCATACAAACAGTGGCATTATTAATAAA 720
CP001177        CGTTACACTGGTTCAAGTGATAACGGTGGGGTTCATACAAACAGTGGCATTATTAATAAA 720
CP000227        CGTTACACTGGTTCAAGTGATAACGGTGGGGTTCATACAAACAGTGGCATTATTAATAAA 720
CP004069        CGTTACACTGGTTCAAGTGATAACGGTGGCGTTCATACAAACAGCGGCATTATTAATAAA 720
L77763          CGTTACACTGGTTCAAGTGATAACGGTGGCGTTCATACAAACAGCGGCATTATTAATAAA 720
BTKM4           CGTTACACTGGTTCAAGTGATAACGGTGGTGTTCATACAAACAGCGGCATTATTAATAAA 720
M83910          CGTTACACTGGTTCAAGTGATAACGGTGGTGTTCATACAAACAGCGGCATTATTAATAAA 720
CP001176        CGTTACACTGGTTCAAGTGATAACGGTGGTGTTCATACAAACAGCGGCATTATTAATAAA 720
CP003763        CGTTACACAGGTTCAAGTGATAACGGTGGCGTTCATACAAACAGTGGTATTATTAACAAA 720
CP001186        CGTTACACAGGTTCAAGTGATAACGGTGGCGTTCATACAAACAGTGGTATTATTAACAAA 720
CP000903        CGTTACACTGGTTCAAGTGATAACGGTGGCGTTCATACAAACAGTGGTATTATTAACAAA 720
                ******** ******************** ************** ** ******** ***

AP007209        CAAGCTTATTTATTAGCGAATGGCGGTACGCATTACGGTGTAACTGTAACTGGTATTGGT 780
CP001177        CAAGCTTATTTATTAGCGAATGGCGGTACGCATTACGGTGTAACTGTAACTGGTATTGGT 780
CP000227        CAAGCTTATTTATTAGCGAATGGCGGTACGCATTACGGTGTAACTGTAACTGGTATTGGT 780
CP004069        CAAGCTTATTTATTAGCAAATGGCGGTACGCATTACGGTGTAACTGTAAATGGTATCGGC 780
L77763          CAAGCTTATTTATTAGCAAATGGCGGTACGCATTACGGTGTAACTGTAAATGGTATCGGC 780
BTKM4           CAAGCTTATTTATTAGCAAATGGTGGTACGCATTATGGTGTAACTGTAACTGGTATCGGG 780
M83910          CAAGCTTATTTATTAGCAAATGGTGGTACGCATTATGGTGTAACTGTAACTGGTATCGGG 780
CP001176        CAAGCTTATTTATTAGCAAATGGTGGTACGCATTATGGTGTAACTGTAACTGGTATCGGT 780
CP003763        CAAGCTTATTTATTAGCAAATGGTGGTACGCATTCTGGTGTAACTGTAACTGGTATTGGT 780
CP001186        CAAGCTTATTTATTAGCAAATGGTGGTACGCATTCTGGTGTAACTGTAACTGGTATTGGT 780
CP000903        CAAGCTTATTTATTAGCAAATGGTGGTACGCATTCTGGTGTAACTGTAACTGGTATCGGT 780
                ***************** ***** **********  ************* ****** **

AP007209        AAAGATAAGTTAGGTGCAATTTATTACCGTGCAAACACACAATATTTCACGCAATCTACT 840
CP001177        AAAGATAAGTTAGGTGCAATTTATTACCGTGCAAACACACAATATTTCACGCAATCTACT 840
CP000227        AAAGATAAGTTAGGTGCAATTTATTACCGTGCAAACACACAATATTTCACGCAATCTACT 840
CP004069        AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
L77763          AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
BTKM4           AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
M83910          AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
CP001176        AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
CP003763        AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
CP001186        AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
CP000903        AAAGATAAATTAGGTGCGATTTACTACCGTGCAAATACACAGTATTTCACGCAATCTACT 840
                ******** ******** ***** *********** ***** ******************
```

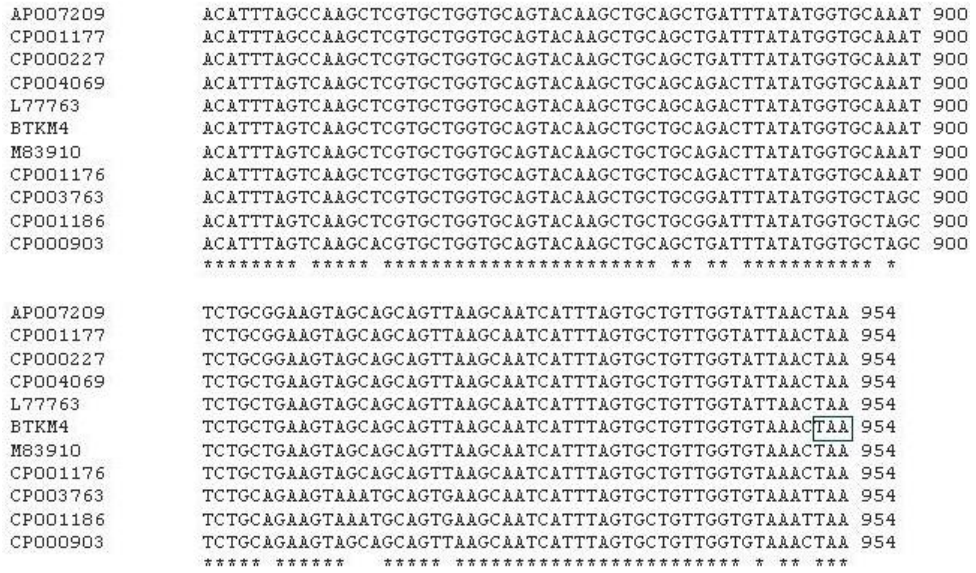

Fig. 4.20 Multiple sequence alignment of the partial protease gene sequences of strain BTKM4 with other similar sequences. Some of the consensus sequences are boxed (blue). The start codon and stop codon of the 888bp ORF are shown in red and green boxes respectively

The partial protease gene sequence of the strain BTKM4 is represented as BTKM4. Other sequences used in the alignment are represented with their GenBank accession numbers AP007209 (*Bacillus cereus* NC7401 genomic DNA, neutral protease), CP001177 (*Bacillus cereus* AH187, neutral protease), CP000227 (*Bacillus cereus* Q1, bacillolysin), CP004069 (*Bacillus thuringiensis* serovar *kurstaki* str. HD73, bacillolysin), L77763 (*Bacillus thuringiensis* neutral protease A (nprA) gene, complete cds), M83910 (*Bacillus cereus* neutral protease (NPRC) gene, complete cds), CP001176 (*Bacillus cereus* B4264, neutral protease), CP003763 (*Bacillus thuringiensis* HD-789, neutral protease), CP001186 (*Bacillus cereus* G9842, neutral protease) and CP000903 (*Bacillus weihenstephanensis* KBAB4, thermolysin).

4.4.3.2 Phylogenetic analysis of the protease gene of strain BTKM4

Unrooted tree was constructed based on neighbor joining method to determine the phylogenetic interrelationship of partial protease gene of BTKM4 and its ten close identities (Fig. 4.21). From the phylogenetic analysis it is clear that the partial protease gene of strain BTKM4 showed similarity towards the neutral protease gene of *Bacillus cereus* (M83910) as they formed a single cluster supported by a bootstrap value of 99%.

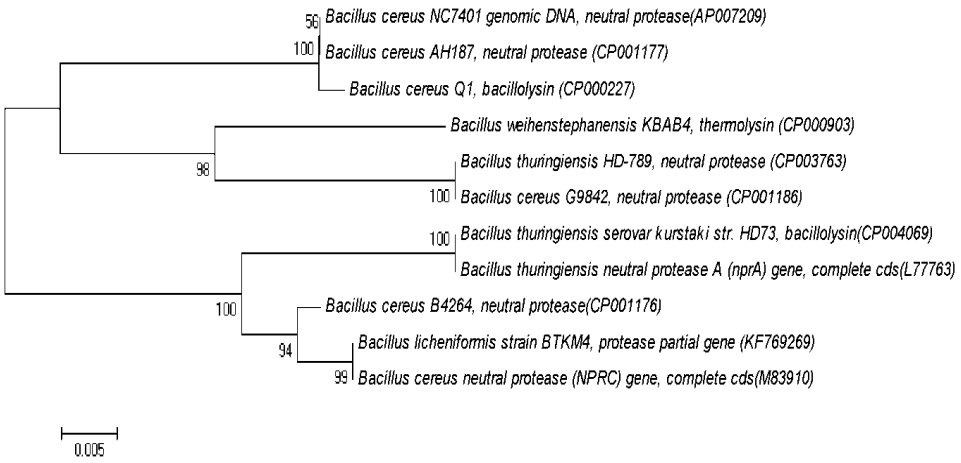

Fig. 4.21 The tree shows the phylogenetic interrelationships of protease gene nucleotide sequences of strain BTMK4 with ten other similar sequences. The Unrooted tree was constructed by the neighbor-joining method. Bootstrap values are given at the branching points and scale bar shows sequence divergence. Accession numbers are given in parentheses.

4.4.4 Characterisation of deduced amino acid sequence of partial protease gene of clone BTM106

4.4.4.1 BLAST analysis of deduced amino acid sequence of partial protease gene of clone BTM106

The 977 bp nucleic acid sequence of the partial protease gene of clone BTM106 was translated *in-silico* by Expasy (http://web.expasy.org/translate) into

its corresponding amino acid sequence. An orf consisting of 957 bp was identified and the deduced amino acid sequence of 324 amino acids, were compared with those available from GenBank using online BLAST tool- blastp (http://www.ncbi.nlm.nih.gov/blast). The GenBank accession numbers and description of ten hits with maximum identity after protein blast of deduced amino acid sequences of the protease gene of BTM106 is as detailed in Table 4.4.

Table 4.4. Protein blast results for the deduced amino acid sequence of partial protease gene of clone BTM106 showing the first ten hits with maximum identity and their GenBank accession numbers

Accession No.	Description	Identity
WP000730385	*Bacillus cereus* bacillolysin	95%
NP830419	*Bacillus cereus* ATCC 14579 bacillolysin	95%
WP000730386	*Bacillus* sp. bacillolysin	95%
WP000730380	*Bacillus cereus* bacillolysin	95%
WP000730381	*Bacillus cereus* bacillolysin	94%
WP000730376	*Bacillus cereus* bacillolysin	94%
WP000730400	*Bacillus thuringiensis* bacillolysin	94%
WP000730371	*Bacillus cereus* bacillolysin	94%
WP000730359	*Bacillus cereus* bacillolysin	94%
WP016089357	*Bacillus cereus* bacillolysin	94%

The amino acid sequence showed 95% identity with bacillolysin of *Bacillus cereus* (WP000730385) and *Bacillus cereus* ATCC 14579 (NP830419). The BLAST analysis confirmed that the amino acid sequence obtained was similar to that encoding the bacillolysin of *Bacillus cereus.*

4.4.4.2 Multiple sequence alignment of the deduced amino acid sequence of partial protease gene of clone BTM106

The genetic variation of the deduced amino acid sequence of the partial protease gene from the clone BTM106 was determined by its multiple sequence alignment with ten hits with maximum identity obtained by BLAST analysis. No

signal peptide was detected in the sequence. Residues that are identical among all six proteases are indicated with an asterisk. Residues that are conserved are indicated with a dot (Fig. 4.22).

```
WP000730381     KAVVKPLVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
WP000730400     KAVVKPLVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
WP000730376     KAVVKPLVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
BTM106          KAVVKPMVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
WP000730386     KAVVKPMVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
NP830419        KAVVKPMVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
WP000730385     KAVVKPMVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
WP000730380     KAVVKPMVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRTTLPG 60
WP000730371     KAVVKPLVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPG 60
WP000730359     KAVVKPLVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPG 60
WP016089357     KAVVKPLVTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPG 60
                ******:**************************************************:****

WP000730381     TLWVDADNVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730400     TLWVDADNVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730376     TLWVDADNVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
BTM106          TLWVDADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730386     TLWVDADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
NP830419        TLWVDADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730385     TLWVDADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730380     TLWVDADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730371     TLWADADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP000730359     TLWADADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
WP016089357     TLWADADNVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNN 120
                ***.**********************:**********************************

WP000730381     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
WP000730400     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIF 180
WP000730376     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
BTM106          AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLVYQNESGALNEAISDIF 180
WP000730386     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
NP830419        AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
WP000730385     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
WP000730380     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
WP000730371     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
WP000730359     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
WP016089357     AFWNGSQMVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSDLIYQNESGALNEAISDIF 180
                ****************************************:*:*****************

WP000730381     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
WP000730400     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
WP000730376     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
BTM106          GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
WP000730386     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
NP830419        GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
WP000730385     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
WP000730380     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGTSDNGGVHT 240
WP000730371     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGSSDNGGVHT 240
WP000730359     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGSSDNGGVHT 240
WP016089357     GTLVEFYDNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGSSDNGGVHT 240
                **************************************************:********
```

```
WP000730381     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730400     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730376     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
BTM106          NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730386     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
NP830419        NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730385     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730380     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730371     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP000730359     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
WP016089357     NSGIINKQAYLLANGGTHSGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAA 300
                ************************************************************

WP000730381     ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730400     ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730376     ADLYGATSAEVAAVKQSFSAVGIN 324
BTM106          ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730386     ADLYGATSAEVAAVKQSFSAVGIN 324
NP830419        ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730385     ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730380     ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730371     ADLYGATSAEVAAVKQSFSAVGIN 324
WP000730359     ADLYGATSAEVAAVKQSFSAVGVN 324
WP016089357     ADLYGANSAEVAAVKQSFSAVGIN 324
                ******.***************:*
```

Fig. 4.22 Multiple sequence alignment of the deduced amino acid sequence of the partial protease gene of BTM106 with other similar sequences.

The deduced amino acid sequence of the partial protease gene of BTM106 is represented as BTM106. Other amino acid sequences used in the alignment are represented by their accession numbers as WP000730381 (*Bacillus cereus* bacillolysin), WP000730400 (*Bacillus thuringiensis* bacillolysin), WP000730376 (*Bacillus cereus* bacillolysin), WP000730386 (*Bacillus* sp. bacillolysin), NP830419 (*Bacillus cereus* ATCC 14579 bacillolysin), WP000730385 (*Bacillus cereus* bacillolysin), WP000730380 (*Bacillus cereus* bacillolysin), WP000730371 (*Bacillus cereus* bacillolysin), WP000730359 (*Bacillus cereus* bacillolysin) and WP016089357(*Bacillus cereus* bacillolysin).

4.4.4.3 Phylogenetic analysis of deduced amino acid sequence of partial protease gene of clone BTM106

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

of the partial protease gene of BTM106 and its ten close identities (Fig. 4.23). From the phylogenetic analysis it is clear that the deduced amino acid sequence of partial protease gene of BTM106 showed similarity to bacillolysin from *Bacillus cereus* as they clustered together to form a major clade. This major clade showed divergence from the amino acid sequence of *Bacillus cereus* neutral protease.

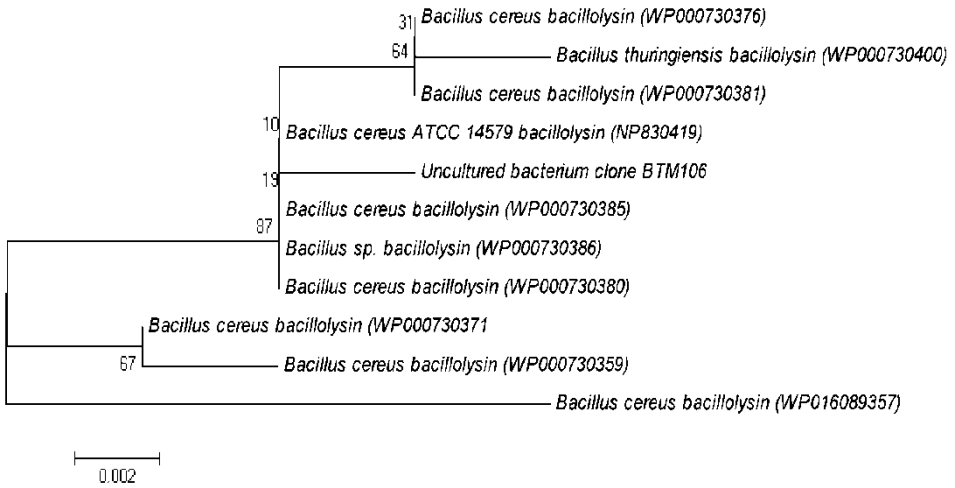

Fig. 4.23 The phylogenetic interrelationships of protease gene amino acid sequences of clone BTM106 with ten other similar sequences. The unrooted tree was constructed by the neighbor-joining method. Bootstrap values are given at the branching points and scale bar shows sequence divergence. Accession numbers are given in parentheses

4.4.5 Characterisation of amino acid sequence of partial protease gene of strain BTKM4

4.4.5.1 BLAST analysis of deduced amino acid sequence of partial protease gene of strain BTKM4

The 954 bp nucleic acid sequence of the partial protease gene of strain BTKM4 was translated *in-silico* by Expasy (http://web.expasy.org/translate) into its corresponding amino acid sequence. An orf consisting of 888bp was identified

and the deduced amino acid sequence of 317 amino acids sequence was compared with those available from GenBank using online BLAST tool- blastp (http://www.ncbi.nlm.nih.gov/blast). The GenBank accession numbers and description of ten hits with maximum identity after protein blast of deduced amino acid sequences of the protease gene of BTKM4 is as detailed in Table 4.5.

Table 4.5. Protein blast results for the deduced amino acid sequence of protease gene of strain BTKM4, showing the first ten hits with maximum identity and their GenBank accession numbers

Accession No.	Description	Identity
1NPCA	*Bacillus cereus* neutral protease	99%
1ESPA	Neutral protease mutant E144s	99%
WP000730369	*Bacillus cereus* bacillolysin	99%
YP002365398	*Bacillus cereus* B4264 neutral protease Npr599	99%
YP002336730	*Bacillus cereus* AH187 neutral protease Npr599	99%
WP016716433	*Bacillus cereus* bacillolysin	99%
YP006595138	*Bacillus cereus* FRI-35 neutral protease Npr599	99%
WP002194371	*Bacillus cereus* bacillolysin	99%
WP016112136	*Bacillus cereus* bacillolysin	99%
WP016083571	*Bacillus cereus* bacillolysin	99%

The amino acid sequence showed 99% identity with *Bacillus cereus* neutral protease (1NPCA) and Neutral protease mutant E144s (1ESPA). The BLAST analysis confirmed that the amino acid sequence obtained was that encoding the neutral protease of *Bacillus cereus.*

4.4.5.2 Multiple sequence alignment of the deduced amino acid sequence of partial protease gene of strain BTKM4

The genetic variation of the deduced amino acid sequence of the partial protease gene from the strain BTKM4 was determined by its multiple sequence alignment with ten hits with maximum identity obtained by BLAST analysis. No

signal peptide was detected in the sequence. Residues that are identical among all six proteases are indicated with an asterisk. Residues that are conserved are indicated with a dot (Fig. 4.24).

```
BTKM4           VTGTNKVGTGKGVLGDTKSLNTMLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
YP002365398     VTGTNKVGTGKGVLGDTKSLNTMLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
1ESPA           VTGTNKVGTGKGVLGDTKSLNTMLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
1NPCA           VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
WP000730369     VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
WP016112136     VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
YP006595138     VTGTNKVGTGKGVLGDTKSLNTMLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
WP016716433     VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
WP002194371     VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
YP002336730     VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADAD 60
WP016083571     VTGTNKVGTGKGVLGDTKSLNTTLSGSSYYLQDNTRGATIFTYDAKNRSTLPGTLWADVD 60
                ********************* ************************************.* 

BTKM4           NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
YP002365398     NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
1ESPA           NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
1NPCA           NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
WP000730369     NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
WP016112136     NVFNAAYDAAAVDAHYYAGITYDYYKNTFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
YP006595138     NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
WP016716433     NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
WP002194371     NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
YP002336730     NVFNAAYDAAAVDAHYYAGKTYDYYKATFNRNSINDAGAPLKSTVHYGSNYNNAFWNGSQ 120
WP016083571     NVFNAAYDAAAVDAHYYAGRTYDYYKATFNRNSINDAGAPLKSTVHYGSKYNNAFWNGSQ 120
                ******************* ****** ***********************:********** 

BTKM4           MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
YP002365398     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
1ESPA           MVYGDGDGVTFTSLSGGIDVIGHSLTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
1NPCA           MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
WP000730369     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
WP016112136     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
YP006595138     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
WP016716433     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
WP002194371     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
YP002336730     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
WP016083571     MVYGDGDGVTFTSLSGGIDVIGHELTHAVTENSSNLIYQNESGALNEAISDIFGTLVEFY 180
                ***********************.************************************ 

BTKM4           DNRNPDWEIGEDIYTPGKAGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
YP002365398     DNRNPDWEIGEDIYTPGKAGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
1ESPA           DNRNPDWEIGEDIYTPGKAGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
1NPCA           DNRNPDWEIGEDIYTPGKAGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
WP000730369     DNRNPDWEIGEDIYTPGKAGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
WP016112136     DNRNPDWEIGEDIYTPGKAGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
YP006595138     DNRNPDWEIGEDIYTPGKPGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
WP016716433     DNRNPDWEIGEDIYTPGKPGDALRSMSDPTKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
WP002194371     DNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
YP002336730     DNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
WP016083571     DNRNPDWEIGEDIYTPGKAGDALRSMSDPAKYGDPDHYSKRYTGSSDNGGVHTNSGIINK 240
                ******************.**********:****************************** 
```

```
BTKM4           QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
YP002365398     QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
1ESPA           QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
1NPCA           QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
WP000730369     QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
WP016112136     QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
YP006595138     QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYYTQSTTFSQARAGAVQAAADLYGAN 300
WP016716433     QAYLLANGGTHYGVTVNGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
WP002194371     QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYYTQSTTFSQARAGAVQAAADLYGAN 300
YP002336730     QAYLLANGGTHYGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
WP016083571     QAYLLANGGTHFGVTVTGIGKDKLGAIYYRANTQYFTQSTTFSQARAGAVQAAADLYGAN 300
                ***********:****.******************:************************

BTKM4           SAEVAAVKQSFSAVGVN 317
YP002365398     SAEVAAVKQSFSAVGVN 317
1ESPA           SAEVAAVKQSFSAVGVN 317
1NPCA           SAEVAAVKQSFSAVGVN 317
WP000730369     SAEVAAVKQSFSAVGVN 317
WP016112136     SAEVAAVKQSFSAVGVN 317
YP006595138     SAEVAAVKQSFSAVGIN 317
WP016716433     SAEVAAVKQSFSAVGIN 317
WP002194371     SAEVAAVKQSFSAVGIN 317
YP002336730     SAEVAAVKQSFSAVGIN 317
WP016083571     SAEVAAVKQSFSAVGIN 317
                ***************:*
```

Fig. 4.24 Multiple sequence alignment of the deduced amino acid sequence of the partial protease gene of BTKM4 with other similar sequences.

The deduced amino acid sequence of the partial protease gene of strain BTKM4 is denoted as BTKM4. Other sequences used for alignment are represented by their Genbank accession numbers YP002365398 (*Bacillus cereus* B4264 neutral protease Npr599), 1ESPA (Neutral protease mutant E144s), 1NPCA (*Bacillus cereus* neutral protease), WP000730369 (*Bacillus cereus* bacillolysin), WP016112136 (*Bacillus cereus* bacillolysin), YP006595138 (*Bacillus cereus* FRI-35 neutral protease Npr599), WP016716433 (*Bacillus cereus* bacillolysin), WP002194371 (*Bacillus cereus* bacillolysin), YP002336730 (*Bacillus cereus* AH187 neutral protease Npr599) and WP016083571(*Bacillus cereus* bacillolysin).

4.4.5.3 Phylogenetic analysis of deduced amino acid sequence of partial protease gene of strain BTKM4

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

of the protease gene of BTKM4 and its ten close identities (Fig. 4.25). From the phylogenetic analysis it is clear that the deduced amino acid sequence of protease genes of BTKM4 showed similarity to neutral protease from *Bacillus cereus* as they clustered together to form a major clade.

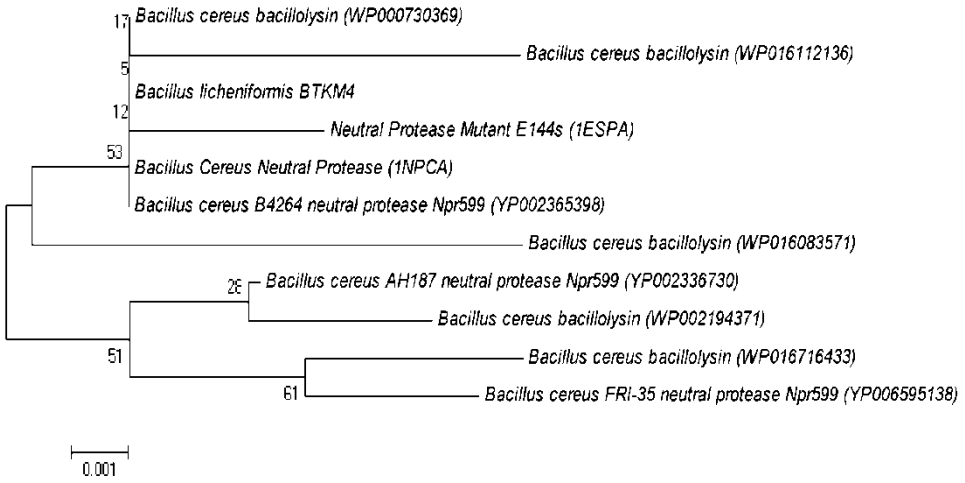

Fig. 4.25 The tree showing the phylogenetic interrelationships of protease gene amino acid sequences of strain BTKM4 with ten other similar sequences. The unrooted tree was constructed by the neighbor-joining method. Bootstrap values are given at the branching points and scale bar shows sequence divergence. Accession numbers are given in parentheses.

4.4.6 Prediction of active site of protease of BTM106 and BTKM4

The active site of the two proteases were elucidated using the conserved domain database of NCBI, based on which the proteases were classified to the MEROPS database. The active site of BTM106 protease was predicted using the deduced 324 amino acid stretch. It was noted that the active site stretched from amino acid 120 to 240, wherein stretching from amino acid 150 to 175, there was a Zinc binding site (Fig. 4.26).

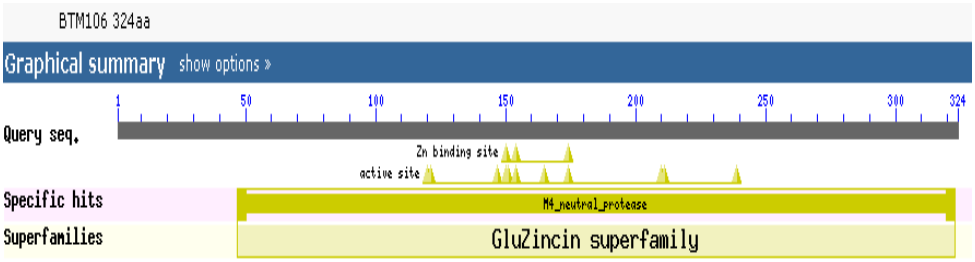

Fig. 4.26 Active site of BTM106 protease

The active site of the BTKM4 protease was also predicted similarly using the deduced 317 amino acids stretch. The active site of the protease of BTKM4 stretched from amino acid 110 to 230, within which there was a Zinc binding site, stretching from amino acid 140 to 170 (Fig. 4.27).

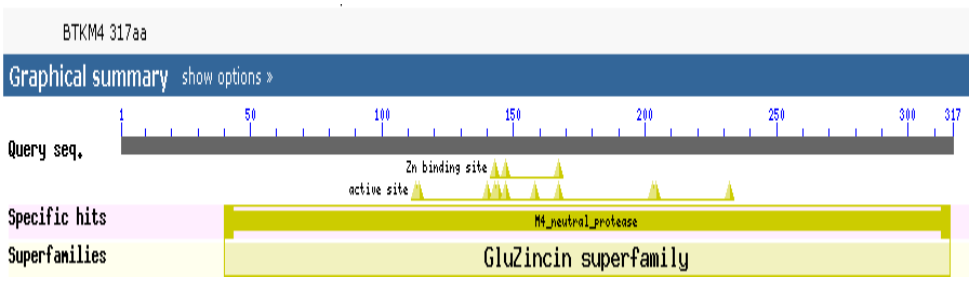

Fig. 4.27 Active site of BTKM4 protease

The two proteases exhibited similar conserved domain architecture, belonging to the M4 neutral protease family of Gluzincin superfamily, which is characteristic of Zinc metalloproteases. Gluzincin superfamily includes several zinc-dependent metallopeptidases such as the M1, M2, M3, M4, M13, M32, M36 peptidases (based on MEROPS classification), and contain two conserved amino acid stretches, represented by their single letter codes as HEXXH and EXXXD (H-Histidine, E-Glutamic acid, D-Aspartic acid and X is any non-conserved amino acid) as part of their active site. All peptidases in this family bind a single catalytic zinc ion which is tetrahedrally co-ordinated by three amino acid ligands, and a water molecule that forms the nucleophile on activation during catalysis. These

enzymes have a two-domain structure with the active site between these two domains. The N-terminal domain contains the HEXXH zinc-binding motif while the helical C-terminal domain, which is unique for the family, carries the third zinc ligand.

4.4.7 Elucidation of conserved motif in the active site of the proteases of BTM106 and BTKM4

The multiple sequence alignment of the 58 amino acid stretch of the active site of the proteases of BTM106 and BTKM4 with others in the database was used for elucidation of the conserved motifs in the active site of the two protease genes**.** The active site of M4 proteases carried two conserved motifs, both being a five-mer, represented by the single letter code amino acid stretches HEXXH and EXXXD (H-Histidine, E-Glutamic acid, D-Aspartic acid and X is any non-conserved amino acid) . These two motifs are observed to be present in the deduced amino acid sequence of both BTM106 and BTKM4 proteases. However, the second motif EXXXD is present as a longer nine-mer GXXNEXXSD (G-Glycine, N-Asparagine, E-Glutamic acid, S-Serine, D-Aspartic acid, and X is any non-conserved amino acid) conserved stretch, instead of being a five-mer stretch. The amino acid sequence of BTM106 and BTKM4 is aligned with the amino acid sequence of other proteases with close similarity belonging to the MEROPS M4 family (Fig. 4.28).

Fig. 4.28 Elucidation of the conserved motifs HEXXH and GXXNEXXSD in the deduced amino acid sequence of BTM106 and BTKM4.

Other M4 proteases used in the alignment are represented by their accession numbers YP003684029 (M4 thermolysin from *Meiothermus silvanus*), ZP04105882 (extracellular protease from *Bacillus thuringiensis*), ZP01132256 (alkaline serine protease from *Pseudoalteromonas tunicata*), ACA34420 (MetalloproteaseA from uncultured bacterium pTW2) and ACA34431 (MetalloproteaseB from uncultured bacterium pTW3).

4.4.8 Phylogenetic analysis of the proteases of BTM106 and BTKM4

The phylogenetic relationship of BTM106 and BTKM4 based on the conserved domains of the protease gene sequence was determined from the unrooted tree constructed based on neighbor joining method. It was clear that the protease gene of BTM106 and BTKM4 are closely related to each other and show identity to M4 family of proteases, which are metalloproteases (Fig. 4.29).

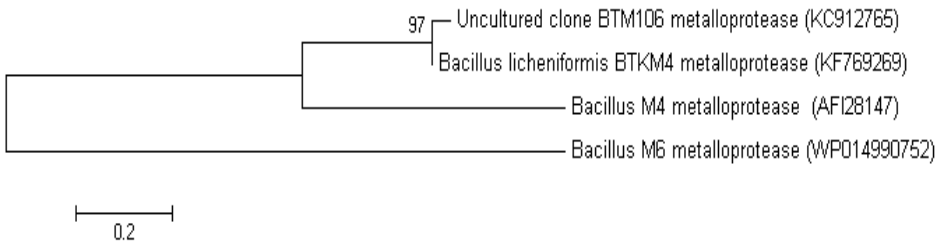

Fig. 4.29 Phylogenetic relationship of protease gene of BTM106 and BTKM4 based on conserved domains in gene sequences

Bootstrap values are given at the branching points and scale bar shows sequence divergence. *Bacillus* M6 metalloprotease was used as the out group. The accession numbers are given in parentheses.

4.4.9 Structure prediction of BTM106 protease

The tertiary structure of the BTM106 protease was predicted using **Phyre**[2] software, from the deduced amino acid sequence (Fig. 4.30). The residues were

modeled by the single highest scoring template and the model was based on the template d1npca (Fold library id) of PDB database. The template had Zincin like fold of the catalytic domain of metalloprotease super family. Out of the deduced 324 amino acid sequence, 317 residues (98% of the sequence) were successfully modeled with 100% confidence on the template d1npca.

Fig. 4.30 Template based homology modeling of BTM106 protease.

The dimensions of the model are given in Angstroms

The secondary structure was also predicted based on the template d1npca (Fig. 4.31). The amino acid sequence was aligned with the predicted secondary structures. The predicted structure consisted of 31% α-helix and 15% β-strand and the model was predicted with 11% disorder.

Fig. 4.31 Secondary structures in BTM106 protease aligned with its amino acid sequence

4.4.10 Structure prediction of BTKM4 protease

Phyre[2] software also helped to successfully model the tertiary structure of BTKM4 protease using the deduced amino acid sequence (Fig. 4.32). The residues

were modeled by the single highest scoring template and the model was also based on the template d1npca (Library id) of PDB database. The template had Zincin like fold of the catalytic domain of metalloprotease super family. 317 residues of the deduced amino acid sequence have been successfully modeled with 100% confidence and 100% coverage.

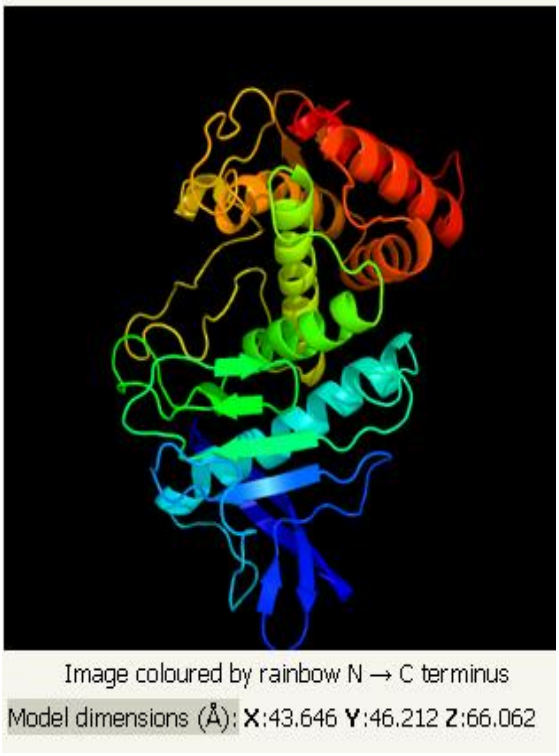

Fig. 4.32 Template based homology modeling of BTKM4 protease.

