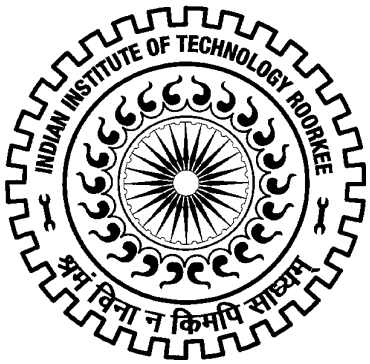
METAGENOMIC APPROACH TO STUDY MICROBES FROM HIMALAYAN GEOTHERMAL SPRINGS

Ph.D. THESIS

by

JITENDRA KUMAR SAHOO



DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
DECEMBER, 2013

METAGENOMIC APPROACH TO STUDY MICROBES FROM HIMALAYAN GEOTHERMAL SPRINGS

A THESIS

Submitted in partial fulfilment of the

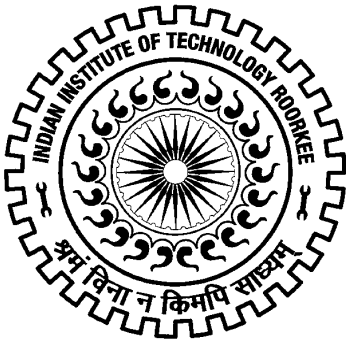
Requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY
in
BIOTECHNOLOGY

by

JITENDRA KUMAR SAHOO



DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
DECEMBER, 2013

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**METAGENOMIC APPROACH TO STUDY MICROBES FROM HIMALAYAN GEOTHERMAL SPRINGS**'' in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from January, 2009 to December, 2013 under the supervision of Dr. Ranjana Pathania, Assistant Professor, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

(Jitendra Kumar Sahoo)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dated: December, 2013 (Ranjana Pathania)
Supervisor

The Ph.D. Viva-Voce Examination of **Mr. Jitendra Kumar Sahoo**, Research Scholar, has been held on..... .

Signature of Supervisor Signature of Chairman, SRC Signature of External Examiner

Signature of Head of the Dept./Chairman, ODC

ABSTRACT

The present study includes a comprehensive analysis of microbial population and their functional properties from two Himalayan geothermal springs through culture independent approach. The first part of the study includes metagenomic DNA isolation and purification from environmental sample of Himalayan thermal springs. Subsequently, total community based bacterial diversity of the thermal springs was assessed. The metagenomic functional studies included cloning and characterization of endoglucanase gene responsible for cellulose hydrolysis. In addition to the above, L-asparaginase-II enzyme from a culturable bacterial isolate was also studied. Real time PCR based transporter gene identification from metagenomic DNA pool of Himalayan hot spring was also carried out.

Microbial communities of the sulphur hot spring at Tattapani geothermal area of Himachal Pradesh state and Tapovan geothermal spring located in Chamoli District of Uttarakhand state, India were analyzed. 16S rRNA gene based microbial identification and phylogenetic affiliation of microorganisms from these hot springs was carried out by culture independent approach. 16S rRNA gene based metagenomic libraries were constructed from the hot spring mat DNA. Restriction Fragment Length Polymorphism (RFLP) patterns of cloned 16S rRNA gene from both the springs were analyzed. About 89 clones from both the constructed libraries were sequenced and their non-chimeric validation was done by using pintail version 1.1. The non-redundant nucleotide sequences were deposited in the GenBank database and have accession numbers JN613324, JN896893 to JN896938, JN934657 to JN934666, JN967771 to JN967773, KC608724 to KC608751. Among Tattapani hot spring derived clones, 35% of the clones belonged to the phyla Proteobacteria which consist of Alpha-proteobacteria (4%), Beta-proteobacteria (7%), Gamma-proteobacteria (11%), and Delta-proteobacteria (13%) respectively. Other phyla identified were 25% of Acidobacteria, 5% of Planctomycetes, 14% of Verrucomicrobia, 4% each of Bacteroidetes, Chloroflexi, Gemmatimonadetes and 9% of total represented unaffiliated clones. These were classified into 8 distinct phyla. Similarly 3 distinct phyla were identified in Tapovan spring 16S rRNA library. 79% of the clones belonged to the phyla Firmicutes, 14% to Proteobacteria and 7% represented unclassified bacteria found in Tapovan library. The Tapovan clone sequences showed similarity to various species of *Anoxybacillus* and *Aneurinibacillus* genera. About 37 clones found collectively in the libraries

showed <97% sequence similarity with the known database of NCBI. Hence, these organisms were affiliated as new species as per the species concepts of microorganisms. Operational Taxonomic Unit was calculated to about 24 for Tattapani and 8 for Tapovan derived clones at 97% similarity cutoff. Shannon-weiner index was calculated as 3.923088 and 1.6705014 for Tattapani and Tapovan springs respectively. The Maximum possible value for species evenness was found to approach towards one i.e. 0.9834805 and 0.8033413 for Tattapani and Tapovan springs respectively. The Neighbor-Joining based phylogenetic tree deduced from both the hot springs showed that microbial signatures differ significantly. The overall microbial diversity of Tattapani spring was higher than Tapovan spring. This is the first report of comprehensive analysis of microbes and their diversity in these Himalayan thermal springs using metagenomics approach.

The aim to conduct functional study of hot spring DNA led to the construction of metagenomic libraries in easily culturable host *Escherichia coli*. Small insert metagenomic libraries were constructed using plasmid pNYL-*rygC*. The Tapovan and Tattapani metagenomic DNA was partially digested with Sau3A I and the vector pNYL-*rygC* was digested with Bam HI. About 2000 clones from each of the libraries were screened for cellulase and xylanase using direct agar plate containing specified substrate. But no visible clone was detected with hydrolysis activity. The predictable cause for this lack of activity may be attributed to difficulty in heterologous gene expression or with the vector system used for the study. To alleviate the limitations imposed by this approach, an alternative method using direct polymerase chain reaction based method was applied.

Metagenomic DNA from Tapovan hot spring was used to identify endo beta- glucanase gene (TM-*cel5A*) using direct PCR. Degenerate primers were constructed for endo beta-glucanase gene and were used to amplify the gene from Tapovan hot spring DNA. The purified PCR product was cloned in pET28(a) expression vector. The codon plus host system was used to minimize codon biasness. The cellulose hydrolysis activity of the functional gene products derived from clones was analyzed on carboxy methyl cellulose (CMC) agar plate and stained with Congo red. Clones with variable hydrolysis pattern were observed on the screening plate. The clone with largest hydrolysis zone was selected. The over expression of gene product in the selected clone was studied using IPTG at variable concentrations and analyzed on PAGE. TM-

Cel5A was found to be stable at 60 °C for 1hour and its activity decreased to to 12% after 2 hours of incubation. TM-Cel5A showed an optimal activity at pH 8.0. TM-Cel5A showed activity in a wide range of pH i.e. 4 to 9. TM-Cel5A also showed the ability to use a wide range of substrates. It has the ability to hydrolyze natural form of storage glucan (Beta- D- glucan). Sequence analysis showed that TM-Cel5A was identical to endoglucanase of *Bacillus* sp. and with *Bacillus licheniformis*. Thus, to the best of our knowledge this is the first report of cloning and characterization of beta-endoglucanase thermostable enzyme from a Himalayan thermal spring.

The community level microbial analysis based on Biolog EcoPlate analysis showed that microbes in both the hot springs were able to utilize Pyruvic Acid Methyl Ester, Tween 40, Tween 80, D- Xylose, Mallic acid, L-Asparagine, L-Threonine and Glycyl-L-glutamic acid. L-Arginine is the amino acid which was exclusively utilized by Tapovan microbes. A bacterial isolate was also identified from Tapovan spring water sample producing L-asparaginase enzyme. The L-asparaginase producer was found to be a closest relative of *Ralstonia* sp. Its L-asparaginase production was confirmed by growth on Modified Czapek Dox medium containing L-asparagine as sole carbon source and by nessler's reagent based biochemical assays. The gene namely L-asparaginase-II responsible for asparaginase activity was cloned and expressed in *Escherichia coli* BL21-CodonPlus cells. L-Asparaginase is the enzyme which is well known for its application in food processing and treatment of acute lymphoblastic leukaemia (ALL).

In addition to the above mentioned functional characterization of microbes in Himalayan springs, a real time PCR based detection of genes responsible for transporters and antibiotic resistance markers was done. Inductively coupled plasma mass spectroscopy (ICPMS) and ionic chromatography (IC) analysis of Tapovan water sample showed presence of high loads of sulphur and heavy metals. Hence, bacterial efflux systems were studied that can respond to such stressful environmental conditions. These transporter/efflux systems play dual role in efflux of heavy metals as well as modern antibiotics leading to multidrug resistance in microorganisms. In order to detect genes responsible for efflux systems in Tapovan metagenomic DNA, real time PCR were carried out using degenerate primers. The qPCR results showed detectable amounts of transporter genes *cusA*, *acrB* and *acrD*. CusA is an integral part of CusCFBA system and is responsible for efflux of copper/silver. AcrB, AcrD are part of Resistance –Nodulation- Cell

division (RND) family of transporters, involved in efflux of amphiphilic substances. No Human mitochondrial DNA (mtDNA) and *E. coli* plasmid pET23a-GFP DNA were found in the metagenomic DNA. This indicates that the sample was free from anthropogenic and modern routine laboratory contaminations. Analysis and quantification of genes was done using threshold (CT) values. Low CT value of Acriflavin resistance protein B (AcrB) was found. The *acrB* gene was cloned in pTZ57R/T vector and the insert was sequenced. The sequence analysis showed highest identity with acrB gene of *Escherichia coli* APEC O78 and Acriflavin resistance protein of *Escherichia coli* P12b. The presence of AcrB, AcrD is also supported by the factor that they are the well known members of hydrophobe/amphiphilic efflux-1 (HAE-1) family of transporters. Their overexpression during nutrient deficient condition and involvement in efflux of currently used antibiotics, disinfectants, dyes, detergents further confirms their presence. The presence of AcrB, AcrD and CusA in Himalayan springs can be attributed to the presence of heavy metals like Al, As, Cu, Fe, B, Se and nutrient deprived conditions.

PUBLICATIONS

- *Bhatia, A., Ali, M., **Jitendra, S, K**., Madan, S., Pathania, R., Ahmed,N., Kazmi. A.A. (2012) Microbial diversity during rotary drum and windrow pile composting. ***Journal of Basic Microbiology***. Vol. 52, 5–15.
- *Bhatia, A., Madan, S., **Jitendra, S, K**., Ali, M., Pathania, R., Kazmi, A,A. (2012) Diversity of bacterial isolates during full scale rotary drum composting. ***Waste Management***. Vol. 33, 1595- 1601.
- **Jitendra, S, K.,** Ghosh I, Dhiman A, Sharma R, Navani N K, Pathania R. Assessment of bacterial diversity in a Himalayan geothermal spring of India by culture independent approach and its community level metabolic profiling. (Communicated in ***Biodiversity and conservation***).

Poster or talk presented in national and international conferences:

- **Jitendra, S, K.,** Dhiman, A., Ghosh, I., Navani, N, K., Pathania, R. Assessment of Microbial diversity of a Himalayan Geothermal Spring (Tattapani) through culture independent approach. International conference on "Microbial World: Recent Innovations and Future Trends" in Association of Microbiologist of India, Bhubaneswar, Odisha, 22-25 Nov, 2012. Poster Presented.
- *Dhiman, A., **Jitendra, S, K.**, Navani, N, K., Pathania, R. Assessment of microbial diversity of river Ganga and subsequent isolation of a psychrotroph with biocontrol properties. International conference on Microbial World: Recent Innovations and Future Trends in Association of Microbiologist of India, Bhubaneswar, Odisha, 22-25 Nov, 2012. Poster Presented.
- *Bhatia, A., **Jitendra, S, K.,** Ali, M., Madan,S., Pathania,R., Ahmed, N. and Kazmi, A.A. Diversity of Bacterial Isolates during High rate Rotary Drum composting, Proceedings of International Interdisciplinary Science Conference , May 30, 2011, Jamia Islamia University, Delhi (India), Oral.

- *Bhatia, A., Ali, M., **Jitendra, S, K**., Pathania, R., Ahmed, N., Kazmi. A.A. Microbial communities during rotary drum and pile composting. IWA conference of Microbes in Waste water and waste treatment, Jan 2011, Goa India. poster.
- *Manasi,G., Tarun, S, K., **Jitendra, S, K.**, Naveen, N, K., Pathania, R. A Microbe Isolated from Uttarakhand Degrades Multiple Pesticides. 5thUttarakhand State Science & Technology Congress, Doon University, Dehradun (India) poster. 10-12 November, 2010.
- *Bhattacharyya, T. Deepak, S. Sharma, R**. Jitendra S, K.** Ghosh, I. and Pathania, R. A novel thermophile isolated from soil exhibits antimicrobial activity against a wide spectrum of bacteria. 5thUttarakhand State Science & Technology Congress, Doon University, Dehradun (India) poster. 10-12 November, 2010.

NCBI sequence submission:

- **Jitendra, S, K**., Ghosh, I., Sharma, R., Navani, N. K., Pathania, R. 16 S rRNA based sequences submitted for Tattapani geothermal spring (Himachal Pradesh) derived microbial clone library for **61** sequences at NCBI with accession numbers **JN613324, JN896893 to JN896938, JN934657 to JN934666, and JN967771 to JN967773**.
- **Jitendra, S, K**., Dhiman, A., Navani, N. K., Pathania, R. **46** Other 16 S rRNA based sequences submitted at NCBI with accession numbers **JN860152- JN860168**, **JN786083 – JN786084, KC608724-KC608751.**

*This work is not included in this thesis.

ACKNOWLEDGEMENT

With obeisance to the Almighty, I express my immense gratitude to Him for bestowing blessings of strength that inspired me, leading to successful compilation of my thesis. It is my privilege to express my sincere thanks and indebtedness to my Ph.D. advisor **Dr. Ranjana Pathania** for guiding me at every step, correcting my mistakes, giving me sufficient time, and support and for encouraging me constantly. I have no words to express my gratitude for her for letting me conduct all the experiments the way I wanted, which really helped me develop my scientific temperament.

Besides my advisor, I would like to thank **Dr. Naveen Kumar Navani** for his encouragement and insightful comments. Without his encouragement and constant guidance, it was not possible for me to complete my work successfully.

I also appreciate the advice of the Student Research Committee (SRC) members **Dr. Partha Roy** (DRC Chairman), **Dr. P. Jeevanandam** (Assoc. Prof Dept. of chemistry) for their critical comments, which enabled me to notice weakness of my dissertation and make the necessary improvements according to their comments. I am grateful to former chairman Student Research Committee (SRC) **Prof. G. S. Randhawa** for his valuable and timely suggestions. I am also grateful to the faculty members Prof. Ritu Barthwal, Prof. R. P. Singh, Prof. R. Prasad, Dr. Vikas Pruthi, Dr. Bijan Choudhury, Dr. Sanjay Ghosh, Dr. Pravindra Kumar, Dr. Shailly Tomar and Dr. Maya Nair for their support and encouragement.

I would also like to give special thanks to my lab mates Rajnikant Sharma, Dr. Piyush Kalra, Santosh Kumar Srivastava, Tarun Sharma, Paramesh R. Lambadi and for their helpful discussion and support. I am thankful to Abhijeet Dhiman, Akanksha Bhatia, Manasi Gupta, Supriya Deepak Patil, Tapas Bhattacharya, Atin Sharma, Timsy Bhando, Anjul Saini, Rekha Sharma and Tamoghna Ghosh for their absolute support and cooperation while working in lab. I would like to thank my lab attendants Akshya and Mohan for their assistance during my work. I am grateful to all those who have helped me directly or indirectly in the successful completion of this thesis. I would like to thank all the members of office staff Mr. Lokesh Kumar, Mr.Ved Pal Singh Saini, Mrs. Shashi Prabha and Mr. Jain for their help and support.

Last but not the least; I want to thank my entire family for their love and understanding throughout this journey. I cannot express enough gratitude to my parents, who have given me

everything of themselves while asking nothing in return. Without their inspiration and pat, I would not have been able to tread even a single stride. They have supported and stood by me at every point of my subsistence and I owe the entire success of my endeavor to their love and affection. So, this thesis is for them.

Finally, I am thankful to Council of Scientific and Industrial Research, Government of India, Uttarakhand state Biotechnology for providing the financial support.

(Jitendra Kumar Sahoo)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ATP	Adenosine triphosphate
BLAST	Basic local alignment search tools
bp	Base pairs
CTAB	Cetyl trimethyl ammonium bromide
°C	Degrees celsius
DNA	Deoxyribonucleic acid
ddH_2O	Deionized distilled water
eDNA	Environmental DNA
EDTA	Ethylenediamine tetra-acetic acid
et al.	Et alia (and others)
EtBr	Ethidium bromide
g	Gram
h	Hour
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilo dalton
LB	Luri bertani
M	Molar
MW	Molecular weight
mg	Milligram
min	Minutes
mM	Millimolar
mL	Milliliter
NCBI	National centre for biotechnology information
ng	Nanogram

OD	Optical density
PCR	Polymerase chain reaction
pmol	Picomol
qRT	Quantitative real time
RNase	Ribonuclease
RNA	Ribonucleic acid
rpm	Rotation per minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
sec	Second
µg	Microgram
µl	Microlitre
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
UV	Ultra violet
w/v	Weight per volume
16S rRNA	16S ribosomal RNA

CHAPTER 1

INTRODUCTION

Microorganisms are known to be one of the most productive sources of structurally unique and biologically useful functional gene products. The first step in studying microbial process is obtaining microbial axenic culture. However, in nature as many as 99% of the microorganisms are not cultivable using standard techniques [39, 67]. Unculturable microorganisms consist of majority of the planet's living organisms. They are the promising source of many industrially important secondary metabolites and enzymes used for human welfare [9, 49]. Hence, culture-independent methods are important to understand the microbial genetic diversity, their ecological role and their ability to produce valuable products. Metagenomics or culture-independent based genomic analysis of microorganisms has the potential to understand the fundamental questions in microbial biotechnology and their relationship with the ecosystem dynamics. Microbial census from rare environments and their diversity is a rewarding area to study. Unlike animals and plants microbial diversity and its conservation has been overlooked because of the misconception that microbial life and their diversity are not under the threat of anthropogenic activity and climatic changes [16]. Microbes from extreme environments have attracted considerable attention due to their peculiar metabolism and physiology. Microbes living in extreme environments have been less studied and became a challenge as compared to microbes in other habitats due to lack of appropriate culture techniques. Geothermal springs in the Himalayan ranges are such rare environments which provide both cold and hot temperature conditions [87].

In the present study, microbial communities of the sulphur hot springs at Tattapani geothermal area of Himachal Pradesh state and Tapovan geothermal spring located in Chamoli District of Uttarakhand state, India were analyzed. Both being geothermal springs in the Indian Himalayan Region are particularly significant providing both extreme 'cold' and 'hot' temperature for exploring microbial diversity. The analysis was based on culture independent strategies to get a first insight into the microbial communities in these sulphur hot spring ecosystems. The study emphasizes on the manual method of isolation of high molecular weight (HMW) and pure eDNA directly from environmental samples. Subsequently, 16S rRNA gene based amplification, cloning into surrogate host organism (*Escherichia coli* DH5α) and sequencing of 16S rRNA gene products was done. 16S rRNA gene containing clones were pre-

screened by restriction fragment length polymorphism (RFLP) to identify unique types of polymorphisms which were then subjected to DNA sequencing. Biostatistical and bioinformatics tools were utilized for analyzing microbial community. Our overall approach was to achieve high coverage of 16S rRNA genes from the two different microbial communities. While eliminating the inclusion of artifacts on ribotype diversity, the richness of microbial 16S rRNA gene sequences was calculated using Rarefaction curves. The direct sequencing (metagenomic sequence analysis) of environmental samples provided a valuable insight into the census of culturable and unculturable organisms in various environmental niches. Microbial species richness and operational taxonomic units (OTUs) are important parameters in determination of microbial biodiversity, its role and functions. Therefore it is quite important to determine the exact microbial diversity from environmental samples. In this study, OTUs were calculated and compared in terms of rarefaction curve for both the thermal springs. The results generate baseline data on the microbial community structure and their diversity in both the hot springs of Himalayan geothermal belt and also provided information about efforts to be made for monitoring and conservation of such fragile ecosystems.

There is a great need of various structurally and functionally different gene and gene products from microorganisms of extreme environments. Such gene products show peculiar properties in terms of their activity and stability. Metagenomics promises the discovery of such uncharacterized gene products by its functional studies. Geothermal springs are home to many such thermophilic microorganisms. Microbes from such environments could be the hopeful choice for novel thermostable functional gene discovery. Direct PCR based approach for study of gene and gene products was used. Metagenomic nucleic acid extract has been earlier used for accessing biodegradative potential of microbes by the discovery of chlorocatechol dioxygenase, phenol hydroxylase, catechol 2,3-dioxygenase genes [36]. In the present study, thermostable endo beta-glucanase (*cel5A*) gene responsible for cellulose degradation was demonstrated in Tapovan metagenomic DNA. Degenerate primers were constructed for endo beta-glucanase (*cel5A*) gene. Tapovan metagenomic DNA was used for direct PCR based amplification of the gene. Subsequently, the gene was cloned using expression vector pET28(a). BL21-CodonPlus cells were used for expression of the cloned endo beta-glucanase gene by minimizing codon biasness [180]. The cloned sequence showed significant differences with endo beta-glucanase gene of *Bacillus licheniformis* and other known relatives. The cloned endo beta-glucanase protein showed an optimum temperature activity at 60 ºC indicating its thermal stability. The protein also

showed a wide range of pH stability. To the best of our knowledge this is the first report of cloning and characterization of endo beta-glucanase thermostable enzyme from Tapovan thermal spring.

Enzyme like L-asparaginase has greater potential for treating acute lymphoblastic leukemia (ALL) [28, 170]. L- asparaginase from *Escherichia coli* and *Erwinia carotovora* has been extensively studied earlier [7]. Whereas, L- asparaginase from *Escherichia coli* is toxic to human [122]. In the present study, enrichment of culturable microbes from Tapovan spring water was carried out using Biolog EcoPlates [61]. The singe carbon source based enrichment of microbes lead to the identification of L-asparaginase-II enzyme from Tapovan sample. A bacterial isolate was derived from the spring water found to produced L-asparaginase enzyme. The L-asparaginase producer was found to be a close relative of *Ralstonia sp.* Its L-asparaginase production was confirmed by growth on Modified Czapek Dox medium containing L-asparagine as sole carbon source and by nessler's reagent based biochemical assays.

Metagenomic approach of studying genes responsible for combating extreme environmental stress conditions provides us with the basic information for microbial existence and their role in the functioning of ecosystem. Tapovan, the presently studied Himalayan geothermal spring is such an extreme environment with high load of ions like sulphur, ammonium and heavy metals like aluminum, selenium, boron, arsenic. Hence, we studied the bacterial efflux systems that respond to such environmental conditions. These transporter/efflux systems play a dual role in efflux of heavy metals as well as modern antibiotics leading to multidrug resistance in microorganisms. In order to detect genes responsible for efflux system in Tapovan metagenomic DNA, real-time PCR was carried out using degenerate primers. The qPCR showed detectable amount of transporter genes *cusA* and *acrD*. CusA is an integral part of CusCFBA system and is responsible for efflux of copper/silver [104]. AcrD is a part of resistance –Nodulation- Cell division (RND) family of transporters, involved in efflux of amphiphilic substances [140]. No Human mitochondrial DNA (mtDNA) and *E. coli* plasmid pET23a-GFP DNA were found in the metagenomic DNA. This indicates that the sample was free from anthropogenic and modern routine laboratory contaminations. Analysis and quantification of genes was done using threshold value (C_T). Acriflavin resistance protein B (AcrB) is one such transporter whose lowest C_T value was observed indicating higher gene titer. The sequence analysis showed highest identity with *acrB* gene of *Escherichia coli* APEC O78

and Acriflavin resistance protein of *Escherichia coli* P12b. The presence of AcrB, AcrD is again supported as it is a well-known member of the hydrophobe/amphiphile efflux-1 (HAE-1) family of transporters involved in efflux of currently used antibiotics, disinfectants, dyes, and detergents. The presence of AcrB, AcrD and CusA in Himalayan springs can be attributed the presence of heavy metals like Al, As, Cu, Fe, B, Se and nutrient deprive conditions.

CHAPTER 2

REVIEW OF LITERATURE

2.1. Metagenomics:

Microbes are the promising source of structurally unique and biologically useful gene products for human welfare. Culture based approach for assessing microbes and their gene products are rewarding and have been in use since centuries [40, 64]. One of the greatest limitations for exploring such novel genes and gene products from nature's backyard is their cultivation using standard microbiological methods [39]. It is now widely accepted by the scientific community that as much as 99% of the microorganisms present in nature cannot be cultivated by the standard laboratory techniques [5, 39, 67, 146]. However, now it is possible to extract microbial genetic material directly from natural environment. Introduction of this previously inaccessible genetic material into easily culturable bacterial hosts, provides a means to access the biosynthetic potential of unculturable microbes. This study is called metagenomics [19].

The term "metagenomics" was first coined by Jo Handelsman and her associates in the University Of Wisconsin and was first published in 1998 [68]. The term metagenome referenced to as a collection of genes sequenced from environment, that could be analysed in a way analogous to the study of a single genome. In Greek, meta means "transcendent" In its research approach, metagenomics transcends individual genes and genomes, allowing scientists to study whole of the genomes in a community [45]. Recently, researchers in University of California (Kevin Chen and Lior Pachter) described metagenomics as "the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the method of isolation and lab cultivation of individual species" [31]. An example in case is the inventor project, taken out in Sargasso Sea which exposed DNA based recognition of about 2000 different species, with 148 types of bacteria that had been never seen before [169]. Metagenomic analysis of environmental DNA (eDNA) can be classified into sequence driven and function driven metagenomic studies [67, 146].

In sequence based metagenomic studies, major focus lies in finding the entire genetic information and the DNA sequence of target environment. The retrieved sequences can be

analyzed in many different ways. Overlapping based sequence information is useful in retrieving complete genome of an individual species in a metagenomic sequence of a particular community. Population ecology and evolution of microbial species in a community can also be studied using the metagenomic sequences reads [150]. Function based metagenomics studies include discovery of new gene products that the microbes in a community can produce. Metagenomic libraries can be constructed using the genetic material isolated directly from environmental samples. Such libraries can be screened for various functions such as enzymes, antibiotics and vitamins etc [47, 98]. Another recent approach of function driven metagenomics, made practicable by use of recent technological advances is enabling direct extraction and identification of novel proteins and secondary metabolites from a microbial community [10, 83, 179].

2.2. Extreme environments and their metagenomic studies:

Currently there is a huge interest in extreme environments, which are the natural habitats of rare extremophilic microorganisms [62, 123]. Microorganisms living in such environments have attracted considerable attention due to their peculiar metabolism, physiology and production of valuable industrial products [62, 109]. Studies on extreme microorganisms help in defining the edges of life and source of interesting enzymes having uncommon and desired properties. Hot springs in particular and their microbial flora are of greater concern today as they are the source of many valuable biocatalysts for industrial and biotechnological importance [93]. An example include development of polymerase chain reaction (PCR) technique due to discovery of taq polymerase from *Thermus aquaticus* [74]. Microbes in such extreme environments are unique in terms of their growth conditions. Many of them are solfataric and thermoacidophilic microbes belonging to bacteria and archaebacteria [3]. Those that are capable of growing at elevated temperatures have been identified as *Bacillus caldolyticus, Thermoplasma acidophilum , Thermoactinomyces vulgaris, Thermomicrobium roseum, Thermoanaerobacter ethanolicu , Geobacillus stearothermophilus Thermus aquaticus, Thermus thermophilus, Thermodesulfobacterium commune, Sulfolobus acidocaldarius, Thermococcus littoralis* and *Methanopyrus kandleri* [143].

Microbes habituated to extreme environments (high or low temperatures, pH, salt, pressure, acidic conditions) are less well studied as compared to other consortia due to their difficulties in growing under standard laboratory conditions [11, 62].

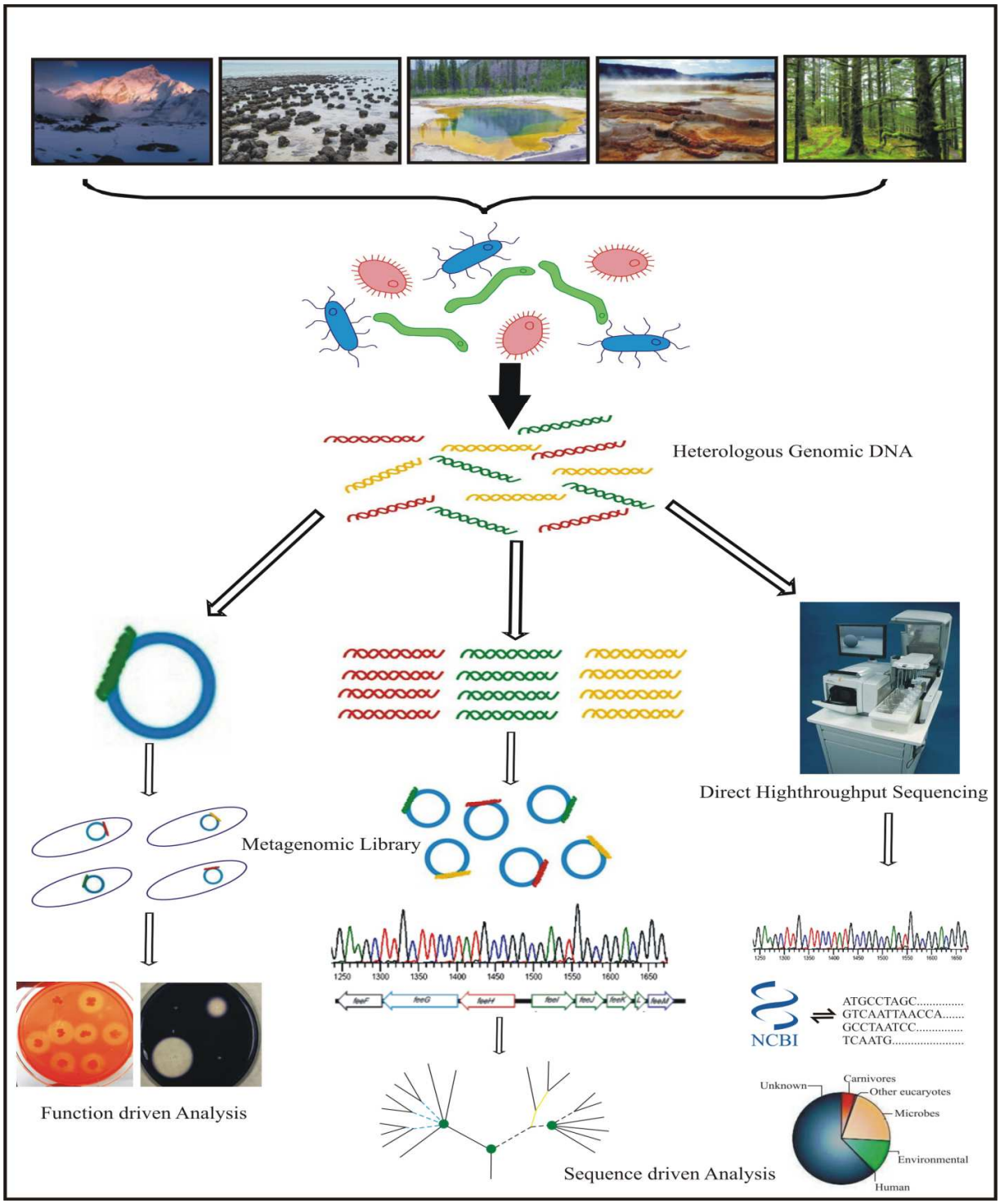


Figure 2.1. Schematic representation of methods for metagenomic sample analysis. *Study of metagenomic DNA from extreme environments using functional library construction, 16S rRNA gene library and by direct sequencing.*

Therefore, the culture independent approach of studying microbes in thermal environment is the best method of choice [93]. Microbial species identified in thermal mats include phylum of bacteria chloroflexi, filamentous cyanobacteria and purple sulfur bacteria [136, 175]. Microbes in mesophilic mats from sulfide and/or sulfur rich environments have been previously studied [48]. However, detailed analysis of the microbial community in hot spring mats using modern culture-independent method has been subject of few studies. The schematic representation of metagenomic study is presented in figure 2.1.

2.2.1. Himalayan geothermal springs and their importance:

The Himalayas represents one of the largest areas of geothermal systems. The Indian Himalayan geothermal belt extends from north east (State of Assam) to north-west (State of Himachal Pradesh) over a length of 1500 kilometers and is home to about 150 geothermal manifestations [70]. Such hot springs are proved to be generated by the emergence of geothermally heated groundwater from Earth's crust and has the connection with the volcanic/ core of the earth [70].

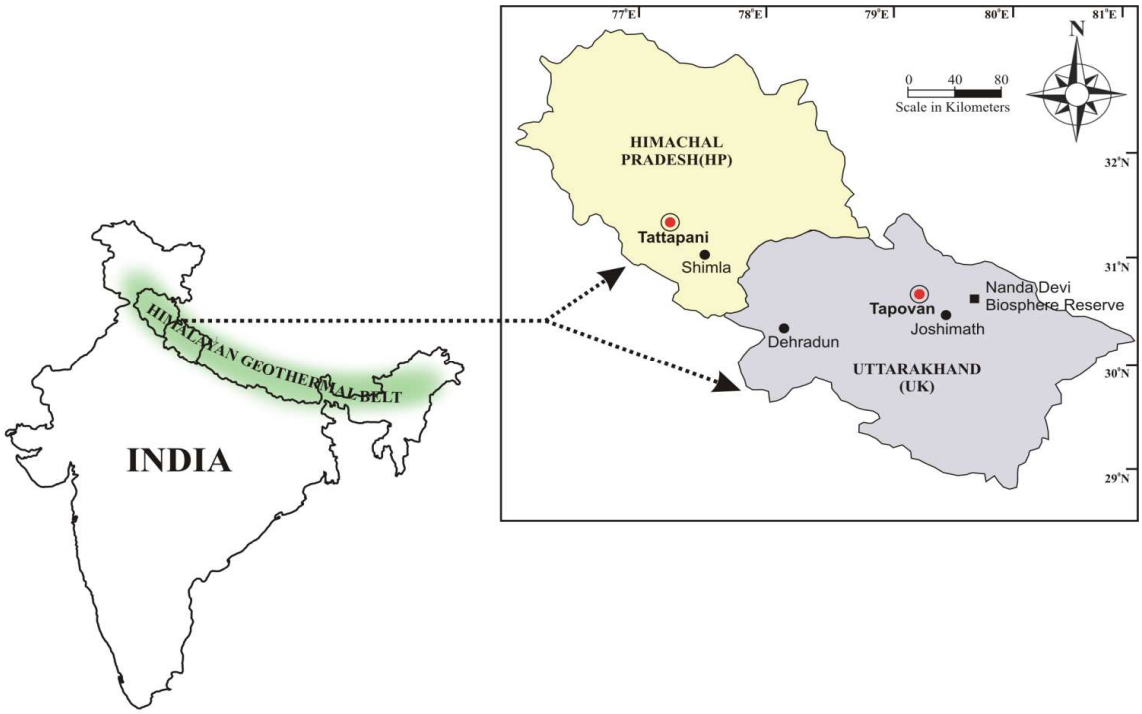


Figure 2.2. Sampling sites and location of the studied geothermal springs. *Tattapani and Tapovan geothermal springs in the Himalayan range of states Uttarakhand and Himachal Pradesh, India. Sampling sites are marked with red centered circles.*

Such extremely hot and surrounding cold environment develops a unique ecosystem inhabited by specialized microbial communities linked with the root of ancient bacterial lineages [87]. Tattapani and Tapovan are two major geothermal springs essentially supplied with deep fresh, slightly alkaline groundwater with a high load of sulphur ion. They are separated from each other by 300 km of air distance and are located in separate states of India in the Himalayan range (figure 2.2). These springs are well popular for their disease curative properties. Tapovan spring harbors an extremely high temperature of 95 °C with boiling water and vapours. Soil and mat in this spring area is siliceous type. There are more than five such thermal manifestations within 5 km radius of the central Tapovan spring. Whereas, Tattapani spring has an average temperature of 65°C with a comparatively more organic environment. This type of habitat are quite rare and no studies till date have been conducted on these environments using culture independent approaches.

2.3. Metagenomic DNA from environmental samples:

Many standard protocols have been developed to extract DNA directly from soil samples. They includes lysis of environmental microbes with bead-beating [115, 132], sonication [42], freeze-thawing [82], grinding in liquid nitrogen [77, 171] or enzymatically, using enzymes such as proteinase K [183], lysozyme [139]. A combination of these methods may also be used [162, 167]. Many literature have assured the quality of isolated metagenomic DNA by minimizing the degree of co-precipitants like humic acid, fulvic acid, phenolic compounds [19, 167]. However maintaining DNA integrity as well as concentration and keeping it free from any co-precipitants is a major challenge for the advance applications of metagenomic DNA.

Extreme environments, with low microbial load become a challenge for metagenomic DNA isolation. Hydrothermal springs, which are the extreme environments host the thick cell walled and spore forming bacteria [141]. Culturing of microbes from such environments is quite difficult under laboratory conditions. Hence, direct DNA isolation based metagenomic study is the most preferred way to assess the genome of such extremophillic microbes. Nucleic acid based analysis of thermal springs in particular has provided illuminating discoveries in functional as well as microbial diversity study [39]. Environmental DNA is quite important as they are the promising source of novel secondary metabolites like enzymes, drugs, small molecule [92, 126].

The major challenge in soil metagenomic project is the co-purification of contaminants such as humic substances, fulvic acids and humin, etc. Direct metagenomic DNA extraction from environmental sample results in the co-extraction of contaminants like humic and fulvic acids. Humic and fulvic acids in soil are formed by biodegradation of dead organic matter in soil [94]. They are functionally dibasic or tribasic acids, which precipitate along with alcohol or isopropyl alcohol [13]. On the basis of its solubility in acids and alkalis, humic substances can be divided into three main types. They are humic acid, which is soluble in alkali and insoluble in acid, ii) fulvic acid, soluble in alkali and acid and iii) humin, insoluble in both alkali and acid [151, 84]. However, complete removal of these contaminants is not possible following classical DNA extraction protocols, such as proteinase, detergent and phenol-chloroform treatments. Therefore the adopted physical method of purification of crude DNA is the most efficient and vital step while working with metagenomic DNA. It is important to note that there is no single method suitable for metagenomic DNA isolation from all environmental samples. Researcher should optimize the protocol that best suits for their DNA isolation.

2.4. Sequence based metagenomic analysis:

Sequence based studies of metagenomic DNA includes either direct sequencing of metagenome using pyrosequencing or by generating clone libraries [106, 107]. Direct sequencing of gene stretches from metagenome use functional primers for pyrosequencing. Most widely used approach for evaluation of microbial species present in natural environments are by their 16S/18S small subunit ribosomal RNA genes analysis. As every organism contains these genes, they are commonly used as phylogenetic markers [166]. 16S rRNA gene can be amplified and cloned in suitable cloning vectors, the sequence comparison with 16S ribosomal RNA gene databases allow us for phylogenetic classification [145].

Environmental samples are much more complex than single organisms in terms of whole genome sequencing, as they contain uncountable number of distinct species. Hence, are considered unsuitable for high throughput sequencing [125, 166]. The first large scale environmental shotgun sequencing was done in acid mines and Sargasso Sea. Direct sequencing from Sargasso Sea resulted 1.6 Gb of DNA reads from 7 independent libraries, The data revealed the presence of genes encoding iron-sulphur proteins (Fe-S Proteins), proteorhodopsins, electron-transport proteins, chitinases etc [169]. Many of these functional genes retrieved are

highly divergent from known family member. Sequence based screening lead to the discovery of genes encoding novel enzymes, such as dioxygenases, [Fe-Fe]-hydrogenases, dimethylsulfoniopropionate-degrading enzymes, nitrite reductases, [NiFe] hydrogenases, hydrazine oxidoreductases, chitinases, and glycerol dehydratases derived from different environments like soil, freshwater and sediment [153,130, 166]

2.4.1. Diversity of microbes through metagenomic approach:

It is believed that one gram of soil contains an estimated of 4×10^7 prokaryotic cells, whereas one gram of fertile agricultural soil contains an estimated of 2×10^9 prokaryotic cells [124]. Based on DNA analysis, number of distinct prokaryotic genomes in one gram of soil estimated as 2,000 to 18,000 [39]. These numbers might be neglected as genomes representing rare and unculturable species might have been excluded from these analyses. Therefore, the microbial diversity associated with one gram of soil might exceed to the known catalogue of prokaryotes in NCBI. The culture independent estimation of microbial diversity includes DNA reassociation and 16S rRNA gene sequencing [71]. The DNA reassociation analysis has the limitations of defining species and other diversity parameters like frequency distribution of species in DNA pool. Interpretation of the reassociation curve is complicated due to lack of proper control. Whereas, the 16S rRNA gene for microbial diversity analysis include the use of universal bacterial primers and has the potential to define diversity parameters like determination of operational taxonomic units (OTUs) [148, 154].

2.4.1.1. 16S rRNA gene and phylogenies:

16S rRNA gene is a part of 30S subunit of prokaryotic ribosome. The 16S rRNA gene is most commonly used molecular marker for prokaryotic identification. The gene contains hyper variable regions that help in species specific signature sequence identification. It has both rapidly and slowly evolving regions. The fast evolving regions are consist of variable regions useful for determination of closely related species. Whereas, the slow evolving region are useful for determination of distant species (figure 2.3) [105, 113]. Advantage of using 16S rRNA gene is due to lack of horizontal gene transfer and presence of a large database (1,613,063) sequences available in National Center for Biotechnology Information (NCBI) and Ribosomal Database Project (RDP) [177]. The method of analyzing these genes include PCR based amplification by the use of universal bacterial primers. Though originally 16S rRNA genes were used for bacterial

identification, subsequently they were capable of classifying bacteria into completely new species and even genera. They have the potential to classify bacteria that have never been successfully cultured.

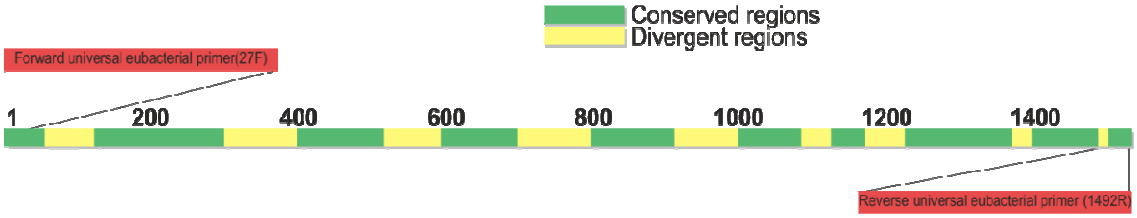


Figure 2.3. 16S rRNA gene of prokaryotes. *16S rRNA gene of prokaryotes consisting of conserved and divergent regions useful for microbial identification and classifications. Eubacterial specific forward primer (27F) and reverse primer (1492R) are shown.*

2.4.1.2. 16S rRNA gene based microbial diversity in metagenomic sample:

Over last two decades, the extensive use of molecular method of amplification, cloning, sequencing of 16S rRNA from environmental DNA have increased the knowledge of prokaryotic diversities tremendously. Databases like GenBank, Ribosomal Database Project (RDP), and Greengenes are commonly used for sequence comparison. Usually, cloned sequences similarity cut-off values of 80, 85, 90, 92, 94, or 97% are designated as phylum, class, order, family, subfamily, or species respectively [43]. The direct sequencing of metagenomic nucleic acids helped us in studying culturable and uncultured organisms from rare extreme environments [2, 76, 127]. Microbial species richness, operational taxonomic units (OTUs) are important parameters in determination of microbial biodiversity, its role and functions [18, 71, 110]. Therefore it is quite important to determine the exact microbial diversity from environmental samples. The parametric and nonparametric empirical models are adopted for species frequency distribution estimation [147]. The development of such statistical models by Schloss and Handelman lead to the estimation of 2000 and 5000 bacterial species in Alaska and Minnesota soil sample [148]. The most convenient way of obtaining such 16S rRNA gene based metagenomic library includes the use universal bacterial primers, their amplification and cloning in cloning vectors like pTZ57R/T or pGEM (figure 2.4).

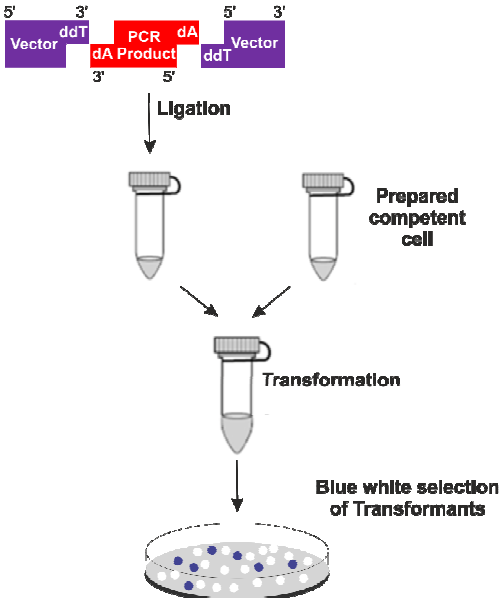


Figure 2.4. Schematic representation of methods for cloning of 16S rRNA gene. *T/A cloning of PCR product of 16S rRNA gene from environmental environment DNA using pTZ57R/T vector and Escherchia coli host.*

2.5. Function based metagenomics analysis:

Function based metagenomic analysis is possible by the construction of metagenomic libraries using suitable vector and host system. Suitable host and vector system has to be chosen for the successful construction and expression of a metagenomic library [67, 137]. Targeted metagenomics based functional study is also possible by the use of degenerate primers.