The dimensions of the model are given in Angstroms

The secondary structure of BTKM4 was predicted based on the template d1npca. It consisted of 34% α-helix and 16% β-strand and the model was predicted with 11% disorder. The amino acid sequence was aligned with the predicted secondary structures as depicted in Fig. 4.33.

Fig. 4.33 Secondary structures in BTKM4 protease aligned with its amino acid sequence

4.5 CHARACTERIZATION OF THE PROTEASE ENZYMES OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

The protease enzyme obtained from clone BTM106 was designated as P106 and the protease obtained from strain BTKM4 was designated as P4.

4.5.1 Extraction of protease P106 by clone BTM106

The cell free supernatant of clone BTM106 in culture medium was assayed for proteolytic activity and partial purification was done by ammonium sulphate precipitation. The ammonium sulphate precipitate of the cell free supernatant of clone BTM106 as well as that of the control (*E.coli* DH5α transformed with pUC18) was electophoresised and the protein profile was analysed on SDS-PAGE as in Fig. 4.34.

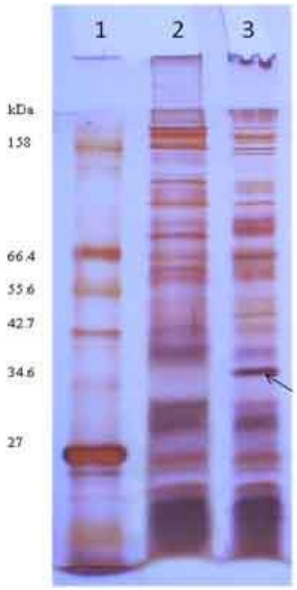

Fig. 4.34 SDS- PAGE showing protein profile of control and clone BTM106

Lane 1- Protein marker (NEB), Lane 2- Protein fraction of control, Lane 3- Protein fraction of clone BTM106. The molecular weight of the different bands of the protein marker is indicated along the side of the gel. The arrow indicates the additional protein band.

The protein profile of clone BTM106 showed the presence of an additional protein band corresponding to ~34kDa which was not found in the protein profile of the control, *E.coli* DH5α transformed with pUC18. It is suggested that this may be the band corresponding to the protease.

4.5.2 SDS PAGE of P106 and P4

Ammonium sulphate precipitation of the cell free supernatant of BTKM4 was also done and compared with that of BTM106. For both proteases, 30-60% ammonium sulphate saturated protein fraction exhibited protease activity. The active fractions were subjected to SDS-PAGE and visualized by silver staining technique (Fig. 4.35).

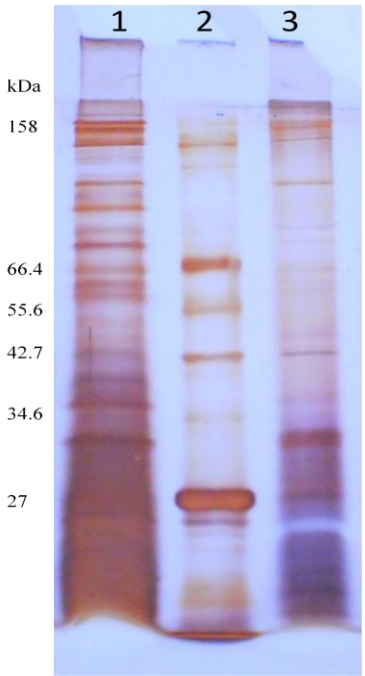

Fig. 4.35 SDS-PAGE of ammonium sulphate fractions (30-60%) of P106 and P4

Lane 1 - P106, Lane 2 - Marker (NEB), Lane 3 - P4 The molecular weight of the different bands of the protein marker is indicated along the side of the gel.

The protein profile of the active fraction was analysed and seemed to have many proteins which were indicated by numerous bands on the gel.

4.5.3 Zymogram

Following ammonium sulfate fractionation, gel filtration was carried out for further purification of the two enzyme proteins. The active fraction obtained after gel filtration chromatography was analysed by SDS PAGE and the protein band responsible for activity showed clearing on zymogram (Fig. 4.36) on gelatin gel.

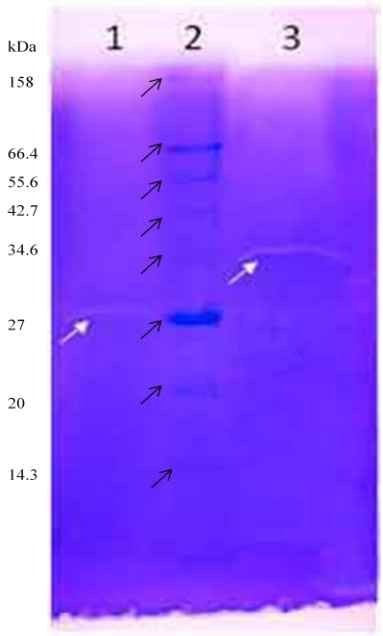

Fig. 4.36 Zymogram of P4 (BTKM4) and P106 (BTM106)

Lane 1 - P4, Lane 2 - Marker (NEB), Lane 3 - P106 The white arrows represent the clearing of protease bands and black arrows denote the bands of the protein marker. The molecular weight of the different bands of the protein marker is indicated along the side of the gel.

The approximate size of the purified proteases were analysed by Gelatin-PAGE. There was only a single clearing zone visible on the gelatin gel indicating a single type of extracellular protease in each case. Purified P106 showed clearing of a band corresponding to ~34 kDa and the purified protease P4 showed clearing of a band corresponding to ~29 kDa when compared with the electrophoretic mobility of protein marker (NEB).

4.5.4 Fold of purification of proteases

The proteases were purified employing standard protein purification steps which involved ammonium sulphate precipitation followed by dialysis and gel filtration chromatography.

4.5.4.1 Fold of purification of P106

The fold of purification of P106 after each step of purification is as summarized in Table 4.6.

Table. 4.6 Fold of purification of P106

Purification step	Volume (mL)	Activity (U/mL)	Protein conc. (mg/mL)	Specific activity (U/mg)	Fold of purification
Crude extract	100	249	3.26	76.38	1*
Ammonium sulphate (30-60%)	2	3822	13.5	283.11	3.7
Gel filtration	0.5	525	0.64	820.31	10.74

*Value taken arbitrarily

Ammonium sulphate required to precipitate the protease from the crude extract was standardized. 30-60% ammonium sulphate saturation precipitated protease P106 with 3.7 fold increase in specific activity compared with that of the crude extract. This fraction upon further purification by gel filtration chromatography yielded pure protease with 10.74 fold increase in specific activity.

4.5.4.2 Fold of purification of P4

The fold of purification of P4 at each step of purification is summarized in Table 4.7.

Table. 4.7 Fold of purification of P4

Purification step	Volume (mL)	Activity (U/mL)	Protein conc. (mg/mL)	Specific activity U/mg	Fold of purification
Crude extract	100	40.33	2.74	14.72	1*
Ammonium sulphate (30-60%)	2	420	11.5	36.52	2.48
Gel filtration	0.5	86.66	0.6	144.43	9.81

*Value taken arbitrarily

Ammonium sulphate required to precipitate the protease from its crude extract was standardized. 30-60% ammonium sulphate saturation precipitated protease, giving a 2.48 fold increase in specific activity of P4 compared with that of the crude extract. This fraction upon further purification by gel filtration chromatography yielded pure protease with 9.81 fold increase in specific activity.

4.5.5 PHYSICOCHEMICAL CHARACTERIZATION OF THE ENZYMES

4.5.5.1 Determination of isoelectric point of P106 and P4

The isoelectric point (pI) of the two proteases P106 and P4 was determined by isoelectric focusing (Fig. 4.37).

(a) (b)

Fig. 4.37 IPG strip showing isoelectric point of (a) P106 and (b) P4

The isoelectric points of proteases P106 and P4 was determined using the IPG strip and was found to be 8.5 and 8.0 respectively.

4.5.5.2 Determination of optimum pH for enzyme activity of P106 and P4

The optimum pH for maximum enzyme activity of proteases P106 and P4 were determined by studying enzyme activity at different pH ranging from 2 to 13 (Fig. 4.38).

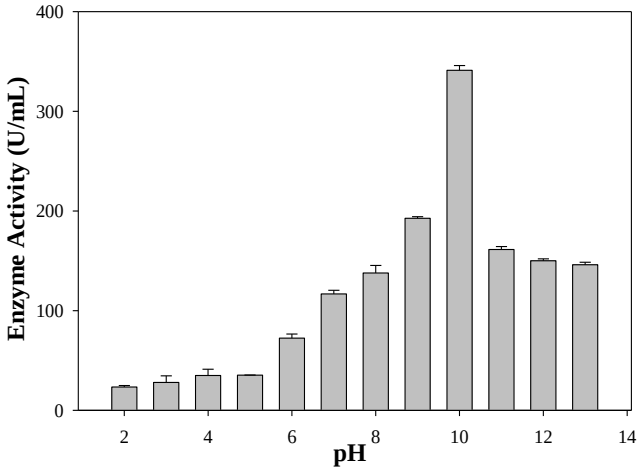

Fig. 4.38(a) Effect of pH on enzyme activity of P106

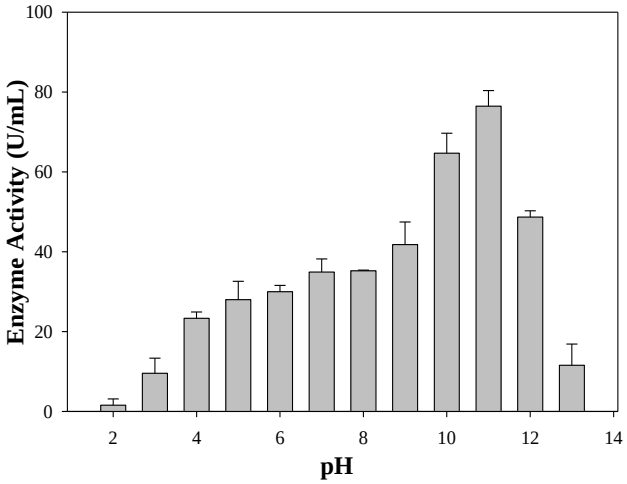

Fig. 4.38(b) Effect of pH on enzyme activity of P4

It is observed that for both P106 and P4 the enzyme activity increases as the pH increases towards the alkaline, peaking at pH 10 and pH 11 respectively, and declining thereafter. The optimum pH for maximum activity of protease P106 was at pH 10, while that of protease P4 was pH 11. At pH 10, protease P106 exhibited activity of 341 ± 4 U/mL. At pH 11, protease P4 showed its maximum activity of 76 ± 6 U/mL. From the results it is clear that both the proteases showed their highest activity at alkaline pH and therefore they belong to the group of alkaline proteases. Also, it could be stated that the protease P4 is more active at a higher alkaline pH than the protease P106.

However, it may be noted that P106 showed higher activity than P4 at all acidic and the alkaline pH tested.

4.5.5.3 Determination of pH stability of protease enzymes

The stability of the two proteases P106 and P4 at different pH conditions ranging from 2 to 13 was also tested. The relative activity was determined by comparing the activity of the enzymes at different pH with the sample having maximum activity (Fig. 4.39).

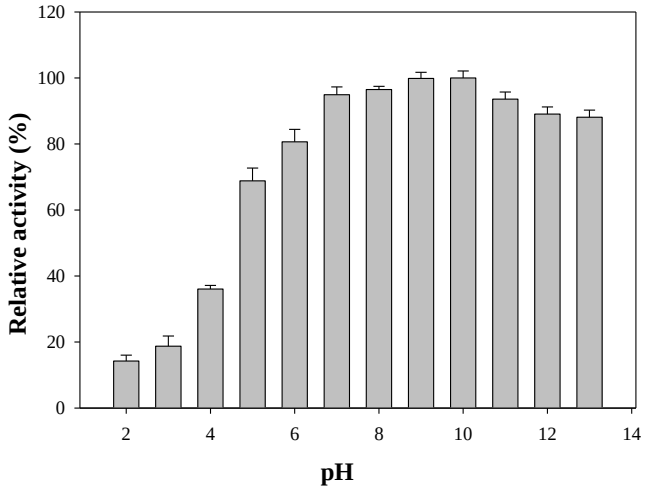

Fig. 4.39(a) Stability of P106 at different pH

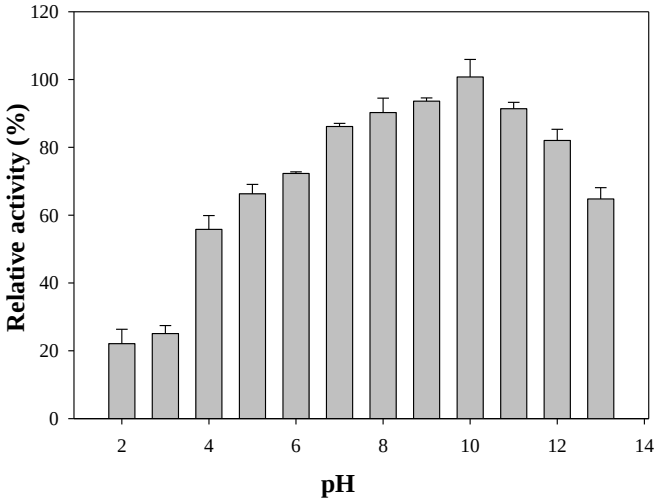

Fig. 4.39(b) Stability of P4 at different pH

Both proteases were stable over a wide pH range, but primarily more stable in the alkaline range. The protease P106 showed maximum stability at pH

10, with more than 80% of relative activity in the pH ranging from 6 to 13. The protease P4 also showed more stability at pH 10 and it demonstrated more than 80% of residual activity in a pH ranging from pH 7 to 12. These results clearly indicate that the protease P106 has more stability than the protease P4 over a wide range of pH.

4.5.5.4 Determination of optimum temperature for protease activity of P106 and P4

The optimum temperature for maximum activity of the proteases was determined by calculating the enzyme activity over temperatures ranging from 30ºC to 100ºC (Fig. 4.40).

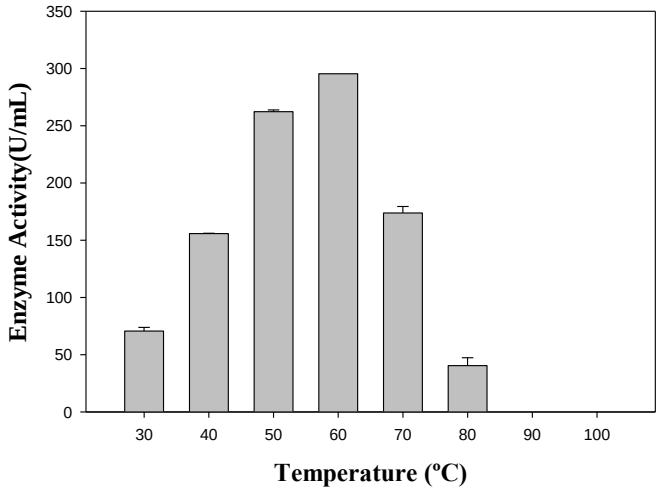

Fig. 4.40(a) Effect of temperature on enzyme activity of P106

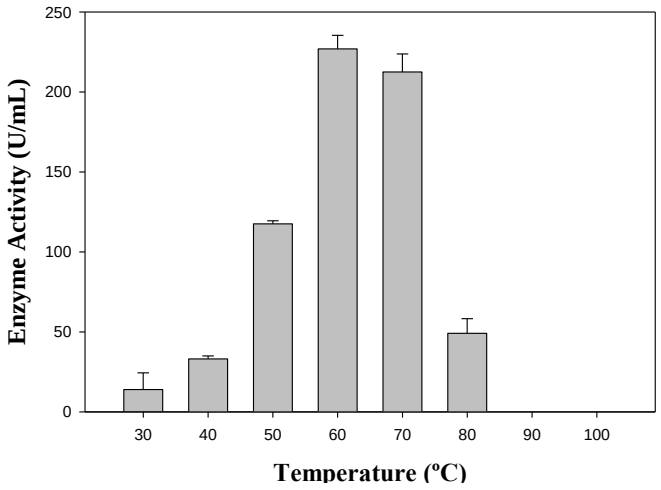

Fig. 4.40(b) Effect of temperature on enzyme activity of P4

Both proteases were active over a broad range of temperature, with maximum enzyme activity at 60ºC. The maximum activity of protease P106 at that temperature was 295 ± 0.2 U/mL while that of protease P4 was 226 ± 0.9 U/mL. There was no observable activity for both proteases when they were incubated at temperatures beyond 80ºC.

4.5.5.5 Determination of temperature stability of protease enzymes

The stability of the proteases at different temperatures ranging from 30ºC to 100ºC was also determined. The relative activity was calculated by comparing with the sample with maximum activity (Fig. 4.41).

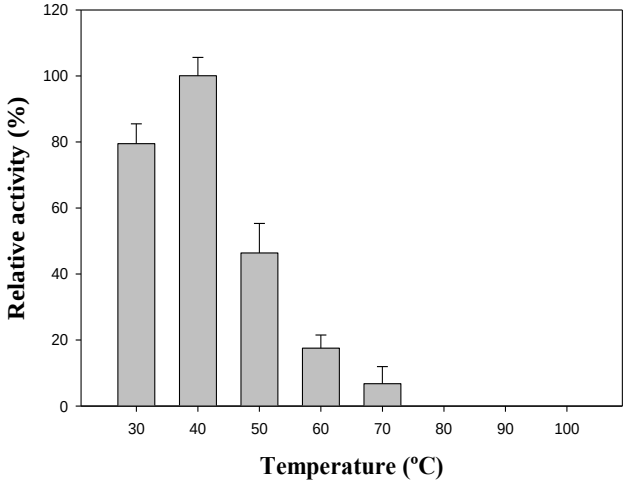

Fig. 4.41(a) Stability of P106 at different temperature

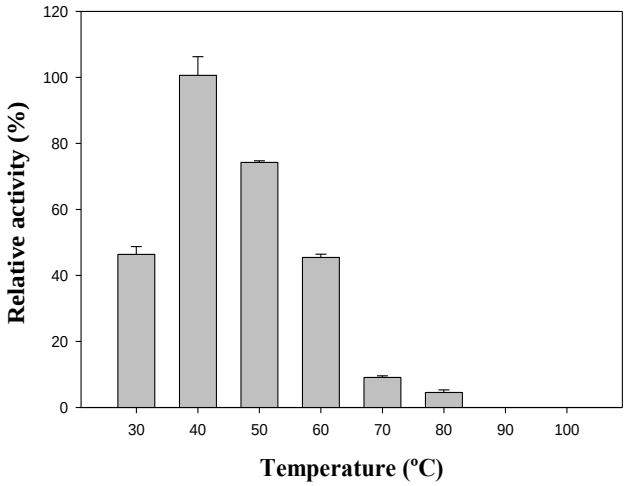

Fig. 4.41(b) Stability of P4 at different temperature

At 40ºC, both the proteases showed maximum stability. The protease P106 exhibited more than 70% of the relative activity at 30ºC and 40ºC whereas

the protease P4 showed more than 70% of its relative activity at 40ºC and 50ºC. The protease P106 was stable upto 70ºC with relative activity lowered to 6%. The enzyme P4 was stable upto 80ºC with relative activity reduced to 4%.

4.5.5.6 Effect of inhibitors on enzyme activity of P106 and P4

The effect of four different protease inhibitors, aprotinin (1mM), EDTA (10mM), Iodoacetamide (10mM) and PMSF (10mM) on the enzyme activity of the two proteases was determined. The residual activity was estimated by comparing the enzyme activity with that of control (Fig. 4.42).

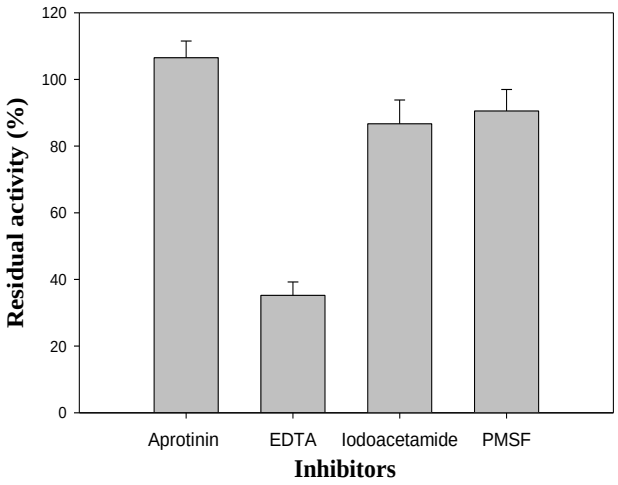

Fig. 4.42(a) Effect of inhibitors on enzyme activity of P106

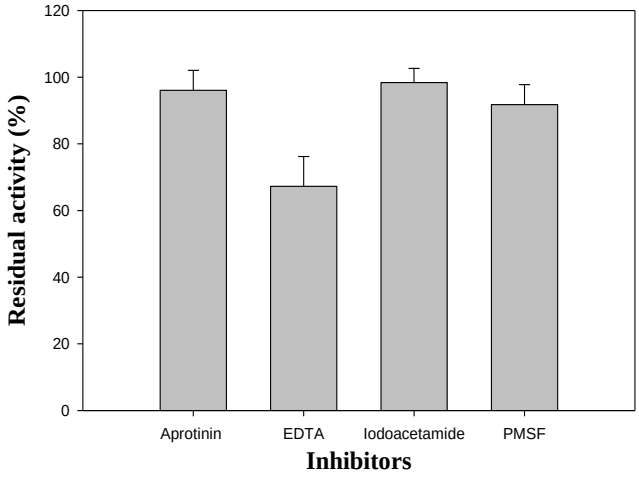

Fig. 4.42(b) Effect of inhibitors on enzyme activity of P4

Among the four different inhibitors tested, EDTA inhibited the activity of both proteases. The enzyme activity of protease P106 was inhibited to 35% residual activity with 10 mM EDTA whereas the activity of P4 was inhibited to 67%. Aprotinin and PMSF did not inhibit the activity of either protease to less than 90% and hence seemed to have no significant effect on protease activity. However, iodoacetamide inhibited the activity of P106 to 86%. These observations clearly indicate that the proteases may be metalloproteases, as their proteolytic activity could be inhibited by the metal chelating effect of EDTA.

4.5.5.7 Determination of substrate specificity of P106 and P4

The ability of the proteases to hydrolyse different proteinaceous substrates (1%) like bovine serum albumin (BSA), casein, gelatin and haemoglobin was determined by calculating the enzyme activity (Fig. 4.43). Both proteases, P106 and P4, demonstrated more specificity towards casein as a substrate followed by haemoglobin. The protease P106 had least preference for gelatin as substrate whereas BSA was the least preferred substrate for the protease P4.

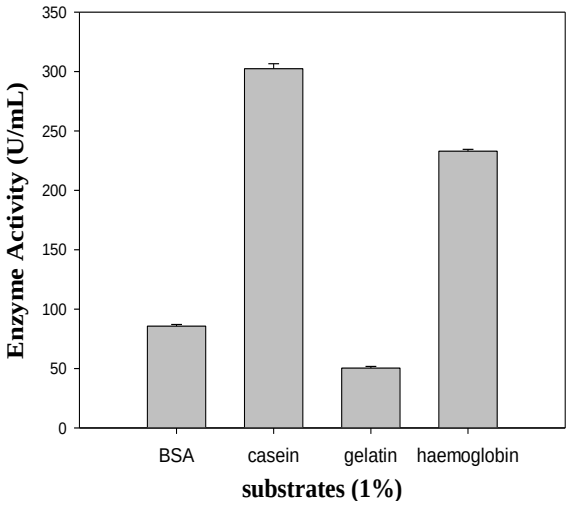

Fig. 4.43(a) Determination of substrate specificity of P106

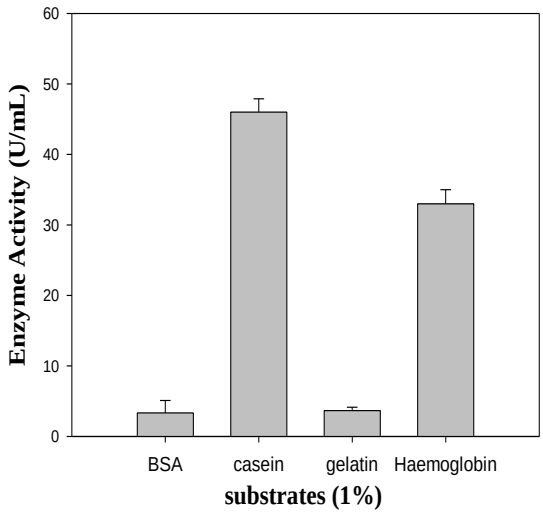

Fig. 4.43(b) Determination of substrate specificity of P4

4.5.5.8 Determination of kinetic parameters - Km and Vmax

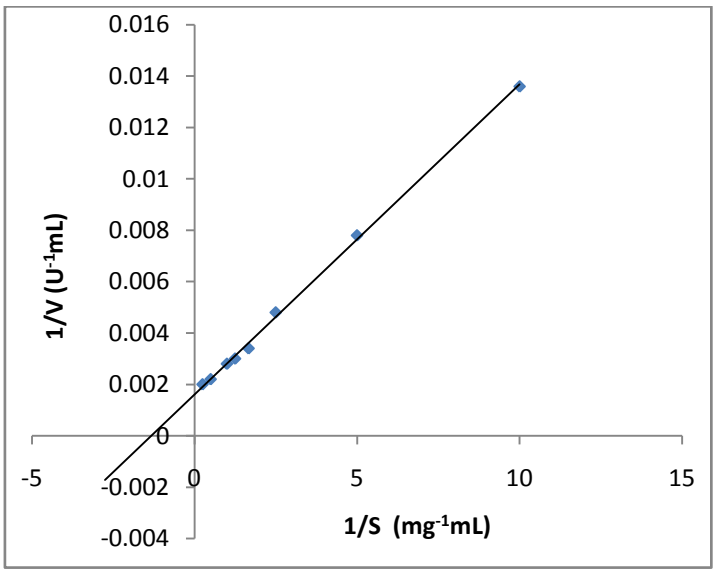

Fig. 4.44(a) Determination of kinetic parameters- Km and Vmax of P106

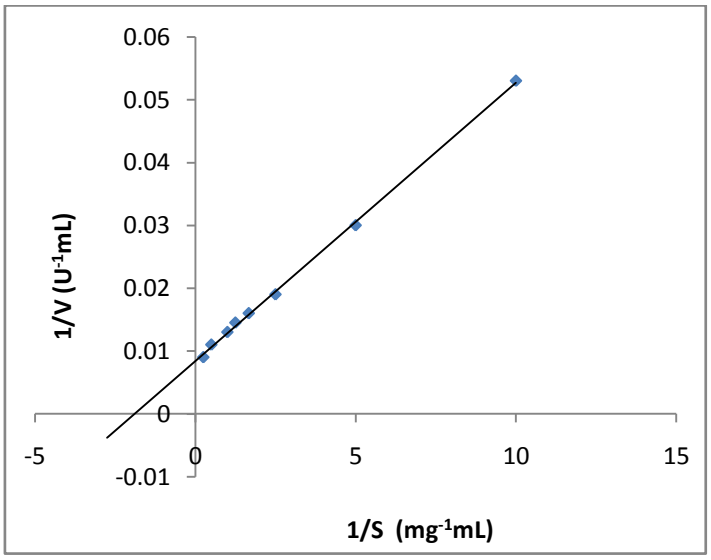

Fig. 4.44(b) Determination of kinetic parameters- Km and Vmax of P4

The kinetic parameters of the proteases were 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 the Km and Vmax of protease P106 were determined as 0.66 mg/mL and 625 U/mL respectively and that of protease P4 was 0.53mg/mL and 122U/mL respectively (Fig. 4.44).

4.5.5.9 Effect of various metal ions on enzyme activity of P106 and P4

The effect of twelve different metal ions on protease activity was determined and the residual activity was estimated (Fig. 4.45, 4.46).

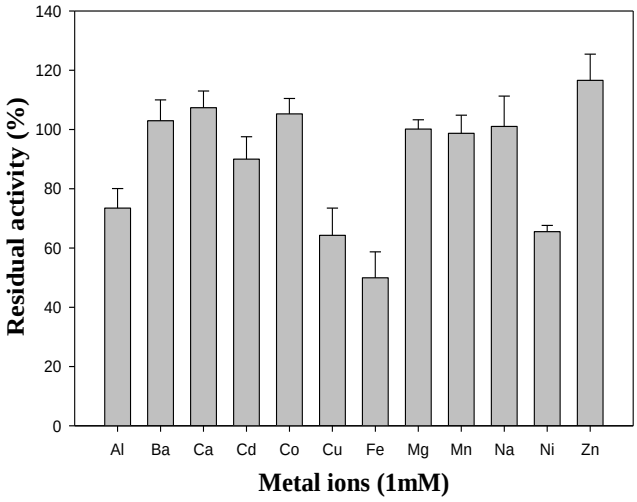

Fig. 4.45 Effect of various metal ions on enzyme activity of P106

Out of the twelve metal ions tested, Zn^{2+} enhanced protease activity of P106, increasing it to 116% residual activity. Ca^{2+} and Co^{2+} also enhanced the activity, but only marginally. However Ba^{2+}, Na^{+} and Mg^{2+} did not have any significant effect on activity of P106. On the other hand, in the presence of Mn^{2+} and Cd^{2+}, the residual activity was slightly reduced, whereas Ni^{2+}, Cu^{2+} and Al^{3+} significantly lowered the enzyme activity. Fe^{3+} inhibited the activity to 49% residual activity.

In the case of enzyme P4, Co^{2+} enhanced the proteolytic activity resulting in residual activity increasing to 151%. This activity enhancement was also observed in the presence of metal ions Mg^{2+}, Na^{+}, Ni^{2+}, and Zn^{2+}. Mn^{2+} enhanced the activity marginally, whereas Ba^{2+} and Ca^{2+} reduced the activity slightly. Al^{3+}, Cd^{2+} and Fe^{3+}significantly reduced the enzyme activity whereas Cu^{2+} reduced the activity to 43%.

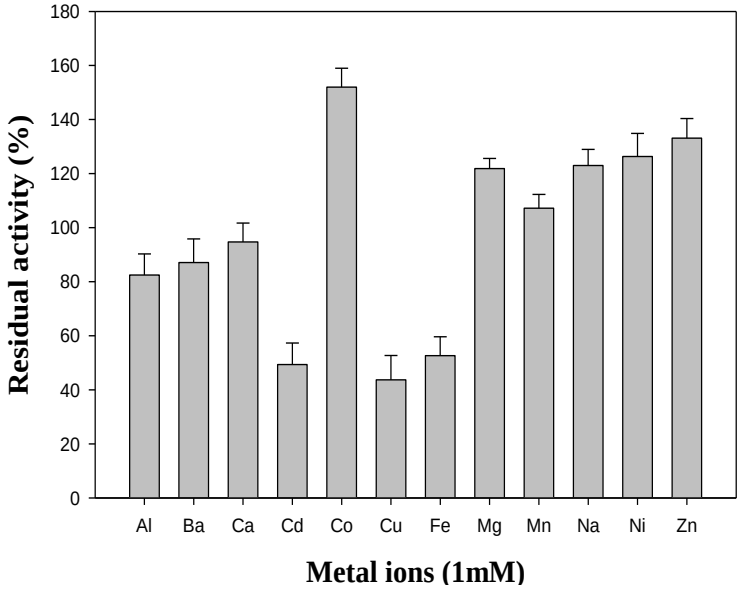

Fig. 4.46 Effect of various metal ions on enzyme activity of P4

4.5.5.10 Effect of various detergents on enzyme activity of P106 and P4

The effect of various detergents like Tween 20, Tween 80, Triton X-100, CTAB and SDS (1%) was determined and their effect on the proteolytic activity of the enzymes is represented in Fig. 4.47.

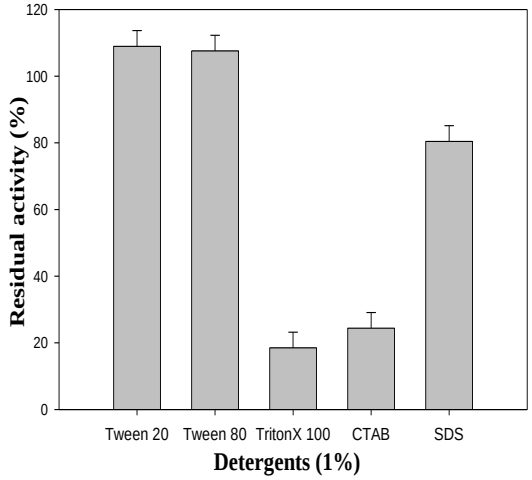

Fig. 4.47(a) Effect of various detergents on enzyme activity of P106

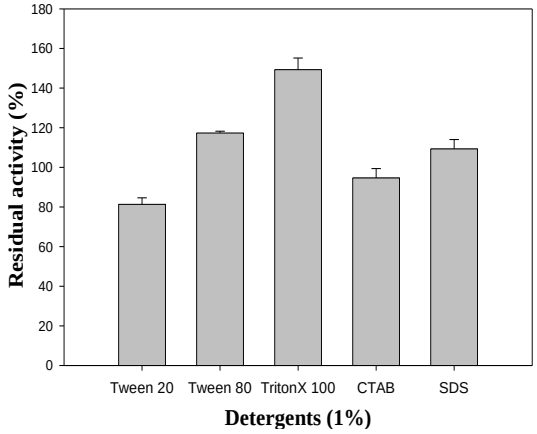

Fig. 4.47(b) Effect of various detergents on enzyme activity of P4

The detergents Tween 20 and Tween 80 marginally enhanced the protease activity of P106. SDS inhibited the enzyme to 80% residual activity whereas CTAB and TritonX 100 reduced the activity to 24% and 18% residual level respectively.

TritonX 100 significantly enhanced the protease activity of P4 to 149%. SDS and Tween 80 also increased the activity marginally. CTAB and Tween 20 reduced the residual activity to 94% and 81% respectively.

4.5.5.11 Effect of DMSO as oxidizing agent on enzyme activity of P106 and P4

The effect of different concentrations (1-5%) of DMSO as oxidizing agent on protease activity was determined and the residual activity is represented in Fig. 4.48.