2.5.1. Metagenomic libraries and their studies:

The basic steps in metagenomic DNA library construction includes size selection of eDNA, cloning of size selected DNA into appropriate vector and host system. Study of several metagenomic library and their screening have been carried out till today [8, 20, 34, 164]. The approach of studying such library are broadly similar, although selection of vector and host systems has varied (table 2.1). Large and small insert metagenomic library are constructed for finding novel secondary metabolites and enzymes. As average size of the genes for most enzymes are 1 to 2 kb, the construction of small insert metagenomic libraries with high copy number plasmid vector are useful. Large insert metagenomic library are useful for screening of antibiotics and other complex operon systems [21, 159]. But, the limitation for cloning in

Bacterial artificial chromosome is that, large size metagenomic DNA is not readily isolated from soil and well preserved in the process of metagenomic library construction.

Table 2.1. Environmental metagenomics studies and their importance in biotechnology and ecology.

Soil type/ Condition/ Location	Host and vector system used	Screening type	Findings	Features	References
Agricultural soil, Madison US	*Escherichia coli*, BAC	16S rRNA gene, Functional	Finding of clones displaying heterologous antibacterial, lipase, amilase and nuclease.	Heterologous expression and 16S rRNA gene libraries are studied	[138]
Agricultural soil, Madison US	*Escherichia coli*, BAC	Functional, color production	Triaryl antibiotics turbomycin A and B	Heterologous expression of agricultural soil DNA	[57]
Uncultivated soil, New England, US	*Escherichia coli*, BAC	Functional, Antibacterial activity	Soil derived clones found producing indirubin	60 MB of soil DNA was screened to find the antileukemic drug indirubin	[95]
Forest soil	*Escherichia coli*, Fosmid	Functional, Antibacterial activity	Identification of antifungal *E. coli* clones	A study of antifungal activity of metagenome	[34]
Forest soil	*Escherichia coli*, Cosmid	Functional, Antibacterial activity	identification of coloured clones in cosmid clones conferring violecine	violecine induces apoptosis in fibroblast cells and active against gram positive cells.	[20]
Forest soil	*Escherichia coli*, Cosmid	Functional, Antibiotic activity	Metagenomic derived clones are identified that were antagonistic to *Bacillus subtilis*	N-acyl-L-tyrosine antibiotics discovered	[21]
Meadow sugar field, cropland soil samples	*Escherichia coli*, plasmid (pSK+) Functional, Carbonyl formation	Functional Carbonyl production	15 of the recombinant clones namly pAK201–215 produced a stable carbonyl	Carbonyl are produced from short-chain polyols like 1,2-ethanediol, 2,3-butanediol.	[85]
Forest soil	*Escherichia coli* Plasmid (pUC19)	Functional, Amylase formation	amylase gene (amyM) environmental library was over expressed and purified .	Amylase produced was used to hydrolyse soluble strach, cyclodextrin,	[182]
Agricultural field soil	*E. coli* TOP10, Plasmid (pZero-2)	Functional, Amidases	Clone was found producing highly active penicillin amidase.	Found amylase was a promissing biocatalyst for synthesis of beta lactum antibiotics	[53]
Rice straw compost	*Escherichia coli* BL21 (DE3), pET20 b(+)	Functional, Cellulase	A 4.7 kb gene fragment was identified identified and subcloned from metagenomic source.	Purified cellulase enzyme was stable over a wide range of pH, temperature	[181]
Forest soil, cow rumen, rooted tree	*Escherichia coli*, Lambda zap expression vector.	Functional, Cellulase	Five endo beta 1, 4 glucanase clones were identified from metagenomic libraries.	Two from cow rumen , one from soil and two from elephant dung active clone from libraries showed cellulase hydrolytic activity.	[50]
Jae Sawn hot spring, Thailand	*E. coli* TOP10, pZErO-2	Functional, Lipolytic	patatin-like phospholipase (PLP) and esterase (Est) were identified from clones of metagenomic library	PLP and Esterase gene product were stable at pH 3.0 to 12 at 70 ° C.	[164]
Pond water sample, University of Delhi, India	*E. coli* DH10B, pUC19 vector	Functional, Lipolytic activity	lipolytic protein identified was shown similarity to yet uncharacterized a/b hydrolase protein family.	532 Mb of community DNA was trapped in 30 K clones.	[134]

2.5.1.1. Vector systems for metagenomic library:

Vectors used for metagenomic library construction is an important parameter for expression of heterologous gene. Selection of appropriate vector depends on the desired insert size, copy number of plasmid. In addition to that, quality of environmental DNA affects the selection of vector systems. Use of pET and pBluescript vector have the advantage that they maintain a high copy number in *Escherichia coli* facilitating even the weak heterologous expression for screening of functional gene products [59]. Active expression of the environmental genes depends on the native promoter. This promoter biasness can be minimized by using cis-acting promoters. The limitations of using plasmid vector is the small insert size hence large number of clones need to be screened for desired gene product. This method limits the finding of single gene products or small operons. To alleviate such problems large insert based vector systems are in use. They include cosmid, Fosmid, vectors which are able to clone 20 to 40 kb insert DNA and BAC vectors able to clone 20 to 200 kb fragments [103]. The advantage of making large insert libraries is the ability to study large fragments of genes, operon or recovering entire metabolic pathways. Fosmid vectors are commonly used for an improved method of cloning and stable maintenance of cosmid-sized (35–45 Kbp) inserts in *E. coli*. The successful detection of novel activities of the heterologously cloned metagenomic based genes depends on its expression. This depends on several factors like the cloned insert gene should carry the complete gene of a particular protein to be studied. The promoter and ribosomal binding sites are essential to be functional in the chosen host system for constructing metagenomic libraries.

2.5.1.1.1. pNYL-*rygC* vector:

pNYL-*rygC* vector is a negative selection vector based on a novel toxic peptide IbsC present in *Escherichia coli* genome [52, 114]. The peptide is located in the opposite orientation of SibC small RNA. The vector has ColE1 origin of replication, toxic peptide IbsC which is regulated by TetO/R [tetracycline operator-repressor]. Tet promoter based on/off regulatory system is one of the strongest and tightly controlled gene regulatory system. TetO/R tightly controls the expression of toxic peptide in *E. coli* DH5αZ1. Strain *E. coli* DH5αZ1 encodes tetracycline repressor [TetR] in the genome and is constitutively synthesized. TetO/R activity is thus controlled via TetR and anhydrotetracyline (ATC) [analogue of tetracycline]. Non-recombinant cells expresses the toxic peptide leads to death of cell. Insertion of a DNA fragment into the multiple cloning site disrupts the gene coding for the toxic peptide, allowing direct positive

selection of clone. This vector system have multiple cloning sites [BamH I, Afe II, Hae II, Sal I and EcoR I] that can be used to clone the desired DNA fragment in pNYL-*rygC* vector.

2.5.1.2. Host systems for metagenomic library:

Escherichia coli is the most common host of choice for construction of environmental libraries[37]. The major problem include in heterologous gene expression is codon bias, inability to use regulatory elements like ribosome binding site and promoters. From many of the metagenomic library study showed that most of the soil bacteria come under bacteria: α, β, γ-Proteobacteria, Acidobacteria, and Actinobacteria. Jeffrey *et.al.* have used *Agrobacterium tumefaciens* (α-proteobacteria), *Caulobacter vibrioides* (α-proteobacteria), *Burkholderia graminis* (β-proteobacteria), *Ralstonia metallidurans* (β-proteobacteria) *Escherichia coli* (γ-proteobacteria), *Pseudomonas putida* (γ-proteobacteria) and from the phylum Proteobacteria for metagenomic library construction. Hence, taking *Escherichia coli* as a choice fulfills the requirements of a easily screenable host [149]. Codon usage also plays a major role in the expression of heterologous genes. Cells with CodonPlus properties have extra copies of t-RNAs which allows reorganization of a wide range of heterogeneous codon and hence a high level of expression occurs. In this study the BL21-CodonPlus strain was used, which were engineered to carry extra copies of tRNAs that facilitates a heterologous expression of proteins in *Escherichia coli* [180].

2.5.2. Screening of metagenomic libraries:

Screening of metagenomic libraries can be achieved either by enzymatic activity of recombinant clones or by identifying the sequences they harbor. Agar plate or broth based simple and clear strategies are suitable for screening huge number of clones. Screening of recombinant clones can also be achieved by using tetrazolium indicator plates. Tetrazolium is colorless and water soluble in its oxidized form but turns into purple to deep red in its reduced state. Positive clones cause oxidation of the substrate on assay plate, tetrazolium gets reduced and the deep red formazan is accumulated in the cell. For detection of hydrolytic enzymes like lipase and esterase activity, agar plates containing tributyrin and tricaproin can be used [117]. Positive clones become visible by the zone of clearance around the colonies. Similarly, cellulase enzyme can also be detected on agar plate containing carboxy methyl cellulose (CMC), and stained with congo-red followed by washing with NaCl for the appearance of clear hallow unstained zone around the positive clones.

Apart from the above plate based assays, high-throughput screening methods have been developed using automated robotic system to combat screening of large number of clones in a very short time period.

Till today several of the secondary metabolites, enzymes, natural products retrieved through this approach are violacein, deoxyviolacein, Long-Chain N-Acyl Amino Acid Antibiotics, turbomycin A, Amylase, Xylanase, Lipase, Esterase, Protease, Cellulase, [67]. Study of extremophilic environmental DNA either by sequence or by functional screening of lead to the discovery of thermostable and thermolabile secondary metabolites with biological potential. A recent example of thermostable enzyme from high temperature was a reverse transcriptase used in RT-PCR, thermostable, alkalophilic beta D galactosidase was discovered from hot spring [65]. Similarly cold adapted enzymes discovered from cold temperature environment metagenomes include lipase [69], protease, amylase, cellulase [49].

2.6. Challenges in metagenomic libraries studies:

Potential difficulties in constructing metagenomic libraries for the expression of bioactive compounds in *E. coli* include, gram negative bacterium may not have the appropriate genetic background for heterologous expression of genes in their active form. Only 40% of the heterologous genes expresses in its active form [49, 91]. These problems can be overcome by initial cloning and maintenance of metagenomic DNA in *Escherichia coli* and subsequent transfer to different suitable host for expression and screening [1]. Metagenome being a complex mixture of genetic material from a great variety of microorganisms, the surrogate host's machineries cannot recognize the transcriptional and/or translational signals within the metagenome [33]. So far, metagenomic approaches have only been able to reveal the surface of the genomic, phylogenetic diversity and novel gene stored in environmental microbes. Strategies are needed to be improved for heterologous gene expression and production of functional recombinant proteins.

2.7. Direct PCR based targeted metagenomics:

Direct amplification of known gene functions is a more direct way to discover gene clusters from metagenomic DNA preparation. Successful amplification using the degenerate primers is a key factor for targeted metagenomics. Many a times the amplification depends on the titer of gene of interest in the metagenomic DNA preparations. Sample pre-information could help in designing

the primers. Pretreatment and enrichment of sample using specified substrates enriches the microorganisms in the sample and subsequent direct PCR based targeted metagenomics lead to the discovery of novel genes [75]. Braker and et. al. have developed the method to amplify nitrate reductase genes nirK and nirS genes from environmental sample, where the degenerate primers were used for amplification of reductase genes [22, 23]. Now a day's researcher are using the degenerate primer to amplify particular gene of interest, followed by direct pyrosequencing to obtain the gene information. Degenerate primers for aromatic dioxygenase genes from eDNA has also been studies like ways [22]. The major challenge of such approach of study is designing appropriate primers which need existing sequence information. Further only a fragment of structural gene gets amplified using gene specific primers, therefore, additional steps are required to obtain the full length genes. In the present study, the degenerate primer against beta endo-glucanase gene responsible for cellulose degradation were constructed and were used for PCR amplification, following touchdown procedure. The criticality of amplification was performed by using Highfidility HotStarTaq DNA polymerase enzyme.

2.7.1. Cellulase enzyme:

Cellulase consist of a set of enzymes produced mainly by bacteria, fungi, and protozoans that catalyze the hydrolysis of cellulose [165]. Other eukaryotic cellulase producers are the termites. Several types of cellulase are known based on their mechanism of action. They are endocellulase randomly cleaves internal regions at amorphous sites of cellulose to create new chain ends. Mechanistically, cellulase consist of three set of enzymes, endo-(1,4)-β-D-glucanase, exo-(1,4)-β-D-glucanase and β-glucosidases.

Cellulase degrades into cellobiose and cellotetrose by exocellulase. The cellobiose then converts into simpler reduced sugar form by the action of β- glucosidase. Endoglucanase (EG) arbitrarily cleaves the O-glycosidic bonds present internally, generating glucan subunit of different lengths resulting in tetrasaccharides, disaccharides such as cellobiose [86]. The exoglucanase releases cellobiose and cellotriose as it acts on the ends of the cellulose chain and β-glycosidases act specifically on the exocellulase product into individual reduced sugar like glucose (figure 2.5) [8, 16].

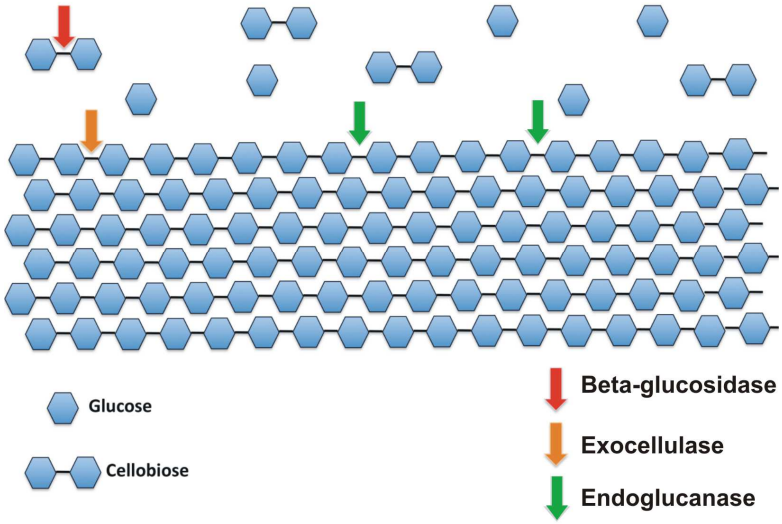


Figure 2.5. The action of endocellulase, exocellulase and beta- glucosidase. *The crystalline structure of cellulose undergoes hydrolysis to simpler cellulose by cellulase enzyme.*

2.7.2. Cellulase from microorganisms:

Cellulase are synthesized by a variety of microorganisms including bacteria and fungi. Their optimum production has been studied by growing them on cellulosic materials. These organisms are classified into aerobic, anaerobic, mesophilic or thermophilic based on their growth conditions. Microbial genera including *Cellulomonas, Clostridium, Trichoderma, Thermomonospora* and fungus *Aspergillus* sp. are most extensively studied for cellulase production and their kinetics [79, 135, 158]. Fungal cellulase are structurally simpler than bacterial cellulase. The bacterial cellulase enzyme composed of multiple domains with separate functions. Functional ability of specific domain has been identified by proteolytic truncation experiments or by genetic deletion analysis. The domains are classified into Catalytic domains (CD), Cellulose-binding domains (CBD) [56]. Cellulases are also produced by other symbiont like microbes in intestine of termites [24]. Many structurally and mechanistically different cellulases has been reported in literature. Cellulolytic enzymes from fungi mesophilic filamentous fungi *Trichoderma reesei* have been studied extensively [108, 135]. Only a limited number of studies are available about the metagenomic approach of finding cellulase out of thermal environments.

2.7.3. Cellulase from metagenomic sample:

The essence of metagenomic approach to study microbes from eDNA involve the use of PCR amplification based on taxonomic group and functional gene analysis. Genes responsible for cellulase production are rarely amplified directly from environmental DNA. However there are very few examples of cellulase discovery through this approach. An alternate strategy to get a preliminary idea about cellulolytic bacteria in a metagenomic DNA pool includes its PCR based 16S ribosomal RNA sequencing. For example, subgroup of genera clostridia, cellulomonas fibrobacters, bacillus and have been studied for their cellulase production [44]. But, the problem resides with their isolation and cultivation in axenic form from environment.

Carboxymethyl cellulose (CMC) is the substrate most commonly used for cellulase detection in solid agar plate or in liquid based assayes [161]. Congo red is the dye that has noncovalent affinity towards cellulose fibers. Hence, congo red is commonly used for screening of cellulose in solid agar plate. Agar plates containing 1% CMC when cultured with cellulase producing organisms and overlaid with congo red solution followed by washing with 1% NaCl results into a zone of hydrolysis on the plate.

2.7.4. Importance and industrial applications of cellulase:

Table 2.2: List of applications of cellulase in agriculture, food and industries.

Industry	Applications
Agriculture	Improves soil quality, plant disease control, enhances seed germination and root development.
Bioconversion	Transformation of cellulosic substances to organic acids and ethanol.
Fermentation	Involve in primary fermentation and improves quality of beer, clarification in wine production.
Food	Improves starch and protein extraction from food, improves texture, flavor and volatile properties of vegetables and fruits, clarification of fruit juices.
Pulp and Paper	Enzymatic deinking, biomechanical pulping, improves fiber brightness.
Textile	Biopolishing of textiles, softening of garments, improves fabrics quality.

Cellulases have been in use for more than 30 years in the sector of food, animal feed, pulp and paper, agriculture, wine and brewing making, biomass refining, textile, and laundry [14, 15, 73, 88] listed in table-2.2. This enzyme have been taken up for research in academic and

industries [14]. Cellulase in combination with xylanase used for deinking of paper wastes. Cellulose bezoar formation in human stomach is also treated using cellulase.

2.8. Asparaginase and its importance:

Asparaginase is an enzyme that catalyzes the conversion of L-asparagine to L-aspartic acid by hydrolysis process. They are naturally produced from bacteria and fungus. Asparaginase from *Erwinia carotovora* are in use for treatment of acute lymphoblastic leukemia (ALL) [170]. L-asparaginase catalyses the hydrolysis of free L-asparagine in blood, hence selectively drives the leukemic cells to death as these cells cannot synthesize the asparagine in its own. L-asparaginase is the most preferred drug of choice for treatment of acute lymphoblastic leukemia because of its non-toxic, biodegradable nature and can be administered at the local sites quite easily. L-asparaginase are used in food processing as they reduce the acrylamide formation during starchy food processing. Acrylamide are formed in fried and oven-cooked foods, potato chips due to the reaction of free asparagine and reducing sugars [17] . L-Asparaginase hinders the interaction between L-asparagine and reducing sugars, therefor acrylamide formation is inhibited [51]. Mineralization of nitrogen in soil is also carried out by L-asparaginase. Major sources of L-Asparaginase are animal, plant and microbes [51]. Hence, there is an urgent need of rapid and greater production of the enzyme to meet the human demand. For clinical trials, apart from *Escherichia coli* and *Erwinia carotovora* this enzyme has been obtained from *Serratia marcescens*, bakers' yeast, guinea pig serum, and chicken livers. Serological studies with the antibodies sensitive to the L-asparagenase from microorganisms proved that L-asparaginase differs immunologically. The availability of two or more different L-asparaginases would be advantageous in clinical trials and in human use.

2.8.1. Screening of Asparaginase:

The modified Czapek Dox medium is used for production and detection of L-asparaginase in cultured isolates. Modified Czapek Dox medium containing L-asparagine is the only C and N source and containing phenol red or bromothymol blue as pH indicator with color changes on the agar plate. Czapek Dox medium with phenol red when grown with L-asparaginase producing organisms, turns into re from yellow. Similarly, bromo thymol blue containing plates under above conditions turns into blue in color due to pH change of the medium. Released ammonia during the reaction lead to an increase in the pH of reaction mixture (figure 2.6). Dye bromo

thymol blue is yellow at acidic pH and turns into blue at alkaline pH. The ammonia (NH3+) production in the reaction mixture containing enzyme L-asparaginase and substrate asparagine can be detected by Nessler's reagent treatment which gives a deep yellow color to the reaction mixture.

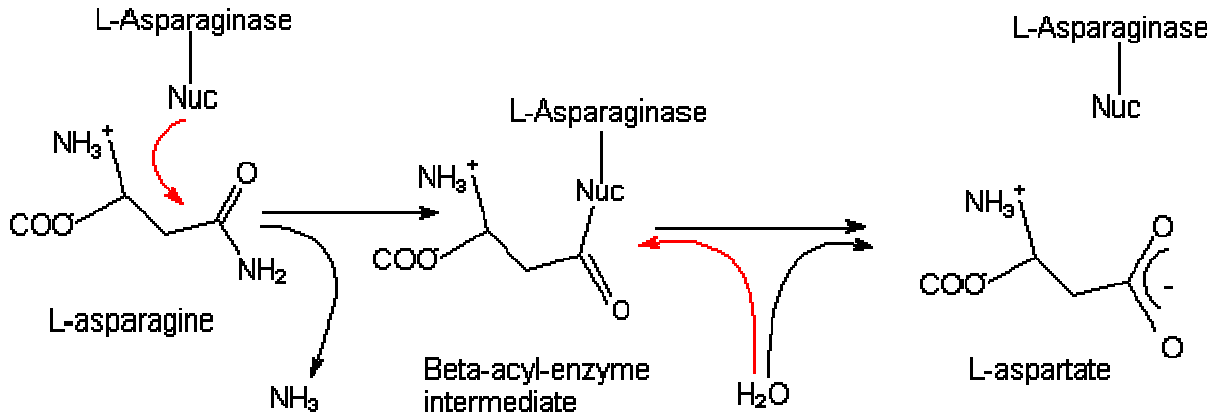


Figure 2.6. Conversion of L-asparagine to L-aspartate in presence of L-asparaginase. *Amino acid L-asparagine converts into L-aspartic acid (Aspartate) in presence of enzyme L-asparaginase. Ammonium ions forms further converts into ammonia in the reaction.*

L-asparaginase from different sources varies with their physico-chemical properties. L-asparaginase from *Escherichia coli* designated as EC-1 and EC-2 [27]. Only EC-2 posses the anti-leukemic activity. EC-1 shows rapid decrease in activity below pH-8.4. L-asparaginase from *Erwinia carotovora* and *Bacillus* sp. shows the optimum activity at pH 8.0. The enzyme activity decreases sharply at 40 ºC from such species [144]. Literature shows that L-asparaginase from many microbes are not a metalloprotein, hence they do not require cofactors. Presence of chelating agents like EDTA and other thiol protecting groups such as 2-betamercaptoethanol, glutathione, dithiothreitol increases the activity of the enzyme. [133, 144].

2.9. Real-time PCR based detection of genes in metagenomic DNA:

2.9.1. Targeted metagenomics based study on transporters and antibiotic resistance genes from eDNA:

Quantitative polymerase chain reaction (qPCR) is a highly sensitive method for the detection of low titer gene in template DNA. Real-time PCR is the most sensitive way to determine rare and low concentrated gene candidates in environmental samples [90, 101]. Researcher have

developed the methods to study such low abundant genes from natural environment using quantitative PCR [99, 155]. Louis and et. al. have developed the real-time PCR based method for determination of butyryl- coenzyme A (CoA) gene from metagenomic DNA preparations [99]. The real-time PCR based approach is also used for determination of glycopeptides, tetracycline and β-lactam antibiotics from Beringian permafrost sediments [38]. Most vital step for studying the PCR based approach to metagenomic DNA include the assessment of origin of studied genes. Use of human mitochondrial DNA, chloroplast DNA of plant and *E. coli* plasmid pET23a-GFP for contamination assessment assay is also reported [6, 121].

2.9.2. Efflux pumps that help combat with extreme environmental conditions:

The Himalayan thermal springs, which are originated from the emergence of geothermally heated ground water is a hotspot for high stress environment. The high temperature, high load of sulphur, high concentration of certain heavy metals makes the existence of prokaryotic organisms challenging [54]. Microorganisms in this stressful environment have developed adaptations and are not only surviving but also nurturing [30, 81]. Bacterial efflux pumps play the vital role in survival and development of specialized microorganisms [119]. List of bacterial transporters are given in table 2.3. The RND family of transporters were responsible for heavy metal resistance in *R. metallidurans* [120]. Such transporters which were involve in efflux of hydrophobe/amphiphilic substances like acriflavin dye, detergents and multiple antibiotics.

Table 2.3: List of different family of transporters:

S. no	Family of transporter	Metal ions	Name and composition of transporter	References
1	RND	Cu^{2+}, Ag^{+}	CusCFBA transporter	[96]
2	RND	ethidium bromide,benzalkonium bromide, Triton X-100, sodium dodecyl sulfate.	AcrB transporter	[100]
3	P-type	Mg^{2+},Mn^{2+},Ca^{2+},K^{+},Cu^{2+},Zn^{2+},Cd^{2+},Pb^{2+},Ag^{+}	One membrane protein	[156]
4	A-type	As^{2+}	One membrane integral protein and a dimeric ATPase subunit	[119]
5	MIT	Most cations	CorA	[25]
6	RND	sodium dodecyl sulfate and deoxycholate	AcrD transporter	[4]

2.9.3. Quantitative PCR based gene detection in metagenomic DNA:

Efflux pumps: CusA, AcrB, AcrD:

Gram-negative bacteria use tripartite efflux system of resistance-nodulation-cell division (RND) family to expel various toxic compounds from cell [129]. The efflux system CusCBA is responsible for efflux of toxic heavy metals from cell cytoplasm to outside. CusA is an integral part of CusCFBA and is responsible for efflux of copper/silver [96]. CusA is a part of heavy metal efflux (HME) subfamily of resistance –Nodulation- Cell division (RND) efflux pumps [120]. CusA is capable of dynamically picking metal ions from cytoplasm and efflux out. Acriflavin resistance protein B (AcrB) is a transporter involve in efflux of antibiotics, disinfectants, dyes, and detergents.

CHAPTER 3

MATERIALS AND METHODS

3.1. Study sites and sample collection:

Soil samples were collected from Tattapani hot spring (31.2489° N and 77.0861° E, height 655 meter from sea level) and Tapovan (30° 54' 16.4916" N, 79° 10' 22.7064" E, and height 1828 meter from sea level) of Himalayan ranges (figure 2.2). These Himalayan thermal springs are known to be originated by the emergence of geothermally heated groundwater. Soil sample (mat), water was collected from the bed of hot springs. Growths of algal mats were recorded at both the sites. The site of sample collection is exposes to a wide range of temperature in different seasons. The Hot Springs maintain a uniform temperature of 65°C and 85°C, pH was recorded about 8.34 and 8.06 respectively. Tattapani soil consists of low organic content, whereas Tapovan spring soil consists of mostly siliceous soil with inorganic environment (Table 4.1). The soil and water samples were brought to the laboratory and soil was directly processed for metagenomic crude DNA preparation without storing it for longer time as prolonged storage under laboratory conditions destroys the natural microbial flora of samples.

3.2. Water chemistry analysis:

Physical and chemical parameters of water were analyzed. Temperature was recorded at the site using thermometer. pH was measured using digital pH meter (Micro pH analytica). Ionic concentration of both the hot springs water were determined using ionic chromatography (IC) and heavy metals present in water sample were determined by inductively coupled plasma mass spectrometry (ICPMS).

3.2.1. Ionic chromatography:

Elemental ions of both the spring water sample were determined using 792 Basic IC-Metrohm. Water sample was used at two times diluted milliQ water. Diluted water sample was filtered with 0.22 μm syringe filter (millipore). Sample was injected through the injector of IC-Metrohm. Ionic concentration were determined for anions Fluoride (F^-), Chloride (Cl^-), Nitrite (NO_2^-), Bromide (Br^-), Nitrate (NO_3^-), Phosphate (HPO_4^{2-}), Sulphate (SO_4^{2-}) and cations Lithium (Li^+), Sodium (Na^+), Ammonium (NH_4^+), Potassium (K^+), Magnesium (Mg^{2+}), Calcium (Ca^{2+}).

3.2.2. Inductively coupled plasma mass spectrometry (ICPMS):

Elemental heavy metals were analyzed using ICPMS (ELAN DRC-e, PerkinElmer SCIEX) for Al, As, B, Co, Cu, Fe, Hg, Mn, Sb, Cr, Cd, Ni, Zn, Se (Table 4.1). The digested water sample was filtered with 0.22 µm syringe filter (Millipore, Germany).

3.3. Metagenomic DNA isolation and purification from hot spring mat:

The method adopted for metagenomic DNA isolation from spring mat was same as previously reported in the literature, with certain modifications for superior quality DNA isolation.

3.3.1. Manual method of metagenomic DNA isolation:

The metagenomic DNA was extracted on a large scale. About 50 gms of sieved fine hot spring soil was taken in a 250 ml nalgene bottle and was treated with 75 ml of lysis buffer. After proper manual mixing, 250 µl of proteinase K (10 mg/ml) was added. Nalgene bottle was incubated at 65 °C for 2 hours. The sample in each bottle was mixed by hand thoroughly after every 15 minutes. After lysis, the lyses soup was allowed to cool at room temperature. RNase A solution was added to the soup at a final concentration of 100µg/ml and incubated at 37 °C for half an hr. The samples were centrifuged at 5000g for 10 mins. The supernatant was transferred into a sterile nalgene bottle. The lysate was treated with 0.7 volumes of Isopropanol and incubated at room temperature for 1 hour. Precipitation of eDNA was carried out at 10000 rpm for 30 mins at 4°C. Supernatant was drained off without disturbing the pellet. The brown pellet was washed with 70% ethanol at 10000 rpm for 10 minutes at 4°C three times. The eDNA pellet was dried at room temperature and dissolved in 1 ml of nuclease free water. The pellet was heated at 50°C for 10 mins and eDNA was completely dissolved by incubating at 4°C overnight. Crude nucleic acid extracts from soil sample had a yellow-brown color. The DNA concentration was estimated spectrophotometrically. The molecular weight and quality of DNA was analyzed by running 10 µl of eDNA in 0.8% agarose with Ethidium Bromide (EtBr) and visualized under UV transilluminator.

0.5M Na EDTA: Dissolve 18.61 gms of Na EDTA in 40-50 ml of DDW, the pH was adjusted to 8.0 by adding 10 M NaOH. The final volume was made to 100 ml. Prepared solution was autoclaved at 15 psi, 121°C for 15 minutes.

Lysis buffer composition:

Tris Cl	100 mM (pH 8.0)
EDTA	100 mM (pH 8.0)
NaCl	1.5 M
CTAB	1%
SDS	1%
Poly clar resin AT	1%

Lysozyme (100mg/ml) stock preparation: 100 mg of lysozyme was dissolved in 1 ml of DDW sterilized through 0.22 μm syringe filter (Millipore, Germany). Stored at -20°C.

Proteinase K (10mg/ml) stock solution preparation:

Glycerol – 5 ml
1M Tris Cl (pH-7.5) - 500 μl
$CaCl_2$- 145 mg
Proteinase K- 100 mg
Final volume made into 10 ml with DDW.

RNase A (10mg/ml) stock preparation: 100 mg of RNase in 200μL of 0.5M Tris Cl and 30 μL of 5 M NaCl remaining DDW to make it 10 mL. Heat at 100 ºC for 15 min, allowed to cool at room temperature and stored at -20 ºC.

3.3.2. DNA isolation trough commercial kit (HiPurA™ Soil DNA purification kit, Himedia, India):

500 mg of soil samples were added to 750 μl lysis solution in HiBead containing tube (Himedia, India). The tube were vortexed at maximum speed for 10 minutes. At the end of vortexing the tubes were centrifuged at 13000 g for 1 minutes to separate out the soil particles. The supernatant soup was taken in a fresh eppendorf tube and 250 μl of inhibitor removal solution (IRSH) was added to the soup. The content was mixed properly and incubated at 4ºC for 5 minutes. The precipitate was separated at 12000 rpm for 1 minutes at room temperature. The supernatant soup was taken in a 2 ml clean collection tube. To which, 1.2 ml of binding solution (SB) (Himedia, India) was added and vortexed. 650 μl of the lysate was taken in HiElute miniprep spin column and centrifuged for 1 minute at 12000 rpm. This step was repeated for all the amount of lysate to pass through the column. The column was washed with 500 μl of diluted wash solution. The washing step was repeated once again. An empty spin was given at 12000 rpm for 2 minutes to

remove extra residual ethanol. The column was transferred to a new collection tube and retained DNA was eluted with 100 µl of nuclease free water (NFW).

3.3.3. Purification of metagenomic DNA using electro-elution method:

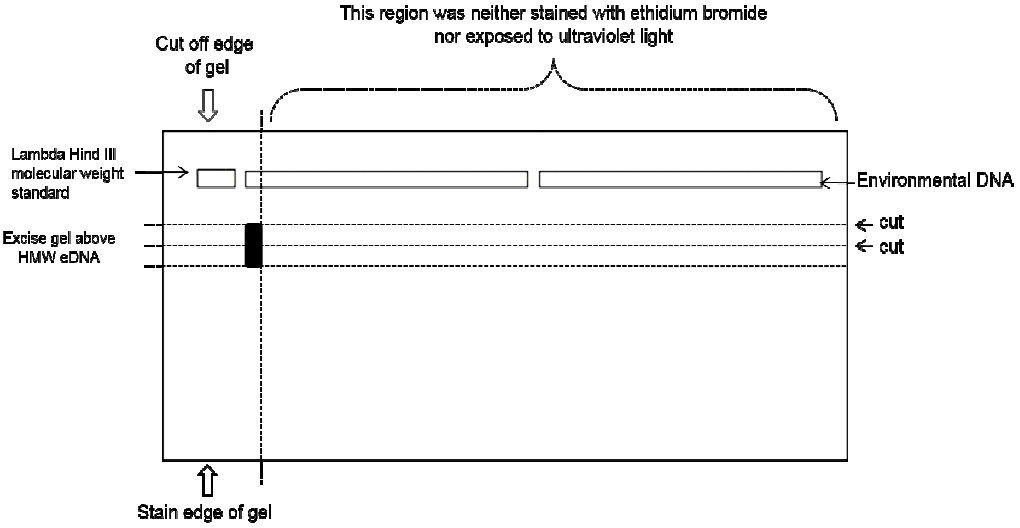


Figure 3.1: Manual setup for agarose gel during large scale electro-elution and purification of metagenomic DNA.

The DNA isolated by the above method carries many inhibitors like humic acids and formic acid which inhibit the subsequent biomolecular applications. 200µl (10µg) of eDNA was loaded onto 1% agarose gel without EtBr, along with lambda Hind III marker. The gel was run at 100V for 1 hour and then voltage was reduced to 60V for 3 hrs. At the end of the run, the gel portion carrying lambda Hind III marker and a portion of eDNA was excised with the help of a new razor. The gel slice was stained with 0.5µg/ml of EtBr for 15 mins. High molecular weight eDNA was visualized under UV light in the stained portion of gel slice. With the help of a razor top and bottom of the high molecular weight eDNA corresponding to 23 kb lambda marker was marked. This was a 1 cm smear around ¼ th of the way down the gel. The gel slice was re aligned with the unstained portion of the gel away from UV light. High molecular weight (HMW) eDNA in the unstained gel slice was excised and placed in 1X TAE. The gel slice along with 1X TAE was placed in the pre activated dialysis membrane (heated in boiling water for 5 mins). The dialysis membrane had a molecular weight cutoff (MWCO) of 1 kDa. Both sides were clamped avoiding air bubbles inside the bag. Dialysis bag was oriented in the gel apparatus as that it was parallel to electrodes (perpendicular to the current) and covered with 1X TAE running

buffer. Electro-elution was done for a period of 3 hours at 100 V. At the end of electro-elution 1X TAE in the bag was carefully removed and placed into 15 ml Amicon centrifugal concentrator (MWCO 30K). Volume of DNA was reduced to 250-500 µl by centrifuging at 4500g for 15 mins. The washing step was repeated again with 1X TAE to collect the unwashed DNA portion. eDNA quality was determined by running 10µl of concentrated DNA along with lambda hind III marker. The quality was found to be satisfactory as it runs along with 23 kb band of lambda hind III MW standard. Quantification of DNA was carried out spectrophotometrically at A_{260} nm.

Quantification of DNA:

1 OD = 50 µg DNA/ ml

Concentration of DNA (µg/ml) = (OD A_{260}) x (dilution factor) x (50 µg DNA/ml)/ (1 OD_{260} unit)

Preparation of EtBr solution: 100 mg of EtBr to 10 mL ddH_2O, mixed well using votexer and stored.

50X TAE stock preparation: 242 gm of Tris base (pH – 8.0), Glacial acetic acid- 57.1 gm, 0.5 M EDTA (pH-8.0) was dissolved in 700ml of DDW and then made upto 1 lit. Prepared solution was autoclaved at 15 psi, 121°C for 15 minutes.

3.3.4. Quantification and quality analysis of DNA using PCR and restriction digestion:

The quantification of purified DNA was determined spectrophotometrically at 260nm and concentration of contaminates protein and humic acid were determined by taking the optical density spectrum from 220 to 350nm. The quality of genomic DNA isolated was determined using DNA to protein ratio and DNA to humic acid ratio at 260/280 and 260/230 respectively.

Quality of DNA was further analyzed by PCR based approach using eubacterial specific 16S rRNA gene primers the 27F and 1492R.

PCR conditions:

Reagents	**Stock concentration**	**Working concentration**	**Amount in 50µL reaction**
Hotstar master mix	10X	1X	5Ml
Dntps	2 Mm	0.2 mM	5µL
Forward primer	10 µM	0.4 µM	2µL
Reverse Primer	10 µM	0.4 µM	2µL
Hot star taq DNA	5 U/µL	2.5 U	0.5µL

polymerase			
Template DNA	50 ng/µL	100 ng/µL	2µL
Nuclease free water			33.5µL

PCR Program: The following PCR program was used for amplification of 16S rRNA gene from Tattapani metagenomic DNA.

Step 1 (Initial denaturation)	94°C for 15 min
Step 2 (Denaturation)	94°C for 1 min
Step 3 (Annealing)	47°C for 30 sec
Step 4 (Extension)	72°C for 1.5 min
Step 5	Go to Step 2 for 5 times
Step 6 (Denaturation)	94°C for 1 min
Step 7 (Annealing)	49°C for 30 sec
Step 8 (Extension)	72°C for 1.5 min
Step 9	Go to Step 2 for 25 times
Step 10 (Final extension)	72°C for 10 minutes
Step 11	Hold at 4°C forever

The amplified PCR product was electrophoresed in 1% agarose gel containing EtBr. The PCR product was visualized under UV light.

Quality of the purified metagenomic DNA was assessed by partial digestion of DNA with Sau3AI enzyme.

Restriction digestion conditions:

Reagent used	**Volume**
Plasmid DNA	5 µL
Buffer Sau3AI (10X)	2 µL
Sau3AI (10U)	0.5 µL
Nuclease free water	12.5 µL
Total	20 µL

Restriction digestion was performed for a period of 1 hour at 37 °C. At the end of incubation reaction was inactivated by adding 1mM of EDTA. The extent of digestion was analyzed by electrophoresis of the digested product with the undigested metagenomic DNA.

The crude DNA, the manually electro-eluted DNA and the commercial kit eluted DNA were compared spectrophotometrically by taking spectrum of OD from 220nm to 350nm. OD at 230, 260, 280nm were retrieved from the spectrum data and the 260/280 ratio for protein contamination and 260/230 ratio for humic acid contamination were determined individually for all three category of isolated DNA.

3.4. Microbial diversity analysis of hot springs through metagenomic approach:

The eDNA isolated and purified metagenomic way from both Tattapani and Tapovan spring was subjected to 16S rRNA gene library construction and sequencing for a comprehensive microbial diversity analysis.

3.4.1. 16S rRNA gene based metagenomic library construction:

3.4.1.1. Amplification of 16S rRNA gene from hot spring DNA:

A set of eubacterial specific primers were used to amplify about 1465 base pair section of the 16S rRNA genes from total metagenomic DNA from both the springs, as described by Acinas et al, 2004. Total environmental DNA were amplified by polymerase chain reaction (PCR) in a reaction mixture (50µl) containing 1X Hotstar master mix (Qiagen), and 50 ng of template metagenomic DNA, 0.5 µMol of each of forward 27f (5'-AGAGTTTGATCCTGGCTCAG-3') and reverse primer 1492r (5'-GGTTACCTTGTTACGACTT-3') and 2.5 units of HotStar hifidelity taq DNA polymerase (Qiagen). The reaction mixtures were incubated in a thermal cycler at 94°C for 4 min for initial denaturation followed by 5 cycles of 60 s at 94°C, 60 s at 47°C, 90 s at 72°C and 25 cycles of 60 s at 94°C, 60 s at 49°C, 90 s at 72°C and a final extension of 10 min at 72 °C. The amplified products from the replicate PCRs were pooled together and analyzed on 1 % agarose gel. The product of 1.5kb size was then selected and eluted from agarose gel using QIAquick® gel extraction kit (Qiagen, USA). The quality and quantity of eluted DNA from gel was analyzed by running a fraction of it on 1% agarose gel as earlier.

3.4.1.2. Gel purification of amplified 16Sgene by QIAquick® gel extraction kit (Qiagen, USA):

The 1.5 kb PCR product was electrophoresed on 1% agarose gel at 70 V/cm^2 for 1 hour along with Gene Ruler 1 kb DNA ladder (Fermentas). The fraction of PCR product was excised out using new razor.

Table 3.1: Primers used in the study for microbial identification, phylogenetic analysis and sequencing.

S. No	Sequence (forward /Reverse)	Sequence (5'-3')	Size of amplicon	Organism type amplified
1	27f	5'-AGAGTTTGATCATGGCTCAG-3'	1465 bp	Bacterial primer
	1492r	5'-GGTTACCTTGTTACGACTT-3'		
2	pUC/M13F	5'-GTTTTCCCAGTCACGAC-3'	Variable	Sequencing
	pUC/M13R	5'-CAGGAAACAGCTATGAC-3'		

3 volume QG* Buffer was added to 1 gel volume (1 mg gel ~ 1 µL) and sample was incubated at 70°C for 10 min or until the gel slice has completely dissolved. Subsequently 1 gel volume of isopropanol (Bio basic Inc, India) was added to the sample and mixed by vortexing. For binding of DNA, sample was applied to the QIAquick® column (Qiagen, USA) and centrifuged for 1 min at 13000 rpm. Flow-through was discarded and the column was placed back into the same tube. For washing, 0.75 mL PE* Buffer was added to column and centrifuged for 1 min at 12000rpm. Column was centrifuged for another 2 min at 12,000 rpm to remove residual wash buffer. QIAquick® column was then placed into a clean 1.5 mL eppendorf tube (Tarson, India). For elution of bound DNA, 30 µl Buffer EB* was added to the center of the QIAquick® column and the column was centrifuged for 1 min at 13000 rpm.

*Buffer QG (Qiagen) – Guanidine thiocyanate

*Buffer PE (Qiagen) – Sodium perchloride and ethanol

*Buffer EB (Qiagen) – 10 mM Tris-Cl, pH 7.5

3.4.1.3. Quantification and ligation of gene of interest into pTZ57R/T vector:

The eluted insert DNA was quantified by spectrophotometer at 260 nm. 450 ng of insert DNA was used for cloning with 150 ng of T/A cloning vector pTZ57R/T (InsT/A clone^{TM} MBI Fermentas, Germany) DNA. 10 µL ligation reaction was set as follows.

Table 3.4: Ligation reaction into pTZ57R/T cloning vector:

Reagents	Final Conc. (In 10µL)	Amount (In 10 µL)
10X T4 ligase buffer*(Fermentas, USA)	1X	1 µL
T4 DNA ligase (10 U µL-1)	5U	0.5 µL
ATP (10 mM)	0.5 mM	0.5 µL
pTZ57R/T (55ng/ µL)	150ng	2.7 µL
Amplified 16S rRNA gene (˜ 85ng/µL)	450ng	5.3 µL

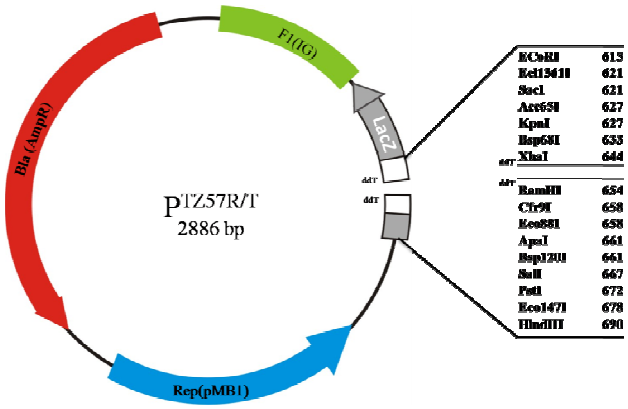


Figure 3.2 : Cloning vector pTZ57R/T (InsT/A clone™ MBI Fermentas, Germany).

Ligation reaction was carried out for 2 hours at 22 ºC. At the end of ligation the mixture heat inactivated at 70 ºC for 15 mins.