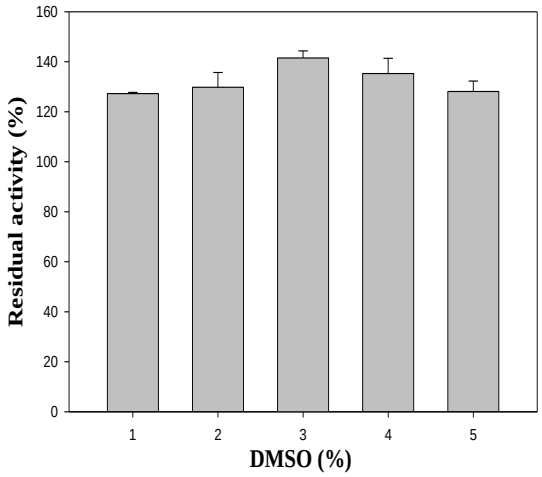

Fig. 4.48(a) Effect of DMSO on enzyme activity of P106

DMSO at a concentration of 3% had maximum effect on protease activity of both P106 and P4. For P106, 3% DMSO improved the enzyme residual activity to 141% while it enhanced the protease residual activity to150% for P4.

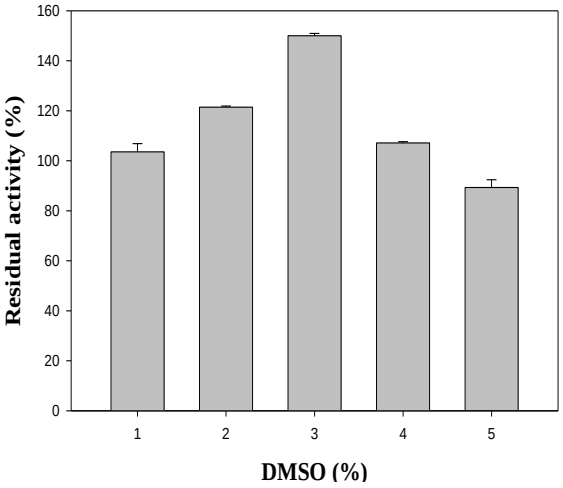

Fig. 4.48(b) Effect of DMSO on enzyme activity of P4

4.5.5.12 Effect of β-mercaptoethanol as reducing agent on enzyme activity of P106 and P4

The effect of different concentrations (0.5mM, 1mM, 5mM, 10mM and 15mM) of the reducing agent β-mercaptoethanol (β-ME) on protease activity of P106 and P4 was determined and the residual activity is represented in Fig. 4.49.

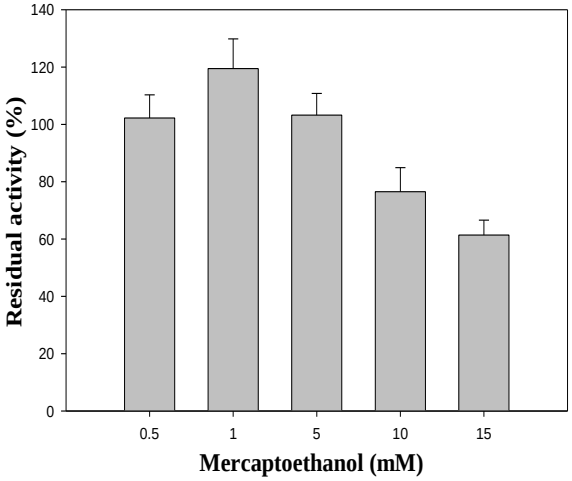

Fig. 4.49(a) Effect of β-mercaptoethanol on enzyme activity of P106

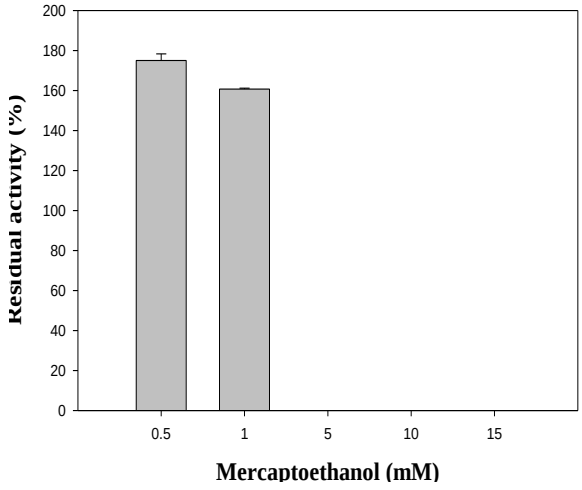

Fig. 4.49(b) Effect of β-mercaptoethanol on enzyme activity of P4

The enzyme activity of P106 was enhanced by β-mercaptoethanol upto a concentration of 1mM. The residual activity increased to 119% with 1mM β-ME, but the activity reduced at higher concentrations. The protease activity was inhibited beyond 5mM β-ME.

In the case of P4, 0.5mM concentration of β-mercaptoethanol increased the protease residual activity to 175%, however higher concentrations of β-ME reduced enzyme activity. The protease activity was completely inhibited beyond 1mM β-ME.

4.6 APPLICATION STUDIES OF PROTEASES

4.6.1 Commercial detergent compatibility of the P106 and P4

The compatibility of the two proteases P106 and P4 with different commercial detergents like Ariel®, Surf Excel®, Sunlight®, Tide® and Wheel® (7mg/mL) was studied, wherein the residual enzyme activity in the presence of the detergents was determined, the outcome is as represented in Fig. 4.50.

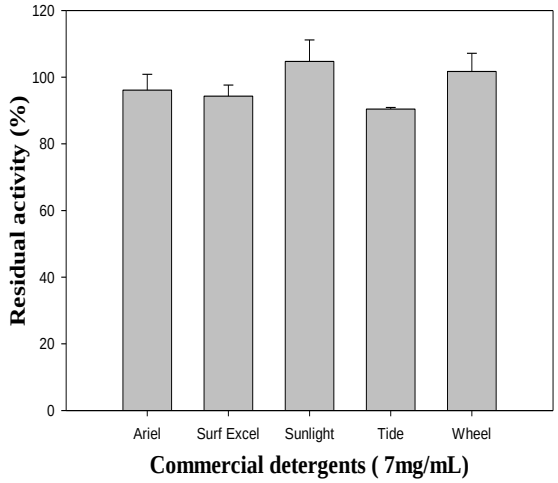

Fig. 4.50(a) Commercial detergent compatibility of P106

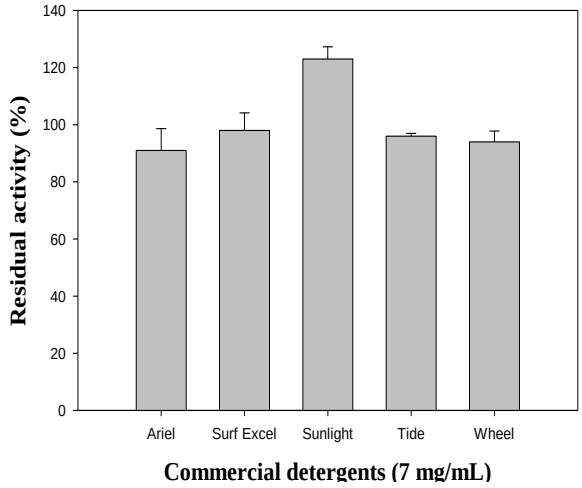

Fig. 4.50(b) Commercial detergent compatibility of P4

The enzyme protein of the two proteases P106 and P4 were stable in the presence of almost all the commercial detergents tested. More than 90% of the activity was retained for both enzymes with all the detergents. Only marginal reduction of enzyme activity of P106 was observed in the presence of Ariel, Surf

Excel and Tide. In the case of P4, the protease activity was enhanced to 123% when incubated along with Sunlight detergent.

4.6.2 Wash performance studies

The effectiveness of the proteases P106 and P4 in stain removal from cloth pieces stained with human blood was determined in order to study their wash performance (Fig. 4.51).

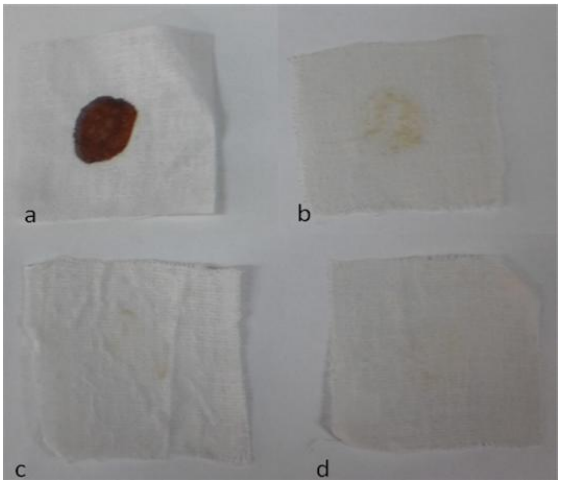

Fig. 4.51 Wash performance of proteases

a. Cotton cloth stained with human blood, b. Stained cloth washed in detergent, c. Stained cloth washed in detergent + P106, d. Stained cloth washed in detergent + P4

Visual examination of the washed cloth pieces indicated that the proteases P106 and P4 supported the detergent in the stain removal process. When the blood stained cloth was washed with detergent, blood stain was not removed completely, with traces of stain retained on the cloth and visible. But, when proteases P106 and P4 were supplemented along with detergents in the washing process, blood stain was removed almost completely from the cloth pieces.

4.6.3 Decomposition of gelatin layer of X-ray film

The hydrolytic activity of the proteases on the gelatin layer of the X-ray film is an important step in the extraction of silver from the X-ray film. The hydrolytic activity was determined by calculating the total protein released into the solution at different time intervals from 0 to 4 hours (Fig. 4.52).

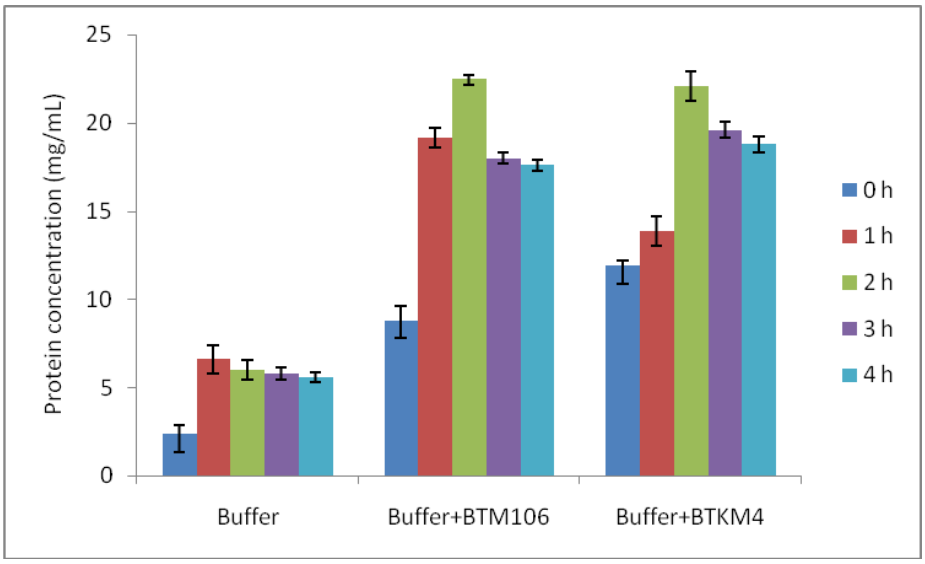

Fig. 4.52 Protein content of the supernatant after treatment of the X-ray film

The degradation of gelatin from the X-ray film by the action of the proteases was evidenced from the protein content of the solution (Fig. 4.52), which was observed to increase upon incubation, as the gelatin was released from the X-ray film into the solution. This is also indicated by the clearing of X-ray film (Fig. 4.53).

When the X-ray film was incubated in buffer alone, there was no observed increase in the protein content of the solution, indicating that no gelatin was released to the solution. Also, there was no clearing of the X-ray film.

In the case of protease P106, the protein content of the solution increased to 19.14 mg/mL after 1h incubation. The gelatin was almost completely degraded from the X-ray film which is evidenced by the clearing of the X-ray film (Fig. 4.53). After 2h incubation, the protein content of the solution increased to 22.46 mg/mL. The concentration of protein did not increase after 2h incubation, indicating that most of the proteins were hydrolysed by 2 hours of incubation.

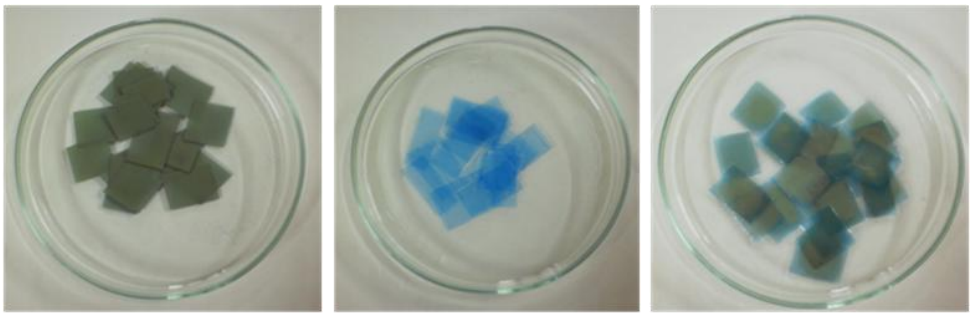

Fig. 4.53 Degradation of gelatin layer of X-ray film

When the X-ray film was incubated with the protease P4, the protein content increased to 13.86mg/mL after 1h incubation and to 22.07 mg/mL after 2h. The degradation of gelatin was more after 1h. Also, complete clearing of the X-ray film was not observed as in the case of P106 at 1h (Fig. 4.53).

Two proteases P106 and P4 isolated by culture independent and dependent approaches were characterized in the present study. The culture independent method, i.e., metagenomics, proved to be powerful tool for the discovery of protease enzyme as well as for analyzing the phylogenetic diversity of the microbial community in the mangrove sediment. The pH stable and thermotolerant nature of the proteases find potential application in detergent formulations. The ability of the proteases to degrade gelatin from used X-ray film is also highly significant.

Chapter - 5

DISCUSSION

Proteases are enzymes involved in the hydrolysis of peptide bonds and are present in all living organisms, with wide applications in pharmaceutical, food, textile, leather and waste management industries and also in the diagnosis of illness (Gupta *et al*., 2002). Microbial proteases are more abundant than those of plant or animal origin, accounting for two-thirds of the commercially available proteases (Kumar and Takagi, 1999). Microbial proteases characterized to date are typically isolated from culturable organisms, which represent less than 1% of the microbial world. This leads to high rediscovery frequency, thereby limiting the discovery of novel enzymes (Strohl, 2000). Hence culture-independent methods may be used to mine the complex microbial communities and to explore their valuable bioresources. The metagenomic approach to identify the enormous number of novel genes for diverse biomolecules of commercial importance is a recent trend. The present study reports the characterization of two protease enzymes obtained by culture independent and dependent approaches from a mangrove sediment sample. Also the phylogenetic diversity of the microbes present in the mangrove sediment was analysed based on their 16S rDNA sequences.

5.1 EXTRACTION OF METAGENOMIC DNA OF MICROBIAL COMMUNITIES IN MANGROVE SEDIMENTS

The first and crucial step in any metagenomic study is the isolation of total community DNA with good quantity and quality. Since DNA is isolated from soil and sediment samples which is a heterogenous mixture containing vast amount of humic contaminants, isolated DNA is invariably complexed with various impurities. Several methods are being followed for effective metagenome

extraction, each method having advantages and disadvantages. Hence three widely used methods were compared to obtain optimum quantity of DNA with maximum purity.

The metagenomic DNA was isolated from three different mangrove sediments collected from three different locales-Kannamaly mangrove (MGK), Kannamaly mangrove associated aquafarm (AQK) and Mangalavanam mangrove (MGM), using three different protocols. The DNA yielded by Protocol I was more than that from the other two methods. Samplewise, more DNA was obtained from the Kannamaly mangrove (MGK) sediment (57.6 ± 2.26 µg/g sediment) than from the other two samples, with Kannamaly mangrove associated aquafarm (AQK) yielding 45 ± 1.41 µg/g, while 41.2 ± 1.7 µg/g was obtained from Mangalavanam mangrove (MGM). The variation in the yield may not only be due to the differences in the sample types, but also due to the differences in the physical and chemical processes employed in various stages in the extraction protocols.

These three protocols are extensively used for the isolation of DNA from soils that have very complex nature. All three methods are based on direct extraction methods, wherein the microbial cells are lysed *in situ* and this method is known to yield maximum DNA. In addition, the DNA isolation is completed within 6-7 hours. In each case, the yield and purity of the isolated DNA depends on various factors like the nature of soil sample, agents used for cell lysis, DNA precipitation method, and so on.

The purity of DNA isolated from the different sediment samples by these three protocols was also compared. The DNA isolated according to only Protocol II was more pure compared to that obtained by protocols I and III, wherein protein contamination was indicated by a higher OD_{260}/OD_{280} ratio.

Protocol I was also used previously to isolate DNA from soil samples collected from hot springs in Himachal Pradesh, India, with a DNA yield of 1.94 µg/µL (Sharma *et al*., 2007). However the purity of the isolated DNA was poor and it was therefore further purified using Q-Sepharose. DNA isolated from soil samples collected from the Apharwat Mountain in the northwestern Himalayas by the same method yielded 100 ng/g of soil (Sudan and Vakhlu, 2013) and since the purity was low, it was not used for cloning purpose. In addition, DNA isolated from eight different soils of diverse composition yielded only crude DNA, with yield in the range from 2.5 to 26.9 mg of DNA per gram soil sample (Zhou *et al*., 1996).

The OD_{260}/OD_{230} ratio of DNA sample isolated from mangrove sediment according to Protocol II was in the range 1.33-1.52, indicating that the sample was devoid of humic contaminants compared to the DNA obtained using the other two methods in this study. For DNA isolated using protocols I and III the OD_{260}/OD_{230} ratio was below 1.1, indicating humic contamination. Protocol II uses aluminium sulphate and this may be responsible for removing the humic contaminants from the samples, a plausible reason for the reasonable purity of the DNA isolated by this method.

Isolation of DNA from compost soil of the rhizosphere of *Clivia miniata* using protocol II resulted in removal of humic substances to a greater extent as aluminium sulphate complexed with them and aided in its removal (Dong *et al*., 2006). But the disadvantage of the method was low DNA yield which might be due to the coprecipitation of DNA along with humic acids, as DNA and humic acid has similar physical and chemical characteristics. The protocol was later used in combination with two others, wherein the sample was treated with aluminum sulfate prior to DNA extraction (Litthauer *et al*., 2010).

Protocol III was developed for DNA isolation of subsurface soil samples from a gas manufacture site in Southern California (SC, USA) and sediment samples from a settling pond in Oak Ridge, Tenn. (ORT). DNA yielded by this method was of good quantity and with less shearing (Tsai and Olson, 1991). In this method lysozyme was used for cell wall lysis in combination with gentle freezing and thawing, hence consequentially DNA obtained was less sheared when compared to other methods. DNA isolated from composite soil samples following the same method was used for metagenomic library construction (Kumar, 2011)

The commercial kits essentially rely on silica gel spin columns for purification of the DNA. The MoBio™ kit yielded a suitable DNA extract from a peat sample (Schneegurt *et al*., 2003) and the procedure only took less than 2hours.

The yield and purity of DNA from Kannamaly mangrove (MGK) obtained by the three different methods used was different. It was amply clear that comparatively more quantity of DNA but with less purity was obtained by Protocol I. The DNA isolated by Protocol II was more pure, even though the DNA yield was much less than that obtained by Protocol I, besides not being enough for downstream processes. Since the yield and purity of DNA isolated by the three different methods followed in this study was not satisfactorily good, DNA isolation of Kannamaly mangrove sediment (MGK) was also tried with the commercially available kit (MoBio™ kit). But, the DNA yield with the kit was very low. However the quality of the DNA obtained was far superior to that obtained by the three protocols. Since the yield of DNA by protocol I was highest, DNA isolated using protocol I was purified using the kit, resulting in significant DNA yield (46.34 ± 1.62 µg/g sediment). Also, the OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios (1.98 and 1.26 respectively) of the DNA isolated and purified by this combination method indicated the DNA purity. Since purity of DNA is an important criteria for PCRability and for library construction, the DNA sample

isolated from Kannamaly mangrove (MGK) using protocol I and purified using MoBio^{TM} kit, was used for all downstream processes.

Yang *et al*., (2007) compared three protocols for total community DNA extraction and purification from compost. Protocol PL, a modified protocol of Zhang *et al*. (2003) and Zhou *et al*. (1996) yielded more DNA (83 ± 4.1 µg/g), followed by protocol LL which is a modified protocol of LaMontagne *et al*. (2002) yielding 62 ± 3.6 µg/g DNA. Protocol UL was a modified protocol of Krsek and Wellington (1999) and yielded least DNA (55 ± 2.9 µg/g). Hence enzymatic extraction in high-salt buffer (protocols LL and PL) was approved to be a more effective method for cell lysis, and mechanical lysis (protocol UL) was insufficient for lysing the cells in a complex system such as compost. This is probably because the enzymes in the high-salt buffer could meet cells better than glass beads, and therefore, they could fully come into play, while the mixture was being incubated and shaken on an orbital shaker. DNA purification using PEG8000 and spin bind cartridge produced pure DNA with $A_{260/280}$ ratios of 1.7–1.8 and were PCR amplifiable.

Leff *et al*., (1995) also reported comparison of three methods, Ogram method (Ogram *et al*., 1987), Jacobsen method (Jacobsen and Rasmussen, 1992) and Tsai method (Tsai and Olson, 1991) for extraction of DNA from Stream Sediments. The Ogram method (1.74 µg/g) had the greatest total DNA yield, and the Jacobsen method (0.35 µg/g) had the lowest. Although DNA yield was greatest for the Ogram method, the DNA was highly fragmented because of shearing by the bead beater. The Jacobsen method also provided fragmented DNA, perhaps because of digestion of DNA during the initial steps of the procedure prior to EDTA addition. The Tsai method provided DNA with the least fragmentation. In all cases, DNA samples were of low purity, apparently because of contamination with humic materials, but to different degrees; the Jacobsen

samples had the lowest amount of contamination, and the Tsai samples had the greatest.

Each protocol used for DNA isolation has its advantages as well as disadvantages. The protocol with high yield may not give highly pure DNA. The highly pure DNA obtained by a certain protocol may be extremely fragmented. Hence the choice of DNA isolation protocol has to be based on the goals of the study. The protocol giving highly pure DNA may be used for cloning purpose; the protocol with fragmented DNA may be used for hybridization studies and so on.

5.2 ANALYSIS OF PHYLOGENETIC DIVERSITY OF MANGROVE METAGENOME BASED ON 16S rDNA SEQUENCES.

Several studies have shown the uniqueness of mangrove sediments with respect to their microbial composition. Studies on microbial diversity in the mangrove sediments are important to understand the process of biogeochemical cycling and pollutants removal. The molecular phylogenetic analysis reveals the occurrence of the bacterial 16S rRNA gene sequences that are unique and sequences that are previously reported in other mangrove sediments.

In the present study, a total of 126 clones were classified based on their taxonomic hierarchy into 8 major phylum of bacterial domain. Majority of the clones belonged to phylum *Proteobacteria* (54%), followed by *Bacteroidetes* (19%) and *Firmicutes* (11%). Phylum *Chloroflexi* and *WS3* phylum account for 4% and 2% respectively. *Planctomycetes*, *Spirochaetes* and *Chlorobi* represented 1% each. Unclassified group accounted for 7% of the total number of clones.

The dominant phylum *Proteobacteria*, comprising of 68 clones (54%) was represented by classes *Gammaproteobacteria* (50%), *Deltaproteobacteria* (46%), *Alphaproteobacteria* (3%) and *Epsilonproteobacteria* (1%). Class *Gammaproteobacteria*, comprising of 34 clones, was represented by orders

Alteromonadales, Xanthomonadales, Methylococcales, Oceanospirillales and *Enterobacteriales* including members of the genus *Marinobacterium, Thiohalomonas, Thiohalophilus, Escherichia/Shigella, Thioprofundum, Oceaniserpentilla, Methylohalomonas, Haliea* and *Methylocaldum*. The class *Deltaproteobacteria*, with 31 clones, was represented by the orders *Desulfobacterales, Desulfuromonadales* and *Syntrophobacterales,* including members of the genus *Desulfuromonas, Desulfobacterium, Pelobacter, Desulfobulbus and Desulfosarcina.*

Most of the clones belonging to class *Gammaproteobacteria* (*Thiohalomaonas, Haliea*) were related to bioconversion of S-containing organic molecules (S-oxidisers), while most of the *Deltaproteobacteria* clones (*Desulfosarcina, Delsuphobulbus* and *Desulphobacterium*) were sulphate reducing bacteria. Members of the genus *Sulfurimonas* (class *Epsilonproteobacteria*) grow chemolithoautotrophically using zero valent sulfur or reduced sulfur compounds as electron donors. Sulfur-oxidising bacterial strains play an important role in detoxification of sulphide in marine sediments whereas 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 deposits characteristic of mangrove forests. Identification of sulfur- oxidising, and sulphur and sulphate reducing bacterial clones point towards the anaerobic conditions prevailing in the mangrove sediments and at the possible maintenance of the biogeochemical cycle.

In a previous report, two 16S rDNA gene libraries were constructed from Sundarban mangrove sediments and partial sequencing of the selected clones revealed the diversity of bacterial strains in the sediment (Ghosh *et al*., 2010). At least 8 different bacterial phyla were detected; the major divisions of detected

bacterial phyla were *Proteobacteria* (Alpha, Beta, Gamma, and Delta), *Flexibacteria* (CFB group), *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Firmicutes*, *Planctomycetes* and *Gammatimonadates*. Among *Proteobacteria*, *Gammaproteobaceria* was the most abundant, with the most abundant sequence types showing similarity to genus *Methylophaga*. Many of them were involved in the S-cycle while some showed similarity to the oil degrading bacterial populations. Among the *Deltaproteobacteria* clones, some of them showed similarity with *Desulfosarcina* sp. and *Desulfuromonas* sp.

Proteobacteria was also the most abundant phylum among the 16S rRNA gene clones derived from Oklahoma tall-grass prairie soil (Spain *et al*., 2009). The orders *Desulfuromonadales* and *Syntrophobacterales* of class *Deltaproteobacteria* and the orders *Xanthomonadales* and *Enterobacteriales* of the class *Gammaproteobacteria* were reported in the library. The order *Sphingomonadales* of class *Alphaproteobacteria* was also reported.

In the present study, the class *Alphaproteobacteria*, with single clone, was represented by the order *Sphingomonadales* which was also reported from the 16S rDNA library derived from Oklahoma tall-grass prairie soil (Spain *et al*., 2009). The members of phylum *Betaproteobacteria* was not reported among the clones. In the present study also, none of the clones showed similarity to members of class *Betaproteobacteria*.

The class *Epsilonproteobacteria*, with a single clone was represented by the order *Campylobacterales*, belonging to the genus *Sulphurimonas*. One clone, affiliated with the *Epsilonproteobacteria* was detected from the marine sediments from Sagami Bay and Tokyo Bay, Japan (Urakawa *et al*., 1999). The clone was most similar to an ectosymbiont of a polychaete, *Alvinella pompejana*, found at deep-sea hydrothermal vents. Moyer *et al*. (1995) reported the existence of

Epsilonproteobacteria in a microbial mat community at Pele's Vent, a hydrothermal vent on Loihi Seamount, Hawaii.

Among the non-proteobacterium group, *Bacteroidetes* was the prominent phylum, with 24 clones (19%) represented by the orders *Sphingobacteriales* and *Flavobacteriales*, comprising members of the genus *Ekhidna*, *Fulvivirga*, *Lewinella*, *Haliscomenobacter* and *Zeaxanthinibacter*. Fourteen clones (11%) belonged to the phylum *Firmicutes*, represented by a single order *Bacillales* including members of the genus *Falsibacillus* and *Bacillus*. Five clones (4%) belonged to the phylum *Chloroflexi*, among which one belonged to class Dehalococcoidetes and to the genus *Dehalogenimonas* while the other four clones could not be classified further and are grouped as unclassified *Chloroflexi*. Two clones belonged to the WS3 candidate phylum and could not be classified further. Two clones represented the phylum *Chlorobi*, both of them belonging to order *Ignavibacteriales* and to the genus *Ignavibacterium*. The phyla *Spirochaetes* and *Planctomycetes* were represented by a single clone each belonging to the orders *Spirochaetales* and *Planctomycetales* respectively. Nine clones (7%) could not be classified into any phylum and hence denoted as unclassified group.

The non proteobacterial phyla identified from Sundarban mangroves (Ghosh *et al*., 2010) include *Flexibacteria* (CFB group) (1 clone), *Actinobacteria* (2 clones), *Acidobacteria* (1 clone), *Chloroflexi* (2 clones), *Firmicutes* (1 clone), *Planctomycetes* (4 clones) and *Gemmatimonadates* (2 clones). The Gram-positive high G+C content subdivision include Clostridia and related organisms, while *Planctomycetes* and related organisms were the non-proteobacterial phyla reported by Gray and Herwig (1996) from marine sediments. Urakawa *et al*. (1999) identified Gram-positive bacteria and the division *Verrucomicrobia*, of the non-proteobacterial lineages from the marine sediments from Sagami Bay and Tokyo Bay, Japan.

To understand the bacterial diversity and species richness of the Kannamaly mangrove sediment clone library, rarefaction curve constructed with 126 sequences showed saturation at genetic distances of 10% and 20% indicating that the sampling effort covered almost the full extent of taxonomic diversity at these genetic distances, at the class and phylum level. At 3% sequence divergence, 116 different OTUs were observed, with the 126 sequences, indicating the species richness of the sample. Ninety OTUs were identified among the 115 sequences from the Arabian sea oxygen minimum zone (AS-OMZ) clone library (Divya *et al*., 2011) whereas 74 OTUs were detected from 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). However, in the present study, the coverage was relatively higher and hence it can be suggested that the sampling was representative of the *in situ* bacterial community present in the mangrove sediment.

5.3 SCREENING FOR PROTEASE PRODUCTION BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

The metagenomic shotgun library was constructed using the DNA isolated from Kannamaly mangrove sediment in vector pUC18 and host *E. coli* DH5α. 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*., 2010). Screening of 210 clones in the library for protease production was carried out and the clone, BTM106, which showed consistent protease activity, was selected for the study. Inspite of the abundance of new enzymes isolated by metagenomic approaches, there are comparatively few data concerning metagenome-derived proteases. Identification of protease gene from metagenomic library was earlier unsuccessful (Rondon *et al*. 2000) or resulted in false-positive clones (Jones *et al*. 2007). However, very few functional metalloproteases were identified through

metagenomic approach and characterized which include fibrinolytic metalloprotease from a deep sea sediment metagenomic library (Lee *et al.,* 2007), two metalloproteases from soil metagenomic libraries (Waschkowitz *et al.,* 2009), an alkaline serine protease from goat skin surface metagenome (Pushpam *et al.,* 2011) and two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts (Neveu *et al.,* 2011).

Simultaneously screening of the laboratory isolates obtained by culture dependent approach from the mangrove sediment was carried out and the strain BTKM4, identified as *Bacillus licheniformis*, was selected as the culture dependent counterpart. Most of the commercial proteases are produced by organisms belonging to the genus *Bacillus* and most of the potential alkaline protease producing *Bacilli* are strains of *Bacillus alcalophilus*, *B. amyloliquefaciens*, *B. licheniformis* and *B. subtilis* (Kumar and Takagi, 1999; Rao *et al.,* 1998; Gupta *et al.,* 2002).

5.4 CHARACTERISATION OF THE PROTEASE GENE OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

Proteases from *Bacillus* sp. have been extensively studied and are found to be highly diverse with respect to their function and characteristics. Even a single organism harbored different types of proteases. Despite the great diversity of their nucleotide sequences, the structures of the active site for proteases with similar function are highly conserved (Chen *et al.,* 2004). There are reports on partial to complete protease gene sequence from different *Bacillus* sp including 1326 bp alkaline protease gene of *B. subtilis* (Sadeghi *et al*., 2009), 1137 bp ORF of subtilisin Carlsberg of *B. licheniformis* (Jacob *et al.,* 1985), 1146 bp ORF of subtilisin DFE gene of *B. amyloliquefaciens* (Peng *et al.,* 2004), 1080 bp ORF of

fibrinolytic metalloprotease (Lee *et al*., 2007) and 1149 bp gene alkaline serine proteinase of *Bacillus pumilus* (Aoyama *et al.*, 2000).

In the present study, partial protease gene was amplified from the protease positive metagenomic clone BTM106 using degenerate primers designed based on the consensus amino acid sequence of the active site of neutral and serine proteases (Chen *et al.*, 2004). The protease gene from the strain *Bacillus licheniformis* BTKM4 was amplified using another set of protease primers specific for mature peptide of proteases (Bach *et al*., 1999). In both cases, the amplicons were cloned and sequenced using vector specific primers so as to get maximum number of nucleotide sequences. The nucleotide sequences and the deduced amino acid sequences of proteases P106 and P4 were compared with other protease sequences in the NCBI database to facilitate comparison and characterization of the protease gene.

The 977 bp long sequence of protease amplicon of BTM106 showed 99% identity to the mature peptide of bacillolysin of *Bacillus thuringiensis* BMB171 (CP001903) and *Bacillus cereus* ATCC 14579 (AE016877) when compared with sequences in the NCBI database. All of the first ten BLAST hit results showed identity to proteases of *Bacillus* sp. indicating that the protease could be of *Bacillus* origin. Multiple sequence alignment of the nucleotide sequence of BTM106 protease gene with other proteases' genes, clearly indicated the similarity between them and their conserved regions. The phylogenetic tree constructed based on the nucleotide sequences also revealed the relatedness towards the bacillolysin of *Bacillus thuringiensis* (CP001903) and *Bacillus cereus* (AE016877) as they all formed a single cluster.

The 954 bp long sequence of protease amplicon from BTKM4 showed 99% identity with the neutral protease gene of *Bacillus cereus* (M83910). All of the NCBI BLAST hit results showed similarity to proteases of *Bacillus* origin,

confirming its identity. From the phylogenetic analysis it is further evident that the partial protease gene of strain BTKM4 showed similarity towards the neutral protease gene of *Bacillus cereus* (M83910) as they form a single cluster supported by a bootstrap value of 99%.

Hence in both proteases, P106 and P4, the nucleotide sequence responsible for the mature peptide got amplified and its length is in correlation with the length of other neutral proteases like the 951 bp long neutral protease NPRC of *Bacillus cereus,* 947 bp sequence of *B. stearothermophilus* (Bach *et al*., 1999) and 942 bp neutral protease *nprB* of *B. subtilis* (Tran *et al*., 1991). Also, multiple sequence alignment of nucleotide sequences indicates that the conserved regions present in BTM106 protease is less than that of BTKM4 protease. This might be due to its culture independent origin and also adds to the novelty in its nucleotide sequence, which is significant; ratifies the aim and purpose of this study, i.e. to look for novelty.