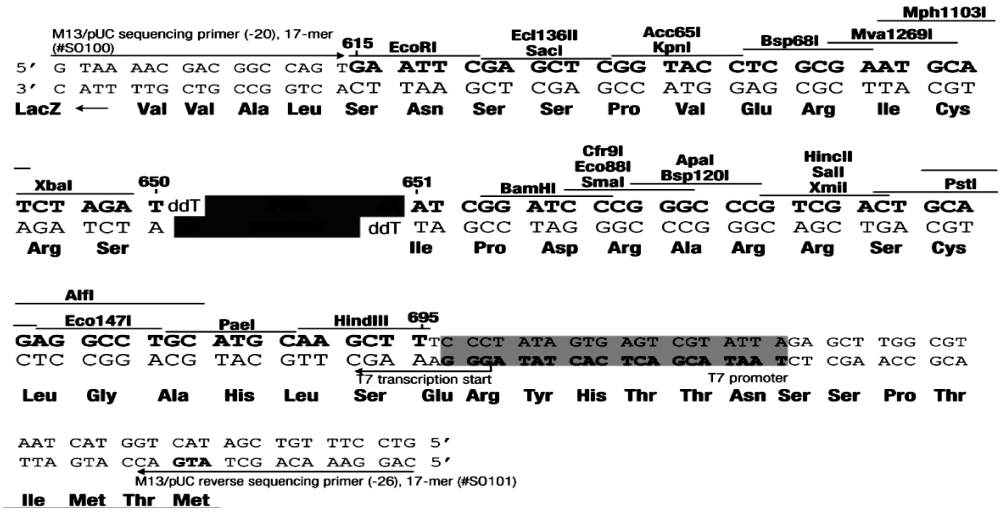


Figure: 3.3. Multiple cloning site of cloning vector pTZ57R/T .

3.4.1.4. Transformation of ligated product into *Escherichia coli* host:

The ligated product in pTZ57R/T vector was subjected to electro transformation in electro competent cell of *Escherichia coli* DH5α.

Electro-competent cell (*Escherichia coli* DH5α) preparation:

Escherichia coli DH5α cell was pre inoculated in 10 ml LB. On the day of competent cell preparation 1% of inoculum was sub cultured in fresh LB and grown till OD reached 0.5 at A_{600}. 400 ml cell were harvested in 2 nalgene tubes by centrifugation at 3K rpm for 20 min at 4 ºC. Cell were suspended in 100 ml of 10% glycerol and kept in ice for 20 min. The cell were harvested at 2500 rpm for 30 mins. The cell was again suspended in 50 ml 10% glycerol. Cell were again centrifuged at 2500 rpm and pellet was dissolved in 1 ml 10% glycerol each.

Harvested cells followed by centrifugation was dissolved in 150 µl of 10% glycerol and aliquated to 50µL each and quick frizzed in liquid nitrogen and stored at -80 ºC.

The heat inactivated ligated product in pTZ57R/T vector was subjected to electro transformation in electro-competent cell of *Escherichia coli* DH5α. 1.5µl of ligated product was mixed with the electro-competent cells on ice. They were mixed gently by tapping on the finger tip. The competent cell was then transferred into pre cooled electroporation cuvette eppendorf – 1mm (Eppendorf, Germany) and an electric pulse of 1700V for 5 m seconds was given to the cells (Using electroporater Eppendorf , Germany). Immediately the cell was transferred to 1 ml SOC media (Himedia, India; 0.5% Yeast Extract; 2% Tryptone; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM Glucose). The cells were enriched at 37 ºC at 200 rpm for an hour. The cells were pooled out by centrifugation and were selected on LB plates containing 100 µg/ml of ampicillin, 40µg/ml of XGal and 0.1 mM IPTG. Single white colonies were re-streaked on LB Ampicillin plates. Plasmid were isolated from cultures raised from these individual clones. Purified plasmids served as templates for 16S rRNA gene sequence determination.

Ampicillin stock solution preparation (100mg/ml): 100 mg of sodium salt of ampicillin (Sigma) was dissolved in 1ml DDW and filtered in 0.22 µm millex syringe filter (Millipore UK). Stored at -20 ºC.

XGal stock solution preparation (20mg/ml): 200 mg of X-Gal (USB) was dissolved in 8ml of Dimethylformamide (DMF) final volume was made upto 10 ml by adding remaining DMF. Solution was stored at -20ºC.

IPTG stock solution preparation (100mM): 0.238 gm of IPTG (USB) was dissolved in 8 ml of DDW and final volume was made upto 10 by adding remaining DDW. Filtered through 0.22 μm millex syringe filter (Millipore UK). Stored at -20 ºC.

10% Glycerol preparation: 50 ml of glycerol (Merck) was diluted with 450 ml of water and was autoclaved at 15 psi, 121°C for 15 minutes.

Preparation of SOC medium: 3.41gm of SOC (Himedia, India) was dissolved in 100 ml of DDW. The solution was autoclaved at 15 psi, 121°C for 15 minutes.

3.4.2. Screening of transformants and restriction pattern analysis:

Randumly selected clones from each library were picked for Restriction Fragment Length Polymorphism (RFLP) analysis. Plasmid were isolated from the clones using GeneJET Plasmid Miniprep Kit (Fermentas germany).

Plasmid DNA isolation:

Overnight culture of individual clones in ampicillin (100 μg/ml) were used for plasmid DNA isolation. 1.5 ml cells were harvested at 8000 rpm centrifugation for 3 minutes. The pelleted cells were suspended in 250 μl of the resuspension buffer. The bacteria were resuspended completely by vortexing or pipetting up and down till no cell clumps remains (RNase A added to the resuspension solution previously). 250μl of lysis solution was added and was inverted for 4 to 5 times sothat the solution becomes viscous and slightly clear. To the lyset mixture 350 μl of neutralization solution was added and mixed thoroughly by inverting the tubes 5 to 6 times. The content in the tube was centrifuged for 5 mins at 12000 rpm to pelletdown the cell debris and chromosomal DNA. The supernatant was transferred to GeneJET spin column, centrifuged for 1 minutes. The column was washed with 500 μl of the diluted wash solution by centrifugation at 12000 rpm. The wash step was repeated once again and the flow through was discarded. The GeneJET spin column was placed in a new 1.5 ml microcentrifuge tube and 50 μl of the Elution Buffer was added to the center of column and elution was done at 12000 rpm centrifugation for 2 mins. The plasmids from libraries were double-digested with HindIII and EcoRI (MBI Fermentas, Germany) restriction enzymes and AluI enzymes in two separate sets. Fragments were vizualized through elecrophoresis on 2% agarose gel. Plasmid clones showing different polymorphism in either of the digestion sets were identified by eye and were considered to be unique. Such plasmids 61 clones from Tattapani and 28 clones from Tapovan which showed

fragment polymophism were were sent for sequencing to SciGenom Labs Private Ltd (Kochi, India).

Double digestion of pTZ57R/T containing 16S rRNA gene with Hind III and EcoRI:

Reagent used	**Volume**
Plasmid DNA	5Ml
Buffer R (10X)	2 µL
Hind III	0.5 µL
EcoRI	0.5 µL
Nuclease free water	12µL
Total	**20µL**

Digestion of pTZ57R/T containing 16S rRNA gene with AluI:

Reagent used	**Volume**
Plasmid DNA	5µL
Buffer Tango (10X)	2 µL
Alu I (10U/µL)	0.5 µL
Nuclease free water	12.5 µL
Total	**20 µL**

Digested product from both the libraries were analyzed for Restriction Fragment Length Polymorphism (RFLP) analysis. Fragments were vizualized through elecrophoresis on 2% agarose gel. Plasmid clones showing different polymorphism in either of the digestion sets were identified by eye and were considered as unique.

3.4.3. Sequencing of clone libraries:

Selected clones from RFLP analysis were sent for sequencing at SciGenom Labs Private Ltd (Kochi, India) using M13 forward primer. Sequencing was performed using ABi 3730Xl (Sanger sequencing methodology).

3.4.4. Affiliation and validation of sequenced clones in NCBI:

The sequenced 16S rDNA fragments were analyzed for chimeric sequence elimination using pintail 1.1 version. The non-chimeric and non-redundant nucleotide sequences from this study were deposited in the GenBank database and got the valid accession numbers.

3.4.5. Relative abundance analysis of microbes:

The non-chimeric sequences were subjected to classify into phylum, class, order, family or species of microorganisms based on their identities profile from Basic Local Alignment Search Tool (BLAST) search. Microbial abundance bar diagram was calculated for both the spring sequenced clone libraries.

3.4.6. Microbial diversity analysis of clone libraries:

Shannon-weiner index and microbial richness were calculated for the clones online FastGroupII online tool (http://fastgroup.sdsu.edu/cal_tools.htm).

3.4.6.1. OTUs and rarefaction analysis for clone libraries:

The 16S rRNA based sequences were analyzed for calculating Operational Taxonomic Units (OTU) and the Rarefaction curve was generated at 95% confidence of Interest (CI) values using Analytical Rarefaction version 1.3 available at UGA Stratigraphy Lab (http://www.uga.edu/strata/software/index.html). Rarefaction curves were generated and compared for both the geothermal springs.

3.4.6.2. Phylogenetic analysis of clone libraries and phylogenetic tree construction:

The phylogenetic relation among clone sequences from both the libraries separately were determined by the construction of phylogenetic tree. Phylip-3.69 version was used for the phylogenetic tree construction. The non-chimeric sequences form the clone libraries were selected and multiple aligned through MUSCLE. The aligned sequence were saved in phylip format. The DNADIST based execute file was operated taking phylip formatted align file. The output of DNADIST was taken as an input file for Neighbor Joining (NJ) distance matrix at 1000 replicator as bootstrap value. The out tree file was visualized through using visualizing software TreeView 1.6.6.

3.5. Microbial culture based isolates and their analysis through Biolog EcoPlates based assay:

The microbial community structure and their metabolic ability was analyzed by Biolog EcoPlates. 500 ml of water sample from both the spring were filtered through 0.22µm filters (Millipore). The filters were shifted to 10 ml of normal saline and were vertexed vigorously for 20 minutes. The assay was set by taking 100 µl water sample of both the hot springs in separate

Biolog plates and were incubated at 37 °C for 24 hours. The sole carbon utilization assay was performed in triplicate. The color change was visualized after proper incubation and its OD was taken at 590nm in MicroStation™ System/MicroLog plate reader USA.

1 ml of water sample from Tapovan spring was inoculated in 2X concentration of nutrient broth and were enriched at 70 °C for 12 hours. 100µl of enriched culture and from 10^{-4}, 10^{-5} dilution were spread on 2% nutrient Gelangum Gelright agar (Applied biosciences, India) plates and plate were kept at 55 °C. Single and distinct bacterial colonies were sub cultured and maintained as axenic culture. Their colony morphology and grams nature were visualized in Nikon, Eclipse Ti-S compound microscope at 100X magnification. The genomic DNA was isolated from all the three axenic cultures using standard protocols. Identification of the organisms were carried out by using 16S rRNA gene. Further characterization of isolates were carried out using Biolog GN microplates and their read was taken at MicroStation™ System/MicroLog, USA. A phylogenetic tree was constructed using Phylip 3.69 version, taking 16S rRNA genes from all the isolates and their relative organisms from BLAST search. Fatty acid methyl ester (FAME) profile of thermophilic isolate was carried out at royal life sciences Pvt. Ltd Hyderabad India.

3.6. Metagenomic functional library construction from hot springs DNA:

Both Tattapani and Tapovan spring metagenomic DNA was subjected to metagenomic library construction for functional screening of enzymes. Manually isolated and purified eDNA was used for plasmid based small inert library construction.

3.6.1. Partial restriction digestion of hot spring DNA:

Metagenomic DNA from both the spring were used for partial digestion using Sau3AI (MBI Fermentas, Germany). The partial digestion reaction was carried out for 5 10, 15 minutes in order to obtain maximum of the partially digested DNA into 2 to 10 kb range. 10 mins of partial digestion was found suitable for cloning in plasmid vector pNYL-*rygC*.

Ragent used	**Volume for 20 µL**	**Volume for 100 µL**
Plasmid DNA	15 Ml	75 Ml
Buffer Sau3AI (10X)	2 µL	10 µL
Sau3AI (10U)	0.5 µL	2.5 µL
Nuclease free water	2.5 µL	12.5 µL

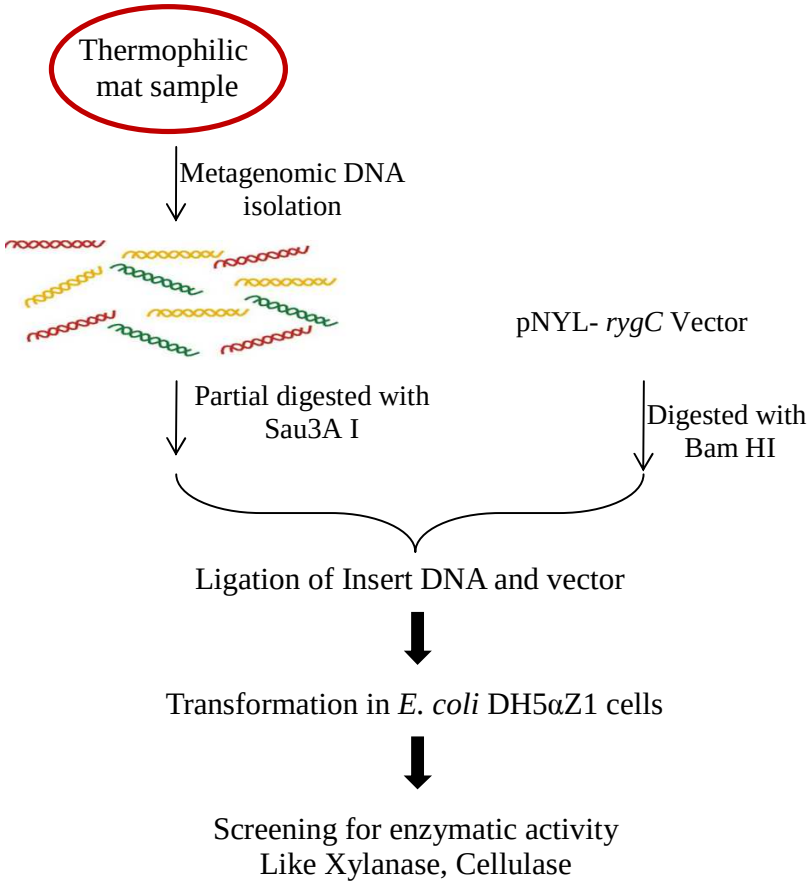


Figure 3.4. Schematic representation of metagenomic functional library construction. *Partial digestion of environmental DNA and their cloning in pNYL-rygC and in Escherichia coli DH5αZ_1 host.*

3.6.2. Size selection and gel purification of partially digested DNA:

The 10 mins partially digested DNA, which found suitable for cloning in plasmid vector was electrophoresed in 1% agarose gel at 80 V/cm^2 for one hour and DNA containing gel slice was excised out in the range of 2 to 9 kb. The sliced gel was treated with 3 gel volume of QG Buffer. Incubated at 70°C for 10 min or until the gel slice has completely dissolved. Subsequently 1 gel volume of isopropanol (Bio basic Inc, India) was added to the sample and mixed by vortexing. The sample was then applied to QIAquick® column (Qiagen, USA) and centrifuged for 1 min at 13000 rpm. Flow-through was discarded and the column was placed back into the same tube. For washing, 0.5 mL PE Buffer was added to column and centrifuged for 1 min at 12000rpm. Column was centrifuged for another 2 min at 13000 rpm to remove residual wash buffer.

QIAquick® column was then placed into a clean 1.5 mL eppendorf tube (Tarson, India). For elution of bound DNA, 30 µl Buffer EB was added to the center of the QIAquick® column, elution was done at 12000 rpm for 1 min.

3.6.3. pNYL-*rygC* plasmid isolation, digestion and its purification:

The pNYL-*rygC* plasmid carrying cells were inoculated in 40 µg/ml of LB Kanamycin broth for overnight growth. 1.5 ml cells were harvested by centrifugation at 6000 rpm for 3 minutes. The pelleted cells were suspended in 250 µl of the resuspension buffer. The bacteria were resuspended completely by vortexing or pipetting up and down till no cell clumps remains (RNase A added to the resuspension solution previously). 250 µl of the lysis solution was added and was mixed thoroughly by inverting the tube 4 to 5 times sothat the solution becomes viscous and slightly clear. 350 µl of the neutralization solution was added to each of the tube and mixed immediately and thoroughly by inverting the tubes 5 to 6 times. The content in the tube was centrifuged for 5 mins at 12000 rpm to pelletdown the cell debris and chromosomal DNA. The supernatant was transferred to GeneJET spin column, centrifuged for 1 minutes. The column was washed with 500 µl of the diluted wash solution by centrifugation at 12000 rpm. The wash step was repeated once again and the flow through was discarded. The GeneJET spin column was placed in a new 1.5 ml microcentrifuge tube and 50 µl of the Elution Buffer was added to the center of column and elution was done at 12000 rpm centrifugation for 2 mins. The isolated plasmid DNA was analysed on 1% agarose gel. Digestion of plasmid DNA with Bam HI (MBI Fermentas, Germany) was carried out as follows. Digestion was performed for a period of 3 hours. At the end of digestion heat inactivation of the reaction mixture was performed at 70 ^{o}C for 15 mins.

Digestion of recombinant pNYL-*rygC* containing metagenomic DNA.

Reagent used	**Volume for 100 µL**
Plasmid DNA	75 µL
Buffer Bam HI (10X)	10 µL
Bam HI (10U)	2.5 µL
Nuclease free water	12.5 µL

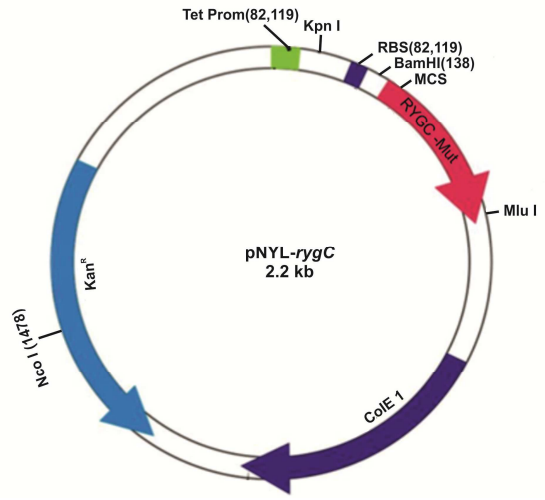


Figure 3.5. pNYL-*rygC* vector. *pNYL-rygC vector used for metagenomic library construction from Tattapani and Tapovan eDNA.*

Whole of the digested product was loaded on to a fused well of 1% agarose gel. The plasmid was electrophoresed in the gel and with help of a new rasor blade 2.2 kb digested plasmid was excised out. Purification of digested plasmid from gel was carried out using QIAquick® gel extraction kit (Qiagen, USA). The quality of the purified vector was analyzed on 1% agarose gel. Quantification of DNA was carried out spectrophotometrically at A_{260} nm.

Quantification of DNA:

1 OD = 50 µg DNA/ ml

Concentration of DNA (µg/ml) = (OD A_{260}) x (dilution factor) x (50 µg DNA/ml)/(1 OD_{260} unit)

3.6.4. Ligation of DNA into pNYL-*rygC* vector and transformation into *Escherichia coli* host:

The prepared partially digested size selected hot springs metagenomic DNA and the Bam HI digested pNYL-*rygC* vector were used for ligation and subsequently transformation in *Escherichia coli* DH5αZ_1 to generate functional metagenomic libraries. 450 ng of partially digested insert DNA was used for cloning with 150 ng of digested pNYL-*rygC* vector. 10 µL ligation reaction was set as follows. Ligation was carried out at 16 °C for 12 hours (Orbitek, Scigenics Biotech). The ligated product was heat inactivated at 70 °C for 15 mins.

Reagents	Final Conc. (In 10µL)	Amount (In 10 µL)
10X T4 ligase buffer (Fermentas, USA)	1X	1 µL
T4 DNA ligase (10 U µL-1)	5U	0.5 µL
ATP (10 mM)	0.5 mM	0.5 µL
pNYL-*rygC* vector (Bam HI digested) (50 ng/µl)	150ng	3 µL
Sau3AI partially digested metagenomic DNA (90 ng/µl)	450ng	5 µL

Ligation was performed at 22 ºC for 2 hours. The ligated mixture was heat inactivated and used for transformation.

Electro-cometent cell (*Escherichia coli* DH5αZ$_1$) preparation:

Escherichia coli DH5αZ$_1$ cell was pre inoculated in 10 ml LB. On the day of competent cell preparation 1% of inoculum was sub cultured in fresh LB and grown till OD reached 0.5 at A_{600}. 400 ml cell were harvested in 2 nalgene tubes by centrifugation at 2500 rpm for 20 min at 4 ºC. Cell were suspended in 100 ml of 10% glycerol and was kept in ice for 20 mins. The cell were harvested at 2500 rpm for 30 mins. The cell was again suspended in 50 ml 10% glycerol. An additional step of wash was given to cell in 1 ml 10% glycerol. Cells were then dissolved in 150 µl of 10% glycerol and aliquated to 50µL each. Quick frizzed in liquid nitrogen and stored at -80 ºC.

The heat inactivated ligated product was electro transformed into electro competent cells of *Escherichia coli* DH5αZ1. 1.5µl of ligated product was mixed with electro-competent cells on ice. They were mixed gently by tapping on the finger tip. The competent cell was then transferred into pre cooled electroporation cuvette-1mm (Eppendorf, Germany) and a electric pulse of 1700V for 5 seconds was given to the cells. Immediately the cell were transferred to 1 ml SOC media (Himedia, India; 0.5% Yeast Extract; 2% Tryptone; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM Glucose). The cells were enriched at 37 ºC at 200 rpm for an hour. The cells were pooled out by centrifugation and were spread on LB plates containing 40 µg/ml kanamycin and 20 ng/ml of Anhydrotetracycline. Plates were incubated at 37 ºC overnight. The transformant clones are further analyzed for insert.

Kanamycin stock solution preparation (50mg/ml): 50 mg of kanamycin (Sigma, USA) was dissolved in 1 ml of DDW and filtered in 0.22 μm millex syringe filter (Millipore UK). Stored at -20 ºC.

3.6.5. Screening of libraries for transformant:

Clones from both Tattapani and Tapovan library were selected randomly for transformant screening. Clones were inoculated in 40 μg/ml of LB kanamycin broth and plasmid were isolated using GeneJET spin column plasmid isolation kit (Thermoscientific bio). Initially plasmid were vizualised in 1% agarose gel. They were quantified spectrophotometrically at A_{260} nm. Isolated plasmid were restriction digested with Kpn I (NEB) and Mlu I (NEB), the reaction mixture is mentioned below. Restriction digestion was performed for 3 hours and was heat inactivated at 70 ºC for 15 mins. The digested product was analysed in 2% agarose gel stained with ethidium bromide. The isolated plasmid was double digested with Kpn I and Mlu I as follows.