An ORF of 957 bp was detected in the nucleotide sequence of BTM106 protease and a 324 amino acid sequence was deduced from its nucleotide sequence. In BTKM4 protease the detected ORF was 888bp long and 317 amino acids sequence was deduced from its nucleotide sequence. The gene encoding subtilisin Carlsberg from *B. licheniformis* have been cloned in pBR322 vector by Jacob *et al.,* (1985) that has a 1137 bp open reading frame (ORF) encoding 379 amino acids (Jacobs *et al.*,1985). Peng *et al* (2004) cloned the subtilisin DFE gene from *B. amyloliquefaciens* DC-4 in pGEM-T plasmid that has 1146 bp ORF encoding 382 amino acids. A fibrinolytic metalloprotease gene with 1080 bp ORF encoding 359 amino acids was reported from deep sea sediment metagenomic library (Lee *et al*., 2007).

The identity of deduced amino acid sequences of both P106 amd P4 showed similarity with that of protease amino acid sequences upon pBLAST

analysis with the NCBI database. The deduced amino acid sequence of BTM106 protease gene showed 95% identity with bacillolysin of *Bacillus cereus* (WP000730385), and in the phylogenetic tree it clustered together with bacillolysins of *Bacillus cereus.* Upon multiple sequence alignment of the deduced amino acid with other reported proteases, majority of the regions remained conserved. Thus the identity of the protease obtained by culture independent approach can be confirmed to be similar to bacillolysin of *Bacillus cereus* or said to be conserved with bacillolysin of *Bacillus cereus.* Here the significance of the conserved protein domain sequences in serving to identify the origin of the protein based on phylogenetic interrelatedness comes to the fore.

Bacillolysins are metalloendopeptidase belonging to EC 3.4.24.28 of IUBMB Enzyme Nomenclature. Other names of bacillolysins include *Bacillus* metalloendopeptidase; *Bacillus subtilis* neutral proteinase; anilozyme P 10; *Bacillus* metalloproteinase; *Bacillus* neutral proteinase and megateriopeptidase. Variants of this enzyme have been found in species of *Bacillus* including *B. subtilis*, *B. amyloliquefaciens*, *B. megaterium*, *B. mesentericus*, *B. cereus* and *B. stearothermophilus* (IUBMB, 1992). They are included in peptidase family M4 of MEROPS database (Rawlings *et al*., 2012).

The deduced amino acid sequence BTKM4 protease gene showed 99% identity with *Bacillus cereus* neutral protease (1NPCA) and Neutral protease mutant E144s (1ESPA). The phylogenetic analysis based on deduced amino acid sequence also confirmed its identity as it clustered along with the neutral proteases of *Bacillus* sp. Most of the aminoacids were conserved when aligned with reported proteases of NCBI database. Hence it can be stated that the mature peptide sequence of neutral protease of *Bacillus licheniformis* BTKM4 is conserved with that of *Bacillus cereus*, which also adds to their phylogenetic relatedness. Tran and coworkers (1991) confirmed the novelty of a *Bacillus subtilis* neutral protease B,

nprB, by the multiple sequence alignment of its mature peptide with other reported proteases.

The conserved domains present in the amino acid sequences of BTM106 and BTKM4 was elucidated using NCBI conserved domain database and both proteases showed similar domain architecture with respect to their active site. A zinc binding site was present within the active site of both proteases, indicating that they may be zinc-dependent metalloprotease. Both proteases exhibit similar conserved domain architecture characteristic of zinc metalloproteases and were classified to the M4 neutral protease family of Gluzincin superfamily included in the MEROPS database, which is a database specifically dedicated to peptidases (Rawlings *et al*., 2012).

Gluzincin superfamily consists of several zinc-dependent metallopeptidases such as the M1, M2, M3, M4, M13, M32, M36 peptidases (based on MEROPS classification), and contain two conserved amino acid stretches, represented by their single letter codes as HEXXH and EXXXD (H-Histidine, E-Glutamic acid, D-Aspartic acid and X is any non-conserved amino acid) as part of their active site. All peptidases in this family bind a single catalytic zinc ion which is tetrahedrally co-ordinated by three amino acid ligands, and a water molecule that forms the nucleophile on activation during catalysis (Marchler-Bauer *et al*., 2013).

This peptidase M4 family includes several endopeptidases such as thermolysin, aureolysin (the extracellular metalloproteinase from *Staphylococcus aureus*), neutral protease from *Bacillus cereus* and protealysin. These enzymes have a two-domain structure with the active site between the domains. The N-terminal domain contains the HEXXH zinc-binding motif while the helical C-terminal domain, which is unique for the family, carries the third zinc ligand. Most

of these secreted proteases degrade extracellular proteins and peptides for bacterial nutrition, especially prior to sporulation (Marchler-Bauer *et al*., 2013).

The multiple sequence alignment of the 58 amino acid stretch of the active site of the proteases of BTM106 and BTKM4, with the amino acid sequence of other proteases with close similarity belonging to the MEROPS M4 family, elucidated the conserved motifs in the active site of the two protease genes**.** The two five-mer conserved motifs HEXXH and EXXXD (H-Histidine, E-Glutamic acid, D-Aspartic acid and X is any non-conserved amino acid) are observed to be present in the deduced amino acid sequence of both BTM106 and BTKM4 proteases. The first motif HEXXH is present as HELTH (H-Histidine, E-Glutamic acid, L-Leucine and T-Threonine) in both proteases. A novel fibrinolytic metalloprotease was reported from a metagenomic library with a His-Glu-Phe-Gly-His sequence in the active site (Lee *et al*., 2007). Bode *et al*. (1993) reported that astacins, metalloprotease, and snake venom exhibited identical zinc-binding regions (His-Glu-X-X-His-X-X-Gly-X-X-His) and this was also a consensus sequence in metalloprotease disintegrins, another member of the zinc metalloprotease superfamily (Poindexter *et al*., 1999).

However, the second motif EXXXD is present as a highly conserved longer nine-mer GXXNEXXSD (G-Glycine, N-Asparagine, E-Glutamic acid, S-Serine, D-Aspartic acid, and X is any non-conserved amino acid) conserved stretch, instead of being a five-mer stretch. Due to this highly conserved nature of the active site of both proteases they can surely be classified as Zinc metalloproteases. The protein sequences of metalloproteases MprA and MprB obtained from metagenomic library with highly conserved zinc-binding motif HEXXH and the third zinc ligand motif GXXNEXXSD (Waschkowitz *et al*., 2009) were most similar to the proteases of the current study.

Neveu and others (2011) classified two serine proteases obtained from desert sand metagenomic library, based on the alignment of the amino acids in their active site with other serine proteases, into S8A subfamily of serine proteases. The active site of this subfamily is characterized by a catalytic triad in the order Asp–(Thr/Ser)–Gly, His–Gly–Thr–His, and Gly–Thr–Ser–Met–Ala–Xaa–Pro (Neurath, 1989).

The phylogenetic relationship of both BTM106 and BTKM4 enzyme protein was analysed and was found to be closely related among themselves and with another *Bacillus* M4 metalloprotease supported by a bootstrap value of 97% on the phylogenetic tree.

In an attempt to analyse the structure of proteases, the secondary as well as tertiary structures of the proteases BTM106 and BTKM4 were predicted with the deduced amino acid sequences using PHYRE^2 software. Both proteases were modeled based on the template d1npca of PDB database which is a neutral protease of *Bacillus cereus*, strain dsm 3101. They are included in Class d (Alpha and beta proteins) with Zincin like fold which contains mixed beta sheet with connection over free side of the sheet. They belong to the superfamily metalloproteases and 'thermolysin like' family which includes a alpha-helical C-terminal domain, characteristic for the family. The domains are complexed with four calcium and a zinc (Fox *et al*., 2013). In the structural model of EAP protease, four calcium binding sites were predicted by the GGv1.0 program (Jasmin *et al*., 2010). The X-ray crystallographic data of matrix metalloproteinase-9 (MMP-9) reveals that the enzyme contains two Zn^{2+} and three to five Ca^{2+} binding sites, of which one Zn^{2+} (coordinated by three histidine residues of HEXXHXXGXXH motif) is unambiguously demonstrable to exhibit the catalytic function. Due to their relative locations on the protein surface, far removed from the active site, non-catalytic Zn^{2+} and all Ca^{2+} appear to exhibit structural rather than catalytic roles (Tobwala and Srivastava, 2013). Involvement of one Zn^{2+} in

the catalysis was obvious by the conserved domain analysis of both proteases. However elucidation of the complete structural model of the proteases would be possible only with the complete protease gene sequence.

5.5 CHARACTERIZATION OF THE PROTEASE ENZYMES OBTAINED BY CULTURE INDEPENDENT AND CULTURE DEPENDENT METHODS

The physicochemical characterization of proteases or for that matter any enzyme is a prerequisite for its various applications. For this, the proteases were extracted from their producers, purified and studied.

The cell free supernatant of the protease positive clone BTM106 was compared with that of control (*E.coli* DH5α transformed with pUC18). When analyzed by SDS-polyacrylamide gel electrophoresis, an extra protein band with a molecular mass of approximately 34 kDa was observed in the culture supernatant of the clone. This protein was absent in the supernatant from the control suggesting that this may be the band corresponding to the protease activity. Similarly an extra protein band was seen on the SDS-PAGE of culture supernatant of neutral protease WB30 (pNPRB) which was absent in the supernatant of its control WB30(pUB18) (Tran *et al*., 1991).

The purification of proteases P106 and P4 was done by fractionation of proteins based on ammonium sulphate saturation followed by gel filtration of the concentrated fraction by sephadex G-75 column. In the case of both P106 and P4, 30-60% ammonium sulphate saturated protein fraction showed protease activity. This active fraction was later concentrated and passed through the gel filtration column to further purify the proteases based on their molecular weight. As the purity of the proteases increased with each purification step, their specific activity also increased during the process. The protease P106 was 3 fold purified after ammonium sulphate fractionation. The specific activity of protease P4 increased

two fold after ammonium sulphate precipitation. After gel filtration chromatography, the protease P106 became almost ten folds purified whereas protease P4 got about nine folds purified.

Various purification methods are followed for the purification of extracellular proteases. The first step in any purification step 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). The alkaline protease from *Bacillus mojavensis* A21was purified from the culture supernatant to homogeneity using 60–80% (v/v) acetone precipitation, Sephadex G-75 gel filtration and CM-Sepharose ion exchange chromatography, with a 6.43-fold increase in specific activity and 16.56% recovery (Haddar *et al*., 2009). The EAG-2 protease from *Bacillus subtilis* was purified by 65-80% ammonium sulphate fraction followed by ultrafiltration and then by passing through DEAE-Sepharose Fast Flow column. These procedures aided in the 11 fold purification of the enzyme (Ghafoor and Hasnain, 2010). Two alkaline proteases AS and HS from *Bacillus* sp. strain GX6638 (ATCC 53278) were precipitated by the addition of 80% (vol/vol) ice-cold acetone and then brought to 80% saturation by addition of solid ammonium sulfate. Further purification of the proteases were brought about by passing through Sephadex G-25 column and then through quaternary ammonium-cellulose (QA52) column (Durham *et al*., 1987).

The neutral protease of *Bacillus subtilis* (ATCC 6051a) was purified from the crude extract by 30-50% ammonium sulfate precipitation, DEAE-cellulose ion exchange chromatography, and CM-cellulose ion exchange chromatography sequentially, and the enzyme became 11 fold purified (Han and Lee, 1997). The thermostable alkaline protease from alkalophilic *Bacillus pumilus* was extracted from the culture supernatant by 40-80% saturation of ammonium sulphate and applied on a DEAE-Sepharose CL-6B column. The protease was not bound to the support signifying it may be a basic protein. These unadsorbed protease fraction

was then applied to column of CM-Sepharose CL-6B and was finally loaded on a FPLC Sephacryl S-200 gel filtration column. After the final purification step, the enzyme was purified to about 36 fold (Kumar, 2002).

Zymography helps to specifically pinpoint the protein band responsible for the protease activity. For this the purified enzyme protein was electrophoresed on SDS-PAGE containing 0.1% gelatin and stained used Coomassie staining solution. A clearing was seen at the band responsible for activity due to degradation of gelatin in that region. For protease P106, a single clearing band at approximately 34 kDa was observed, which is in correlation with the extra protein band observed in the cell free extract of the clone when compared with the control. In the case of P4, a single clearing band was observed corresponding to approximately 29 kDa. Presence of a single clearing band indicates that a single extracellular protease is present in both cases.

Generally, the molecular weights of alkaline proteases from microorganisms range between 15 and 36 kDa (Ghafoor and Hasnain, 2010). The molecular weight of the protease P4 (29 kDa) is higher than alkaline protease A21 (20 kDa) (Haddar *et al*., 2009), *Bacillus subtilis* EAG-2 protease (27 kDa) (Ghafoor and Hasnain, 2010) Subtilisin Carlsberg (27.3 kDa), Subtilisin BPN' (27.5 kDa) (Horikoshi, 1990) serine protease AS (27.5 kDa) (Durham *et al*., 1987), alkaline proteases from *B. subtilis* PE-11 (15 kDa) (Adinarayana *et al*., 2003) and *Kurthia spiroforme* sp. (8 kDa) (Steele *et al*., 1992). The molecular weight of protease P4 is even higher than that of alkaline protease from *B. mojavensis* (30 kDa) (Beg and Gupta, 2003). However, both proteases are smaller than the serine protease HS (36 kDa) from alkalophilic *Bacillus* sp. Strain GX6638 (Durham *et al*., 1987).

The isoelectric points of proteases P106 and P4 were determined by isoelectric focusing. The isoelectric point (pI) of P106 was determined as 8.5 and

that of P4 was 8.0, both in the alkaline range. The isoelectric point of two serine proteases HS and AS from alkalophilic *Bacillus* sp. Strain GX6638 was ~4.2 and 5.2 respectively. The isoelectric point of subtilisin Carlsberg was observed to be 9.4 (Markland and Smith, 1971), while that of BPN' was 7.8 (Ottesen and Svenden, 1971). The pI of *Bacillus subtilis* neutral protease was determined to be 9.0 by cross partitioning method (Han and Lee, 1997). Two commercial preparations Esperase and Savinase T (Novo Industry), produced by alkalophilic *Bacillus* spp., have very high isoelectric points (pI 11.0) (Rao *et al*., 1998). Protease IV produced by *Pseudomonas aeruginosa* has an isoelectric point of 8.70 (Engel *et al*., 1998).

The optimum pH for maximum enzyme activity of proteases P106 and P4 at different pH ranging from 2 to 13 were determined. The stability of the proteases at different pH was also studied. For both proteases the activity increased towards alkaline pH reaching a maximum at pH 10.0 for protease P106 and at pH 11.0 for protease P4. Both proteases showed maximum stability at pH 10.0. Protease P106 retained more than 80% of its activity at a wide range of pH from 6 to 13 while protease P4 retained more than 80% of its activity at a pH ranging from 7 to 12 indicating that P106 is more stable than P4. Hence the pH studies indicate that both proteases can be grouped as alkaline proteases.

The temperature stability and the optimum temperature of the proteases P106 and P4 over temperatures ranging from 30ºC to 100ºC were determined. Both proteases were active over a broad range of temperature, with maximum enzyme activity at 60ºC. The maximum activity of protease P106 at that temperature was 295 ± 0.2 U/mL while that of protease P4 was 226 ± 0.9 U/mL. Both the proteases showed its maximum stability at 40ºC. The protease P106 exhibited more than 70% of the relative activity at 30ºC and 40ºC and was stable upto 70ºC. The protease P4 showed more than 70% of its relative activity at 40ºC and 50ºC was stable upto 80ºC.

The proteases P106 and P4 were stable over a wide range of pH and temperature, protease P106 being more stable at different pH while P4 was more stable at different temperatures. The alkaline and thermotolerant nature of alkaline proteases find their application in detergent industry.

The optimum pH range of alkaline proteases is generally between pH 9 and 11, with a few exceptions (Rao *et al*., 1998). The optimum temperature of alkaline proteases ranges from 50 to 70ºC (Gupta *et al.*, 2002). The alkaline protease of *Bacillus* sp. B18 is the enzyme with the highest optimal temperature value (85°C) among alkaline protease-producing species from the genus *Bacillus* (Kumar and Takagi, 1999). Ferrero et al. (1996) reported an alkaline protease from *B. licheniformis* with optimum activity at 60°C and stability for 10 min between 30 and 60°C. The alkaline serine-protease from *Bacillus mojavensis* A21was highly active over a wide range of pH from 7.0 to 13.0, with an optimum at pH 8.5 and exhibited maximal activity at 60ºC (Haddar *et al*., 2009). The alkaline protease from alkalophilic *Bacillus pumilus* was optimally active at pH 11.5 and temperature of 55 to 60°C (Kumar, 2002). The serine proteases AS and HS from alkalophilic *Bacillus* sp. Strain GX6638 had optimal proteolytic activities over a broad pH range (pH 8 to 12) and exhibited temperature optima of 65°C (Durham *et al*., 1987). The alkaline protease from *Bacillus clausii* I-52 has an optimum pH of around 11 and optimum temperature of 60°C (Joo *et al*., 2003). The optimum proteolysis of extracellular protease EAG-2 from *Bacillus subtilis* was observed at pH 8.5 and temperature 65°C (Ghafoor and Hasnain, 2010).

The fibrinolytic metalloprotease obtained from a metagenomic library of deep sea sediment showed optimal activity at 50°C for 1 h and pH 7.0 (Lee *et al*., 2007). The serine proteases DV1 from metagenomic library of the Gobi and Death Valley deserts showed optimum activity at pH 8.0 and temperature 55°C, while another protease M30 had an optimum pH at 11.0 and optimal activity at 40°C (Neveu *et al*., 2011). The pH and temperature optima for an alkaline serine

protease isolated from goat skin surface metagenome were 10.5 and 42°C respectively (Pushpam *et al*., 2011).

The effect of four different protease inhibitors, aprotinin (1 mM), EDTA (10 mM), Iodoacetamide (10 mM) and PMSF (10 mM) on the enzyme activity of the two proteases was determined. Among the four different inhibitors tested, EDTA inhibited the activity of both proteases P106 and P4 to 35% and 67% residual activity respectively. These observations clearly indicate that these proteases are metalloproteases, as their proteolytic activity could be inhibited by the metal chelating effect of EDTA. EDTA binds divalent ions such as Ca^{2+} and Zn^{2+} present in the structural model of metalloproteases (Rao *et al*., 1998).The proteolyitc activity of fibrinolytic metalloprotease obtained from deep sea sediment metagenomic library was inhibited by 1 mM EDTA (Lee *et al*., 2007) and that of two proteases MprA and MprB from soil metagenomic library were inhibited by 10 mM EDTA (Waschkowitz *et al*., 2009).

The ability of the proteases to hydrolyse different proteinaceous substrates (1%) like bovine serum albumin (BSA), casein, gelatin and haemoglobin was determined. Both proteases, P106 and P4, demonstrated more specificity towards casein as a substrate followed by haemoglobin. The protease P106 had least preference for gelatin as substrate whereas BSA was the least preferred substrate for the protease P4.

The alkaline serine protease from goat skin surface metagenome showed relatively high activity with 0.1% (w/v) casein when compared with other substrates, BSA and gelatin (Pushpam *et al*., 2009). Protease IV obtained from *Pseudomonas aeruginosa* was able to digest chromogenic substrates, Chromozym PL (tosyl-Gly-Pro-Lys-4-nitroanilide) and Val-Leu-Lys-4-nitroanilide, suggesting that this enzyme cleaves on the carboxyl side of lysine residues (Engel *et al*., 1998). EAG-2 protease exhibited broad substrate activity relationship and rapid

proteolysis was showed with different natural substrate like casein, gelatin, ovalbumin and chicken albumin, except for haemoglobin (Ghafoor and Hasnain, 2010). *B. pumilus* alkaline protease efficiently hydrolysed only two chromogenic peptide nitroanilides, Glu-Gly-Ala-Phe-pNA and Glu-Ala-Ala-Ala-pNA suggesting that the protease has a preference for aromatic and hydrophobic amino acids at the P1 position (Kumar, 2002).

The kinetic parameters Km and Vmax of the proteases were estimated by plotting the initial velocity as the function of the concentration of casein as substrate. The Lineweaver-Burk plot was constructed and the Km and Vmax of protease P106 were determined as 0.66 mg/mL and 625 U/mL respectively and that of protease P4 was 0.53 mg/mL and 122 U/mL respectively.

The calculated Vmax and Km of metagenome derived serine alkaline protease for azocasein were 366 U/mg and 0.13 mg/ml respectively (Pushpam *et al*., 2011). The Km values of *B. pumilus* alkaline protease observed with synthetic substrates Glu-Gly-Ala-Phe-pNA and Glu-Ala-Ala-Ala-pNA were 1.1 mmol/L and 3.7 mmol/L respectively (Kumar, 2002). The Km and Vmax of *Psedomonas aeruginosa* derived protease IV for tosyl-Gly-Pro-Lys-paranitroanilide (Chromozym PL) were 319 mM and 1.33 mM/min and for Val-Leu-Lys-paranitroanilide were 727 mM and 0.74 mM/min respectively (Engel *et al*., 1998).

The effect of twelve different metal ions Al^{3+}, Ba^{2+}, Ca^{2+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mg^{2+}, Mn^{2+}, Na^{+}, Ni^{2+} and Zn^{2+} (1 mM) on protease activity of P106 and P4 was determined. Metal ions of Zn, Ca and Co increased the protease activity of P106. The proteolytic activity of P4 was enhanced in the presence of metal ions of Co, Mg, Na, Ni, Mn and Zn indicating that these metal ions are required as cofactors for enzyme activity. Metal ions of Fe, Al and Cu significantly reduced the enzyme activity of both proteases.

Zn and Ca provide structural as well as thermal stability to metalloproteases (Rao *et al*., 1998). The presence of zinc binding site has been elucidated from the conserved domains of both proteases P106 and P4. The enzymatic activity of fibrinolytic metalloprotease was enhanced in the presence of metal ions of Ca and Co (1 mM), indicating that these metal ions are important cofactors for the enzyme activity (Lee *et al*., 2007). Zn sometimes inhibited the activity when present at higher millimolar concentrations as in the case of fibrinolytic metalloprotease and some other metalloproteases (Larsen and Auld, 1991, Luciano *et al*., 1998) as additional binding of metal ions to the binding site lowers the activities (Lee *et al*., 2007). The alkaline serine protease was activated by metal ions of Co and Mn and inhibited by Fe (Pushpam *et al*., 2011). The enzymatic activities of many proteases were enhanced by Ca (Neveu *et al*., 2011, Ghafoor and Hasnain, 2010, Kumar, 2002, Haddar *et al*., 2009), Mg (Kumar, 2002, Haddar *et al*., 2009) and Mn (Kumar, 2002, Tekin *et al*., 2012). Metal ions of Fe inhibited the proteolytic activity of *Bacillus cohnii* APT5 alkaline protease (Tekin *et al*., 2012).

The effect of various non-ionic detergents such as Triton X-100, Tween 80, Tween 20, and ionic detergents like SDS and CTAB on the proteolytic activity of P106 and P4 was studied. The detergents Tween 20 and Tween 80 marginally enhanced the protease activity of P106. SDS inhibited the enzyme to 80% residual activity whereas CTAB and Triton X-100 reduced the activity to 24% and 18% residual level respectively. Triton X-100 significantly enhanced the protease activity of P4 to 149%. SDS and Tween 80 also increased the activity marginally. CTAB and Tween 20 reduced the residual activity to 94% and 81% respectively. The stability of the proteases in the presence of some of the ionic and non-ionic detergents is highly significant as they can be used in detergent formulations. Many other alkaline proteases are also found to be stable in the presence of ionic and non-ionic detergents (Tekin *et al*., 2012, Joo *et al*., 1998, Haddar *et al*., 2009).

The increase in enzyme activity in the presence of detergents may be due to the effect of surfactants on the unfolding of the substrate moiety (Vita *et al*., 1985, Chaphalkar and Dey, 1998).

The effect of different concentrations (1-5%) of DMSO as oxidizing agent on protease activity of P106 and P4 was determined. 3% DMSO enhanced the protease activity of both P106 and P4. There is little published literature available concerning the stability studies of protease towards DMSO. However the serine protease M30 from metagenomic library was stable at 2 mM and 5 mM DMSO (Neveu *et al*., 2011). There are also some reports on the stability of proteases in the presence of oxidizing agents like hydrogen peroxide and sodium perborate (Joo *et al*., 1998, Haddar *et al*., 2009)

The effect of different concentrations (0.5 mM, 1 mM, 5 mM, 10 mM and 15 mM) of the reducing agent β-mercaptoethanol (β-ME) on protease activity of P106 and P4 was determined. Protease P106 showed stabilty upto 5 mM β-ME. Protease P4 was stable upto 1 mM β-ME, but it was completely inhibited beyond that concentration. The stability of the proteases may be due to the absence of disulphide bonds in their structure, as no cysteine moiety is present in the amino acid sequence. However, proteases containing cysteine moiety in their sequence are readily inactivated in the presence of reducing agents like β-ME and DTT (Neveu *et al*., 2011, Pushpam *et al*., 2011, Engel *et al*., 1998)

5.6 APPLICATION STUDIES OF PROTEASES

The suitability of an enzyme preparation for use in detergents 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. In the present study, proteases P106 and P4 showed excellent stability and compatibility in the presence of the laundry detergents

tested like Ariel®, Surf Excel®, Sunlight®, Tide® and Wheel®. More than 90% of the activity was retained for both enzymes with all the detergents when incubated for 40ºC for 1 hour.

An alkaline protease enzyme produced from *B. cereus* retained more than 80% of its activity in seven commercial detergents tested (Banik and Prakash, 2004). A thermostable alkaline protease from *Bacillus licheniformis* RP1 retained more than 83% activity with all the detergents tested even after 1 h incubation at 40ºC (Sellami *et al*., 2008). Singh *et al*. (2001) reported a serine alkaline protease from *Bacillus* sp. SSR1 showing nearly 70–80% of activity in most of the detergents at 40ºC. They also reported that above 40ºC, addition of additives like $CaCl_2$ is required for stability of enzyme in detergents. Also Banerjee *et al*. (1999) and Bhosale *et al*. (1995) reported that protease enzyme retained high activity in commercial detergents after supplementation of additives $CaCl_2$ and glycine. Comparison of these results with the present study indicates that proteases P106 and P4 are superior as they retain more than 90% of enzyme activity in presence of laundry detergents at higher temperature and without supplementation of any additives.

Various alkaline proteases have long been incorporated as biobuilders into detergents to hydrolyse and remove proteinaceous materials from stained cloth (Ito *et al*., 1998, Kobayashi *et al*., 1995). The proteases P106 and P4 were found to be effective in stain removal from cloth pieces stained with human blood when used along with detergent. Detergent stable proteases capable of removing blood stain from cotton fabrics have been reported from *Arthrobacter ramosus* and *Bacillus alcalophilus* (Kanekar *et al*., 2002). The alkaline and thermostable nature along with detergent compatibility, as well as the stain removal efficiency of the proteases, P106 and P4, suggest them as suitable candidates in detergent formulations.

Used X-ray film contains about 1.5 to 2.0 % (by weight) silver in its gelatin layers. Alkaline proteases find potential application in the bioprocessing of used X-ray films for silver recovery. The conventional practice of burning film for silver recovery causes a major environmental pollution problem. Thus, the enzymatic hydrolysis of the gelatin layers on the X-ray film enables not only the silver recovery, but also recycling of the polyester film base.

In the present study the hydrolytic activity of the proteases on the gelatin layer of the X-ray film was determined by calculating the total protein released into the solution at different time intervals. The protein content of the solution was observed to increase upon incubation, as the gelatin was released from the X-ray film into the solution. This was also evidenced by the clearing of X-ray film. In the case of protease P106, the gelatin was almost completely degraded from the X-ray film after 1h incubation which is evidenced by the clearing of the X-ray film. Partial clearing of the X-ray film was observed in the case of P4 after 1h, and complete clearing after 2 hours incubation. Hence it could be concluded that both proteases P106 and P4 were capable of degrading gelatin from used X-ray film and P106 is faster in the process. Alkaline proteases from *Bacillus* sp. B21-2 (Fujiwara and Yamamoto, 1987) *Bacillus* sp. B189 (Fujiwara *et al*., 1991) and *B. coagulans* PB-77 (Gajju *et al*., 1996) were also found to be capable of gelatin degrading from used X-ray films and thus enabling silver recovery.

The present work was aimed at characterization of two proteases obtained by an innovative culture independent approach as well as by a traditional culture dependent approach. The potential of metagenomics in the discovery of biomolecules as well as phylogenetic diversity analysis was proved by the study. The presence of the 16S rRNA phylogenetic anchor in the clone with DNA insert showing protease activity was not detected. This may be due to the small size of the insert. The attempt to analyse the microbial diversity based on 16Sr DNA sequences therefore, helped to understand the phylogenetic makeup of the

sediment sample. However characterization studies revealed that the partial protease gene obtained via the culture independent methods was much similar to that produced by various other *Bacillus* species. This only strengthend the conviction that the enzyme obtained by culture independent methods may indeed be from a *Bacillus* species. Both proteases were pH stable, thermotolerant and detergent compatible and hence can find application in laundry industry. They were also found to be efficient in gelatin degradation from used X-ray film.

Chapter - **6**

SUMMARY AND CONCLUSION

Microorganisms dominate the biosphere and are ubiquitously present by virtue of their wealth of physiologies and molecular adaptations which enable them to survive in almost all environmental niches, some being so very harsh and inhospitable that no other life form co exists. Microbial diversity represents our planet's greatest, but least utilized resource for biotechnologically important products and processes. In the milieu of traditional approaches failing to represent the scope of our microbial diversity, culture independent approaches like metagenomics gain importance. Metagenomics helped to reveal the amazing treasures within the microbial world and to untap their vast genetic and biochemical potential for biotechnological innovation.

The metagenomic DNA was isolated from three different mangrove sediments using three commonly used protocols with minor modifications. The highest DNA yield was with protocol I whereas DNA of maximum purity was obtained by protocol II, irrespective of the sample type. Among the three mangrove sediment samples, DNA isolated from Kannamaly mangrove (MGK) was of higher concentration than the other two samples.

Since the concentration and purity of DNA obtained by the three methods was not satisfactory or good enough for downstream processing, metagenomic DNA isolation from the Kannamaly mangrove sediment was also attempted using MoBio Ultraclean^{TM} soil DNA isolation kit. The metagenomic DNA isolated using protocol I, but purified by the MoBio kit was satisfactorily good with respect to quantity and quality. Hence this DNA was used for all further processes in the study.

The metagenomic investigations included the phylogenetic as well as the functional analysis. Towards this end, two libraries were constructed, a phylogenetic library using the 16S rRNA gene amplicons and a shotgun library for functional analysis. To analyse the phylogenetic diversity of the mangrove sediment a phylogenetic library was constructed with 126 phylogenetic clones. The clones were classified based on their taxonomic hierarchy using RDP Naïve Bayesian rRNA Classifer into 8 major phylum of bacterial domain. Majority of the clones belonged to phylum *Proteobacteria* (54%), followed by *Bacteroidetes* (19%) and *Firmicutes* (11%). Phylum *Chloroflexi* and *WS3* phylum account for 4% and 2% respectively. *Planctomycetes*, *Spirochaetes* and *Chlorobi* represented 1% each. Unclassified group accounted for 7% of the total number of clones.

The dominant phylum *Proteobacteria*, comprising of 68 clones (54%) was represented by classes *Gammaproteobacteria* (50%), *Deltaproteobacteria* (46%), *Alphaproteobacteria* (3%) and *Epsilonproteobacteria* (1%). Class *Gammaproteobacteria*, comprising of 34 clones, was represented by orders *Alteromonadales, Xanthomonadales, Methylococcales, Oceanospirillales* and *Enterobacteriales* including members of the genus *Marinobacterium, Thiohalomonas, Thiohalophilus, Escherichia/Shigella, Thioprofundum, Oceaniserpentilla, Methylohalomonas, Haliea* and *Methylocaldum*. The class *Deltaproteobacteria*, with 31 clones, was represented by the orders *Desulfobacterales, Desulfuromonadales* and *Syntrophobacterales,* including members of the genus *Desulfuromonas, Desulfobacterium, Pelobacter, Desulfobulbus and Desulfosarcina.*

Most of the clones belonging to class *Gammaproteobacteria* (*Thiohalomaonas, Haliea*) were related to bioconversion of S-containing organic molecules (S-oxidisers), while most of the *Deltaproteobacterium* clones (*Desulfosarcina, Delsuphobulbus* and *Desulphobacterium*) were sulphate reducing bacteria. Members of the genus *Sulfurimonas* (class *Epsilonproteobacteria*) grow

chemolithoautotrophically using zero valent sulfur or reduced sulfur compounds as electron donors.

The class *Alphaproteobacteria* was represented by the order *Sphingomonadales* and the class *Epsilonproteobacteria* was represented by the order *Campylobacterales* and genus *Sulphurimonas*. Identification of sulfur-oxidising and sulphur and sulphate reducing bacterial clones indicates the anaerobic condition in the mangrove sediment and a possible maintenance of the biogeochemical cycle.

Bacteroidetes was the prominent phylum among the non-proteobacterial group , with 24 clones (19%) represented by two orders *Sphingobacteriales* and *Flavobacteriales*, which comprised members of the genus *Ekhidna, Fulvivirga, Lewinella, Haliscomenobacter* and *Zeaxanthinibacter*. Fourteen clones (11%) belonged to the phylum *Firmicutes*, represented by a single order *Bacillales* including members of the genus *Falsibacillus* and *Bacillus.* Five clones (4%) belonged to the phylum *Chloroflexi,* among which one belonged to class *Dehalococcoidetes* and to the genus *Dehalogenimonas* while the other four clones could not be classified further and are grouped as unclassified *Chloroflexi*. Two clones belonged to the WS3 candidate phylum and could not be classified further. Two clones represented the phylum Chlorobi, both of them belonging to order Ignavibacteriales and to the genus *Ignavibacterium*. The phyla *Spirochaetes* and *Planctomycetes* were represented by a single clone each, belonging to the orders *Spirochaetales* and *Planctomycetales* respectively. Nine clones (7%) could not be classified into any phylum and hence denoted as unclassified group.

The phylogenetic relationship of the clones were analysed by constructing the phylogenetic tree of proteobacterial and non-proteobacterial clones. The 16S rDNA clones belonging to each phylum clustered together showing their phylogenetic relatedness.

To understand the bacterial diversity and species richness of the Kannamaly mangrove sediment clone library, rarefaction curve was constructed with 126 sequences and showed saturation at genetic distances of 10% and 20% indicating that the sampling effort covered almost the full extent of taxonomic diversity at these genetic distances, at the class and phylum level. At 3% sequence divergence, 116 different OTUs were observed, with the 126 sequences, indicating the species richness of the sample.

The metagenomic shotgun library comprising 210 clones was constructed in vector pUC18 and host *E. coli* DH5α using the DNA from Kannamaly mangrove sediment. . Screening of the 210 clone library for protease activity helped to segregate clone BTM106 which showed consistent protease production. This was selected for subsequent study. Strain BTKM4 a protease producer, identified as *Bacillus licheniformis* and obtained from the mangrove sediment using culture based methods was selected as the culture dependent counterpart.

The recombinant plasmid was isolated from the clone BTM106 and the protease gene insert was amplified using protease specific degenerate primers. The amplified partial protease gene (~1kb) was cloned in pGEMT and sequenced. The 977 bp long nucleotide sequence showed 99% identity to the mature peptide of bacillolysin of *Bacillus thuringiensis* BMB171 (CP001903) and *Bacillus cereus* ATCC 14579 (AE016877) in the NCBI database. The multiple sequence alignment with similar sequences in the NCBI database pointed out the conserved regions. The phylogenetic tree also revealed the relatedness among towards the bacillolysin of *Bacillus thuringiensis* (CP001903) and *Bacillus cereus* (AE016877), as they all formed a single cluster.

NCBI Orf finder helped detect an ORF in the BTM106 protease gene sequence with 957 bp, in the reading frame 2 and in 5'-3' direction. The deduced amino acid sequence of BTM106 protease gene also showed 95% identity with bacillolysin of *Bacillus cereus* (WP000730385), and in the phylogenetic tree

clustered together with *Bacillus cereus* bacillolysins, which are metalloendopeptidases found in species of *Bacillus* including *B. subtilis, B. amyloliquefaciens, B. megaterium, B. mesentericus, B. cereus* and *B. stearothermophilus.*

The mature peptide gene was amplified using protease specific primers from strain *Bacillus licheniformis* BTKM4. The (~1kb protease amplicon cloned in pGEMT was sequenced using vector specific primers. The 954 bp sequence showed 99% identity with the neutral protease gene of *Bacillus cereus* (M83910) in NCBI database. The multiple sequence alignment of the nucleotide sequence demonstrated conserved regions within the sequences. From the phylogenetic analysis it was further evident that the partial protease gene of strain BTKM4 showed similarity towards the neutral protease gene of *Bacillus cereus* (M83910) as they formed a single cluster supported by a bootstrap value of 99%.

An Open Reading Frame of 888 bp was detected in the reading frame 1 in 5'-3' direction of BTKM4 protease gene by the NCBI ORF finder. The deduced amino acid sequence of BTKM4 protease gene showed 99% identity with *Bacillus cereus* neutral protease (1NPCA) and neutral protease mutant E144s (1ESPA). The phylogenetic analysis based on deduced amino acid sequence also confirmed its identity, seeing that it clustered along with the neutral proteases of *Bacillus* sp.

The conserved domains present in the amino acid sequences of BTM106 and BTKM4 was elucidated using NCBI conserved domain database and both proteases showed similar domain architecture with respect to their active site. A Zinc binding site was observed within the active site of both proteases, indicative of their being Zinc dependent metalloprotease. Both exhibited similar conserved domain architecture characteristic of Zinc metalloproteases and were classified to the M4 family of Gluzincin superfamily included in the MEROPS database.

The two also contained two conserved amino acid stretches, represented by their single letter codes as HEXXH and EXXXD (H-Histidine, E-Glutamic

acid, D-Aspartic acid and X is any non-conserved amino acid) as part of their active site, which are characteristic of several zinc-dependent metallopeptidases in the gluzincin superfamily., wherein a single catalytic zinc ion is tetrahedrally co-ordinated by three amino acid ligands, and a water molecule that forms the nucleophile on activation during catalysis.

The two five-mer conserved motifs HEXXH and EXXXD (H-Histidine, E-Glutamic acid, D-Aspartic acid and X is any non-conserved amino acid) were observed in the deduced amino acid sequence of both BTM106 and BTKM4 proteases. The first motif HEXXH was present as HELTH (H-Histidine, E-Glutamic acid, L-Leucine and T-Threonine) in both proteases. The second motif EXXXD was present as a highly conserved longer nine-mer GXXNEXXSD (G-Glycine, N-Asparagine, E-Glutamic acid, S-Serine, D-Aspartic acid, and X is any non-conserved amino acid) conserved stretch, instead of being a five-mer stretch. This highly conserved nature of the active site of both proteases also ensures their classification as Zinc metalloproteases. The phylogenetic analysis showed them to be closely related among themselves and with another *Bacillus* M4 metalloprotease supported by a bootstrap value of 97% on the phylogenetic tree.

PHYRE^2 software was used to predict the secondary as well as tertiary structures of the proteases BTM106 and BTKM4 based on the template d1npca of PDB database, a neutral protease of *Bacillus cereus*, strain dsm 3101. The models predicted that they are included in Class d (Alpha and beta proteins) with Zincin like fold which contained mixed beta sheet with connection over free side of the sheet. The domains complex with four calcium and a zinc. They belong to the superfamily metalloproteases and 'thermolysin like' family which includes alpha-helical C-terminal domain characteristic for the family. Involvement of one Zn^{2+} in the catalysis was obvious by the conserved domain analysis of both proteases.

The crude extract of the proteases was prepared from the culture broth for characterization. The crude extract of clone BTM106 when analysed on SDS-

PAGE showed the presence of an additional protein band, the extra band at 34 kDa corresponding to the protease activity of clone BTM106.

The proteases P106 and P4 were purified from the crude extract by 30-60% ammonium sulphate saturation, followed by gel filtration through Sephadex G-75 column. The enzyme activity and specific activity of the purified proteases were calculated and their fold of purification was determined. The specific activity of P106 was five times more than that of P4. After gel filtration chromatography, the protease P106 was ten folds purified whereas protease P4 was more than nine folds purified.

Zymogram of the purified proteases on Gelatin-PAGE showed a single clearing zone on the gel corresponding to 34 kDa for protease P106 and 29 kDa for protease P4. The clearing zone of P106 at 34 kDa was in correlation with the extra protein band observed in the crude extract of the clone when compared with the control. The isoelectric points were determined as 8.5 for P106 and 8.0 for P4.

Both proteases were active at alkaline pH, with maximum activity of P106 at pH 11.0 and of P4 at pH 10.0. P106 was stable at pH ranging from 6 to 13, while P4 was stable in pH from 7 to 12. Hence P106 and P4 can be alluded to as alkaline proteases and P106 shows more pH stability than P4.

The proteases were stable over a wide range of temperature indicating that they were also thermotolerant. Both proteases exhibited optimum activity at 60ºC, and showed maximum stability at 40ºC. P4 was more thermotolerant than P106 as it was stable upto 80ºC while P106 was stable upto 70ºC.

The two proteases were inhibited by 10mM EDTA. They showed more substrate specificity towards casein. The Km and Vmax of P106 was 0.66 mg/mL and 625 U/mL respectively, and that of P4 was 0.53 mg/mL and 122 U/mL respectively indicating that P4 has more affinity than P106 for the substrate casein.

Metal ions of Zn, Ca and Co enhanced P106 while metal ions of Co, Mg, Na, Ni, Zn and Mn activated P4. The protease P106 was inhibited by metal ions of Cd, Ni, Cu, Al and Fe whereas P4 was inhibited by metal ions of Ba, Al, Cd, Fe and Cu.

P106 was stable in the presence of non-ionic detergents like Tween 20 and Tween 80 while P4 was stable in non-ionic detergents like Tween 80, TritonX-100 and ionic detergent SDS. P106 was stable upto 5% DMSO and its proteolytic activity was enhanced in the presence of the oxidant. P4 was stable upto 3% DMSO, beyond that concentration its activity reduced. P106 was stable in upto 5 mM β-mercaptoethanol while P4 got deactivated beyond 1 mM β-mercaptoethanol.

The two proteases P106 and P4 showed excellent stability and compatibility in the presence of the laundry detergents tested like Ariel®, Surf Excel®, Sunlight®, Tide® and Wheel®. More than 90% of the activity was retained for both enzymes with all the detergents, when incubated for 40ºC for 1hour.

The efficiency of the proteases P106 and P4 to be used as biobuilders along with detergents to hydrolyse and remove proteinaceous materials from stained cloth was proved. Both proteases were effective in stain removal from blood stained cloth pieces, along with detergent at 40ºC for 30 minutes.

The hydrolytic activity of the proteases P106 and P4 on the gelatin layer of used X-ray film was determined and the degradation of gelatin layer was evidenced by increase in the protein content of the solution. This was also supported by the visual examination of the X-ray film as the X-ray film was completely cleared upon incubation. The protease P106 was more effective in the process taking only 1hour for complete clearing of X-ray film while protease P4 took 2 hours for the same.

CONCLUSIONS

The present study targeted the use of two methods-the culture independent and dependent approaches in the search for enzymes with novel properties and also to peek into the soil microbial diversity. The problems associated with DNA isolation from soil communities were highlighted by the use of three different protocols as well as a commercially available kit. The metagenomic DNA isolation was standardized as a combination of two protocols. The phylogenetic diversity of the bacterial community in the mangrove sediment based on 16S rDNA was revealed eight different phyla.

The commercial demand for different types of proteases especially alkaline proteases with novel and robust properties is increasing day by day. The present study was a humble attempt towards characterization of two proteases and to check their efficiency in industrial applications. Both proteases, by virtue of their interesting features like pH stability, thermotolerance and detergent compatibility, proved to be excellent candidates for various applications and can be commercialized after elaborate study.

Metagenomic approach paves the way to explore the structural and functional diversity of the microbial world. Regardless of the vast resource of molecular diversity present in metagenomic DNA, the number of actual discoveries is often restricted by technological and logistic constraints. However the discovery rate and robustness of the approach have been increased to a very large extent by the most modern next-generation sequencing (NGS).

REFERENCES

Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Völker U, Ellingsen TE, Valla S (2009): A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS Microbiology Letters **296**, 149-158.

Abhrajyoti G, Nirmalya D, Amit B, Amit T, Sathyaniranjan KB, Kalyan C, Dhrubajyoti C (2010): Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. Saline Systems **6**, 1-11.

Adinarayana K, Ellaiah P, Prasad DS (2003): Purification and partial characterization of thermostable serine alkaline protease from a newly isolated *Bacillus subtilis* PE-11. American Association of Pharmaceutical Scientists Pharmaceutical Science and Technology **4**, 1–9.

Aehle W, Sobek H, Amory A, Vetter R, Wilke D, Schomburg D (1993): Rational protein engineering and industrial application: structure prediction by homology and rational design of protein-variants with improved "washing performance" – the alkaline protease from *Bacillus alcalophilus*. Journal of Biotechnology **28**, 31–40.

Ahan YS, Kim CJ, Choi SH (1990): Production of protease by the extreme halophile, *Halobacterium* sp. Journal of Korean Chemical Society **33**, 247–251.

Aksornkoae S, Arroyo C, Blasco F, Burbridge PR, Tuck CH, Cintron G, Davie JDS, Dixon JA, Hamilton LS, Heald E, Hegerl E, Lal P, Luna LA, Pannier F, Ramdial B, Saenger P, Schaeffer-Novelli Y, Schweithelm J, Snedaker SC, Srivastava PDL, Weidenbach R, Yokel B, Dixon RG, Eong OJ, Saifullah SM (1984): Handbook for mangrove area management. United

Nations Environment Program and East-West Center, Environment and Policy Institute, Honolulu, Hawaii.

Alongi DM (1988): Bacterial productivity and microbial biomass in tropical mangrove sediments. Microbial Ecology **15**, 59-79.

Alongi DM, Christoffersen P, Tirendi F (1993): The influence of forest type on microbial-nutrient relationships in tropical mangrove sediments. Journal of Experimental Marine Biology and Ecology **171**, 201–223.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. Journal of Molecular Biology **215**, 403-410.

Amann RI, Ludwig W, Schleifer KH (1995): Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiological Reviews **59**, 143–169.

Aminin ALN, Warganegara FM, Aditiawati P, Akhmaloka (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.

Andersen LP (1998): Method for dehairing of hides or skins by means of enzymes. US Patent 5,834,299.

Andrade LL, Leite DCA, Ferreira EM, Ferreira LQ, Paula GR, Maguire MJ, Hubert CRJ, Peixoto RS, Domingues RMCP, Rosado AS (2012): Microbial diversity and anaerobic hydrocarbon degradation potential in an oil-contaminated mangrove sediment. BMC Microbiology **12**, 186.

Angelov A, Mientus M, Liebl S, Liebl W (2009): A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. Systematic and Applied Microbiology **32**,177–185.

Anissa H, Ali B, Rym A, Alya SK, Moncef N (2009): A novel surfactant-stable alkaline serine-protease from a newly isolated *Bacillus mojavensis* A21. Purification and characterization. Process Biochemistry **44**, 29–35.

Anne MS, Lee RK, Mostafa SE (2009): Abundance, composition, diversity and novelty of soil Proteobacteria. The ISME Journal **3**, 992–1000.

Aoyama M, Toma C, Yasuda M, Iwanaga M (2000): Sequence of the Gene Encoding an Alkaline Serine Proteinase of *Bacillus pumilus* TYO-67. Microbiology and immunology **44**, 389-393.

Argos P (1987): A sensitive procedure to compare amino acid sequences. Journal of Molecular Biology **193**, 385-396.

August PR, Grossman TH, Minor C, Draper MP, MacNeil IA, Pemberton JM, Call KM, Holt D, Osburne MS (2000): Sequence analysis and functional characterization of the violacein biosyntheticpathway from *Chromobacterium violaceum.* Journal of Molecular Microbiology and Biotechnology **2**, 513-519.

Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (1987): Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley Interscience, New York.

Avneet KS, Jyoti V (2013): Isolation of a thioesterase gene from the metagenome of a mountain peak, Apharwat, in the northwestern Himalayas. Biotechnology **3**, 19–27.

Bach HJ, Errampalli D, Leung KT, Lee H, Hartmann A, Trevors JT, Munch JC (1999): Specific detection of the gene for the extracellular neutral protease of Bacillus cereus by PCR and blot hybridization. Applied and Environmental Microbiology **65**, 3226-3228.

Bajorath J, Hinrichs W, Saenger W (1988): The enzymatic activity of proteinase K is controlled by calcium. European Journal of Biochemistry **176**, 441–447.

Baker BJ, Banfield JF (2003): Microbial communities in acid mine drainage. FEMS Microbiology Ecology **44**, 139–152.

Baker GC, Smith JJ, Cowan DA (2003): Review and re-analysis of domain-specific 16S primers. Journal of Microbiological Methods **55**, 541–555.

Bakken LR (1985): Separation and purification of bacteria from soil. Applied and Environmental Microbiology **49**, 1482–1487.

Bakken LR, Lindahl V (1995): Recovery of bacterial cells from soil. In Van Elsas JD, Trevors JT (Eds): Nucleic Acids in the Environment: Methods and Applications. Springer-Verlag, Heidelberg, Germany, pp. 9–27.

Banerjee R, Agnihotri R, Bhattacharyya BC (1993): Purification of alkaline protease of *Rhizopus oryzae* by foam fractionation. Bioprocess Engineering **9**, 245–248.

Banerjee UC, Sani RK, Azmi W, Sani R (1999): Thermostable alkaline protease from *Bacillus brevis* and its characterisation as a laundry detergent additive. Process Biochemistry **35**, 213–219.

Banik RM, Prakash M (2004): Laundry detergent compatibility of the alkaline protease from *Bacillus cereus*. Microbiological Research **159**, 135-140.

Bano N, Nisa MU, Khan N, Saleem M, Harrison PJ, Ahmed SI, Azam F (1997): Significance of bacteria in the flux of organic matter in the tidal creeks of the mangrove ecosystem of the Indus River delta, Pakistan. Marine Ecology Progress Series **45**, 1-12.

Barett AJ (1994): Proteolytic enzymes: serine and cysteine peptidases. In Methods in Enzymology. Vol. 244. Academic Press, New York-London, **244**, 1–15.

Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994): Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proceedings of the National Academy of Sciences **91**, 1609-1613.

Barthomeuf C, Pourrat H, Pourrat A (1992): Collagenolytic activity of a new semi-alkaline protease from *Aspergillus niger*. Journal of Fermentation and Bioengineering **73**, 233–236.

Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA (2009): Synthesis of methyl halides from biomass using engineered microbes. Journal of the American Chemical Society **131**, 6508-6515.

Bech LM, Branner S, Breddam K, Groen H (1993): Oxidation stable detergent enzymes. U.S. Patent No. 5,208,158. U.S. Patent and Trademark Office, Washington, DC.

Beg QK, Gupta R (2003): Purification and characterization of an oxidation-stable, thiol-dependent serine alkaline protease from *Bacillus mojavensis*. Enzyme and Microbial Technology **32**, 294–304.

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL (2000): Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science **289**, 1902-1906.

Bell DJ, Hoare M, Dunnill P (1983): The formation of protein precipitates and their centrifugal recovery. Advances in Biochemical Engineering/ Biotechnology **26**, 1–72.

Benlloch S, López-López A, Casamayor EO, Øvreås L, Goddard V, Daae FL, Smerdon G, Massana R, Joint I, Thingstad F, Pedrós-Alió C, Rodríguez-Valera F (2002): Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. Environmental Microbiology **4**, 349-360.

Bennett-Lovsey RM, Herbert AD, Sternberg MJE, Kelley LA (2008): Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins **70**, 611–625.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The protein data bank. Nucleic Acids Research **28**, 235-242.

Bernholdt HF (1975): Meat and other proteinaceous foods. In Gerald Reed (Ed): Enzymes in Food Processing. New York, Academic Press, pp. 473-491.

Berry AE, Chiocchini C, Selby T, Sosio M, Wellington EMH (2003): Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiology Letters **223**, 15-20.

Bhosale SH, Rao MB, Deshpande VV, Srinivasan MC (1995): Thermostability of high-activity alkaline protease from *Conidiobolus coronatus* (NCL 86.8.20). Enzyme Microbial Technology **17**, 136–139.

Birnboin HC, Doly J (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic acids research **7**, 1513-1523.

Blank CE, Cady SL, Pace NR (2002): Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. Applied and Environmental Microbiology **68**, 5123–5135.

Blum H, Beier H, Gross HJ (1987): Silver staining of proteins in polyacrylamide gels. Electrophoresis **8**, 93-99.

Böckle B, Galunsky B, Müller R (1995): Characterization of a keratinolytic serine proteinase from *Streptomyces pactum* DSM 40530. Applied and Environmental Microbiology **61**, 3705–3710.

Bode W, Gomis-Rüth FX, Stöckler W (1993): Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins.' FEBS Letters **331**, 134–140.

Bowman JP, McCammon SA, Brown MV, Nichols DS, McMeekin TA (1997): Diversity and association of psychrophilic bacteria in Antarctic sea ice. Applied and Environmental Microbiology **63**, 3068–3078.

Bradford MM (1976): A dye binding assay for protein. Analytical Biochemistry **72**, 248–254.

Brady SF (2007): Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. Nature Protocols **2**, 1297–1305.

Brady SF, Chao CJ, Clardy J (2002): New natural product families from an environmental DNA (eDNA) gene cluster. Journal of the American Chemical Society **124**, 9968-9969.

Brady SF, Chao CJ, Handelsman J, Clardy J (2001): Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Organic Letters **3**, 1981-1984.

Brady SF, Clardy J (2000): Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. Journal of American Chemical Society **122**, 12903-12904.

Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, Murzin AG (1995): A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. Nature **378**, 416–419.

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F (2002): Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of Sciences of the United States of America **99**, 14250–14255.

Brenner S (1988): The molecular evolution of genes and proteins: a tale of two serines. Nature **334**, 528–530.

Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, Barrell BG, Cole ST (1998): Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infection and Immunity **66**, 2221–2229.

Brown MV, Bowman JP (2001): A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). FEMS Microbiology Ecology **35**, 267–275.

Bruce KD, Hiorns WD, Hobman JL, Osborn AM, Strike P, Ritchie DA (1992): Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. Applied and Environmental Microbiology **58**, 3413–3416.

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**, 7–20.

Buyer JS, Sasser M (2012): High throughput phospholipid fatty acid analysis of soils. Applied Soil Ecology **61**, 127–130.

CASP8 Proceedings (2009): Proteins: Structure, Function, and Bioinformatics **77**, 1-228.

Chakrabarti SK, Matsumura N, Ranu RS (2000): Purification and characterization of an extracellular alkaline serine protease from *Aspergillus terreus* (IJIRA 6.2). Current Microbiology **40**, 239-244.

Chakraborty R, Srinivasan M (1993): Production of a thermostable alkaline protease by a new *Pseudomonas* sp. by solid substrate fermentation. Journal of Microbiology and Biotechnology **8**, 7–16.

Chandrasekaran S, Dhar SC (1983): A low cost method for the production of extracellular alkaline proteinase using tapioca starch. Journal of Fermentation Technology **61**, 511–514.

Chaphalkar SR, Dey S (1994): Some aspects of production of extracellular protease from *Streptomyces diastaticus*. Journal of Microbiology and Biotechnology **9**, 85–100.

Chaphalkar SR, Dey S (1998): Thermostable alkaline metalloprotease from newly isolated alkalophilic *Streptomyces diastaticus* strain SS1. Indian Journal of Biochemistry and Biophysics **35**, 34-40.

Chellappan S, Jasmin C, Basheer SM, Elyas KK, Bhat SG, Chandrasekaran M (2006): Production, purification and partial characterization of a novel protease from marine *Engyodontium album* BTMFS10 under solid state fermentation. Process Biochemistry **41**, 956–961.

Chen FC, Shen LF, Chak KF (2004): A facile analytical method for the identification of protease gene profiles from *Bacillus thuringiensis* strains. Journal of Microbiological Methods **56**, 125-132.

Cheng H, Grishin NV (2005): DOM-fold: A structure with crossing loops found in DmpA, ornithine acetyltransferase, and molybdenum cofactor-binding domain. Protein Science **14**, 1902–1910.

Cheng SW, Hu HM, Shen SW, Takagi H, Asano M, Tsai YC (1995): Production and characterization of keratinase of a feather-degrading *Bacillus licheniformis* PWD-1. Bioscience Biotechnology and Biochemistry **59**, 2239-2239.

Cheryan M (1986): Ultrafiltration Handbook. Technomic Publishing Company, Lancaster, PA.

Chung EJ, Lim HK, Kim JC, Choi GJ, Park EJ, Lee MH, Chung YR, Lee SW (2008) Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. Applied and Environmental Microbiology **74**, 723–730.

Claesson MJ, Cusack S, O'Sullivan O (2001): Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proceedings of the National Academy of Sciences of the United States of America **108**, 4586–4591.

Coenye T, Vandamme P (2003): Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiology Ecology **228**, 45–49.

Cole C, Barber JD, Barton GJ (2008): The Jpred 3 secondary structure prediction server. Nucleic Acids Research **36**, W197–W201.

Cole JR, Wang Q, Chai B, Tiedje JM (2011): The Ribosomal Database Project: sequences and software for high-throughput rRNA analysis. Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches. John Wiley & Sons, Hoboken, NJ, pp. 313-324.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM (2014): Ribosomal Database Project: data and tools for high throughput rRNA analysis Nucleic Acids Research **41**, D633-D642.

Connon SA, Giovannoni SJ (2002): High-throughput methods for culturing microorganisms in very low nutrient media yield diverse new marine isolates. Applied and Environmental Microbiology **68**, 3878–3885.

Coolbear T, Monk C, Peek K, Morgan HW, Daniel RM (1992): Laboratory-scale investigations into the use of extremely thermophilic proteinases for cleaning ultrafiltration membranes fouled during whey processing. Journal of Membrane Science **67**, 93–101.

Cottrell MT, Moore JA, Kirchman DL (1999): Chitinases from uncultured marine microorganisms. Applied and Environmental Microbiology **65**, 2553–2557.

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS, August PR, Nalin R, Guerineau M, Jeannin P, Simonet P, Pernodet JL (2003): Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Applied and Environmental Microbiology **69**, 49–55.

Couto GH, Glogauer A, Faoro H, Chubatsu LS, Souza EM, Pedrosa FO (2010): Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. Genetics and Molecular Biology **9**, 514-523.

Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF (2010): Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse Proteobacteria. Applied and Environmental Microbiology **76**, 1633–1641.

Curtis TP, Sloan WT, Scannell JW (2002): Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America **99**, 10494–10499.

Dalev P (1990): An enzyme-alkaline hydrolysis of feather keratin for obtaining a protein concentrate for fodder. Biotechnology Letters **12**, 71-72.

Danovaro R, Luna GM, Dell'Anno A, Pietrangeli B (2006): Comparison of two fingerprinting techniques, terminal restriction fragment length polymorphism and automated ribosomal intergenic spacer analysis, for determination of bacterial diversity in aquatic environments. Applied and Environmental Microbiology **72**, 5982–5989.

De Filippo C, Cavalieri D, Di Paola M (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 of the United States of America **107**, 14691–14696.

De Marco AC, Dick AJ (1978): Aminopeptidase I activities in several microorganisms. Canadian Journal of Biochemistry **56**, 66–71.

Deane SM, Robb FT, Woods DR (1987): Production and activation of a SDS-resistant alkaline serine exoprotease of *Vibrio alginolyticus*. Journal of General Microbiology **133**, 391–398.

Debette J (1991): Isolation and characterization of an extracellular proteinase produced by a soil strain of *Xanthomonas maltophila*. Current Microbiology **22**, 85–90.

Degrange V, Bardin R (1995): Detection and counting of *Nitrobacter* populations in soil by PCR. Applied and Environmental Microbiology **61**, 2093–2098.

Delmont TO, Robe P, Clark I, Simonet P, Vogel TM (2011): Metagenomic comparison of direct and indirect soil DNA extraction approaches. Journal of Microbiological Methods **86**, 397–400.

DeSantis GZZ, Greenberg WA, Wong K (2002): An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. Journal of the American Chemical Society **124**, 9024–9025.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006): Greengenes, a chimera-checked

16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology **72**, 5069–5072.

Dewar K, Sabbagh L, Cardinal G, Veilleux F, Sanschagrin F, Birren B, Levesque RC (1998): *Pseudomonas aeruginosa* PAO1 bacterial artificial chromosomes: strategies for mapping, screening, and sequencing 100 kb loci of the 5.9 Mb genome. Microbial and Comparative Genomics **3**, 105–117.

Dhandapani R, Vijayaragavan R (1994): Production of a thermophilic extracellular alkaline protease by *Bacillus stearothermophilus* AP-4. World Journal of Microbiology and Biotechnology **10**, 33–35.

Diaz-Perez SV, Alatriste-Mondragon F, Hernandez R, Birren B, Gunsalus RP (1997): Bacterial artificial chromosome (BAC) library as a tool for physical mapping of the archaeon *Methanosarcina thermophila* TM-1. Microbial and Comparative Genomics **2**, 275–286.

Divya B, Ammini P, Loka Bharathi PA, Shanta N (2011): 16S rRNA-based bacterial diversity in the organic-rich sediments underlying oxygen deficient waters of eastern Arabian Sea. World Journal of Microbiology and Biotechnology **27**, 2821-2833.

Dodson G, Wlodawer A (1998): Catalytic triads and their relatives. Trends in Biochemical Sciences **23**, 347–352.

Don RD, David BS, Stellwag EJ (1987): Novel Alkaline- and Heat-Stable Serine Proteases from Alkalophilic *Bacillus* sp. Strain GX6638. Journal of Bacteriology **169**, 2762-2768.

Donaghy JA, McKay AM (1993): Production and properties of an alkaline protease by *Aureobasidium pullulans*. Journal of Applied Bacteriology **74**, 662–666.

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**, 217–222.

Donovan WP, Tan Y, Slaney AC (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**, 2311-2317.

Dorado J, Field JA, Almendros G, Alvarez RS (2001): Nitrogen removal with protease as a method to improve the selective delignification of hemp stemwood by white-rot fungus *Bjerkandera* sp. strain BOS55. Applied Microbiology and Biotechnology **57**, 205–211.

Durham DR (1987): Utility of subtilisin GX as a detergent additive. Journal of Applied Bacteriology **63**, 381–386.

Durham DR, Stewart DB, Stellwag EJ (1987): Novel alkaline and heat-stable serine proteases from alkalophilic *Bacillus* sp. strain GX6638. Journal of Biotechnology **169**, 2762–2768.

Dutton RJ, Turnbaugh PJ (2012): Taking a metagenomic view of human nutrition. Current Opinion in Clinical Nutrition and Metabolic Care **15**, 448–454.

Edgcomb VP, McDonald JH, Devereux R, Smith DW (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**, 1516–1523.

Eisen JA (2007): Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. PLoS biology **5**, 384-388.

Ekici OD, Paetzel M, Dalbey RE (2008): Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. Protein Science **17**, 2023–2037.

Ekkers DM, Cretoiu MS, Kielak AM, van Elsas JD (2012): The great screen anomaly—a new frontier in product discovery through functional metagenomics. Applied and Environmental Microbiology **93**, 1005–1020.

El-Shanshoury AR, El-Sayed MA, Sammour RH, El-Shouny WA (1995): Purification and partial characterization of two extracellular alkaline

proteases from *Streptomyces corchorusii* ST36. Canadian Journal of Microbiology **41**, 99–104.

Elsied H, Stokes HW, Nakamura T, Kitamura K, Fuse H, Maruyama A (2009): Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. Environmental Microbiology **9**, 2298-2312.

Engel LS, Hill JM, Caballero AR, Green LC, O'Callaghan RJ (1998): Protease IV, a Unique Extracellular Protease and Virulence Factor from *Pseudomonas aeruginosa*. The Journal of Biological Chemistry **273**, 16792–16797.

Englard S, Seifter S (1990): Precipitation Techniques. Methods in enzymology **182**, 285-300.

Estell DA, Graycar TP, Wells JA (1985): Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. Journal of Biological Chemistry **260**, 6518–6521.

Eyers L, George I, Schuler L, Stenuit B, Agathos SN, Fantroussi SE (2004): Environmental genomics: exploring the unmined richness of microbes to degrade xenobiotics. Applied and Environmental Microbiology **66**, 123–130.

Faegri A, Torsvik VL, Goksoyr J (1977): Bacterial and fungal activities in soil: separation of bacteria and fungi by a rapid fractionated centrifugation technique. Soil Biology & Biochemistry **9**, 105–112.

Fan X, Liu X, Wang K, Wang S, Huang R, Liu Y (2011): Highly soluble expression and molecular characterization of an organic solvent-stable and thermotolerant lipase originating from the metagenome. Journal of Molecular Catalysis B: Enzymatic **72**, 319–326.

Felicioli R, Garzelli B, Vaccari L, Melfi D, Balestreri E (1997): Activity staining of protein inhibitors of proteases on gelatin-containing polyacrylamide gel electrophoresis. Analytical Biochemistry **244**, 176-179.

Felix F, Brouillet N (1966): Purification and properties of two peptidases from brewer's yeast. Biochimica et Biophysica Acta **122**, 127–144.

Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VA, Strompl C, Elborough K, Jarvis G, Neef A, Yakimov MM, Timmis KN, Golyshin PN (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environmental Microbiology **7**, 1996–2010.

Ferrero MA, Castro GR, Abate CM (1996): Thermostable alkaline protease of *Bacillus licheniformis* MIR29: isolation, production and characterization. Applied Microbiology and Biotechnology **45**, 327-332.

Fogarty WM, Griffin PJ, Joyce AM (1974): Enzymes of *Bacillus* species—Part 2. Process Biochemistry **9**, 27–35.

Fox NK, Brenner SE, Chandonia JM (2013): SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Research **42**, D304-D309.

Frederiks WM, Mook ORF (2004): Metabolic mapping of proteinase activity with emphasis on in situ zymography of gelatinases: Review and protocols. Journal of Histochemistry and Cytochemistry **52**, 711–722.

Freeman SA, Peek K, Prescott M, Daniel R (1993): Characterization of a chelator-resistant proteinase from *Thermus* strain Rt4A2. Biochemical Journal **295**, 463–469.

Friedrich J, Gradisar H, Mandin D, Chaumont JP (1999): Screening fungi for synthesis of keratinolytic enzymes. Letters in Applied Microbiology **28**:127–130.

Frostegard A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P (1999): Quantification of bias related to the extraction of DNA directly from soils. Applied and Environmental Microbiology **65**, 5409–5420.

Fujimaki M, Yamashita M, Okazawa Y, Arai S (1970): Applying proteolytic enzymes on soybean. 3. Diffusable bitter peptides and free amino acids in peptic hydrolysate of soybean protein. Journal of Food Science **35**, 215–218.

Fujinaga M, Cherney MM, Oyama H, Oda K, James MN (2004): The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. Proceedings of the National Academy of Sciences of the United States of America **101**, 3364–3369.

Fujiwara N, Masui A, Imanaka T (1993): Purification and properties of the highly thermostable alkaline protease from an alkaliphilic and thermophilic *Bacillus* sp. Journal of Biotechnology **30**, 245–256.

Fujiwara N, Tsumiya T, Katada T, Hosobuchi T, Yamamoto K (1989) Continuous recovery of silver from used X-ray films using a proteolytic enzyme. Process Biochemistry **24**, 155–156.

Fujiwara N, Yamamoto K (1987): Decomposition of gelatin layers on X-ray films by the alkaline protease from *Bacillus* sp. Hakkokogaku Kaishi-Journal of the Society of Fermentation Technology **65**, 531–34.

Fujiwara N, Yamamoto K, Masui A (1991): Utilization of a thermostable alkaline protease from an alkalophilic thermophile for the recovery of silver from used X-ray film. Journal of Fermentation and Bioengineering **72**, 306–308.

Gabor EM, Alkema WB, Janssen DB (2004a): Quantifying the accessibility of the metagenome by random expression cloning techniques. Environmental Microbiology **6**, 879–886.

Gabor EM, de Vries EJ, Janssen DB (2003): Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. FEMS Microbiology Ecology**44**, 153–163.

Gabor EM, de Vries EJ, Janssen DB (2004b): 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**, 948–958.

Gajju H, Bhalla TC, Agarwal HO (1996): Utilization of thermostable alkaline protease from *Bacillus coagulans* PB-77 for silver recovery from used X-ray films. Proceedings of 37th Annual Conference of Association of Microbiologists of India. Chennai, India, pp. 79.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003): ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research **31**, 3784-3788.

George S, Raju V, Krishnan MRV, Subramanian TV, Jayaraman K (1995): Production of protease by *Bacillus amyloliquefaciens* in solid-state fermentation and its application in the unhairing of hides and skins. Process Biochemistry **30**, 457-462.

Germano S, Pandey A, Osaku CA, Rocha SN, Soccol CR (2003): Characterization and stability of proteases from *Penicillium* sp. produced by solid-state fermentation. Enzyme and Microbial Technology **32**, 246-251.

Ghafoor AFIA, Hasnain S (2010): Purification and characterization of an extracellular protease from *Bacillus subtilis* EAG-2 strain isolated from ornamental plant nursery. Polish Journal of Microbiology **59**, 107-112.