Reagent used	**Volume for 10 μL**
Plasmid DNA	5 μL
Buffer 2 (10X)	1μL
Kpn I (10U)	0.5 μL
Mlu I (10U)	0.5 μL
Nuclease free water	3 μL

The digested product was visualized on 2% agarose gel for insert analysis.

3.6.6. Screening of libraries for enzymes:

The transformation product was used directly for screening on the agar plate. Clones were screened for cellulase and xylanase.

Screening of cellulase on agar plate:

Transformants product was spread on LB agar plate containing 40 mg/ml of kanamycin, 20ng/ml of anhydrotetracyclin (ATC) and 1% Carboxy Methyl Cellulose (CMC) (Sigma, USA). Plates were incubated at 37 ºC for 12 hours. Plates with the transformed clones were then incubated at 30 ºC for 72 hours for expression of secondary metabolites. The plates were then flooded with 1mg/ml solution of congo red for 15 mins. The plates were then washed off using 1N NaCl for 15 mins.

Screening of Xylanase on agar plate:

Transformants product was selected on LB agar plate containing 40 mg/ml of kanamycin, 20ng/ml of ATC and 1% Xylan (sigma USA). Plates were incubated at 37 °C for 12 hours. Plates with the transformed clones were then incubated at 30 °C for 72 hours for expression of secondary metabolites. The plates were then flooded with 1mg/ml solution of congo red for 15 mins. The plates were then washed off using 1N NaCl for 15 mins.

3.7. Direct PCR based discovery of enzyme (cellulase) from hot spring eDNA:

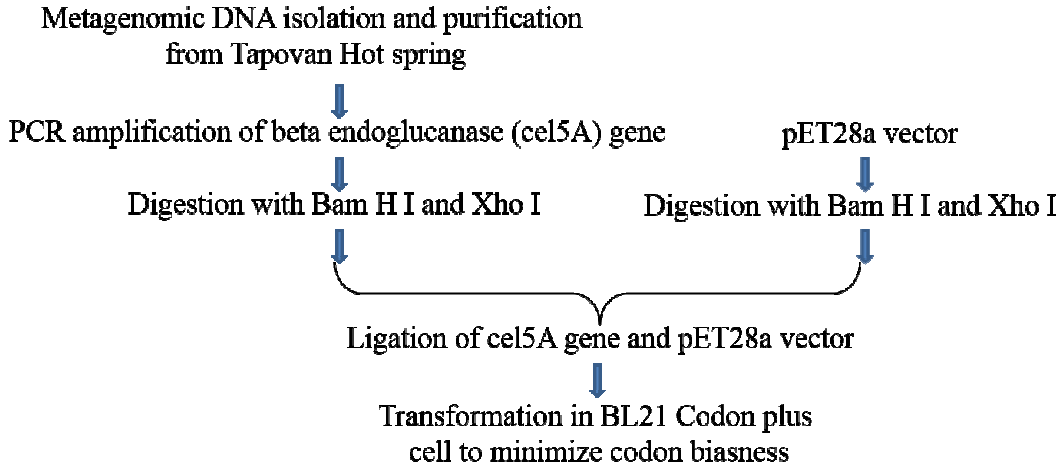


Figure 3.6. Schematic representation of direct PCR based cloning of beta endo glucanase gene. *PCR amplification and cloning of cel5A gene in pET28(a) vector using Escherchia coli BL21- CodonPlus.*

Facing the difficulties in expression of clone libraries in pNYL- *rygC*, an alternate and direct approach of enzyme screening by PCR was performed to discover cellulase gene from Tapovan metagenomic DNA.

3.7.1. Cloning of cellulase (*cel5A*) gene from hot spring DNA:

The Tapovan spring metagenomic DNA was PCR amplified with degenerate primers of beta endoglucanase gene (*cel5A*) responsible for cellulose degradation.

Primer used for cellulase (*cel5A*) gene amplification:

Degenerate primer for cel5A gene was constructed from the consensus sequences of the gene.

Primer name	**Primer sequence 5' to 3'**
cel5A FP with Bam HI	ATTAGGATCCCGTTCCATCTCTGTCTTCAT
Cel5A RP with Xho I	ATTACTCGAGGGCTCATGTCCGAAAAC

3.7.1.1. PCR amplification of cel5A gene:

PCR amplification of cel5A gene was carried out from the purified metagenomic DNA of Tapovan spring.

Reagents	**Stock concentration**	**Working concentration**	**Amount in 50µL reaction**
Hotstar master mix	10X	1X	5µL
dNTPS	2 mM	0.2 mM	5µL
Forward primer (cel5A FP)	10 µM	0.4 µM	2µL
Reverse Primer (cel5A RP)	10 µM	0.4 µM	2µL
Hot star taq DNA polymerase	5 U/µL	2.5 U	0.5µL
Template DNA	50 ng/µL	100 ng/µL	5µL
Nuclease free water			30.5µL

PCR program used for amplification:

Step 1 (Initial denaturation)	94°C for 15 min
Step 2 (Denaturation)	94°C for 1 min
Step 3 (Annealing)	57°C for 30 sec
Step 4 (Extension)	72°C for 1.5 min.
Step 5	Go to Step 2 for 5 times
Step 6 (Denaturation)	94°C for 1 min
Step 7 (Annealing)	59°C for 30 sec
Step 8 (Extension)	72°C for 1.5 min
Step 9	Go to Step 2 for 25 times
Step 10 (Final extension)	72°C for 5 minutes
Step 11	Hold at 4°C forever

3.7.1.2. pET28(a) plasmid DNA isolation and digestion:

The pET28(a) plasmid carrying cells were inoculated in 40 µg/ml of LB Kanamycin broth for overnight growth. Cells were harvested and plasmid was isolated using GeneJET spin column as described in the previous section.

Double digestion of pET28(a) and PCR amplified cel5A gene:

Reagent used	**Volume for 10 µL**
Plasmid DNA	5 µL
Buffer 2 (10X)	1µL
Bam HI (10U)	0.5 µL
Xho I (10U)	0.5 µL
Nuclease free water	3 µL

Double digestion was carriedout for 3 hours at 37 ºC.

3.7.1.3. Purification and ligation of cel5A gene in pET28(a):

Double digested vector pET28(a) and cel5A gene were electrophoresed in 1% agarose gel at 2V/cm. DNA was purified using QIAquick® gel extraction kit (Qiagen, USA) as described above.

Reagents	**Final Conc. (In 10µL)**	**Amount (In 10 µL)**
10X T4 ligase buffer (Fermentas, USA)	1X	1 µL
T4 DNA ligase (10 U µL-1)	5U	0.5 µL
ATP (10 mM)	0.5 mM	0.5 µL
pET28(a) vector (Bam HI and Xho I digested) (50 ng/µl)	150ng	3 µL
cel5A gene digested with Bam HI and Xho I (90 ng/µl)	450ng	5 µL

3.7.1.4. Transformation in *Escherichia coli* BL21- CodonPlus cells:

The overnight culture of *Escherichia coli* BL21- CodonPlus cell was subcultured in 10 ml aliquots of LB broth and cultured till OD at A_{600} reaches 0.5. 1.5 ml of culture was taken in microfuge tubes and centrifuged at 3000 rpm for 5 minutes at 4ºC. Subsequently, supernatant was discarded and pellet was re-suspended in 1 ml of 0.1 M $CaCl_2$ solution and kept in ice for 30 minutes. The cells were then centrifuged again for 5 minutes at 3000 rpm at 4 ºC. The cell pellet was then dissolved in 50 µl of 0.1 M $CaCl_2$ solution. 5 µl of ligated product was mixed into the prepared competent cells. Cells were then kept on ice for 15 mins and then transferred to 42ºC

water bath for 90 sec immediately the competent cell was placed on ice for 3 minutes. The cell were transferred to 1 ml SOC media (Himedia, India; 0.5% Yeast Extract; 2% Tryptone; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM Glucose). The cells were enriched at 37 °C at 200 rpm for an hour. The cells were pooled out by centrifugation and were spread on LB plates containing 40 μg/ml kanamycin. Plates were then incubated at 37°C for 12 hours. Transformed colonies were selected and screened for insert.

3.7.2. Screening of clones for cellulase activity:

LB CMC plates were prepared by adding 1% of Carboxy Methyl Cellulose (CMC) (Sigma USA) onto 2% LB agar. Selected clones from the library were screened for cellulolytic activity by streaking on LB CMC plates. The plates with the transformed clones were then incubated at 37 °C for 12 hours for utilization of CMC on the plates. The plates were then flooded with 1mg/ml solution of congored dye for 15 mins. Plates were washed off using 1N NaCl for 15 mins. Zone of hydrolysis of CMC was visualized on the plates. The hydrolysis patterns of about 46 randomly selected clones were analyzed in the same manner.

3.7.3. Over expression of TM-cel5A gene:

The clone providing highest zone of hydrolysis was selected. Overexpression of cloned TM-cel5A gene was carried out using varying concentration of IPTG. Overnight culture of the selected clone in LB containing 40 μg/ml kanamycin was used as seed culture. 1% of the seed culture was used for sub culturing in LB containing 40 μg/ml kanamycin. The culture was done 200 rpm at 37 °C till the OD of the culture reaches 0.6 at A_{600}. IPTG at a final concentration of 0.4 mM, 0.5 mM, 0.6 mM were used for induction of TM-cel5A gene. Induction was carried out for a period of 5 hours at 37°C with 200 rpm. At the end of induction cells were centrifuged at 12000 rpm for 10 minutes at 4 °C. The supernatant was analyzed for over expression of TM-Cel5A protein.

3.7.4. Polyacrylamide gel electrophoresis for TM-Cel5A protein:

The over expressed TM-Cel5A protein was analyzed on 12% SDS-PAGE* (Bio-RAD gel assembly). *12% SDS-PAGE preparation:

Resolving Gel (pH 8.8, 10 mL):

30% Acrylamide mix (29:1; Acrylamide: Bisacrylamide)	4 Ml
1.5 M Tris-Cl (Himedia, India; pH 8.8)	2.5 Ml
10% (w/v) SDS (Himedia, India)	100 µL
10% (w/v) APS (Himedia, India)	100 µL
TEMED (Himedia, India)	5 µL
Autoclaved water	3.3 Ml

Stacking Gel (pH 6.8, 5 mL):

30% Acrylamide mix (29:1; Acrylamide: Bisacrylamide)	500 µL
1.5 M Tris-Cl (Himedia, India; pH 6.8)	380 µL
10% (w/v) SDS (Himedia, India)	30 µL
10% (w/v) APS (Himedia, India)	30 µL
TEMED (Himedia, India)	4 µL
Autoclaved water	2.1 mL

3.7.5. Purification of recombinant endoglucanase in *Escherichia coli:*

The selected recombinant bacteria was cultivated in 500 ml of LB containing 40µg/ml of kanamycin on rotary shaker 200rpm at 37 ºC for 12 hours. Production of recombinant TM-cel5A was induced at OD_{600} of 0.6 by addition of 0.5mM isopropyl- β- D thiogalactopyranoside (IPTG). As the TM-cel5A gene carries a signal peptide (from sequence date) in its native form, it was found that the protein gets secreted out into the culture media. To harvest the recombinant protein, the cell free supernatant of LB broth was collected by separating cells at 12000 rpm for 10 minutes at 4 ºC. Ammonium sulphate precipitation of the total proteins was carried out at 80% saturation level. The precipitated proteins were collected at 18000g for 10 minutes centrifugation at 4 ºC. The precipitated protein was dissolved in 50 mM sodium phosphate buffer of pH 8.0. The crude protein preparation was dialyzed in 1 kDa cutoff membranes. The protein fraction apart from TM-Cel5A protein carries a very few other protein contamination. They were selectively removed by filtering the dialyzed fraction using 30 kDa centrifugal concentrator (Amicon, 15 ml size).

3.7.6. Phylogenetic tree construction for TM-Cel5A amino acid:

The TM-Cel5A amino acid sequence was retrieved from the nucleotide sequences by using ORF finder software. The amino acid sequences were subjected to BLAST (Basic Local Alignment

Search Tool). The closest relatives were selected and a phylogenetic tree was constructed . The phylogenetic tree was constructed using the phylip file of multiple sequence alignment of protein sequences. Phylip-3.69 version was used for the distance based phylogenetic tree construction. Bootstrap replicator values of 1000 was taken during the tree construction.

3.7.7. Multiple sequence alignment and phylogenetic relation of TM-Cel5A amino acid with closest relatives:

The complete nucleotide sequence of the cloned endoglucanase based TM-cel5A gene was determined, subsequently the open reading frame was determined. A blast search of gene was performed to obtain the homologous genes. Similarly, the obtained ORF was used for the retrieve of homologous amino acid sequences encoding endoglucanase. Amino acid sequence based multiple sequence alignment of studied endoglucanase was done with closest relatives. Highly closest amino acid sequences from cultured and uncultured counter parts were selected for distance based phylogenetic tree construction, bootstrap value of 1000 considered for resampling. Apart from *Bacillus* sp., *Streptomyces* sp., *fungal* sp., *archeal* sp. were considered for the phylogenetic classification. TM-Cel5A amino acid sequence was predicted for active site by multiple alignment search [78, 168].

3.7.8. Standard curve determination for BSA and glucose:

Protein estimation by (bicinchoninic) BCA using BSA standard graph:

Protein estimation was done using BCA protein assay kit (Thermo scientific USA). A standard curve was plotted using bovine serum albumin (BSA). BSA in the range of 0.5 TO 2 mg/ml were taken for the assay. 25µl of standard (BSA) and TM-Cel5A protein were taken in microtiter plate separately. 200 µl of BCA working reagent (WR) added to both the wells. The content was mixed thoroughly. The plate was incubated at 37 °C for 30 mins. Absorbance was taken at A_{562}nm. BSA standard curve was plotted for different dilutions taking absorbance versus protein concentrations. Concentration of TM-Cel5A protein was determined from the linear equation obtained from the standard curve.

Glucose standard curve determined by DNS method:

Glucose (Himedia, India) of 1 to 5 mg/0.5ml concentration was taken in different test tube. They were treated with 1.5 ml DNS (sigma USA) solution. The reaction mixture were kept in boiling water for 5 mins. The reaction mixture was cooled down under running tape water. Reaction was

terminated by addition of 100 µl sodium potassium tartarate. The color development was measured spectrophotometrically at A_{540} nm.

Dinitrosalicillic acid method of estimation of enzymatic activity of TM-Cel5A protein:

Purified TM-Cel5A protein of 50 µl was treated with equal volume of CMC at a final concentration of 1% dissolved in sodium phosphate buffer. The reaction was kept at 60 ºC for 20 mins. The mixture was then treated with 300 µl of DNS (sigma USA) and the mixture was kept on boiling water for 5 mins. The reaction mixture was cooled down under running tape water. Reaction was terminated by addition of 100 µl sodium potassium tartarate. The color development was measured spectrophotometrically at A_{540} nm.

3.7.9. Activity based biochemical characterization of TM-Cel5A protein:

3.7.9.1. Optimum temperature determination for TM-Cel5A:

The optimal temperature of TM-Cel5A protein was determined by measuring endoglucanase activity against CMC with temperature range of 20 to 90 ºC by 10 degree increments. The reaction was set using 1% (w/v) CMC in 50 mM sodium phosphate buffer pH 8.0 at different temperatures for 30 minutes. 100 µl of substrate was taken with 100 µl of purified enzyme and the tubes were kept at specified temperature from 20 to 80 ºC for 30 minutes. Quantification of reduced sugar (glucose) generated due to activity of endoglucanase was determined by using DNS method as described earlier. Thermal stability of TM-Cel5A was determined by incubating enzyme with equal volume of 1% (w/v) CMC from temperature 60 to 80 ºC with 10 degree increment for 2 hour. Their enzymatic activity was measured using DNS method.

3.7.9.2. Optimum pH determination for TM-Cel5A protein:

Optimal pH activity was determined by measuring endoglucanase activity against 1% (w/v) CMC with pH range of 4 to 9. Sodium acetate buffer of concentration 50 mM was used for pH 4 to 6, Sodium phosphate buffer of 50 mM was used for pH 6 to 8 and Tris buffer of 50 mM was used for pH 8 and 9. Equal volume of enzyme preparation and appropriate buffer with pH was incubated at 4 ºC for 16 hour. pH stability of TM-Cel5A was determined using standard assay procedure (1% CMC in 50 mM sodium phosphate buffer, pH 8.0 at 60 ºC, 30 min).

3.7.9.3. Substrate specificity of endoglucanase TM-Cel5A:

Substrate specificity of enzyme TM-Cel5A was carried out at optimal pH and temperature conditions for 20 minutes with 1% polysaccharide substrates. Various substrate used for the study were Beta-D-glucan (sigma), Laminarin (sigma), Avicel (sigma). Whereas for *p*-Nitrophenyl-β-D-Glucopyranoside (pNPG) (sigma), *p*-Nitrophenyl-β-D-xylopyranoside (pNPX) (sigma), were used in a final concentration of 1 mM. Sodium phosphate buffer (50 mM) pH 8.0 was used for dissolving the substrates.

Filter paper assay:

Filter paper assay was performed with purified protein sample of TM-Cel5A. The assay was performed using 50µl of enzyme preparation, 50 mg of Whatman no.1 filter paper at 60 °C and for 1 hour followed by addition of 300 µL dinitrosalicillic acid (DNS). The mixture was heated for 5 minutes at 100 °C. The reaction mixture was cooled under tape water and 100 µl of Roche salt consisting 40% sodium potassium tartarate was added OD was taken at 540 nm.

3.8. Biolog EcoPlate based enrichment and isolation of L-asparaginase producing strain from Tapovan spring:

100 µl of the untreated water from both the springs were taken directly into the Biolog EcoPlate and was incubated at 37°C for 96 hours. Their color development was determined by taking OD at A_{590} nm in MicroStation™ System/MicroLog plate reader USA.

The average metabolic response (AMR) and community level physiological profiling (CLPP) were calculated as follows:

AMR = ∑ (O D of well – O D of negative)/31.

CLPP = ∑ (O D of wells with positive responses – OD of negative)/31.

Preparation of Modified Czapek Dox medium:

Czapek Dox medium is a specially designed minimal media for auxotrophic growth.

Na_2HPO_4	6 g
KH_2PO_4	2 g
NaCl	0.5g
L-asparagine	20 g
Glycerol	2 g
$MgSO_4.7H_2O$	1 Mg

$CaCl_2$. $2H_2O$ 0.005 g
pH was adjusted to 5.5
Final volume to 1 lit with DDW.

Modified Czapek Dox agar medium was prepared by adding phenol red indicator of 0.009% and 2% agar. The broth and plate were inoculated with the positive organism. Broth was incubated at 37 °C, 200 rpm for 24 hours. Plates were kept at static condition to the above. Uninoculated plates were set as control.

3.9. Cloning of L- asparaginase gene from Tapovan isolate:

3.9.1. Design of primer forL- asparaginase gene:

Initially Tapovan isolate was identified through 16S rRNA sequencing. The bacterium was identified as *Ralstonia* sp. Aparaginase-II gene from *Ralstonia* sp. was used for designing primer for L-asparaginase gene .

Primer name	**Primer sequence 5' to 3'**
ASP FP with Bam H I	ATTAGGATCCATGTCCTTGCCTACCATC
ASP RP with Xho I	AGGCCTCGAGTTATATCGATCAATGACTTG

3.9.2. PCR amplification of L-asparaginase gene:

Reagents	**Stock concentration**	**Working concentration**	**Amount in 50µL reaction**
Hotstar master mix	10X	1X	5µL
Dntps	2 Mm	0.2 mM	5µL
Forward primer (cel5A FP)	10 µM	0.4 µM	2µL
Reverse Primer (Cel5A RP)	10 µM	0.4 µM	2µL
Hot star taq DNA polymerase	5 U/µL	2.5 U	0.5µL
Template DNA	50 ng/µL	100 ng/µL	5µL
Nuclease free water			30.5µL

PCR program used for L-asparaginase gene amplification:

Step 1 (Initial denaturation)	94°C for 15 min
Step 2 (Denaturation)	94°C for 1 min
Step 3 (Annealing)	53°C for 30 sec
Step 4 (Extension)	72°C for 1.5 min.

Step 5	Go to Step 2 for 5 times
Step 6 (Denaturation)	94°C for 1 min
Step 7 (Annealing)	55°C for 30 sec
Step 8 (Extension)	72°C for 1.5 min
Step 9	Go to Step 2 for 25 times
Step 10 (Final extension)	72°C for 10 minutes
Step 11	Hold at 4°C forever

3.9.3. Double digestion of pET28(a), PCR amplification of L-asparaginase gene and ligation:

Reagent used	**Volume for 100 µL**
Plasmid DNA	50 µL
Buffer 2 (10X)	10µL
Bam HI (10U)	2 µL
Xho I (10U)	2 µL
Nuclease free water	36 µL

Double digestion was carriedout for 3 hours at 37 ºC.

The vector was treated with 1U of Alkaline phosphatase for 1 hour at 37 ºC.

Both digested vector and insert were purified using QIAquick® column (Qiagen, USA) gel extraction kit. The purified DNA was used for ligation as presented below.

Reagents	**Final Conc. (In 10µL)**	**Amount (In 10 µL)**
10X T4 ligase buffer (Fermentas, USA)	1X	1 µL
T4 DNA ligase (10 U µL-1)	5U	0.5 µL
ATP (10 mM)	0.5 mM	0.5 µL
pET28(a) vector (Bam HI and Xho I digested) (50 ng/µl)	150ng	3 µL
Cel5A gene digested with Bam HI and Xho I (90 ng/µl)	450ng	5 µL

Ligation was performed at 22 ºC for 2 hours. The ligated mixture was heat inactivated at 70 ºC for 15 mins was and used for transformation.

3.9.4. Transformation in *Escherichia coli* BL21- CodonPlus cells:

The overnight culture of *Escherichia coli* BL21- CodonPlus cell was subcultured in 10 ml aliquots of LB broth and cultured till OD at A_{600} reaches 0.5. 1.5 ml of culture was taken in microfuge tubes and centrifuged at 2500 rpm for 5 minutes at 4 °C. Subsequently supernatant was discarded and the cells were suspended in 1 ml of 0.1 M $CaCl_2$ solution and kept in ice for 30 minutes. The cells were then centrifuged again for 5 minutes at 3000 rpm at 4°C. The cell pellet was then dissolved in 50 µl of 0.1 M $CaCl_2$ solution. 5 µl of ligated product was mixed into the prepared competent cells. Cells were then kept on ice for 15 mins and then transferred to 42 °C water bath for 90 sec, immediately the competent cells were placed on ice for 3 minutes. The cells were then transferred to 1 ml SOC media (Himedia, India; 0.5% Yeast Extract; 2% Tryptone; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM Glucose). The cells were enriched at 37 °C at 200 rpm for an hour. The cells were pooled out by centrifugation and spread on LB plates containing 40 µg/ml kanamycin. Plates were then incubated at 37 °C for 12 hours. Transformed colony was selected and screened for insert.