Ghorbel-Frikha B, Sellami-Kamoun A, Fakhfakh N, Haddar A, Manni L, Nasri M (2005): Production and purification of a calcium-dependent protease from *Bacillus cereus* BG1. Journal of Industrial Microbiology and Biotechnology **32**, 186-194.

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**, 1-11.

Gilles AM, Imhoff JM, Keil B (1979): α-Clostripain. Chemical characterization, activity and thiol content of the highly active form of clostripain. Journal of Biological Chemistry **254**, 1462–1468.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002): Isolation ofantibiotics turbomycin A and B from a metagenomic library ofsoil microbial DNA. Applied and Environmental Microbiology **68**, 4301-4306.

Gillespie DE, Rondon MR, Williamson LL, Handelsman J (2005): Metagenomic libraries from uncultured microorganisms. In Osborn AM, Smith CJ (Eds): Molecular Microbial Ecology. Taylor & Francis, New York, pp. 261-279.

Glover FA (1985): Ultrafiltration and Reverse Osmosis for the Dairy Industry. Technical Bulletin. The National Institute for Research in Dairying, pp. 270

Godfrey T, West S (1996): Industrial enzymology. Macmillan Publishers, New York, pp. 3.

Gold AM, Fahrney D (1964): Sulfonyl fluorides as inhibitors of esterases. II. Formation and reactions of phenylmethanesulfonyl α-chymotrypsin. Biochemistry **3**, 783–791.

Govind NS, Mehta B, Sharma M, Modi VV (1981): Protease and carotenogenesis in *Blakeslea trispora*. Phytochemistry **20**, 2483–2485.

Gradišar H, Kern S, Friedrich J (2000): Keratinase of *Doratomyces microsporus*. Applied Microbiology and Biotechnology **53**, 196-200.

Gray JP, Herwig RP (1996): Phylogenetic analysis of the bacterial communities in marine sediments. Applied and Environmental Microbiology **62**, 4049-4059.

Graycar TP (1999): Proteolytic cleavage, reaction mechanism. In Flickinger MC, Drew SW (Eds): Bioprocess technology: fermentation, biocatalysis and bioseparation. Wiley, New York, pp. 2214–2222.

Greene RV, Cotta MA, Griffin HL (1989): A novel, symbiotic bacterium isolated from marine shipworm secretes proteolytic activity. Current Microbiology **19**, 353–356.

Greene RV, Griffin HL, Cotta MA (1996): Utility of alkaline protease from marine shipworm bacterium in industrial cleaning applications. Biotechnology Letters **18**, 759–764.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA (2009): Topographical and temporal diversity of the human skin microbiome. Science **324**, 1190–1192.

Grice EA, Kong HH, Renaud G, Young AC, Bouffard GG, Blakesley RW, Wolfsberg TG, Turner ML, Segre JA (2008): A diversity profile of the human skin microbiota. Genome Research **18**, 1043–1050.

Gupta R, Beg QK, Lorenz P (2002): Bacterial alkaline proteases: molecular approaches and industrial applications. Applied and Environmental Microbiology **59**, 15–32.

Gupta R, Gupta K, Saxena RK, Khan S (1999): Bleach-stable, alkaline protease from *Bacillus* sp. Biotechnology Letters **21**, 135–138.

Haddar A, Bougatef A, Agrebi R, Sellami-Kamoun A, Nasri M (2009): A novel surfactant-stable alkaline serine-protease from a newly isolated *Bacillus mojavensis* A21 Purification and characterization. Process Biochemistry **44**, 29–35.

Hadrys H, Balick M, Schierwater B (1992): Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecology **1**, 55–63.

Hall TA (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**, 95-98.

Hameed A, Keshavarz T, Evans CS (1999): Effect of dissolved oxygen tension and pH on the production of extracellular protease from a new isolate of *Bacillus subtilis* K2, for use in leather processing. Journal of Chemical Technology and Biotechnology **74**, 5-8.

Han J, Park CH, Ruan R (1995): Concentrating alkaline serine protease, subtilisin, using a temperature-sensitive hydrogel. Biotechnology Letters **17**, 851-852.

Handelsman J (2004): Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews **68**, 669-685.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998): Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry and Biology **5**, R245–R249.

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**, 524-534.

Hartley BS (1960): Proteolytic enzymes. Annual review of biochemistry **29**, 45-72.

Healy FG, Ray RM, Aldrich HC, Wilkie AC, Shanmugam KT (1995): Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. Applied and Environmental Microbiology **43**, 667–674.

Hellemond VEW, Janssen DB, Fraaije MW (2007): Discovery of a novel styrene monooxygenase originating from metagenome. Applied and Environmental Microbiology **73**, 5832-5839.

Henne A, Daniel R, Schmitz RA, 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**, 3901–3907.

Henne A, Schmitz RA, Bomeke 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**, 3113–3116.

Herron PR, Wellington EMH (1990): New method for extraction of *Streptomycete* spores from soil and application to the study of lysogeny in sterile amended and non-sterile soil. Applied and Environmental Microbiology **56**, 1406–1412.

Hidetoshi U, Kumiko KT, Kouichi O (1999): Microbial diversity in marine sediments from Sagami Bay and Tokyo Bay, Japan, as determined by 16S rRNA gene analysis. Microbiology **145**, 3305–3315.

Hodgson J (1994): The changing bulk biocatalyst market. Nature Biotechnology **12**, 789–790.

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**, 265-278.

Hooper NM (1994): Families of zinc metalloproteases. FEBS Letters **354**, 1-6.

Horikoshi K (1971): Production of alkaline enzymes by alkalophilic microorganisms. Agricultural and Biological Chemistry **35**, 1783-1791.

Horikoshi K (1990): Enzymes of alkalophiles. In Louis MO (Ed): Microbial enzymes and biotechnology. Elsevier Applied Science, Amsterdam, pp. 275–294.

Hotha S, Banik RM (1997): Production of alkaline protease by *Bacillus thuringiensis* H14 in aqueous two-phase systems. Journal of Chemical Technology and Biotechnology **69**, 5–10.

Hu Y, Zhang G, Li A, Chen J, Ma L (2008): Cloning and enzymatic characterization of a xylanase gene from a soil-metagenomic library with an efficient approach. Applied and Environmental Microbiology **80**, 823-830.

Huang Q, Peng Y, Li X (2003): Purification and characterization of an alkaline serine protease from dehairing function from *Bacillus pumilus*. Current Microbiology **46**, 169-173.

Hugenholtz P, Goebel BM, Pace NR (1998): Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. Journal of Bacteriology **180**, 4765–4774.

Hurt RA, Qiu X, Wu L, Roh Y, Palumbo AV, Tiedje JM, Zhou J (2001): Simultaneous recovery of RNA and DNA from soils and sediments. Applied and Environmental Microbiology **67**, 4495–4503.

Hutadilok-Towatana N, Painupong A, Suntinanalert P (1999): Purification and characterization of an extracellular protease from alkaliphilic and thermophilic *Bacillus* sp. PS719. Journal of Bioscience and Bioengineering **87**, 581-587.

Ichida JM, Krizova L, LeFevre CA, Keener HM, Elwell DL, Burtt Jr EH (2001): Bacterial inoculum enhances keratin degradation and biofilm formation in poultry compost. Journal of Microbiological Methods **47**, 199–208.

Illumina (2012): Metagenomics Research Review, An Overview of Publications Featuring Illumina® Technology.

International Union of Biochemistry (1992): Enzyme nomenclature. Academic Press, Orlando, Florida.

Ish-Horowicz D, Burke JF (1981): Rapid and efficient cosmid cloning. Nucleic acids research **9**, 2898-2989.

Ishikawa H, Ishimi K, Sugiura M, Sowa A, Fujiwara N (1993): Kinetics and mechanism of enzymatic hydrolysis of gelatin layers of X-ray film and release of silver particles. Journal of Fermentation and Bioengineering **76**, 300–305.

Ishoey T, Woyke T, Stepanauskas R, Novotny M, Lasken RS (2008): Genomic sequencing of single microbial cells from environmental samples. Current Opinion in Microbiology **11**, 198–204.

Ito S, Kobayashi T, Ara K, Ozaki K, Kawai S, Hatada Y (1998): Alkaline detergent enzymes from alkaliphiles: enzymatic properties, genetics, and structures. Extremophiles **2**, 185-190.

IUBMB (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology) (1992): Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, Academic Press, Orlando.

Jackson CR, Harper JP, Willoughby D, Roden EE, Churchill PF (1997): A simple efficient method for the separation of humic substances and DNA from environmental samples. Applied and Environmental Microbiology **63**, 4993–4995.

Jacobs M, Eliasson M, Uhlén M, Flock JI (1985): Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Research **13**, 8913-8926.

Jacobsen CS (1995): Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and PCR amplification assay. Applied and Environmental Microbiology **61**, 3347–3352.

Jacobsen CS, Rasmussen OF (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**, 2458–2462.

Jacobson JW, Glick JL, Madello KL (1985): Composition for cleaning drains clogged with deposits containing hairs. US Patent 4,540,506.

Jasmin C, Chellappan S, Sukumaran RK, Elyas KK, Bhat SG, Chandrasekaran M (2010): Molecular cloning and homology modelling of a subtilisin-like serine protease from the marine fungus, *Engyodontium album* BTMFS10. World Journal of Microbiology and Biotechnology **26**, 1269-1279.

Jellouli K, Ghorbel-Bellaaj O, Ayed HB, Manni L, Agrebi R, Nasri, M (2011): Alkaline-protease from *Bacillus licheniformis* MP1: Purification, characterization and potential application as a detergent additive and for shrimp waste deproteinization. Process Biochemistry **46**, 1248-1256.

Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, Kim SJ, Lee JH (2009): Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. Applied Microbiology and Biotechnology **81**, 865–874.

Jian S, Wenyi T, Wuyong C (2011): Kinetics of enzymatic unhairing by protease in leather industry. Journal of Cleaner Production **19**, 325-331.

Jiang W, Bond JS (1992): Families of metalloendopeptidases and their relationships. FEBS Letters **312**, 110-114.

Johnvesly B, Naik GR (2001): Studies on production of thermostable alkaline protease from thermophilic and alkaliphilic Bacillus sp. JB-99 in a chemically defined medium. Process Biochemistry **37**, 139-144.

Jones BV, Sun F, Marchesi JR (2007): Using skimmed milk agar to functionally screen a gut metagenomic library for proteases may lead to false positives. Letters in Applied Microbiology **45**, 418-420.

Joo HS, Kumar CG, Park GC, Paik SR, Chang CS (2003): Oxidant and SDS-stable alkaline protease from *Bacillus clausii* I-52: production and some properties. Journal of Applied Microbiology **95**, 267–272.

Joseph SJ, Hugenholtz P, Sangwan P, Osburne CA, Janssen PH (2003): Laboratory cultivation of widespread and previously uncultured soil bacteria. Applied and Environmental Microbiology **69**, 7210–7215.

Jung HH, Lee CH (1997): Determining isoelectric points of model proteins and *Bacillus subtilis* neutral protease by the cross partitioning using poly (ethylene glycol)/Dextran aqueous two-phase systems. Colloids and Surfaces B: Biointerfaces **9**, 131 -137.

Kaeberlein T, Lewis K, Epstein SS (2002): Isolating 'Uncultivable' microorganisms in pure culture by a simulated natural environment. Science **296**, 1127–1129.

Kakirde KS, Wild J, Godiska R, Mead DA, Wiggings AG, Goodman RM, Szybalski W, Liles MR (2011): Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. Gene **475**, 57–62.

Kalisz HM (1988): Microbial proteinases. Advances in Biochemical Engineering / Biotechnology **36**, 1–65.

Kanehisa K (2000): Woven or knit fabrics manufactured using yarn dyed raw silk. US Patent No. 6080689. US Patent and Trademark Office, Washington DC.

Kanekar PP, Nilegaonkar SS, Sarnaik SS, Kelkar AS (2002): Optimization of protease activity of alkaliphilic bacteria isolated from an alkaline lake in India. Bioresource Technology **85**, 87-93.

Karbalaei-Heidari HR, Ziaee AA, Amoozegar, MA, Cheburkin Y, Budisa N (2008): Molecular cloning and sequence analysis of a novel zinc-metalloprotease gene from the *Salinivibrio* sp. strain AF-2004 and its extracellular expression in *E. coli*. Gene **408**, 196-203.

Kato T, Yamagata Y, Arai T, Ichishima E (1992): Purification of a new extracellular 90-kDa serine proteinase with isoelectric point of 3.9 from *Bacillus subtilis* (natto) and elucidation of its distinct mode of action. Bioscience, Biotechnology, and Biochemistry **56**, 1166–1168.

Kauffmann IM, Schmitt J, Schmid RD (2004): DNA isolation from soil samples for cloning in different hosts. Applied Microbiology Biotechnology **64**, 665-670.

Keay L, Wildi BS (1970): Proteases of the genus *Bacillus*. I. Neutral proteases. Biotechnology and Bioengineering **12**, 179–212.

Kelley LA, Sternberg MJ (2009): Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols **4**, 363-371.

Kerr RA (1997): Life goes to extremes in the deep earth – and elsewhere? Science **276**, 703–704.

Kim CJ, Kim KC, Oh MJ, Choi SH (1991): Production of alkaline proteinase by the moderate halophile, *Halomonas* sp. ES10. Journal of Korean Chemical Society **34**, 307–311.

Kim CJ, Oh MJ, Choi SH (1992): Characteristics of the alkaline proteinase from the moderate halophile, *Halomonas* sp. ES10. Journal of Korean Chemical Society **35**, 237–241.

Kim HK, Hoe HS, Suh DS, Kang SC, Hwang C, Kwon ST (1999): Gene structure and expression of the gene from *Beauveria basiana* encoding bassiasin I, an insect cuticle-degrading serine protease. Biotechnology Letters **21**, 777–783.

Kim JM, Chung HS, Park SJ (1993): Properties of alkaline protease isolated from *Nocardiopsis dassonvillei*. Korean Biochemical Journal **26**, 81–85.

Kim W, Choi K, Kim Y, Park H, Choi J, Lee Y, Oh H, Kwon I, Lee S (1996): Purification and characterization of a fibrinolytic enzyme produced from *Bacillus* sp. strain CK 11-4 screened from Chungkook-Jang. Applied and Environmental Microbiology **62**, 2482–2488.

Kitayama M (1992): New low-temperature alkaline protease. Patent JP4271781.

Knietsch A, Waschkowitz T, Bowien S, Henne A, Daniel R (2003a): Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. Journal of Molecular Microbiology and Biotechnology **5**, 46–56.

Knietsch AT, Waschkowitz S, Henne AB, Daniel R (2003b): Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. Applied and Environmental Microbiology **69**, 1408–1416.

Kobayashi T, Hakamada Y, Adachi S, Hitomi J, Yoshimatsu T, Koike K, Kawai S, Ito S (1995): Purification and properties of an alkaline protease from alkalophilic *Bacillus* sp. KSM-K16. Applied Microbiology and Biotechnology **43**, 473-481.

Kobayashi T, Hakamada Y, Hitomi J, Koike K, Ito S (1996): Purification of alkaline proteases from a *Bacillus* strain and their possible interrelationship. Applied Microbiology and Biotechnology **45**, 63–71.

Kobayashi T, Ogasawara A, Ito S, Saitoh M (1985): Purification and some properties of alkaline proteinase produced by *Pseudomonas maltophila*. Agricultural Biology and Chemistry **49**, 693–698.

Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y, Doolittle WF (2009): Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds. PLoS ONE **24**, e5276.

Kohno D, Sakiyama Y, Lee S, Takiguchi N, Mitsutani A, Kitaguchi H, Kato J (2007): Cloning and Characterization of A Gene Encoding Algicidal Serine Protease from *Pseudoalteromonas* sp. Strain A28. Journal of Environmental Biotechnology **7**, 99–102.

Kong HH, Oh J, Deming C, Conlan S, Grice EA, Beatson MA, Nomicos E, Polley EC, Komarow HD, NISC Comparative Sequence Program, Murray PR, Turner ML, Segre JA (2012): Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. Genome research **22**, 850–859.

Kowalchuk GA, Buma DS, de Boer W, Klinkhamer PG, van Veen JA (2002): Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. Antonie Van Leeuwenhoek **81**, 509–520.

Krsek M, Wellington EMH (1999): Comparison of different methods for the isolation and purification of total community DNA from soil. Journal of Microbiological Methods **39**, 1–16.

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009): Coding sequence determinants of gene expression in *Escherichia coli*. Science, **324**, 255–258.

Kudrya VA, Simonenko IA (1994): Alkaline serine proteinase and lectin isolation from the culture fluid of *Bacillus subtilis*. Applied Microbiology and Biotechnology **41**, 505–509.

Kumar CG (1997): Studies on microbial alkaline proteases for use in dairy detergents. Ph.D. Thesis. National Dairy Research Institute (Deemed University), Karnal, India.

Kumar CG (2002): Purification and characterization of a thermostable alkaline protease from alkalophilic *Bacillus pumilus*. Letters in Applied Microbiology **34**, 13-17.

Kumar CG, Malik RK, Tiwari MP (1998): Novel enzyme-based detergents: an Indian perspective. Current Science **75**, 1312–1318.

Kumar CG, Takagi H (1999): Microbial alkaline proteases: from a bioindustrial viewpoint. Biotechnology Advances **17,** 561-594.

Kumar CG, Tiwari MP (1994): Certain cultural characteristics of alkaline protease producing actinomycetes. Proceedings of Micon International-94. Mysore, India, pp. 93–94.

Kumar CG, Tiwari MP, Jany KD (1997): Purification and characterization of two alkaline proteases from an alkalophilic *Bacillus* sp. Zeitschrift für Ernährungs-Wissenschaft **36**:48-48.

Kumar PP (2011): An improved method for extraction of heterologous DNA from environmental samples for the construction of metagenomic libraries. International Journal of Pharma and Bio Sciences **12**, 368-373.

Kumar S, Dudley J (2007): Bioinformatics software for biologists in the genomics era. Bioinformatics **23**, 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**, 299-306.

Kumar S, Tamura K, Nei M (1994): MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Computer Applications in the Biosciences **10**, 189-191.

Kunitz M (1947): Crystalline soybean trypsin inhibitor II. General properties. The Journal of general physiology **30**, 291-310.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V (1997): The complete genome sequence of the gram positive bacterium *Bacillus subtilis*. Nature **390**, 249–256.

Kuo LC, Shafer JA (1994): Retroviral proteases. Methods in Enzymology **241**, 3–178.

Kurata A, Uchimura K, Kobayashi T, Horikoshi K (2010): Collagenolytic subtilisin-like protease from the deep-sea bacterium *Alkalimonas collagenimarina* AC40T. Applied Microbiology and Biotechnology **86**, 589-598.

Kuske CR, Banton KL, Adorada DL, Stark PC, Hill KK, Jackson PJ (1998): Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. Applied and Environmental Microbiology **64**, 2463–2472.

Kwon YT, Kim JO, Moon SY, Lee HH, Rho HM (1994): Extracellular alkaline protease from alkalophilic *Vibrio metschnikovii* strain RH530. Biotechnology Letters **16**, 413–418.

Laan JCVD, Gerritse G, Mulleners LJ, Hoek VDRA, Quax WJ (1991): Cloning, characterization and multiple chromosomal integration of a *Bacillus*

alkaline protease gene. Applied and Environmental Microbiology 57, 901-909.

Labbe JP, Rebegrotte P, Turpine M (1974): Demonstrating extra-cellular leucine aminopeptidase (EC 3.4.1.1) of Aspergillus oryzae (IP 410): leucine aminopeptidase 2 fraction. Proceedings of the Academy of Sciences, Paris, 278D.

Laemmli UK (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Lakay FM, Botha A, Prior BA (2007) Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. Journal of Applied Microbiology **102**, 265-273.

Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, Rupp S (2007): Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. Journal of Biotechnology **127**, 575-592.

LaMontagne MG, Michel FC Jr, Holden PA, Reddy CA (2002): Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis Journal of Microbiological Methods **49**, 255–264.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985a): Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences of the United States of America **82,** 6955-6959.

Lane DJ, Stahl DA, Olsen GJ, Heller DJ, Pace NR (1985b): Phylogenetic analysis of the genera *Thiobacillus* and *Thiomicrospira* by 5S rRNA sequences. Journal of Biotechnology **163,** 75-81.

Larcher G, Cimon B, Symoens F, Tronchin G, Chabasse D, Bouchara JP (1996): A 33 kDa serine proteinase from *Scedosporium apiospermum*. Biochemical Journal **315**, 119–126.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007): Clustal W and Clustal X version 2.0.Bioinformatics **23**, 2947-2948.

Larsen KS, Auld DS (1991): Characterization of an inhibitory metal binding site in carboxypeptidase A. Biochemistry **30**, 2613–2618.

Lasken RS (2009): Genomic DNA amplification by the multiple displacement amplification (MDA) method. Biochemical Society Transactions **37**, 450–453.

LeCleir GR, Buchan A, Maurer J, Moran MA, Hollibaugh JT (2007): Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. Environmental Microbiology **9**, 197– 205.

Lee DG, Jeon JH, Jang MK, Kim NY, Lee JH, Kim SJ, Kim GD, Lee SH (2007): Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. Biotechnology Letters **29**, 465-472.

Lee SE, James MH, Armando RC, Linda CG, Richard JC (1998): Protease IV, a Unique Extracellular Protease and Virulence Factor from *Pseudomonas aeruginosa*. The Journal of Biological Chemistry **273**, 16792–16797.

Lee SW, Won K, Lim HK (2004): Screening for novel lipolytic enzymes from uncultured soil microorganisms. Applied and Environmental Microbiology **65**, 720–726.

Lee SY, 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**, 3787–3793.

Lee YH, Chang HN (1990): Production of alkaline protease by *Bacillus licheniformis* in an aqueous two-phase system. Journal of Fermentation and Bioengineering **69**, 89–92.

Leff LG, Dana JR, McArthur VJ, Shimkets LJ (1995): Comparison of Methods of DNA Extraction from Stream Sediments. Applied and Environmental Microbiology **61**, 1141-1143.

Li S, Xu LH, Hua H, Ren CA, Lin ZL (2007): A set of UV-inducible autolytic vectors for high throughput screening. Journal of Biotechnology **127**, 647–652.

Li Y, Wexler M, Richardson DJ, Bond PL, Johnston AWB (2005): Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. Environmental Microbiology **7**, 1927-1936.

Liao L, Xu XW, Wang CS, Zhang DS, 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**, 890–901.

Liesack W, Weyland H, Stackebrandt E (1991): Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. Microbial Ecology **21**, 191–198.

Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. Applied and Environmental Microbiology **71**, 7768–7777.

Lindahl V, Bakken LR (1995): Evaluation of methods for extraction of bacterial from soil. FEMS Microbiology Ecology **16**, 135–142.

Linderstrøm-Lang K, Ottesen M (1947): A new protein from ovalbumin. Nature **159**, 807-808.

Litthauer D, Abbai NS, Piater LA, Heerden EV (2010): Pitfalls using tributyrin agar screening to detect lipolytic activity in metagenomic studies. African Journal of Biotechnology **9**, 4282-4285.

Liu CL, Beck CM, Strobel RJ, Overholt JM (1988): Low-temperature active alkaline protease from *Nocardiopsis dassonvillei* and its preparation. Patent Cooperation Treat, Application WO 8803947.

Liu Z, Desantis TZ, Andersen GL, Knight R (2008): Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Research **36**, e120.

Lozupone C, Hamady M, Knight R (2006): UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics **7**, 371.

Lozupone C, Knight R (2005): UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology **71**, 8228–8235.

Lu AYH, Junk KW, Coon MJ (1969): Resolution of the cytochrome P-450 containing ω-Hydroxylation system of liver microsomes into three components. Journal of Biological Chemistry **244**, 3714–3721.

Lu JJ, Perng CL, Lee SY, Wan CC (2000): Use of PCR with universal primers and restriction endonuclease digestions for detection and identification of common bacterial pathogens in cerebrospinal fluid. Journal of Clinical Microbiology **38**, 2076–2080.

Lu Y, Abraham WR, Conrad R (2007): Spatial variation of active microbiota in the rice rhizosphere revealed by *in situ* stable isotope probing of phospholipid fatty acids. Environmental Microbiology **9**, 474–481.

Luciano P, Tokatlidis K, Chambre I, Germanique JC, Geli V (1998): The mitochondrial processing peptidase behaves as a zinc-metallopeptidase. Journal of Molecular Biology **280**, 193–199.

Ludwig W, Strunk O, Westram R, Richter L, Meier H (2004): ARB: a software environment for sequence data. Nucleic Acids Research **32**, 1363–1371.

Lyimo TJ, Pol A, Harhangi HR, Jetten MSM, den Camp HJMO (2009): Anaerobic oxidation of dimethylsulfide and methanethiol in mangrove sediments is

dominated by sulfate-reducing bacteria. FEMS Microbiology Ecology **70**, 483–492.

MaaritNiemi R, Heiskanen I, Wallenius K, Lindstrom K (2001): Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. Journal of Microbiological Methods **45**, 155–165.

Maccaferri S, Biagi E, Brigidi P (2001): Metagenomics: key to human gut microbiota. Digestive Diseases **29**, 525–530.

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D, Osburne MS (2001): Expression and isolation of antimicrobial small molecules from soil DNA libraries. Journal of Molecular Microbiology and Biotechnology **3**, 301–308.

Madigan MT, Marrs BL (1997): Extremophiles. Scientific American **276**, 82-87.

Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR (1996): The Ribosomal Database Project (RDP). Nucleic Acids Research **24**, 1.

Mala M, Srividya S (2010): Partial purification and properties of a laundry detergent compatible alkaline protease from a newly isolated *Bacillus* species Y. Indian J Microbiol **50**, 309-317.

Malathi S, Chakraborty R (1991): Production of alkaline protease by a new *Aspergillus flavus* isolate under solid substrate fermentation conditions for use as a depilation agent. Applied and Environmental Microbiology **57**, 712–716.

Manachini PL, Fortina MG, Parini C (1988): Thermostable alkaline protease produced by *Bacillus thermoruber*—a new species of *Bacillus*. Applied Microbiology and Biotechnology **28**, 409–413.

Manichanh C, Rigottier-Gois L, Bonnaud E (2006): Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut **55**, 205–211.

Manonmani HK, Joseph R (1993): Purification and properties of an extracellular proteinase of *Trichoderma koningii*. Enzyme and microbial technology **15**, 624-628.

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH (2013): CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Research **1**, 41 -52.

Margesin R, Palma N, Knauseder F, Schinner F (1992): Purification and characterization of an alkaline protease produced by a psychrotropic *Bacillus* sp. Journal of Biotechnology **24**, 203–206.

Margesin R, Schinner F (1994): Properties of cold-adapted microorganisms and their potential role in biotechnology. Journal of Biotechnology **33**, 1–14.

Mark AS, Sophia YD, Charles FK (2003): Direct Extraction of DNA from Soils for Studies in Microbial Ecology. Current Issues in Molecular Biology **5**, 1-8.

Markland FS, Smith EL (1971): Subtilisins: primary structure and physical properties. In Boyer PD (Ed): The enzymes. Vol 3. Academic Press, New York. pp 561-608.

Martin AP (2002): Phylogenetic approaches for describing and comparing the diversity of microbial communities. Applied and Environmental Microbiology **68**, 3673–3682.

Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA (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**, 2452–2463.

Masse FWJL, Tilburg R (1983): The benefit of detergent enzymes under changing washing conditions. Journal of the American Oil Chemists' Society **60**, 1672–1675.

Matsubara H, Feder J (1971): The Enzyme. Vol. 3, Academic Press, New York.

Matsui T, Matsufuji H, Seki E, Osajima K, Nakashima M, Osajima Y (1993): Inhibition of angiotensin I-converting enzyme by *Bacillus licheniformis* alkaline protease hydrolysates derived from sardine muscle. Bioscience Biotechnology Biochemistry **57**, 922–925.

Matsumoto K, Yamamoto T, Kamata R, Maeda (1984): Pathogenesis of serratial infection: activation of the Hageman factor-prekallikrein cascade by serratial protease. The Journal of Biochemistry **96**, 739–749.

Matthews BW (1988): Structural basis of the action of thermolysin and related zinc peptidases. Accounts of Chemical Research **21**, 333–340.

McCabe KM, Zhang YH, Huang BL, Wagar EA, McCabe ER (1999): Bacterial species identification after DNA amplification with a universal primer pair. Molecular Genetics and Metabolism **66**, 205–211.

McGuffin JL, Bryson K, Jones DT (2000): The PSIPRED protein structure prediction server. Bioinformatics **16**, 404–405.

Menon AS, Goldberg AL (1987): Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP, Journal of Biological Chemistry **262**, 14929–14934.

Mikami Y, Miyashita K, Arai T (1986): Alkalophilic actinomycetes. Actinomycetes **19**, 76–191.

Miller DN (2001): Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. Journal of Microbiological Methods **44**, 49–58.

Miller DN, Bryant JE, Madsen EL, Ghiorse WC (1999): Evaluation and optimization of DNA extraction and purification procedures for soil and

sediment samples. Applied and Environmental Microbiology **65**, 4715–4724.

Mills DEK, Entry JA, Gillevet PM, Mathee K (2007): Assessing microbial community diversity using amplicon length heterogeneity polymerase chain reaction. Soil Science Society of America Journal **71**, 572–578.

Mirete S, de Figueras CG, Gonza` lez-Pastor JE (2007): Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. Applied and Environmental Microbiology **73**, 6001-6011.

Mizusawa K, Ichishima E, Yoshida F (1964): Studies on the proteolytic enzymes of thermophilic *Streptomyces*. I. Purification and some properties. Agricultural and Biological Chemistry **28**, 884–895.

Mizusawa K, Ichishima E, Yoshida F (1966): Studies on the proteolytic enzymes of thermophilic *Streptomyces.* II. Identification of the organism and some conditions of protease formation. Agricultural and Biological Chemistry **30**, 35–41.

Mocali S, Benedetti A (2010): Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. Research in Microbiology **161**, 497-505.

Moncrief JS, Obis Jr R, Barroso LA, Kling JJ, Wright RL, VanTassell RL, Lyerly DM, Wilkins TD (1995): The enterotoxin of *Bacteroides fragilis* is a metalloprotease. Infection and Immunity **63**, 175–181.

More MI, Herrick JB, Silva MC, Ghiorse WC, Madsen EL (1994): Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Applied and Environmental Microbiology **60**, 1572–1580.

Morihara K (1963): *Pseudomonas aeruginosa* proteinase. 1. Purification and general properties. Biochimica et Biophysica Acta **73**, 113–124.

Morita Y, Hasan Q, Sakaguchi T, Murakami Y, Yokoyama K, Tamiya E (1998): Properties of a cold-active protease from psychrotrophic

Flavobacterium balustinum P104. Applied Microbiology and Biotechnology **50**, 669-675.

Morohoshi T, Oikawa M, Sato S, Kikuchi N, Kato N, Ikeda T (2011): Isolation and characterization of novel lipases from a metagenomic library of the microbial community in the pitcher fluid of the carnivorous plant *Nepenthes hybrida*. Journal of bioscience and bioengineering **112**, 315–320.

Moyer CL, Dobbs FC, Karl DM (1995): Phylogenetic diversity of the bacterial community from a microbial mat at an active hydrothermal vent system, Loichi Seamount, Hawaii. Applied and Environmental Microbiology **61**, 1555-1562.

Mukhopadhyay RP, Chandra AL (1993): Protease of a keratinolytic *Streptomycete* to unhair goat skin. Indian Journal of Experimental Biology **31**, 557–558.

Munson MA, Banerjee A, Watson TF, Wade WG (2004): Molecular analysis of the microflora associated with dental caries. Journal of Clinical Microbiology **42**, 3023–3029.

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995): SCOP: a structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology **247**, 536–540.

Najafi MF, Deobagkar D, Deobagkar D (2005): Potential application of protease isolated from *Pseudomonas aeruginosa* PD100. Electronic journal of biotechnology **8**, 79-85.

Nakanishi T, Matsumura Y, Minamiura N, Yamamoto T (1974): Purification and some properties of an alkalophilic proteinase of a *Streptomyces* species. Agricultural and Biological Chemistry **38**, 37–44.

Nalin R, Normand P, Simonet P, Domenach AM (1999): Polymerase chain reaction and hybridization on DNA extracted from soil as a tool for *Frankia* spp. population distribution studies in soil. Canadian Journal of Botany **77**, 1239–1247.

Nalin R, Robe RP, Tran VV (2001): Method for indirectly extracting non-cultivable DNA organisms and DNA by said method. France, Patent 01/11/2001.

Neelakanta G, Sultana H (2013): The Use of Metagenomic Approaches to Analyze Changes in Microbial Communities. Microbiology Insights **6**, 37–48.

Neklyudov AD, Ivankin AN, Berdutina AV (2000): Properties and uses of protein hydrolysates (Review). Applied Biochemistry and Microbiology **36**, 452-459.

Neurath H (1989): Proteolytic processing and physiological regulation. Trends in biochemical sciences **14**, 268–271.

Neveu J, Regeard C, DuBow MS (2011): Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. Applied Microbiology and Biotechnology **91**, 635-644.

Nielsen MH, Jepsen SJ, Outtrup H (1981): Enzymes for low temperature washing. Journal of the American Oil Chemists' Society **58**, 644–649.

Nilegaonkar SS, Zambare VP, Kanekar PP, Dhakephalkar PK, Sarnaik SS (2007): Production and partial characterization of dehairing protease from *Bacillus cereus* MCM B-326. Bioresource technology **98**, 1238-1245.

Nocker A, Burr M, Camper AK (2007): Genotypic microbial community profiling: A critical technical review. Microbial Ecology **54**, 276–289.

Normurodova KT, Makhsumkhanov AA, Alimova BK, Pulatova OM, Bozorov NI (2010): Isolation, purification, and characterization of highly active protease from *Bacillus subtilis*-150. Chemistry of Natural Compounds **46**, 833-834.

O'Meara GM, Munro PA (1984a): Selection of a proteolytic enzyme to solubilize lean beef meat tissue. Enzyme and Microbial Technology **6**, 181–185.

O'Meara GM, Munro PA (1984b): Effects of reaction variables on the hydrolysis of lean beef tissue by Alcalase. Meat Science **11**, 227–238.

Oberoi R, Beg QK, Puri S, Saxena RK, Gupta R (2001): Characterization and wash performance analysis of an SDS-resistant alkaline protease from a *Bacillus* sp. World Journal of Microbiology and Biotechnology **17**, 493–497.

Oda K (2012): New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases. Journal of Biochemistry **151**, 13–25.

Ogino H, Watanabe F, Yamada M, Nakagawa S, Hirose T, Noguchi A, Yasuda M, Ishikawa H (1999): Purification and characterization of organic solvent-stable protease from organic solvent-tolerant *Pseudomonas aeruginosa* PST-01. Journal of Bioscience and Bioengineering **87**, 61-68.

Ogram A, Sayler GS, Barkay T (1987): The extraction and purification of microbial DNA from sediments. Journal of Microbiological Methods **7**, 57–66.

Ogrydziak DM (1993): Yeast extracellular proteases. Critical Reviews in Biotechnology **13**, 1–55.

Oguma K, Fujinaga Y, Inoue K (1995): Structure and function of *Clostridium botulinum* toxins. Microbiology and Immunology **39**, 161–168.

Ohmiya K, Tanimura S, Kobayashi T, Shimizu S (1979): Application of immobilized alkaline protease to cheese-making. Journal of Food Science **44**, 1584–1588.

Ong PS, Gaucher M (1976): Production, purification and characterization of thermomycolase, the extracellular serine protease of the thermophilic fungus *Malbranchea pulchella* var. *sulfurea*. Canadian Journal of Microbiology **22**, 165–176.

Ono A, Miyazaki R, Sota M, Ohtsubo Y, Nagata Y, Tsuda M (2007): Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. Applied Microbiology and Biotechnology **74**, 501-510.

Orcutt BN, Joye SB, Kleindienst S, Knittel K, Ramette A, Reitz A, Samarkin V, Treude T, 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 **57**, 2008-2021.

Orhan E, Omay D, Guvenilir Y (2005): Partial purification and characterization of protease enzyme from *Bacillus subtilis* and *Bacillus cereus*. Applied Biochemistry and Biotechnology **121**, 183–194

Orsini M, Romano-Spica V (2001): A microwave-based method for nucleic acid isolation from environmental samples. Letters in Applied Microbiology **33**, 17–20.

Ottesen M, Svenden I (1971): The subtilisins. Methods of Enzymology **19**, 199-215.

Oyama H, Kinjoh M, Watari M, Murao S (1997): Purification and characterization of an alkaline proteinase produced by *Pimelobacter* sp. Z-483. Journal of Fermentation and Bioengineering **84**, 351-353.

Pace NR (2009): Mapping the Tree of Life: Progress and Prospects. Microbiology and Molecular Biology Reviews **73**, 565–576.

Pace NR, Stahl DA, Lane DJ, Olsen GJ (1985): Analyzing natural microbial populations by rRNA sequences. ASM News **51**, 4–12.

Paliwal N, Singh SP, Garg SK (1994): Cation-induced thermal stability of an alkaline protease from a *Bacillus* sp. Bioresource Technology **50**, 209–211.

Pan T, Lin S (1991): Fermentative production of alkaline protease as detergent additive. Journal of Chinese Biochemical Society **20**, 49–60.

Pecs M, Eggert M, Schügerl K (1991): Affinity precipitation of extracellular microbial enzymes. Journal of Biotechnology **21**, 137–142.

Pedersen KB, Christiansen M, Lindegaard P (1992): Novel proteases. Patent Cooperation Treaty Application WO 9218622.

Peer VY, Chapelle S, De Wachter (1996): A quantitative map of nucleotide substitution rates in bacterial rRNA. Nucleic Acids Research **24**, 3381–3391.

Pel J, Broemeling D, Mai L, Poon H, Tropini G, Holt RA,Warren R, Marziali A (2009): Non linear electrophoretic response yields a unique parameter for separation of biomolecules. Proceedings of the National Academy of Sciences of USA **106**, 14796–14801.

Peng Y, Yang XJ, Xiao L, Zhang YZ (2004): Cloning and expression of a fibrinolytic enzyme (subtilisin DFE) gene from Bacillus amyloliquefaciens DC-4 in Bacillus subtilis. Research in microbiology 155, 167-173.

Perea A, Ugalde U, Rodriguez I, Serra JL (1993): Preparation and characterization of whey protein hydrolysates: application in industrial whey bioconversion processes. Enzyme and Microbial Technology 15, 418–423.

Petinate SDG, Branquinha MH, Coelho RRR, Vermelho AB, Giovanni-De-SimoneS (1999): Purification and partial characterization of an extracellular serine-proteinase of Streptomyces cyaneus isolated from Brazilian cerrado soil. Journal of Applied Microbiology 87, 557-563.

Phadatare SU, Deshpande VV, Srinivasan MC (1993): High activity alkaline protease from *Conidiobolus coronatus* (NCL 86.8.20): Enzyme production and compatibility with commercial detergents. Enzyme Microbial Technology **15**, 72–76.

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**, 2717–2722.

Poindexter K, Nelson N, DuBose RF, Black RA, Cerretti DP (1999): The identification of seven metalloproteinase-disintegrin (ADAM) genes from genomic libraries. Gene **237**, 61–70.

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 47, 228–235.

Polymenakou PN, Bertilsson S, Tselepides A, Stephanou EG (2005): Bacterial community composition in different sediments from the Eastern Mediterranean Sea: a comparison of four 16S ribosomal DNA clone libraries. Microbial Ecology **50**,447-462.

Porteous LA, Armstrong JL (1991): Recovery of bulk DNA from soil by a rapid, small-scale extraction method. Current Microbiology **22**, 345–348.

Porteous LA, Seidler RJ, Watrud LS (1997): An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. Molecular Ecology **6**, 787–791.

Power PM, Jones RA, Beacham IR, Bucholtz C, Jennings MP (2004): Whole genome analysis reveals a high incidence of non-optimal codons in secretory signal sequences of *Escherichia coli*. Biochemical and Biophysical Research Communications **322**,1038–1044.

Prabavathi R, Mathivanan V, Ambika A (2012): Screening of Protease Enzyme By Construction of Metagenomic Library From Marine Soil Sediments. International Journal of Pharma Sciences and Research (IJPSR) 3, 396-399.

Prakash T, Khandelwal M, Dasgupta D, Dash D, Brahmachari SK (2004): CoPS: comprehensive peptide signature database. Bioinformatics **20**, 2886-2888.

Prakash T, Oshima K, Morita H, Fukuda S, Imaoka A, Kumar N, Sharma VK, Kim SW, Takahashi M, Saitou N, Taylor TD, Ohno H, Umesaki Y, Hattori M. (2011): Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of th17 cell differentiation. Cell host & microbe **10**, 273-284.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W (2007): SILVA: a comprehensive online resource for quality checked and aligned ribosomal

RNA sequence data compatible with ARB. Nucleic Acids Research **35**, 7188–7196.

Purdy KJ, Embley TM, Takii S, Nedwell DB (1996): Rapid extraction of DNA and rRNA from sediments by a novel hydroxyapatite spin-column method. Applied and Environmental Microbiology **62**, 3905–3907.

Pushpam P, Rajesh T, Gunasekaran P (2011):Identification and characterization of alkaline serine protease from goat skin surface metagenome. AMB Express **1**, 3.

Rahaman RS, Chee JY, Cabral JMS, Hatton TA (1988): Recovery of an extracellular alkaline protease from whole fermentation broth using reversed micelles. Biotechnology Progress **4**, 218–224.

Rahman RNZA, Razak CN, Ampon K, Basri M, Yunus WMZW, Salleh AB (1994): Purification and characterization of a heat-stable alkaline protease from *Bacillus stearothermophilus* F1. Applied and Microbiology and Biotechnology **40**, 822–827.

Rai SK, Mukherjee AK (2009): Ecological significance and somebiotechnological application of an organic solvent stablealkaline serine protease from *Bacillus subtilis* strain DM-04. Bioresource Technology **100**, 2642-2645.

Rajamani S, Hilda A (1987): Plate assay to screen fungi for proteolytic activity. Current Science **56**, 1179-1181.

Rajan M (2004) Global market for industrial enzymes to reach $2.4 million by 2009 Business Communications Company, Inc. RC-147U Enzymes for Industrial Applications.

Ramsay AJ (1984): Extraction of bacteria from soil: efficiency of shaking or ultrasonication as indicated by direct counts and autoradiography. Soil Biology and Biochemistry **16**, 457–481.

Ranjan R, Grover A, Kapardar RK, Sharma R (2005): Isolation of novel lipolytic genes from uncultured bacteria of pond water. Biochemistry and Biophysics Research Communications **335**, 57-65.

Rao MB, Tanksale AM, Ghatge MS, Deshpande VV (1998): Molecular and biotechnological aspects of microbial proteases. Microbiology and molecular biology reviews **62**, 597-635.

Rappe MS, Connon SA, Vergin KL, Giovannoni SJ (2002): Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature **418**, 630–633.

Rappe MS, Giovannoni SJ (2003): The uncultured microbial majority. Annual Review of Microbiology **57**, 369–394.

Rattray FP, Bockelmann W, Fox PF (1995): Purification and Characterization of an Extracellular Proteinase from Brevibacterium linens ATCC 9174. Applied and Environmental Microbiology **61**, 3454-3456.

Raut S, Sen SK, Kabir NA, Satpathy S, Raut S (2012): Isolation and characterization of protease producing bacteria from upper respiratory tract of wild chicken Bioinformation **8**, 326-330.

Rawlings ND, Barrett AJ (1993): Evolutionary families of peptidases. Biochemical Journal **290**, 205–218.

Rawlings ND, Barrett AJ (2012): Introduction: aspartic and glutamic peptidases and their clans. In Rawlings ND, Salvesen GS (Eds): Handbook of Proteolytic Enzymes. Elsevier, Amsterdam, pp. 3-19.

Rawlings ND, Barrett AJ, Bateman A (2010): MEROPS: the peptidase database. Nucleic Acids Research **38**, D227-D233.

Rawlings ND, Barrett AJ, Bateman A (2011): Asparagine Peptide Lyases: A seventh catalytic type of proteolytic enzymes. Journal of Biological Chemistry 286, 38321-38328.

Rawlings ND, Barrett AJ, Bateman A (2012): MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Research **40**, 343-350.

Rebeca BD, Pena-Vera MT, Diaz-Castaneda M (1991): Production of fish protein hydrolysates with bacteriaproteases; yield and nutritional value. Journal of Food Science **56**, 309–314.

REFERENCES

Rhee JK, Ahn DG, Kim YG, Oh JW (2005): New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. Applied and Environmental Microbiology **71**,817–825.

Richardson TH, Tan X, Frey G, Callen W, Cabell M, Lam D, Macomber J, Short JM, Robertson DE, Miller C (2002): A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase Journal of Biological Chemistry **277**, 26501–26507.

RiesenfeldCS, Schloss PD, HandelsmanJ (2004): Metagenomics: genomic analysis of microbial communities. Annual Review of Genetics **38**, 525-552.

Rincon-Florez VA, Carvalhais LC, SchenkPM (2013): Culture-Independent Molecular Tools for Soil and Rhizosphere Microbiology. Diversity **5**, 581-612.

Robe P, Nalin R, Capellano C, Vogel TA, Simonet P (2003): Extraction of DNA from soil. European Journal of Soil Biology **39**,183–190.

Robertson DE, Chaplin JA, DeSantis G, Podar M, Madden M, Chi E, Richardson T, Milan A, Miller M, Weiner DP, Wong K, McQuaid J, Farwell B, Preston LA, Tan X, Snead MA, Keller M, Mathur E, Kretz PL, Burk MJ, Short JM (2004): Exploring nitrilase sequence space for enantioselective catalysis. Applied and Environmental Microbiology **70**, 2429–2436.

Rochelle PA, Fry JC, Parkes RJ, Weightman AJ (1992): DNA extraction for 16S rRNA gene analysis to determine genetic diversity in deep sediment communities. FEMS Microbiology Letters **79**, 59–65.

Romero FJ , García LA , Salas JA , Díaz M , Quirós LM (2001): Production, purification and partial characterization of two extracellular proteases

from Serratia marcescens grown in whey. Process Bioochemistry **36**, 507-515.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C (2000): Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Applied and Environmental Microbiology **66**, 2541-2547.

Rondon MR, Raffel SJ, Goodman RM, 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 of USA **96**, 6451–6455.

Roose-Amsaleg CL, Garnier-Sillam E, Harry M (2001): Extraction and purification of microbial DNA from soil and sediment samples. Applied Soil Ecology **18**, 47-60.

Rothschild LJ, Mancinelli RL (2001): Life in extreme environments. Nature **49**, 1092–1101.

Roy S, Hens D, Biswas D, Biswas D, Kumar R (2002): Survey of petroleum-degrading bacteria in coastal waters of Sunderban Biosphere Reserve. World Journal of Microbiology and Biotechnology **18**, 575-581.

Ryu K, Kim J, Dordick JS (1994): Catalytic properties and potential of an extracellular protease from an extreme halophile. Enzyme and Microbial Technology **16**, 266–275.

Saano A, Tas E, Pippola S, Lindstrom K, Van Elsas JD (1995): Extraction and analysis of microbial DNA from soil. In Van Elsas JD, Trevors JT (Eds): Nucleic Acids in the Environment: Methods and Applications. Springer-Verlag, Heidelberg, Germany, pp. 49–67.

Sadeghi HMM, Rabbani M, Naghitorabi M (2009): Cloning of alkaline protease gene from *Bacillus subtilis* 168. Research in Pharmaceutical Sciences **4**, 43-46.

Saeki K, Hitomi J, Okuda M, Hatada Y, Kageyama Y, Takaiwa M, Kubota H et al. (2002): A novel species of alkaliphilic *Bacillus* that produces an oxidatively stable alkaline serine protease. Extremophiles **6**, 65-72.

Saitou N, Nei M (1987): The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4**, 406-425.

Sakiyama T, Toyomasu T, Nagata A, Imamura K, Takahashi T, Nagai T, Nakanishi K (1998): Performance of protease as a cleaning agent for stainless steel surfaces fouled with protein. Journal of Fermentation and Bioengineering **55**, 297–301.

Samal BB, Karan B, Parker C, Stabinsky Y (1991): Isolation and thermal stability of two novel serine proteinases from the fungus *Tritirachium album* Limber. Enzyme and Microbial Technology **13**, 66–70.

Samal BB, Karan B, Stabinsky Y (1990): Stability of two novel serine proteinases in commercial laundry detergent formulations. Biotechnology and Bioengineering **35**, 650–652.

Sambrook J, Fritsch E, Maniatis (2000): Molecular cloning, A laboratory manual. C. S. H. Laboratory, USA.

Sana B, Ghosh D, Saha M, Mukherjee J (2006): Purification and characterization of a salt, solvent, detergent and bleach tolerant protease from a new gamma-Proteobacterium isolated from the marine environment of the Sundarbans. Process Biochemistry **41**, 208-215.

Sarika S, Farrah GK, Ghulam NQ (2010): Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. Applied Microbiology and Biotechnology **86**, 1821–1828

Saurabh S, Jasmine I, Pritesh G. Kumar R (2007): Enhanced productivity of serine alkaline protease by *Bacillus* sp. using soybean as substrate. Malaysian Journal of Microbiology **3**, 1-6.

Schipper C, Hornung C, Bijtenhoorn P, Quitschau M, Grond S, Streit W (2009): Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*. Applied and Environmental Microbiology **75**, 224–233.

Schloss PD (2008): Evaluating different approaches that test whether microbial communities have the same structure. ISME Journal **2**, 265–275.

Schloss PD, Handelsman J (2005): Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Applied and Environmental Microbiology **71**, 1501–1506.

Schloss PD, Handelsman J (2006a): Introducing SONS, a tool that compares the membership of microbial communities. Applied and Environmental Microbiology **72**, 6773–6779.

Schloss PD, Handelsman J (2006b): Introducing TreeClimber, a test to compare microbial community structure. Applied and Environmental Microbiology **72,** 2379–2384.

Schloss PD, Larget BR, Handelsman J (2004): Integration of microbial ecology and statistics: a test to compare gene libraries. Applied and Environmental Microbiology **70**, 5485–5492.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009): Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology **75**, 7537-7541.

Schmidt TM, Bleakley B, Nealson KH (1988): Characterization of an extracellular protease from the insect pathogen *Xenorhabdus luminescens*. Applied and Environmental Microbiology **54**, 2793-2797.

Schmidt TM, DeLong EF, Pace NR (1991): Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. Journal of Biotechnology **173**, 4371–4378.

Schneegurt MA, Dore SY, Kulpa Jr CF (2003): Direct extraction of DNA from soils for studies in microbial ecology. Current issues in molecular biology **5**, 1-8.

Schulz EC, Ficner R (2011): Knitting and snipping: chaperones in β-helix folding. Current opinion in Structural Biology **21**, 232-239.

Seemüller E, Lupas A, Stock D, Löwe J, Huber R, Baumeister W (1995): Proteasome from *Thermoplasma acidophilum*: a threonine protease. Science **268**, 579–582.

Sellami-Kamoun A, Haddar A, Ali N, Ghorbel-Frikha B, Kanoun S, Nasri M (2008): Stability of thermostable alkaline protease from *Bacillus licheniformis* RP1 in commercial solid laundry detergent formulations. Microbiological Research **163**, 299-306.

Shah K, Mody K, Keshri J, Jha B (2010): Purification and characterization of a solvent, detergent and oxidizing agent tolerant protease from *Bacillus cereus* isolated from the Gulf of Khambhat. Journal of Molecular Catalysis B: Enzymatic **67**, 85-91.

Shakila T, Srivastava DK (2013): Cooperative binding of calcium ions modulates the tertiary structure and catalytic activity of Matrix-Metalloproteinase-9. Advances in Enzyme Research **9**, 17-29.

Shankar S, More SV, Laxman RS (2010): Recovery of silver from waste X-ray film by alkaline protease from *Conidiobolus coronatus*. Kathmandu university journal of science, engineering and technology **6**, 60-69.

Sharma PK, Capalash N, Kaur J (2007): An improved method for single step purification of metagenomic DNA. Molecular Biotechnology **36**, 61-63.

Sharma R, Ranjan R, Kapardar RK, Grover A (2005): Unculturable' bacterial diversity: An untapped resource. Current Science **89**, 72-77.

Sharma VK, Kumar N, Prakash T, Taylor TD (2009): MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. Nucleic Acids Research **38**, D468-D472.

Sharma VK, Kumar N, Prakash T, Taylor TD. (2012): Fast and accurate taxonomic assignments of metagenomic sequences using MetaBin. PLoS ONE **7**, e34030.

Shastry S, Prasad MS (2002): Extracellular protease from *Pseudomonas* sp. (CL 1457) active against *Xanthomonas campestris*. Process Biochemistry **37**, 611-621.

Shevchenko LS, Luk'yanov PA, Mikhailov VV (1995): Elastolytic activity of a marine isolate of *Bacillus pumilus*. Mikrobiologia **64**, 642–644.

Shibuya Y, Yamamoto T, Morimoto T, Nishino N, Kambara T, Okabe H (1991): *Pseudomonas aeruginosa* alkaline proteinase might share a biological function with plasmin. Biochimica et Biophysica Acta **1077**, 316–324.

Shih YC, Prausnitz JM, Blanch HW (1992): Some characteristics of protein precipitation by salts. Biotechnology and Bioengineering **40**, 1155–1164.

Shimogaki H, Takeuchi K, Nishino T, Ohdera M, Kudo T, Ohba K, Iwama M, Irie M (1991): Purification and properties of a novel surface-active agent and alkaline-resistant protease from *Bacillus* sp. Y. Agricultural and Biological Chemistry **55**, 2251–2258.

Shivaji S, Bhanu NV, 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 Microbiology Letters **189**, 247-252.

Showell MS (1999): Enzymes, detergent. In Flickinger MC, Drew SW (Eds): Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis and Bioseparation. Vol 2. Wiley, New York, pp. 958–971.

Shrinivas D, Naik GR (2011): Characterization of alkaline thermostable keratinolytic protease from thermoalkalophilic *Bacillus halodurans* JB 99

exhibiting dehairing activity. International Biodeterioration & Biodegradation **65**, 29-35.

Sidler W, Niederer E, Suter F, Zuber H (1986): The primary structure *of Bacillus cereus* neutral proteinase and comparison with thermolysin and *Bacillus subtilis* neutral proteinase. Biological Chemistry Hoppe-Seyler **367**, 643–657.

Sielecki ARM, Fujinaga RJ, James MGM (1991): Refined structure of porcine pepsinogen at 1.8A resolution. Journal of Molecular Biology **219**, 671–692.

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011): Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology **7**,1.

Sigma DS, Moser G (1975): Chemical studies of enzyme active sites. Annual Review of Biochemistry **44**, 889–931.

Simon C, Daniel R (2009): Achievements and new knowledge unraveled by metagenomic approaches. Applied Microbiology and Biotechnology **85**, 265–276.

Simonet P, Capellano A, Navarro E, Bardin R, Moiroud A (1984): An improved method for lysis of *Frankia* with achromopeptidase allows detection of new plasmids. Canadian Journal of Microbiology **30**, 1292–1295.

Simonet P, Grosjean MC, Misra AK, Nazaret S, Cournoyer B, Normand P (1991): *Frankia* genus-specific characterization by polymerase chain reaction. Applied and Environmental Microbiology **57**, 3278–3286.

Singh J, Batra N, Sobti RC (2001): Serine alkaline protease from a newly isolated *Bacillus* sp. SSR1. Process Biochemistry **36**, 781-785.

Sinha R, Singh SP, Ahmed S, Garg SK (1996): Partitioning of a *Bacillus* alkaline protease in aqueous two-phase systems. Bioresource Technology **55**, 163–166.

Sinha U, Wolz SA, Pushkaraj JL (1991). Two new extracellular serine proteinases from *Streptomyces fradiae*. International Journal of Biochemistry **23**, 979–984.

Smacchi E, Fox PF, Gobbetti M (1999): Purification and characterization of two extracellular proteinases from *Arthrobacter nicotianae* 9458. FEMS Microbiology Letters **170**, 327-333.

Solbak AI, Richardson TH, McCann RT, Kline KA, Bartnek F, Tomlinson G, Tan X, Parra-Gessert L, Frey GJ, Podar M, Luginbuhl P, Gray KA, Mathur EJ, Robertson DE, Burk MJ, Hazlewood GP, Short JM, 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**, 9431–9438.

Song JS, Jeon JH, Lee JH, Jeong SH, Jeong BC, Kim SJ, Lee SH (2005): Molecular characterization of TEM-type beta lactamases identified in cold-seep sediments of Edison Seamount (south of Lihir Island, Papua New Guinea). Journal of Microbiology **43**, 172–178.

Sørensen MA, Kurland CG, Pedersen S (1989): Codon usage determines translation rate in *Escherichia coli*. Journal of Molecular Biology **207**, 365– 377.

Spain AM, Krumholz LR, Elshahed MS (2009): Abundance, composition, diversity and novelty of soil Proteobacteria. The ISME journal **3**, 992-1000.

Stahl DA, Lane DJ, Olsen GJ, Pace NR (1985): Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. Applied and Environmental Microbiology **49**, 1379-1384.

Staley JT, Konopka A (1985): Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Review of Microbiology **39**, 321-346.

Steele DB, Fiske MJ, Steele BP, Kelley VC (1992): Production of a low molecular weight, alkaline active, thermostable protease by a novel spiral-shaped bacterium, *Kurthia spiroforme* sp. nov. Enzyme and Microbial Technology **14**, 358–360.

Steffan RJ, Goksoyr J, Bej AK, Atlas RM (1988): Recovery of DNA from soils and sediments. Applied and Environmental Microbiology **54**, 2908–2915.

Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996): Characterization of uncultivated prokaryotes: isolation and analysis of a 40- kilobase-pair genome fragment front a planktonic marine archaeon. Journal of Biotechnology **178**, 591–599.

Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, Breznak JA (2004): New strategies for cultivation and detection of previously uncultured microbes. Applied and Environmental Microbiology **70**, 4748–4755.

Strohl WR (2000): The role of natural products in a modern drug discovery program. Drug discovery today **5**, 39-41.

Su YC, Pan TM (1972): Alkaline protease produced by *Streptomyces griseus*. Chung Kuo Nunh Yeh Nua Hsue Hui Chih (Special Issue), 33–44.

Sudan AK, Vakhlu J (2013): Isolation of a thioesterase gene from the metagenome of a mountain peak, Apharwat, in the northwestern Himalayas. Biotech **3**, 19-27.

Suenaga H, Ohnuki T, Miyazaki K (2007): Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environmental Microbiology **9**, 2289–2297.

Sumandeep, Bhushan B, Beg QK, Hoondal GD (1999): Partial purification and characterization of a thermostable alkaline protease of an alkalophilic *Bacillius* sp. NG-27. Indian Journal of Microbiology **39**, 185-189.

Sundararajan S, Kannan CN, Chittibabu S (2011): Alkaline protease from *Bacillus cereus* VITSN04: Potential application as a dehairing agent. Journal of bioscience and bioengineering **111**, 128-133.

Takagi H, Kondou M, Hisatsuka T, Nakamori S, Tsai YC, Yamasaki M (1992): Effects of an alkaline elastase from an alkalophilic *Bacillus* strain on the tenderization of beef meat. Journal of Agricultural and Food Chemistry **40**, 2364–2368.

Takami H, Akiba T, Horikoshi K (1989): Production of extremely thermostable alkaline protease from *Bacillus* sp. no. AH-101. Applied and Microbiology and Biotechnology **30**, 120–124.

Takami H, Nakamura S, Aono R, Horikoshi K (1992): Degradation of human hair by a thermostable alkaline protease by alkalophilic *Bacillus* sp. AH-101. Bioscience Biotechnology and Biochemistry **56**, 1667-1669.

Tamai Y, Saeki K, Iwata J, Watanabe Y (1994): Purification and characterization of an alkaline protease from *Oerskovia xanthineolytica* TK-1. Journal of Fermentation and Bioengineering **77**, 554–556.

Tamura K, Nei M (1993): Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution **10**,512-526.

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**, 2731-2739.

Tanimoto SY, Tanabe S, Watanabe M, Arai S (1991): Enzymatic modification of zein to produce a non-bitter peptide fraction with a very high Fischer ratio for patients with hepatic encephalopathy. Agricultural Biological Chemistry **55**, 1119–1123.

Tanskul S, Hiraga K, Takada K, Rungratchote S, Suntinanalert P, Oda K (2009): An alkaline serine-proteinase from a bacterium isolated from bat feces: purification and characterization. Bioscience, Biotechnology, and Biochemistry **73**, 2393-2398.

Tebbe CC, 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**, 2657–2665.

Tekin N, Cihan AÇ, Takaç ZS, Tüzün CY, Tunç K, Çökmüş C (2012) : Alkaline protease production of *Bacillus cohnii* APT5. Turkish Journal of Biology **36**, 430-440.

Thangam EB, Rajkumar GS (2000): Studies on the production of extracellular protease by *Alcaligenes faecalis*. World Journal of Microbiology and Biotechnology **16**, 663-666.

Thumar J, Singh SP (2007): Two-step purification of a highly thermostable alkaline protease from salt-tolerant alkaliphilic *Streptomyces clavuligerus* strain Mit-1. Journal of Chromatography B **854**, 198-203.

Tien CC, Chao CC, Chao WL (1999): Methods for DNA extraction from various soils: a comparison. Journal of Applied Microbiology **86**, 937–943.

Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, Pootanakit K, Eurwilaichitr L (2008): Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. Journal of Biotechnology **133**, 42–49.

Tobe S, Takami T, Ikeda S, Horikoshi K (1976): Production and some enzymatic properties of alkaline proteinase of *Candida lipolytica*. Agricultural and Biological Chemistry **40**, 1087–1092.

Tobwala S, Srivastava DK (2013): Cooperative binding of calcium ions modulates the tertiary structure and catalytic activity of Matrix-Metalloproteinase-9. Advances in Enzyme Research **1**, 17-29.

Torsten T, Jack G, Folker M (2012): Metagenomics - a guide from sampling to data analysis. Microbial Informatics and Experimentation **2**, 3.

Tran L, Wu XC, Wong SL (1991): Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*. Journal of bacteriology 173, 6364-6372.

Tsai YC, Juang R, Lin SF, Chen SW, Yamasaki, M, Tamura G (1988). Production and further characterization of an alkaline elastase produced by alkalophilic *Bacillus strain* Ya-B. Applied and Environmental Microbiology **54**, 3156-3161.

Tsai YC, Lin SF, Yamasaki M, Tamura G (1986): Characterization of an alkaline elastase from alkalophilic *Bacillus* Ya-B. Biochimica et Biophysica Acta **883,** 439–47.

Tsai YC, Yamasaki M, Tamura G (1984): Substrate specificity of a new alkaline elastase from an alkalophilic *Bacillus*. International Journal of Biochemistry **8**, 283–288.

Tsai YC, Yamasaki M, Yamamoto-Suzuki Y, Tamura G (1983): A new alkaline elastase of an alkalophilic *Bacillus*. Biochemistry International **7**, 577–583.

Tsai YL, Olson BH (1991): Rapid method for direct extraction of DNA from soil and sediments. Applied and Environmental Microbiology **57**, 1070–1074.

Tsai YL, Olson BH (1992): Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Applied and Environmental Microbiology **58**, 2292-2295.

Tsai YL, Park MJ, Olson BH (1991): Rapid method for direct extraction of mRNA from seeded soils. Applied and Environmental Microbiology **57**, 765–768.

Tsuchiya K, Nakamura Y, Sakashita H, Kimura T (1992): Purification and characterization of thermostable alkaline protease from alkalophilic *Thermoactinomyces* sp. HS682. Bioscience Biotechnology and Biochemistry **56**, 246–250.

Tsujibo H, Miyamoto K, Hasegawa T, Inamori Y (1990): Purification and characterization of two types of alkaline serine proteases produced by an alkalophilic actinomycete. Journal of Applied Bacteriology **69**, 520–529.

Turpin PE, Maycroft KA, Rowlands CL, Wellington EM (1993): An ion-exchange based extraction method for the detection of salmonellas in soil. Journal of Applied Bacteriology **74**, 181–190.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004): Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**, 37-43.

Uchiyama T, Abe T, Ikemura T, Wantanabe K (2005): Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nature biotechnology **23**, 1038–1048.

Uchiyama T, Miyazaki K (2009): Functional metagenomics for enzyme discovery: challenges to efficient screening. Current Opinion in Biotechnology **20**, 616–622.

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**, 3305-3315.

Urtz BE, Rice WC (2000): Purification and characterization of a novel extracellular protease from *Beauveria bassiana*. Mycological Research **104**, 180-186.

Van Elsas JD, Speksnijder AJ, van Overbeek LS (2008): A procedure for the metagenomics exploration of disease-suppressive soils. Journal of Microbiological Methods **75**, 515–522.

Varela H, Ferrari MD, Belobradjic L, Vazquez A, Loperena ML (1997): Skin unhairing proteases of *Bacillus subtilis*: production and partial characterization. Biotechnology Letters **19**, 755–758.

Vasconcellos SPD, Angolini CFF, Garcia INS, Dellagnezze BM, Silva CCD, Marsaioli AJ, Neto EVD, Oliveira VMD (2010): Screening for hydrocarbon biodegraders in a metagenomic clone library derived from Brazilian petroleum reservoirs. Organic Geochemistry **41**, 675–681.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004): Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**, 66-74.

Vidya J, Swaroop S, Singh SK, Alex D, Sukumaran RK, Pandey A (2011): Isolation and characterization of a novel α-amylase from a metagenomic library of Western Ghats of Kerala, India Biologia **66**, 939-944.

Villegas A, Kropinski AM (2008): An analysis of initiation codon utilization in the domain Bacteria—concerns about the quality of bacterial genome annotation. Microbiology **154**, 2559-2561.

Vincent SS, John SB (2009): Buffers: principles and practice. In Richard RB, Murray PD (Eds): Methods in Enzymology. Elsevier Science & Technology Books, Massachusetts, pp. 43–56.

Vita C, Dalzappo D, Fontana A (1985): Limited proteolysis of thermolysin by subtilisin: isolation and characterization of a partially active enzyme derivative. Biochemistry **24**, 1798-1806.

Voget S, Leggewie C, Uesbeck A, Raasch C, Jaeger KE, Streit WR (2003): Prospecting for novel biocatalysts in a soil metagenome. Applied and Environmental Microbiology **69**, 6235–6242.

Volossiouk T, Robb EJ, Nazar RN (1995): Direct DNA extraction for PCR-mediated assays of soil organisms. Applied and Environmental Microbiology **61**, 3972–3976.

Wang C, Meek DJ, Panchal P, Boruvka N, Archibald FS, Driscoll BT, Charles TC (2006): Isolation of poly-3-hydroxybutyrate metabolism genes from

complex microbial communities by phenotypic complementation of bacterial mutants. Applied and Environmental Microbiology **72**, 384-391.

Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J,Meurer G, Saxena G, Andersen RJ, Davies J (2000): Novel natural products from soil DNA libraries in a S*treptomycete* host. Organic Letters **2**, 2401-2404.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007): Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology **73**, 5261-5267.

Ward OP (1985): Proteolytic enzymes. In Blanch HW, Drew S, Wang DI (Eds): Comprehensive Biotechnology-The Principles, Applications and Regulations of Biotechnology in Industry. Vol. 3. Pergamon Press, Oxford and New York, pp. 709–818.

Warren SJ (1992): Method of tenderising meat before slaughtering. European Patent Appl EP 0471470.

Waschkowitz T, Rockstroh S, Daniel R (2009): Isolation and characterization of metalloproteases with a novel domain stucture by contruction and screening of metagenomic libraries. Applied and Environmental Microbiology **75**, 2506–2516.

Watson RR (1976): Substrate specificities of aminopeptidases: a specific method for microbial differentiation. Methods in Microbiology **9**, 1–14.

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). Biogeoscience Discussion **5**, 1127- 1144.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991): 16S ribosomal DNA amplification for phylogenetic study. Journal of Biotechnology **173**, 697–703.

Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L (2003): Database

resources of the National Center for Biotechnology, Nucleic acids research **31**, 28–33.

Wilkesman J, Liliana Kurz (2009): Protease Analysis by Zymography: A Review on Techniques and Patents. Recent Patents on Biotechnology **3**, 175-184.

Wilkinson DE, Jeanicke T, Cowan DA (2002): Efficient molecular cloning of environmental DNA from geothermal sediments. Biotechnology Letters **24**, 155-161.

William PD, Yuping T, Annette CS (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**, 2311–2317.

Williamson KE, Kan J, Polson SW, Williamson SJ (2011): Optimizing the indirect extractionof prokaryotic DNA from soils. Soil Biology and Biochemistry **43**, 736–748.

Wilson SA, Young OA, Coolbear T, Daniel RM (1992): The use of proteases from extreme thermophiles for meat tenderization. Meat Science **32**, 93–103.