Analysis of L-asparaginase insert in the clones:

Selected clones were incubated in LB broth containing 40 µg/ml kanamycin. Overnight cultures were used for plasmid DNA isolation and plasmid were digested with Bam HI and Xho I

Sequence based analysis of L-asparaginase gene:

A phylogenetic tree was constructed from the closest relatives of the cloned L-asparaginase gene using phylip 3.69. Bootstrap values of 1000 was taken during the phylogentic tree construction.

3.10. Sequence based gene targeting and detection in metagenomic DNA:

3.10.1. Real-time PCR based C_T values determination of efflux pumps and antibiotic resistance marker in hot spring eDNA:

Real-time PCR were performed for threshold value (C_T) with the purified metagenomic DNA of Tapovan spring for the detection of genes responsible for efflux system and antibiotic resistance marker. Primers used were listed in the table 3.2.

Table 3.2: List of primers of transporter and drug resistance markers used for real-time PCR in the study.

S. No.	Class of transporter antibiotic	Primers name	Primer sequence 5' -> 3'	Amplicon size (bp)
1	CusCFBA Transporter	cusA F	ATGCCACCGATTACCACCGTTGCGA	200
		cusA R	ATTGGTTCATTCGTGCCCG	
2	Efflux pump	AcrB F	GCGGCGGCCGCAACGCGCGGAACGGCTAGG	180
		AcrB R	GCGAAGCTTTAAAAGAGGACCTCGTGTTTC	
3	Efflux pump	acrD F	GCTCTCGCAGAACTTCCTCGC	175
		acrD R	AGMCKGGGGTCGATCTCGTAG	
4	Glycopeptide antibiotics	vanX F	CCAAGTACGCCACSTGGGACAAC	220
		vanX R	CTTCGTCCGGCCGTCCTCC	
5	Polyketide antibiotics	tetM-1	GTTAAATAGTGTTCTTGGAG	950
		tetM-up	CTGGCAAACAGGTTC	
6	Beta-lactam antibiotic	bla F	CCTCGCCCGGCGCGTCCASTA	155
		bla R	AGMAGGTTGGCCGCSGTGTTGTC	
7	Aminoglycosides	aacA F	CTGGAACGACGCYCCGCCSTAC	300
		aacA R	GGSGCGCCSAGSAGCAGYA	
8	Macrolides	erm F	ATATCCTACGATTGCACCGC	160
		erm R	GGTACCCGTGGAGTCACTGT	
9	Plants (cpDNA)	RbcL h1a F	GGCAGCATTCCGAGTAACTCCTC	138
		rbcL h2a R	CGTCCTTTGTAACGATCAAG	
10	Human (mtDNA)	MtDNA FP	CCCCATGCTTACAAGCAAGT	131
		MtDNA RP	TGGCTTTATGTACTATGTAC	
11	*E. coli* plasmid pET23a-GFP	Gfp FP	CAACATTGAAGATGGAAGCG	193
		Gfp RP	CATGCCATGTGTAATCCCAG	

3.10.2. Real-time PCR amplification of transporter and antibiotic resistance gene:

2X buffer	12.5 µl
Forward primer	0.75 µl
Reverse primer	0.75 µl
Tapovan Metagenomic DNA	1 µl
NFW	10 µl

Real-time PCR program:

Stage-1: 95 ºC -120 Sec

Stage 2: 95ºC -120 Sec, *Annealing - 30 sec, 72 ºC - 30 sec, For 30 cycle

Stage – 3: 72 ºC - 800 sec, 20 ºC - 800 sec

*Annealing temperature for primers set beta-lactam, Macrolides, Glyco-peptide antibiotics were kept at 55 ºC, for Multi drug efflux pump *acrB* and *mexB* were kept at 48 ºC, Tetracycline was at 40 ºC, Plants (cpDNA), Human (mtDNA), E. coli plasmid pET23a-GFP, cusA , acrD.

Cloning of *acrB* gene:

3.10.3. PCR amplification of *acrB* gene:

Reagents	Stock concentration	Working concentration	Amount in 50µL reaction
Hotstar master mix	10X	1X	5µL
dNTPS	2 mM	0.2 mM	5µL
Forward primer (acrB C FP)	10 µM	0.4 µM	2µL
Reverse Primer (acrB RP)	10 µM	0.4 µM	2µL
Hot star taq DNA polymerase	5 U/µL	2.5 U	0.5µL
Template DNA	50 ng/µL	100 ng/µL	5µL
Nuclease free water			30.5µL

PCR program used for acrB gene amplification:

Step 1 (Initial denaturation)	94°C for 15 min
Step 2 (Denaturation)	94°C for 1 min
Step 3 (Annealing)	46°C for 30 sec
Step 4 (Extension)	72°C for 1.5 min.
Step 5	Go to Step 2 for 5 times
Step 6 (Denaturation)	94°C for 1 min
Step 7 (Annealing)	48°C for 30 sec
Step 8 (Extension)	72°C for 1.5 min
Step 9	Go to Step 2 for 25 times
Step 10 (Final extension)	72°C for 10 minutes
Step 11	Hold at 4°C forever

3.10.4. Ligation and transformation of *acrB* gene:

PCR product was gel purified and ligated with TA cloning vector

Reagents	**Final Conc. (In 10µL)**	**Amount (In 10 µL)**
10X T4 ligase buffer (Fermentas, USA)	1X	1 µL
T4 DNA ligase (10 U µL-1)	5U	0.5 µL
ATP (10 mM)	0.5 mM	0.5 µL
pET28(a) vector (Bam HI and XhoI digested) (50 ng/µl)	150ng	3 µL
Cel5A gene digested with Bam HI and XhoI (90 ng/µl)	450ng	5 µL

Ligation was performed at 22 ºC for 2 hours. The ligated mixture was heat inactivated at 70 °C for 15 minutes and used for chemical transformation into *Escherichia coli* DH5α. The positive clones were selected by blue-white screening using LB agar plate containing ampicillin (100µg/ml), 40µg/ml of XGal and 0.1 mM IPTG. Insert from recombinant plasmid was sequenced using ABi 3730Xl (Sanger sequencing technologies).

3.10.5. Sequence identification and phylogeny of *acrB* gene:

The partial sequence of cloned *acrB* gene was identified through BLAST analysis. The closest relatives were selected for a phylogenetic tree construction. Bootstrap replicator values of 1000 were taken during the tree construction.

CHAPTER 4

RESULTS AND DISCUSSION-I

Extraction and purification of high molecular weight (HMW) environmental DNA (eDNA) from Himalayan geothermal springs

4.1. Sampling sites and microscopic views:

Direct photographs were taken at the collection site of Tapovan geothermal spring. As described earlier, the site at Tapovan was siliceous and inorganic type. Deposit of calcium, calcium carbonate, silica and Bacillariophyceae representatives were found at the site (figure 4.1). The algal mat of surrounding hot water flowing area are shown in figure 1(C) and their microscopic view taken at 100x magnification are shown in D and E section (figure 4.1).

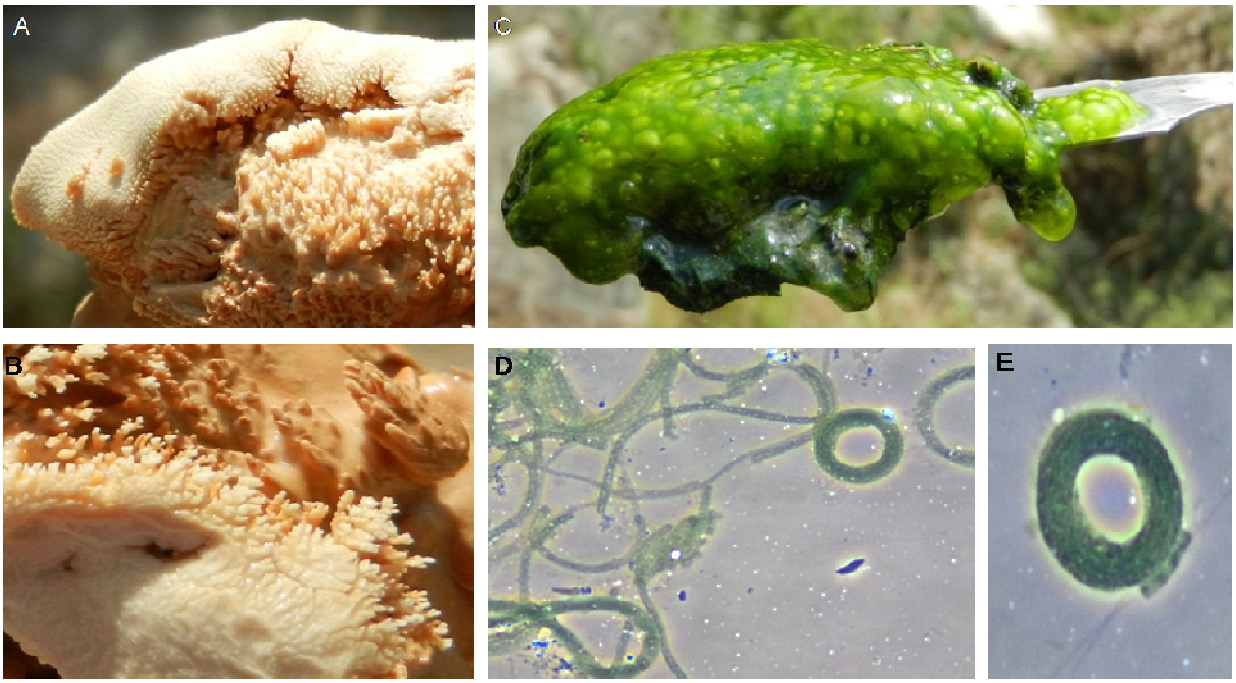


Figure 4.1: Sampling site of Tapovan. *A and B Deposition of calcium, calcium carbonate, silica and Bacillariophyceae found in Tapovan hot spring site. C: Growth of algae (Chlorophyceae). D and E: microscopic view at 100x magnification of algal mat found in Tapovan hot spring site.*

4.2. Water chemistry analysis of hot springs:

The sample collection latitude and height are shown for Tattapani and Tapovan springs (Table 4.1). Temperature of water was at about 65ºC and 95ºC and pH measured about 8.34 and 8.06 for Tattapani and Tapovan springs respectively. Biochemical oxygen demand (BOD) of Tattapani spring was calculated to be about 4.76 and for Tapovan spring was 4.1. The ionic concentrations, which were determined by ionic chromatography were indicated in the table 4.1. The heavy metal concentration of both the springs were determined by Inductively Coupled Plasma Mass Spectroscopy (ICPMS) given in table 4.1.

The heavy metal concentrations determined in terms of parts per billions (ppb) were used to plot the bar digram (figure 4.2), which indicates that aluminum (shown in table 1), Arsenic, Boron, Selenium concentration were beyond the standard biologically allowed concentrations in water bodies [112]. Apart from that Copper, Iron, Nickel, Manganese concentration were found in significant amount in both the hot springs.

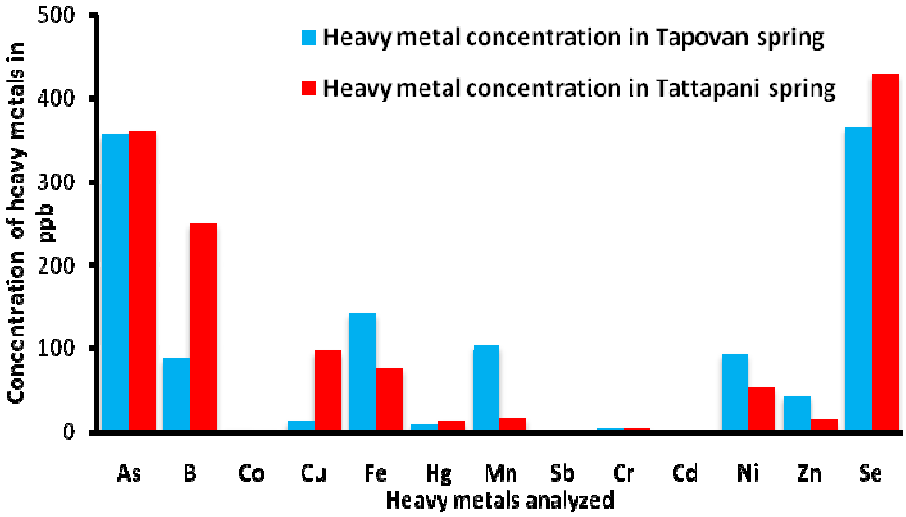


Figure 4.2: Concentrations of Heavy metals in hot spring water. *Heavy metals were measured in parts per billions (ppb) of Tapovan and Tattapani spring water sample determined by ICPMS (ELAN DRC-e, PerkinElmer SCIEX).*

Table 4.1. Water chemistry analysis of hot springs. *Tattapani and Tapovan thermal springs sample collection sites with physical parameters. Water chemistry analysis for ions and heavy metals. Values are presented as milligrams per liter unless indicated otherwise[(a)] °N: Degree North, °E: Degree South, ºC: Degree Centigrade, m: Meter.*

Analyte	Tattapani spring	Tapovan spring
[a]Stations latitude	31.2489°N	30°N
	77.0861°E	79°E
[a]Height (m) in Himalayan range	655	1828
[a]Temperature (ºC)	65	95
[a]Mat sample texture	Soil with low organic content	Siliceous soil
[a]Ph	8.34	8.06
Alkalinity	200	216
BOD	4.76	4.1
Al	0.1718	0.6513
As	0.3613	0.3561
B	0.2505	0.0887
Co	0.0025	0.0037
Cu	0.0969	0.0130
Fe	0.0771	0.1427
Hg	0.0140	0.0093
Mn	0.0169	0.1028
Sb	0.0013	0.0004
Cr	0.0049	0.0041
Cd	0.0006	0.0011
Ni	0.0527	0.0934
Zn	0.0144	0.0428
Se	0.4294	0.3656
Li	2.08	4.93
Na^{+}	NA	436.31
NH_4^{+}	543.27	953.83
K^{+}	3.64	5.5
Ca^{+2}	31.8	17.92
Mg^{+2}	159.16	340.92
SO_4^{-2}	42	285
Cl^{-}	69.5	220
F^{-}	3.25	3.8
PO_4^{-2}	0.08	0.08

4.3. Comparison of physically purified eDNA with commercial kit:

The electrophoresis of unpurified crude eDNA, electro-eluted purified eDNA and commercial kit eluted eDNA inferred that the manually purified eDNA had high molecular weight and appreciably good concentration. The crude and kit eluted eDNA were comparatively low in concentrations, carrying humic substances (figure 4.3).

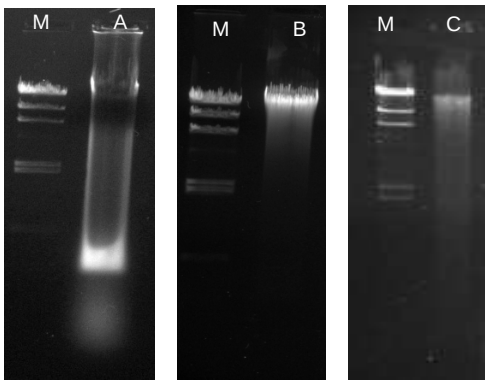


Figure 4.3. Comparison of purified metagenomic DNA. *M: Lambda Hind III Molecular weight marker (A) Crude DNA isolated manually and before purification. (B) Purified DNA using electro-elution method (C) Isolated DNA using HiPura soil DNA isolation kit (Himedia).*

The optical density spectrum from 220nm to 350nm of studied eDNA inferred that there is a significant increase in OD at A_{260} nm for manually purified eDNA. There was a considerable decrease in peak at A_{230} nm which is the measurement for humic substances in case of manually

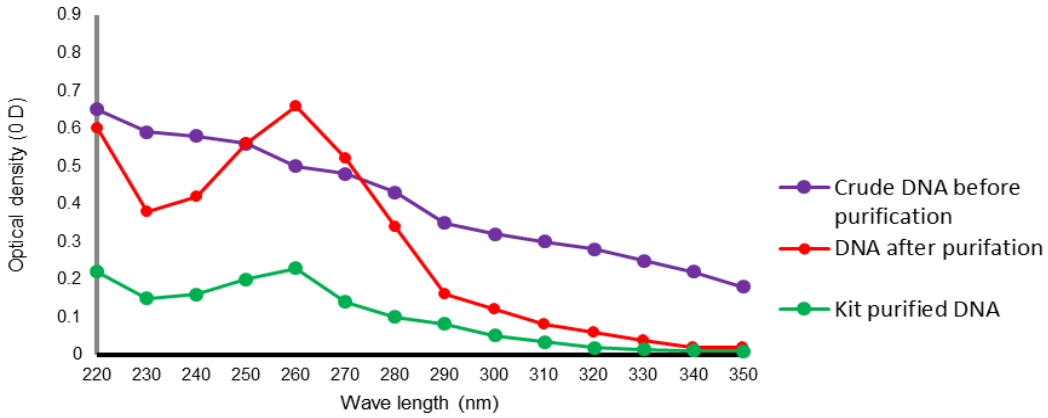


Figure 4.4: Spectrum is shown in the range of A_{220} to A_{350} for metagenomic DNA. *Optical density spectrum of crude, kit purified DNA and electroeluted DNA determined spectrophotometrically.*

purified DNA. There is significant decrease in peak at A_{280} nm indicating its lower protein contaminations (figure 4.4).

The 260/280 ratio for nucleic acid to protein contamination for all the three eDNA preparations indicates that the electro-eluted eDNA has the desired range of ratio. Similarly the 260/230 ratio for humic substances contaminations showed that manually electro-eluted eDNA have lower contamination of humic substances compared to other group of isolated eDNA (Table 4.2).

Table 4.2: Analysis of Crude, kit eluted and electroeluted metagenomic DNA through spectrophotometric method.

Properties	Crude DNA before purification	DNA after purification	Kit eluted DNA
OD at 230 nm	0.59	0.38	0.15
OD at 260 nm	0.56	0.66	0.23
OD at 280 nm	0.43	0.34	0.11
OD 260/280	1.3	1.94	2.09
OD 260/230	0.94	1.73	1.53
Concentration of eDNA	2.5µg/ml	3.3µg/ml	1.15µg/ml

4.4. Quality analysis of extracted eDNA by PCR and restriction digestion:

The complete digestion of metagenomic DNA with Sau3AI inferred that the purified eDNA was free from enzyme inhibitors (figure 4.5).

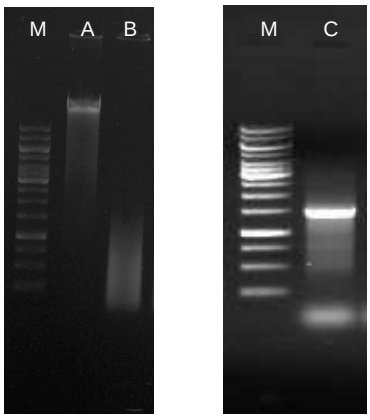


Figure 4.5: Analysis of quality of Tattapani metagenomic DNA. **A:** *Tattapani metagenomic DNA undigested fraction of eDNA* ***B:*** *Restriction digested with Sau3AI enzyme of Tattapani eDNA* ***C:*** *PCR of 16S rRNA from Tattapani hot spring metagenomic DNA.*

The PCR amplification of metagenomic DNA with 16S rRNA gene further confirmed DNA to be free from contaminations. This indicates the suitability of metagenomic DNA for further molecular biological applications (figure 4.5).

4.4. Discussion:

In the present study, we employed a manual method of environmental DNA (eDNA) extraction and purification from Tattapani geothermal spring mat. The basic steps explained here are efficient to extract and purify DNA from diverse soil types [115, 183]. Cell lysis protocol made use of strong detergent like SDS in the presence of high temperature to lyse bacteria in soil. 1% (w/v) Cetyl Trimethyl Ammonium Bromide (CTAB) was used during the DNA extraction procedure which resulted in reduction of humic acid and fulvic acid contaminations [82, 89]. The cell disruption strategies adapted in the method are likely to lyse greater population of bacteria present in the metagenomic sample [66] and therefore, provide a chance to access DNA from more diverse representatives of bacteria in environment. Treatment of lysozyme facilitates lysis of thicker peptidoglycan layer containing (gram positive) group of microorganisms [183]. In contrast, the eDNA purification principle involve in commercial kits is based on silica gel membrane which have limited binding capacity for nucleic acids and thus yield lower concentration of environmental DNA.

The eDNA isolated using present method migrated above or along with 23 kb DNA marker on agarose gel. Environmental DNA isolation results in co-extraction of humic, fulvic acids and phenolic compounds [183]. Removal of these humic substances is difficult from crude DNA preparations. DNA purification is a critical step for their subsequent use in molecular biology applications. Therefore, isolated high molecular weight (HMW) eDNA was subsequently gel-purified using electro-elution method. The electro-elution of DNA generated a greater degree of pure genomic DNA which was free from humic substances. The eDNA extraction and purification method described here is simple, efficient and yields higher amount of DNA. Described extraction method requires approximately 10 hours. PCR reactions were performed to assess the purity of purified eDNA as it is known that DNA polymerase are strongly inhibited by presence of humic substances [176]. In this case successful PCR amplification confirmed the purity of eDNA. Similarly, tetra-cutter based restriction enzyme digestion supported the interpretation [167]. In addition, the optical density (OD) analysis of eDNA showed a favourable range of ratio for nucleic acid to protein and nucleic acid to humic substances.

Tattapani spring belongs to group of Himalayan thermal springs with inorganic and alkaline environments. The spring possess high load of ions like SO_4^{2-}, NH_4^+, Mg^{2+} and heavy metals like As, B, Cu, Se. These heavy metals are present beyond the normal concentration in water bodies [112]. The thermal springs with rare and low microbial load provided us a challenge in isolation and purification of metagenomic DNA.

CHAPTER 5

RESULTS AND DISCUSSION-II

Microbial diversity analysis of Himalayan geothermal springs (Tattapani and Tapovan) through metagenomic approach

5.1. Cloning of 16S rRNA gene from Himalayan hot springs, cloning efficiency:

PCR product of 16S rRNA genes from both the hot springs were cloned in pTZ57R/T vector. Approximately 1000 clones from each of the library were obtained based on blue white selection on agar plate. Out of them randomly selected clones were taken for restriction pattern analysis.

5.1.1. Restriction pattern analysis of 16S rRNA derived clone libraries:

The restriction enzyme digestion pattern of the plasmids harboring the 16S rRNA gene of Tattapani spring eDNA is shown for nine clones in figure 5.1. The RFLP analysis of these randomly chosen clones showed that all the clones except 'A' and 'I' are unique in their restriction signature (figure 5.1). The vector backbone was found aligned in the same plane. Similarly, the RFLP analysis of Tapovan 16S rRNA gene metagenomic library is shown for ten clones named 'A' to 'J' in figure 5.2. The RFLP analysis of Tapovan clone library showed that 'A, B, D, H, I' are unique in their restriction signature (figure 5.2).