Woese CR (1987): Bacterial evolution. Microbiological Reviews **51**, 221–271.

Woese CR, Fox GE (1977): Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proceedings of the National Academy of Sciences of USA **74**, 5088–5090.

Wolff AM, Showell MS, Venegas MG, Barnett BL, Wertz WC (1996): Laundry performance of subtilisin proteases. In Bott R, Betzel C (Eds): Subtilisin Enzymes: Practical Protein Engineering. Plenum Press, New York, pp. 113–120.

Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, Rossmann MG (2009): Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Molecular Cell **34**, 375-386.

Xie F, Chao Y, Yang X, Yang J, Xue Z, Luo Y, Qian S (2010): Purification and characterization of four keratinases produced by Streptomyces sp. strain

16 in native human foot skin medium. Bioresource Technology **101**, 344-350.

Yamagata Y, Ishiki K, Ichishima E (1995): Subtilisin Sendai from alkalophilic *Bacillus* sp. Molecular and enzymatic properties of the enzyme and molecular cloning and characterization of the gene, *aprS*. Enzyme and Microbial Technology **17**, 653–663.

Yang Y, Jiang L, Zhu L, Wu Y, & Yang S (2000): Thermal stable and oxidation-resistant variant of subtilisin E. Journal of Biotechnology **81**, 113-118.

Yang ZH, Xiao Y Zeng GM, Zh Y, Liu SY (2007): Comparison of methods for total community DNA extraction and purification from compost. Applied Microbiology and Biotechnology **74**, 918–925.

Yeates C, Gillings MR, Davison AD, Altavilla N, Veal DA (1997): PCR amplification of crude microbial DNA extracted from soil. Letters in Applied Microbiology **25**, 303-307.

Young CC, Burghoff RL, Keim JG, Minak-BerberoV, Lute JR, Hinton SM (1993): Polyvinyl pyrrolidone–agarose gel electrophoresis purification of polymerase chain reaction amplifiable DNA from soils. Applied and Environmental Microbiology **59**, 1972–1974.

Yum DY, Chung HC, Bai DH, Oh DH, Yu JH (1994): Purification and characterization of alkaline serine protease from an alkalophilic *Streptomyces* sp. Bioscience Biotechnology and Biochemistry **58**, 470–474.

Yun J, Kang S, Park S, Yoon H, Kim MJ, 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**, 7229–7235.

Zalucki YM, Beacham IR, Jennings MP (2009): Biased codon usage in signal peptides: a role in protein export. Trends in Microbiology **17**, 146–150.

Zalucki YM, Jennings MP (2007): Experimental confirmation of a key role for non-optimal codons in protein export. Biochemical and Biophysical Research Communications **355**, 143–148.

Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M (2002): Cultivating the uncultured. Proceedings of the National Academy of Sciences of the United States of America **99**, 15681–15686.

Zhang RF, Cao H, Cui ZHL, Li SHP, Fan B (2003): Extraction and purification of soil microbial total DNA. Acta Microbiologica Sinica **43**, 276–282.

Zhou J, Bruns MA, Tiedje JM (1996): DNA recovery from soils of diverse composition. Applied and Environmental Microbiology **62**, 316–322.

APPENDIX - I

DNA extraction buffer (Protocol I - Zhou et al., 1996)

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

0.1 M phosphate buffer (pH 8.0)

1 M Na_2HPO_4 - 93.2 mL

1 M NaH_2PO_4 - 6.8 mL

To prepare 1 L of 0.1 M sodium phosphate buffer pH 8, the above mixtures should be diluted to 1 L (final volume) with H_2O.

SDS lysis buffer

100 mM NaCl

500 mM Tris [pH 8.0]

10% SDS [w/v]

Lysis solution (Protocol III - Tsai and Olson, 1991)

0.15 M NaCl

0.1 M Na_2EDTA [pH 8.0]

15 mg of lysozyme

1M Tris-HCl

Tris base - 60.57 g

Deionised water - 500 mL

Adjusted to desired pH using concentrated HCl

0.5M EDTA

EDTA - 18.6 g

Deionised water - 100 mL

TE buffer

1 M Tris-Cl (pH 8)	-	10 mL
500 mM EDTA (pH 8.0)	-	2 mL

50X TAE Buffer

Tris base - 121 g

Glacial acetic acid - 28.6 mL

0.5 M EDTA pH 8.0 - 50 mL

Deionised water added to make volume to 500 mL

1X TAE Buffer

50X TAE buffer – 10 mL

Deionised water – 490 mL

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

25 mM Tris - HCl - pH 8.0

50 mM glucose

10 mM EDTA

Autoclaved the solution and stored at 4^{o}C.

Solution II

0.2 N NaOH (freshly diluted from a 10 N stock)

1%SDS

Solution II prepared as fresh and used at room temperature.

Solution III

5.0 M 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.

Carbonate-bicarbonate buffer (pH 10.0)

Solution A : 21.2 g of anhydrous sodium carbonate dissolved in 1000 mL of distilled water

Solution B : 16.8 g of sodium bicarbonate dissolved in 1000 mL of distilled water

.

27.5mL of A + 22.5mL of B, diluted to a total volume of 200 mL with distilled water to obtain buffer of pH 10.0.

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.

8. **Protein staining solution**

Coomassie brilliant blue (0.1%)	-	100 mg
Methanol (40%)	-	40 mL
Glacial acetic acid (10%)	-	10 mL
Distilled water	-	50 mL

9. Destaining solution

Methanol (40%)	-	40 mL
Glacial acetic acid (10%)	-	10 mL
Distilled water	-	50 mL

Protein marker for SDS-PAGE

Broad range molecular weight protein marker mix of New England BioLabs (Ipswich, MA) was used for detecting the approximate size. It is a ready-to-load marker. The protein marker was mixed and 7µL was taken in a tube. It was heated for 5 min at 100ºC. After a quick microcentrifuge spin (Tarsons Spinwin, Kolkota, India) the marker was loaded directly on to a gel. The composition of the marker mix is as given below.

Components		MW in Da
Myosin	-	212,000
MBP-β- galactosidase	-	158,194
β- galactosidase	-	116, 351
Phosphorylase b	-	97,184
Serum albumin	-	66,409
Glutamic dehydrogenase	-	55,561
MBP2	-	42,710
Thioredoxin reductase	-	34,622
Triosephosphate isomerase	-	26,972
Trypsin inhibitor	-	20,000
Lysozyme	-	14,313
Aprotinin	-	6,517
Insulin A	-	3,400
B chain	-	2,340

Reagents for silver staining

1.	**Fixer**		
	Methanol (50%, v/v)	-	50 mL
	Acetic acid (5%, v/v)	-	5 mL
	Milli Q water	-	45 mL
2.	**Washing**		
	Methanol (50%, v/v)	-	50 mL
	Milli Q water	-	50 mL
3.	**Sensitizing solution**		
	Sodium thiosulfate (0.02%, w/v)	-	20 mg
	Milli Q water	-	100 mL
4.	**Silver nitrate solution**		
	Silver nitrate (0.2%, w/v)	-	200 mg
	Milli Q water	-	100 mL
5.	**Developer**		
	Sodium carbonate (6%, w/v)	-	3 g
	Formaldehyde	-	12.5 µL
	Milli Q water	-	100 mL
6.	**Stop solution**		
	Sodium-EDTA	-	1.4 g
	Milli Q water	-	100 mL

Buffers of different pH

Hydrochloric acid- potassium chloride buffer (pH 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 200 mL 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: 0.2 M NaH_2PO4

Solution B: 0.2 M Na_2HPO4

Mixed 39 mL of solution A with 6 mL of solution B and the volume was made up to 200 mL with distilled water, followed by filter sterilization.

Tris (hydroxymethylamino methane buffer system (pH 8 and 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 and 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.

Desired pH	Solution A (mL)	Solution B (mL)
10	27.5	22.5
10.7	45.0	5

Sodium hydroxide - Potassium chloride buffer (pH 12 and 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

APPENDIX - II

LURIA BERTANI BROTH

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. When used as solid agar medium, 2.0% agar (w/v) was added to the medium for agar plate preparation. Final pH is 7.5 ± 0.2.

NUTRIENT MEDIUM

Ingredients		**g/L**
Peptone	-	5
Sodium chloride	-	5
Beef extract	-	1
Yeast extract	-	2

Suspended 13 g of media (Himedia, Mumbai, India) in 1000 mL distilled water. Mixed well, autoclaved at 15 lbs pressure for 15 minutes and cooled to 50-55°C. When used as solid agar medium, 2.0% agar (w/v) was added to the medium. Final pH 7.4 ± 0.2.

ZOBELL MARINE BROTH

Ingredients		**g/L**
Peptic digest of animal tissue	-	5
Yeast extract	-	1
Ferric citrate	-	0.1
Sodium chloride	-	19.45
Magnesium chloride	-	8.8
Sodium sulphate	-	3.24
Calcium chloride	-	1.8
Potassium chloride	-	0.55
Sodium bicarbonate	-	0.16
Potassium bromide	-	0.08
Strontium chloride	-	0.034
Boric acid	-	0.022
Sodium silicate	-	0.004
Sodium fluorate	-	0.0024
Ammonium nitrate	-	0.0016
Disodium phosphate	-	0.008

Suspended 40.25 grams of media (Himedia, Mumbai, India) in 1000 mL distilled water. Heated to boiling to dissolve the ingredients completely. Sterilized by autoclaving at 15 lbs pressure (121°C) for 15 minutes. When used as solid agar medium, 2.0% agar (w/v) was added to the medium for agar plate preparation. Final pH -7.6 ± 0.2.

LIST OF PUBLICATIONS

Publications in peer reviewed journals

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 **3**, 29-32 (ISSN No. 2249-555X).

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**, 1-4 (ISSN: 2347-7407).

Full paper in proceedings of national/ international symposium/ conferences/ seminars

Harisree P Nair, **Helvin Vincent** and Sarita G Bhat (2012): Metagenomic approach for analysis of bacterial diversity of Mangalavanam mangrove soil. Proceedings of National symposium on emerging trends in Biotechnology (ISBN-978-93-80095-39-4)

Posters/Abstracts in national/international symposia

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 (ICGS 2010) organized by MKU, TamilNadu (12-14, November, 2010).

Helvin Vincent, Smitha S and Sarita G Bhat (2009): Isolation of a novel protease gene using metagenomic approach International Symposium on Marine

Ecosystems: Challenges and opportunities (MECOS 2009), Cochin, India (9-12, February 2009).

Helvin Vincent and Sarita G Bhat (2008): Novel amylolytic enzyme from soil metagenome International Conference on Biodiversity Conservation and Management, BIOCAM 2008 organised by Rajiv Gandhi Chair in Contemporary studies, CUSAT, Cochin, India (3-6, February 2008).

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, India (3-7, February 2014)

GenBank Submissions - 247 Nos

KF569952 - KF569972
KF453864 - KF453949
KF483718 - KF483791
KC143083 - KC143099
JX852421 - JX852429
JX465646 - JX465653
JQ936986 - JQ936988
JQ898300 - JQ898308
JQ868595 - JQ868605
JQ805720 - JQ805723
GU904004 - GU904004
KF569950
HM030820
HQ589022
KF769269
KC912765

RESEARCH PAPER | Biotechnology | Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

Community Genomics Involving Culture Independent Approach for Assessing the Phylogenetic Diversity of Mangrove Sediment

KEYWORDS: Metagenomics, 16Sr DNA, ARDRA, Phylogeny

Helvin Vincent	Harisree P. Nair	Dr. Sarita G Bhat
Department of Biotechnology, Cochin University of Science and Technology, Kerala, India.	Department of Biotechnology, Cochin University of Science and Technology, Kerala, India.	Department of Biotechnology, Cochin University of Science and Technology, Kerala, India.

ABSTRACT *Soil community genomics or metagenomics is employed in this study to analyze the evolutionary relatedness of mangrove microbial community. The metagenomic DNA was isolated from mangrove sediment and 16SrDNA was amplified using universal primers. The amplicons were ligated into pTZ57R/T cloning vector and transformed onto E. coli JM109 host cells. The recombinant plasmids were isolated from positive clones and the insert was confirmed by its reamplification. The amplicons were subjected to Amplified Ribosomal DNA Restriction Analysis (ARDRA) using three different tetra cutter restriction enzymes namely Sau3A1, Hha1 and HpaII. The 16SrDNA insert were sequenced and their identity was determined. The sequences were submitted to NCBI database and accession numbers obtained. The phylogenetic tree was constructed based on Neighbor-Joining technique. Clones belonged to two major phyla of the bacterial domain, namely Firmicutes and Proteobacteria, with members of Firmicutes predominating. The microbial diversity of the mangrove sediment was explored in this manner.*

INTRODUCTION

Mangroves are highly productive ecosystems with immense ecological values. The microbial communities in the mangrove sediment represent some of the most complex microbial habitats on the Earth. There may be several thousand species of bacteria in 1g of soil (Torsvik, 1990). Since only a small percentile (<1%) of existing bacteria are cultivable by standard microbiological methods, the others remain largely unexplored. Hence to assess the diversity of an environmental sample, modern molecular tools based on the PCR amplification of 16S rRNA gene, the phylogenetic anchor, are employed (Handlesman, 2004). In metagenomic approach, 16S rDNAs are amplified by PCR from nucleic acids extracted from environmental samples, whereafter the PCR products are cloned and sequenced. Amplified Ribosomal DNA restriction analysis (ARDRA) which is based on DNA sequence variations present in PCR-amplified 16S rRNA genes(Smit, 1997) also serve as an effective tool in phylogenetic analysis.

In the present study, the 16S rDNA clone library was constructed with the metagenomic DNA isolated from mangrove sediment. Clones were clustered according to the dendrograms of ARDRA results and sequenced. The partial 16S rRNA gene sequences were compared with NCBI data base and taxonomic hierarchy was assigned.

MATERIALS AND METHODS

Isolation of metagenomic DNA

The benthic sediment samples collected from mangroves of Kochi, Kerala, India (09°52'43.3"N 76°15'50.6"E), were transported to the laboratory under refrigeration. The metagenomic or total soil DNA was isolated from sediment sample according to Zhou et al (1996) and purified using UltraClean Soil DNA isolation kit (MoBio, CA, USA). The isolated DNA was electrophoresed on 1% agarose gel at 80V for 1 hour, stained using ethidium bromide (Sambrook, 2000) and visualized under ultraviolet illumination and gel pictures were captured using Gel documentation system (Syngene, CA, USA). The DNA was quantified using a UV-Visible spectrophotometer (Shimadzu, Kyoto, Japan).

PCR amplification of 16SrDNA sequences of metagenomic DNA

The 16SrDNA sequences of the metagenomic DNA was amplified using universal 16SrDNA forward (5'AGAGTTTGATCCTGGCTCAG 3') and reverse primer (5' ACGGCTACCTTGTTACGACTT 3') (Shivaji, 2000). 50ng template DNA was used in a 20µL reaction with an initial denaturation for 2 min at 94 °C, 34 cycles of denaturation at 94 °C for 30 sec, annealing at 54 °C for 30 sec and extension at 72 °C for 2 min and with a final extension for 10 min at 72 °C. PCR was performed in a thermal cycler (Biorad, CA, USA). The amplicon was then electrophoresed and visualized using gel documentation system (Syngene). The PCR product was purified using Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) and the purified PCR product was used for cloning.

Construction of 16SrDNA library

The 16SrDNA amplicons were ligated into TA cloning vector pTZ57R/T and transformed onto *E. coli* JM109 host cells using InsTAclone PCR cloning kit (Thermo Scientific, MA, USA). The transformed colonies were screened for α-complementation by using X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (isopropyl-β-D thiogalactoside). All positive clones were confirmed by reamplification of 16S rDNA inserts from the recombinant plasmids, isolated according to alkali lysis method, and were stored on LB agar (HiMedia, Mumbai, India) plates containing ampicillin.

ARDRA profiling of metagenomic clones

The 16SrDNA amplicons were digested using 3 different tetrameric restriction enzymes Sau3A1, Hha1 and HpaII (Thermo Scientific) (Gurtler, 1991) with specific restriction sites, ('GATC, GCG'C and C'CGG respectively). Five units of each enzyme was mixed with 5µL of PCR amplicon in separate reaction mixtures and incubated for 4 hours at 37°C. The digested fragments were analyzed on 2% agarose gel, and visualized using gel documentation system (Syngene). Dendrograms were constructed from the distance matrix by using the Unweighted Pair Group Method with Arithmetic means (UPGMA) (Sneath, 1973). The plasmids with discrete ARDRA banding pattern for the 16SrDNA insert were sequenced.

Sequencing of 16SrDNA amplicons

The 16SrDNA was sequenced from the recombinant plasmids by Sanger's Dideoxy method using ABI 3730 Excel (Applied Biosystems, CA, USA). The identity of the sequences was determined by comparing with the sequences in the NCBI database using BLAST software (Altschul, 1990). The sequences were compiled and aligned using ClustalX program (Thompson, 1997) using BioEdit software (Hall, 1999).

Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

The phylogenetic tree was constructed using the Neighbor-Joining method (Saitou, 1987) with 1000 resampling bootstrap using MEGA software (Tamura, 2007).

RESULTS

Construction of 16SrDNA library

The total community or metagenomic DNA isolated from the mangrove sediment with a concentration of 1.56µg/µL was used as template for 16SrDNA amplification. Cloning of purified 16SrDNA PCR product yielded 20 phylogenetic clones after α-complementation and reamplification. The recombinant plasmids isolated from the positive clones showed distinct band shift when compared with control plasmid (Fig.1).

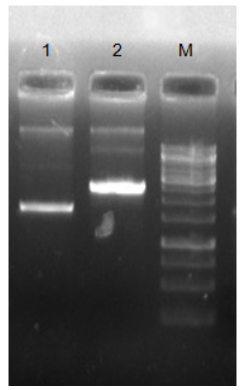

Fig.1: Agarose gel showing band shift of recombinant plasmid

Lane1 – control plasmid pTZ57R/T

Lane2 – recombinant plasmid

Lane M– 1kb marker (Thermo Scientific)

ARDRA profiling of metagenomic clones

The ARDRA profile of 16SrDNA clones obtained with the enzyme Sau3A1 showed distinct patterns on 2% agarose gel (Fig. 2).

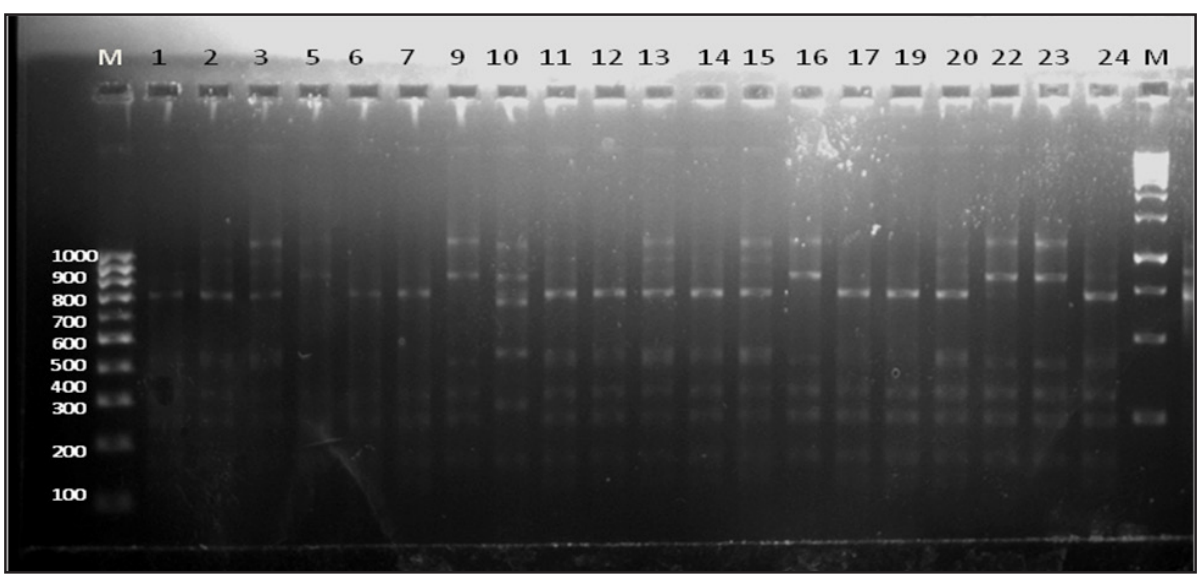

Fig.2: ARDRA profile of 16SrDNA clones with Sau3A1

Lane numbers represent the clone numbers
M represent 1kb DNA marker
Clustering of the clones based on the ARDRA banding pattern using the enzyme Sau3A1 is as depicted in Fig. 3

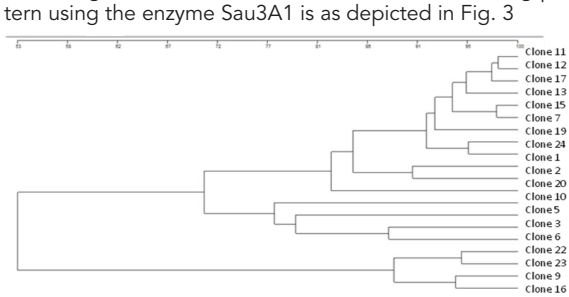

Fig.3. Clustering based on ARDRA pattern using Sau3A1

Most of the clones showed discrete banding pattern with a few exceptions. Clones 11, 12 and 17 showed similar banding pattern and they clustered together. Similarly, clones 22,23,9 and 16 showed similar pattern and they formed a separate cluster.

The ARDRA profile of 16SrDNA clones obtained with the enzyme Hha1 showed similar patterns on 2% agarose gel for most of the clones indicating that the number of Hha1 restriction sites is same in most of the clones. However clone 10 showed a variation in the banding pattern. (Fig. 4)

Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

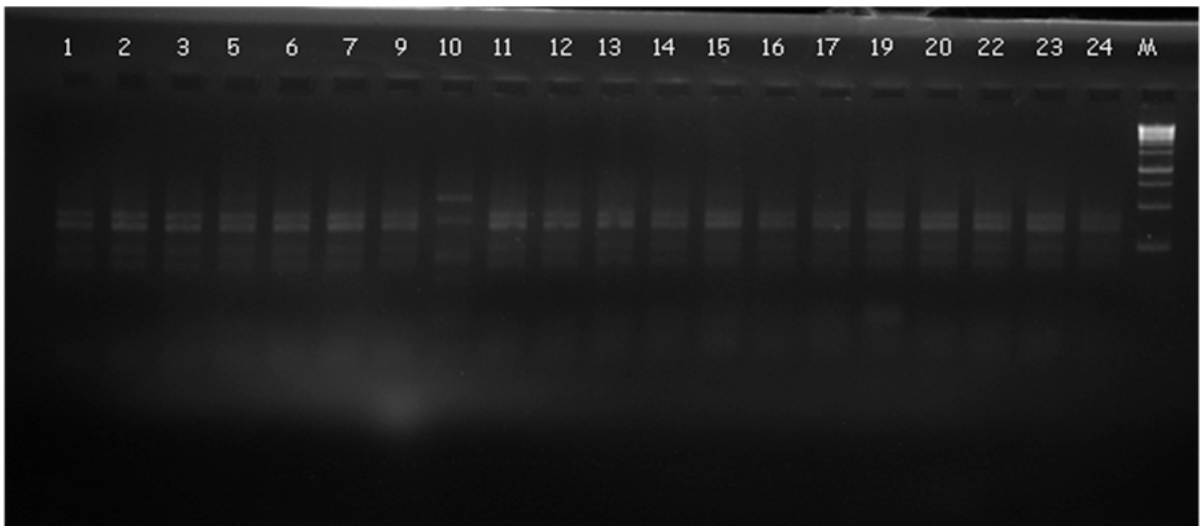

Fig.4: ARDRA profile of 16SrDNA clones with Hha1

Lane numbers represent the clone numbers

M represent DNA marker (Lambda DNA/EcoR1/HindIII double digest)

Clustering of the clones based on the ARDRA banding pattern using Hha1 is depicted in Fig. 5, which clearly showed that while most others clustered together, Clone 10 remained separate from all the others

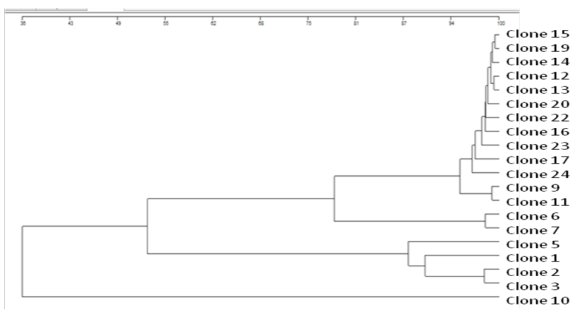

Fig.5. Clustering based on ARDRA pattern using Hha1

The ARDRA profile of 16SrDNA clones obtained with the enzyme HpaII showed varied patterns on 2% agarose gel for most of the clones (Fig. 6).

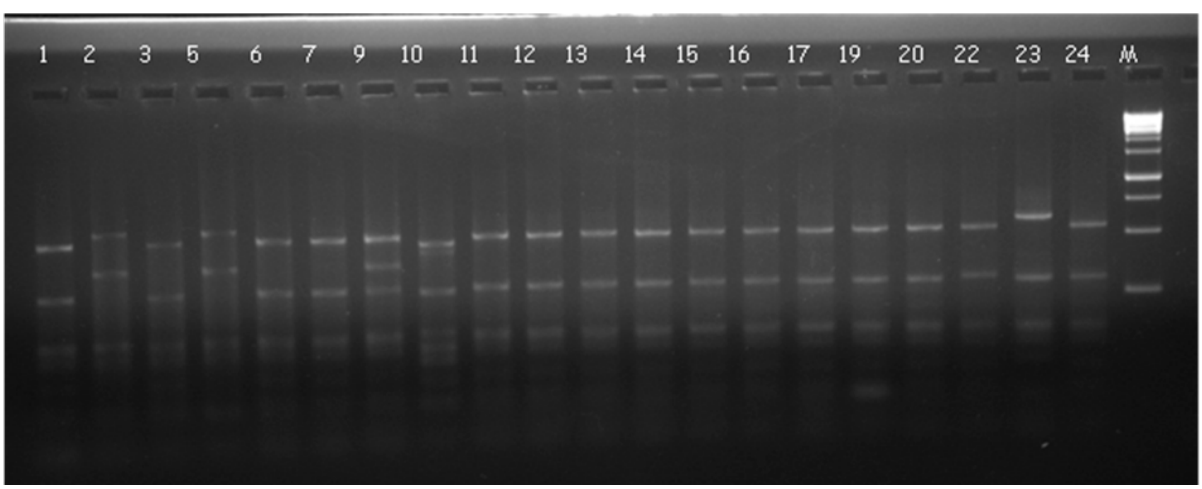

Fig.6: ARDRA profile of 16SrDNA clones with HpaII

Lane numbers represent the clone numbers

M represent DNA marker (Lambda DNA/EcoR1/HindIII double digest)

Clustering of the clones based on the ARDRA banding pattern using the enzyme HpaII is depicted in Fig. 7

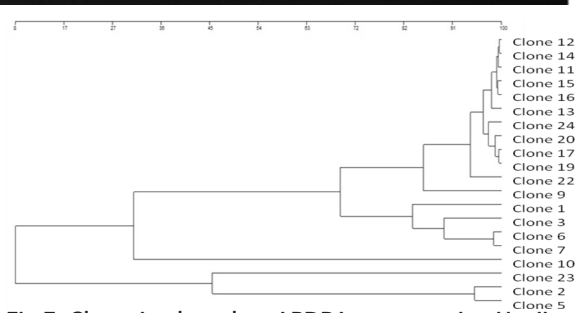

Fig.7. Clustering based on ARDRA pattern using HpaII

Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

Clones 11, 12, 13, 14, 15 and 16 illustrated a particular banding pattern and therefore claded together while clones 2 and 5 showing a different banding pattern formed another cluster.

Sequencing of 16SrDNA amplicons of the clones
The 16SrDNA insert of 15 recombinant plasmids with discrete ARDRA profile were sequenced and the sequences were submitted to NCBI database and accession numbers were assigned (JQ868595 - JQ868601, JQ898300 - JQ898308, JQ805720 - JQ805723).

The phylogenetic tree constructed based on the 16S rDNA sequences of these 15 clones are as shown in figure 8.

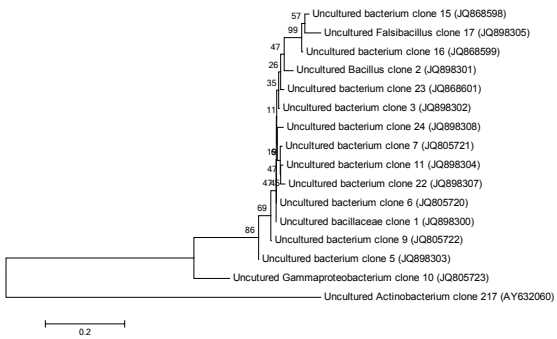

Figure 8: Phylogentic relationship based on partial 16SrDNA sequences of 15 selected clones. *Actinobacterium* clone (accession number AY632060) was used as outgroup. Accession numbers are given in parentheses. The number at the nodes of the phylogenetic tree are percentages indicating the levels of bootstrap support based on a Neighbour-Joining analysis of 1,000 resampled data sets.

The phylogenetic tree in figure 8 is based on the16S rRNA sequences of the 15 clones. Clone 10 showed similarity to *Gamma Proteobacteria*, while most others showed similarity to uncultured bacteria. Clone 1 was most similar to unclassified *Bacillaceae*, while Clone 2 to uncultured *Bacillus*. Clone 17 showed identity to uncultured *Falsibacillus*. They all belong to phylum *Firmicutes* indicating that the predominant group of microbes in the mangrove sediment analyzed belonged to phylum *Firmicutes*.

DISCUSSION

In the present study, 20 different 16SrDNA clones of mangrove sediment microbial community generated using metagenomic approach were analysed to study their phylogenetic diversity. Sequence analysis indicated that they belonged to two major phyla of bacterial domain, namely, *Firmicutes* and *Proteobacteria*. The predominant group was phylum *Firmicutes*. The *Proteobacteria*, especially the *Gammaproteobacteria* were found to be abundant in the Sundarban mangrove(Ghosh, 2010). Many other diverse phylotypes like *Flexibacteria, Actinobacteria, Acidobacteria, Chloroflexi*, *Planctomycetes* and *Gammatimonadates* were also reported previously from different environmental samples like marine sediments(Gray, 1996) and Oklahoma prairie soil(Anne, 2009), etc. Urakawa et al. (1999) reported five major lineages of the domain *Bacteria*: the gamma, delta and epsilon *Proteobacteria*, Gram-positive bacteria and the division *Verrucomicrobia* from Sagami Bay and Tokyo Bay. It can be assumed that the dominant phylum *Firmicutes*, especially the class *Bacillaceae*, may have been carried to the mangroves from the land by rain as they are mainly terrestrial inhabitants.

The restriction enzymes, Sau3A1 and HpaII produced clear ARDRA patterns capable of distinguishing the clones from each other. The restriction pattern formed by Hha1 was not significant. Guan *et al*. (2003) reported the identification of *Lactobacillus* in crops of broilers using ARDRA pattern generated by enzymes, HaeIII and MseI. Wu *et al* (2006) reported the identification of *Bacillus* sp. based on the ARDRA pattern generated using six different enzymes namely, AluI, TaqI, HhaI, MboI, RsaI and MspI.

The sequence analysis as well as the ARDRA pattern analysis indicated the phylogenetic diversity of the mangrove microbial community.

The present study confirms that the mangrove microbial community is highly diverse. Since culture dependent studies are highly limiting, culture independent analysis or metagenomics serves as a valuable tool for assessing the phylogenetic diversity of environmental samples.

CONCLUSION

The present study analyzed the phylogenetic diversity of mangrove sediment microbial community using metagenomic approaches. The analysis revealed that the community comprised of two major phylum of bacterial domain namely, *Proteobacteria* and *Firmicutes* with phylum *Firmicutes* predominating. The study needs to be extended for the understanding of many other different phylogroups present in this unique exosystem.

Acknowledgement
The first author acknowledges Cochin University of Science and Technology, Kerala, India and University Grants Commission, New Delhi, India for supporting the work with necessary facilities as well as financial support in the form of JRF and SRF. The authors also acknowledge the support of the UGC major project to the corresponding author.

REFERENCE
Altschul, S. F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local | alignment search tool. Journal of Molecular Biology(215), 403-410. | Anne, M., Lee, R. and Mostafa, S. (2009). Abundance, composition, diversity and novelty of soil Proteobacteria. ISME Journal(3), 992-1000. | Ghosh, A., Nirmalya, D., Amit, B., Tiwari, A., Sathyaniranjan, K.B., et al. (2010). Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. Saline Systems(6), 1. | Gray, J. P. and Herwig, R. P. (1996). Phylogenetic analysis of the bacterial communities in marine sediments. Applied and Environmental Microbiology(62), 4049-4059. | Guan, L. L., Hagen, K.E., Tannock, G.W., Korver, D.R., Fasenko, G.M., et al. (2003). Detection and identification of Lactobacillus species in crops of broilers of different ages by using PCR-denaturing gradient gel electrophoresis and amplified ribosomal DNA restriction analysis. Applied and Environmental Microbiology, 69(11), 6750-6757. | Gurtler, V., Wilson, V. A. and Mayall, B. C. (1991). Classification of medically important clostridia using restriction endonuclease site differences of PCR-amplified 16S rDNA. Journal of General Microbiology(137), 2673–2679. | Hall, T. A. (1999). Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series(41), 95-98. | Handlesman, J. (2004). Application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews(68), 669-685. | Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution(4), 406-425. | Sambrook, J., Fritsch, E.F. and Maniatis, T. (2000). Molecular Cloning: a laboratory manual. | Shivaji, S., Bhanu, N. V. and 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), 247-252. | Smit, E., Leeflang, P. and Wernars, K. (1997). Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. FEMS Microbiology Ecology (23), 249–261. | Sneath, P. H. A. and Sokal, R. R. (1973). Numerical taxonomy. (W. H. Freeman & Co., San Francisco). | Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution(24), 1596-1599. | Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research (25), 4876–4882. | Torsvik, V., Goksoyr, J. and Daae, F.L. (1990). High diversity in DNA of soil bacteria. Applied and Environmental Microbiology(56), 782-787. | 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. | Wu, X. Y., Walker, M.J., Hornitzky, M. and Chin, J. (2006). Development of a group-specific PCR combined with ARDRA for the identification of Bacillus species of environmental significance. Journal of Microbiological Methods(64), 107–119. | Zhou, J., Bruins, M.A. and Tiedje, J.M. (1996). DNA recovery from soils of diverse composition. Applied and Environmental Microbiology(62), 316-322. |