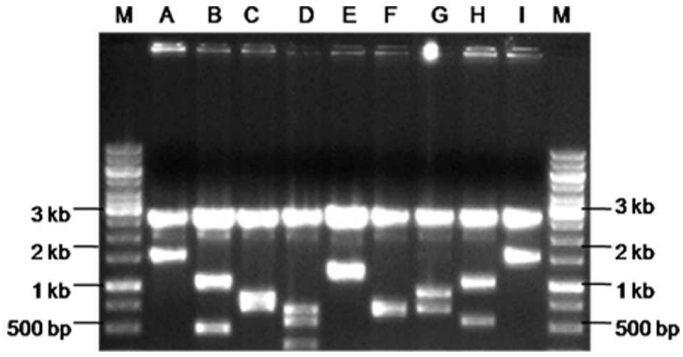


Figure 5.1: RFLP pattern of Tattapani derived cloned plasmids. *(A to I) were restriction fragmented with Hind III and EcoR I. Except Lane A and I all others show different Restriction Fragment Length Polymorphism (RFLP) pattern.*

Based on initial RFLP profile of the plasmids isolated from two libraries, it was clear that a greater microbial diversity inherent with the geothermal springs of Tattapani. The greater microbial diversity of Tattapani thermal spring is due to the high organic rich environment and nutrient status of the site.

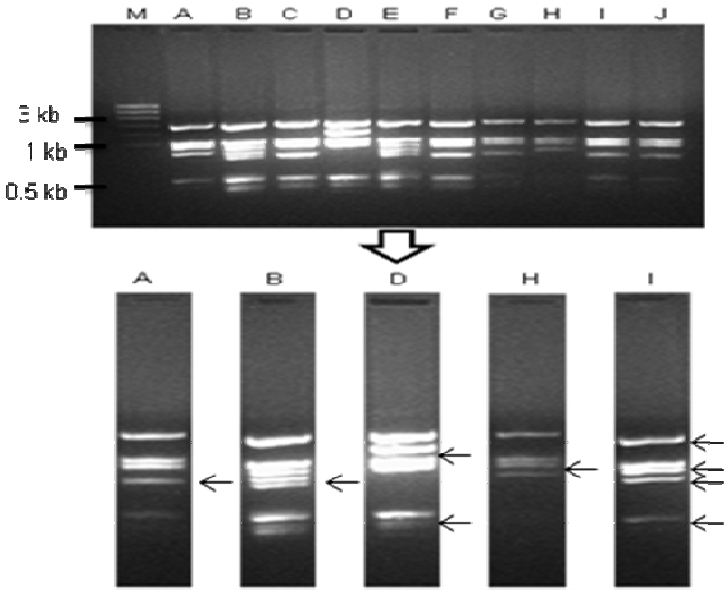


Figure 5.2: RFLP pattern of Tapovan derived cloned plasmids (A to J) were restriction fragmented with Alu I. *Lanes A, B, D H, I shows different Restriction Fragment Length Polymorphism (RFLP) patterns indicated by arrows.*

5.2. Sequence based analysis of clone libraries:

5.2.1. Sequencing of clone libraries and affiliation of clones:

Sequencing of 95 clones were done using ABi 3730Xl (Sanger sequencing technologies). The 16S rRNA genes were partially sequenced. 88 sequences combined from both the libraries were found non chimeric whereas the rest 7 chimeric sequences were eliminated. All 88 sequences were deposited in NCBI and accession numbers were obtained as JN613324, JN896893 to JN896938, JN934657 to JN934666, JN967771 to JN967773, and KC608724 to KC608751.

The BLAST search resulted in identification of probable bacterial genera. Sequences with 97% cut-off values were taken as a demarcation for species level classification. A total of 8 phyla were identified from the Tattapani hot spring mat sample 16S rRNA library.

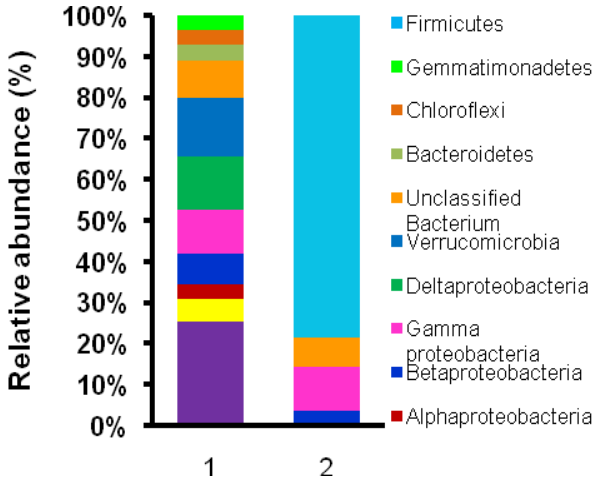


Figure 5.3: Relative abundance of phyla in clone libraries. *Bar 1 represents Relative abundance (%) of microbial phyla in Tattapani mat derived 16S rRNA gene library. Bar 2 represents Relative abundance (%) of microbial phyla in Tapovan mat derived 16S rRNA gene library.*

Out of the total sequenced clones from Tattapani spring, 35% of the clones belonged to the *Proteobacteria* which consist of Alpha-proteobacteria (4%), Beta-proteobacteria (7%), Gamma-proteobacteria (11%), and Delta-proteobacteria (13%) respectively. Other phyla identified are closest to *Acidobacteria* (25%), *Planctomycetes* (5%), *Verrucomicrobia* (14%), Bacteroidetes, *Chloroflexi*, *Gemmatimonadetes* (4%) each, whereas about 9% of total sequences could not be attributed to known group and thus represent unaffiliated clones. The relative abundance of different phyla in respective clone libraries are presented in figure 5.3. The library for Tapovan was dominated by clones affiliated to *firmicutes* (79%) whereas rest of the clones were related to *proteobacteria* (14%) and 7% represented unclassified bacteria. Thus overall Tapovan derived clones were classified into 3 phyla.

The sequences from both the 16S rRNA gene libraries were compared with their closest relatives in BLAST search. About 20 clones from Tattapani derived library and 17 clones from Tapovan library showed less than 97% identities with the known bacterium in the database. They have been presented in the ascending order of their identities (Table 5.1 and 5.2).

Table 5.1 : List of new species found in Tattapani mat derived clones.

Clone name	Phylum	Closest relative (GenBank accession no.)	Percent Identities	Originally Reported habitat	Accession no of clones
TPD 53	Proteobacteria	*Lysobacter* sp. (DQ249997)	90	Soil	JN934662
TPD 13	Proteobacteria	*Desulfocella* sp.(FM176303.1)	91	Stream water	JN896903
TPD 20	Proteobacteria	Alpha proteobacterium (EF417666)	91	Soil	JN934657
TPD 39	Proteobacteria	*Aeromonas* sp. (HM159970)	91	Soil	JN934659
TPD 69	Proteobacteria	Alpha proteobacterium (AY921779)	91	Soil	JN934664
TPD 76	Uncultured bacterium	Uncultured bacterium (EU135158)	91	Soil	JN934666
TPD 40	Chloroflexi	Chloroflexi bacterium (AY360611)	92	Rice field soil	JN896922
TPD 6	Betaproteobacteria	Uncultured beta proteobacterium (AY921981)	93	Soil	JN896896
TPD 37	Verrucomicrobia	Verrucomicrobia bacterium (AY694610)	93	Soil	JN896920
TPD 72	Gemmatimonadetes	Gemmatimonadetes bacterium (EF662585)	93	Soil	JN896938
TPD 19	Planctomycetes	Planctomycete (HQ329068)	94	Soil	JN896906
TPD 21	Proteobacteria	Rhodocyclaceae bacterium (EU359959)	94	Soil	JN896907
TPD 71	Uncultured bacterium	Uncultured bacterium (GQ500696)	94	Biofilm	JN896937
TPD 22	Proteobacteria	*Geobacter* sp. (HM217298)	95	Aquifer	JN896908
TPD 32	Verrucomicrobia	Verrucomicrobia bacterium (FN594679)	95	Rock Biofilms	JN896917
TPD 41	Acidobacteria	Acidobacteria bacterium (JF319291)	95	Soil	JN896923
TPD 42	Acidobacteria	Acidobacteria bacterium clone (JF319291)	95	Cave soil	JN934660
TPD 52	Proteobacteria	*Lysobacter* sp. (FJ711224)	95	Soil	JN934661
TPD 33	Proteobacteria	*Geothermobacter* sp. (FM176316)	96	Mountain soil	JN934662
TPD 35	Unclassified bacterium	Uncultured bacterium (AB426197)	96	Soil	JN896919

Table 5.2 : List of new species found in Tapovan mat derived clones.

Clone Name	Phylum	Closest relative (GenBank accession no.)	Percent Identities	Originally reported habitat	Accession no of clones
TPN 18	Uncultured bacterium	Uncultured bacterium clone ECC289h01 (FJ557175.1)	87	Fecal microbiota	KC608737
TPN 6	Firmicutes	*Anoxybacillus flavithermus* strain D (AY672763.1)	91	Milk powder	KC608729
TPN 23	Uncultured bacterium	Uncultured bacterium clone nbt27b05 (EU535551.1)	93	Human skin microbiota	KC608739
TPN 14	Firmicutes	*Clostridium isatidis* strain WV6 (NR_026347.1)	94	Woad vat	KC608735
TPN 5	Firmicutes	*Bacillus* sp. SP83 (JQ808139)	95	Volcano, Greece	KC608728
TPN 10	Firmicutes	*Aneurinibacillus* sp. Z3 (FJ268961.1)	95	Phyllosphere	KC608732
TPN 24	Proteobacteria	*Pseudoxanthomonas* sp. M1-3 (AB039330.1)	95	Cellulolytic mixed culture	KC608741
TPN 25	Firmicutes	*Aneurinibacillus thermoaerophilus* (EF032876.1)	95	Contaminated soil	KC608742
TPN 31	Firmicutes	*Aneurinibacillus thermoaerophilus* (EF032876.1)	95	Oil spilled soil	KC608748
TPN 4	Firmicutes	*Aneurinibacillus* sp. Z3 (FJ268961.1)	96	Isolated from phyllosphere	KC608727

TPN 12	Firmicutes	*Geobacillus* sp. JAM-FM0901 (AB362274.1)	96	methane-rich sediments	KC608733
TPN 15	Firmicutes	*Anoxybacillus flavithermus* strain D (AY672763.1)	96	Milk powder	KC608736
TPN 28	Firmicutes	*Aneurinibacillus thermoaerophilus* (EF032876.1)	96	Oil spilled soil	KC608745
TPN 30	Firmicutes	*Aneurinibacillus thermoaerophilus* (EF032876.1)	96	Oil spilled soil	KC608747
TPN 31	Firmicutes	*Aneurinibacillus thermoaerophilus* (EF032876.1)	95	Oil spilled soil	KC608748
TPN 32	Proteobacteria	*Pseudoxanthomonas* sp. M1-3 (AB039330.1)	96	Cellulolytic mixed culture	KC608749
TPN 34	Firmicutes	*Aneurinibacillus thermoaerophilus* (EF032876.1)	96	Oil spilled soil	KC608751

5.3. Diversity analysis through OTU and rarefaction calculations:

Rarefaction curve based on obtained OTUs for both Tapovan and Tattapani spring derived libraries were platted against number of clones analyzed. The red color curve represents the obtained OTUs from Tattapani clone library and green curve represents the obtained OTUs from Tapovan library (figure 5.4).

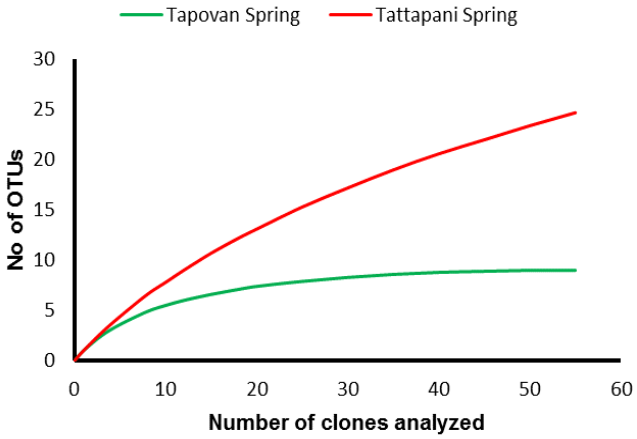


Figure 5.4: Rarefaction curves of Tattapani and Tapovan clone libraries. *Rarefaction curves showing the sampling effort of 16S rRNA libraries in respective springs. The graph has been plotted with 5% CI values.*

Table 5.3: Diversity parameters in mat clone libraries.

Clone library constructed	No of clones analyzed	No of OTUs	No of distinct phyla identified	Shannon-weiner index (97% similarity cutoff)	Maximum possible value for species evenness
Tattapani soil DNA	63	24	8	3.9230	0.9834
Tapovan soil DNA	28	9	3	1.6705	0.8033

OTUs: *Operational Taxonomic Units.*

OTU was calculated to about 24 for Tattapani and 8 for Tapovan derived clones at 97% similarity cutoff values. Shannon-weiner index, were calculated to about 3.9230 and 1.6705 respectively for Tattapani and Tapovan spring respectively. The maximum value for species evenness was found to approach towards one (0.9834) for Tattapani whereas for the Tapovan site species evenness value was 0.8033 (Table 5.3). This indicates the overall microbial diversity of Tattapani spring is quite high as compared to Tapovan spring.

5.4. Phylogenetic tree analysis of Tattapani and Tapovan clone libraries:

All phylogenetic trees were constructed with the sequenced cloned from respective libraries with their closest relatives. Tattapani library derived clones have been grouped in a phylogenetic tree shown in figure 5.5. Most of the clones in Tattapani library showed identities with Proteobacterium, Bacteroidetes, Plactomyces, Verrucomicrobia and Chloroflexi. Similarly Tapovan library derived clones have been grouped in a phylogenetic tree shown in figure 5.6. Most of the clones in Tapovan library showed identities with Firmicutes and Proteobacteria.

Figure 5.5: Neighbor-Joining distance based tree constructed from partial sequences of 16S rRNA gene clones from Tattapani hot spring sample. *The figure shows Phylogenetic relationships of Proteobacteria, Bacteroidetes, Plactomyces, Verrucomicrobia, Chloroflexi. Bootstrap values of 1000 resamplings were given at the branch point. Clusters of bacterial phylum are grouped separately.*

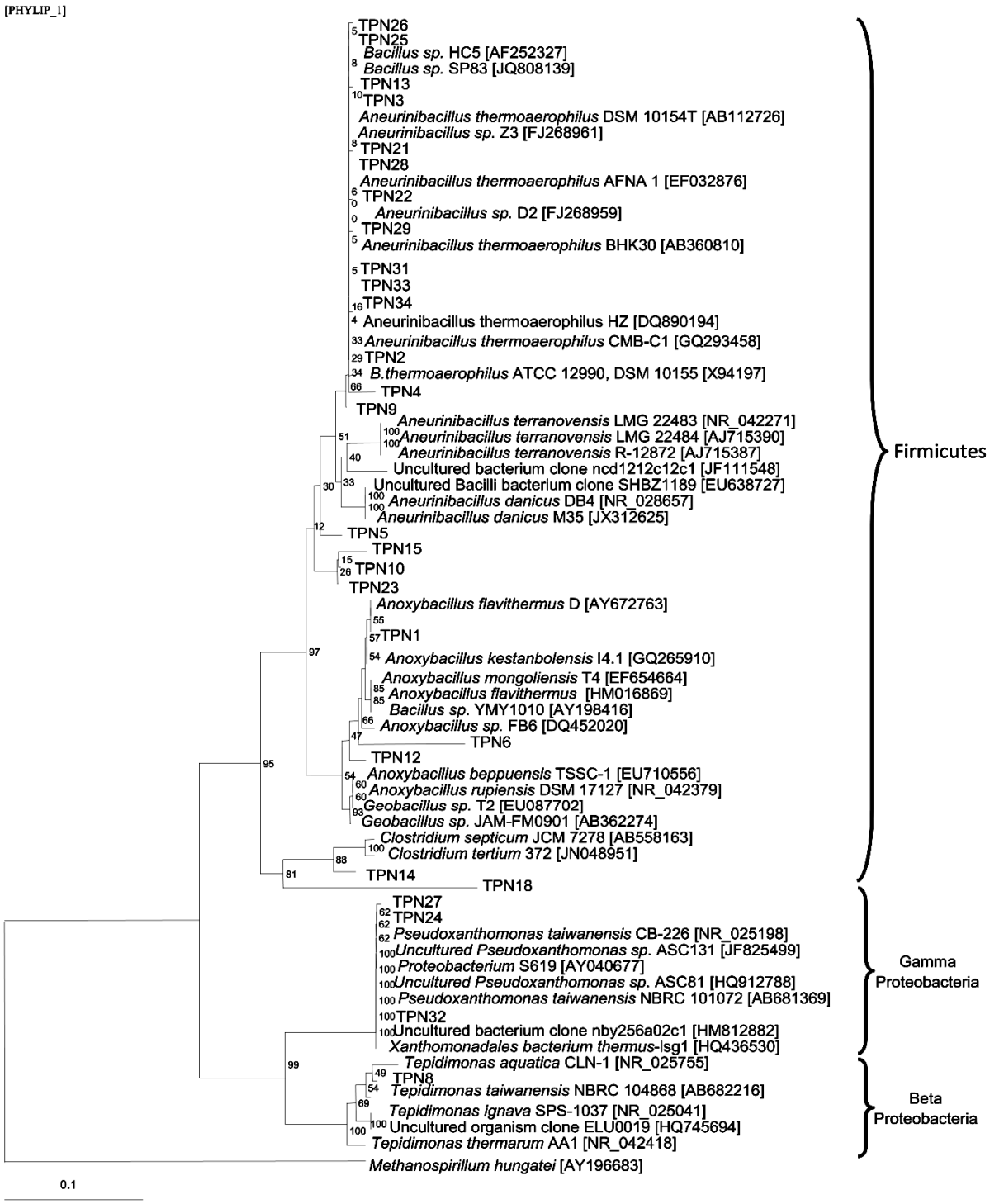


Figure 5.6: Neighbor-Joining distance based tree constructed from partial sequences of 16S rRNA gene clones from Tapovan hot spring sample. *The figure shows Phylogenetic relationships of Firmicutes, Proteobacteria. Bootstrap values of 1000 resamplings were given at the branch point.*

Many of the clones in both the 16S rRNA gene libraries showed identities with uncultured group of bacterium. Clones from both the libraries also showed close relatedness with known thermophilic and thermotolerant strains.

To compare the relatedness of thermophilic clones sequences from both the libraries a combined Neighbor-Joining distance based phylogenetic tree was constructed (figure 5.7). The putative thermophilic and thermotolerant clone sequences from both Tattapani and Tapovan geothermal springs got segregated into two distinct classes showing unique thermal isolates from both ecological niches.

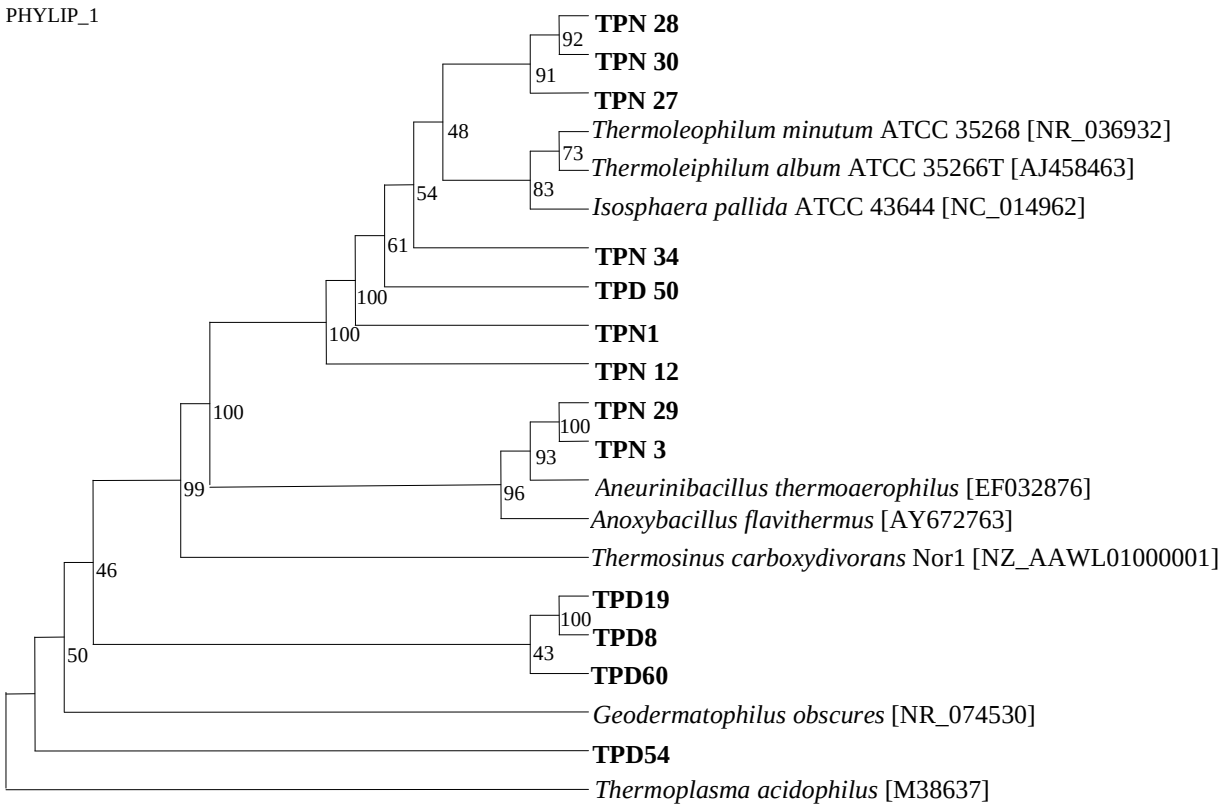


Figure 5.7: Phylogenetic relation of Tattapani and Tapovan derived clones with thermotolerant and thermophilic organisms. *Neighbor-Joining distance based tree deduced from partial sequences of 16S rRNA gene clones from Tattapani and Tapovan hot spring samples. The figures shows phylogenetic relationship to thermotolerant and thermophilic organisms. Bootstrap values of 1000 resamplings were given at the branch point.*

5.5. Culture based microbial diversity in the Himalayan springs:

Culture based microbial studies include their enrichment using Biolog EcoPlates and by enriched media based isolation of single colonies. The Biolog EcoPlates after proper incubation were scanned for both the spring samples (figure 5.8). Microbial single carbon utilization profile were determined by measuring the color development using MicroStation™ System/MicroLog plate reader USA. The metabolic profile of microbes from both spring were given in table 5.4. It was found that the average and community level metabolic profile of Tattapani spring was higher than Tapovan spring. The culture dependent analysis again support the culture independent based phylogenetic analysis.

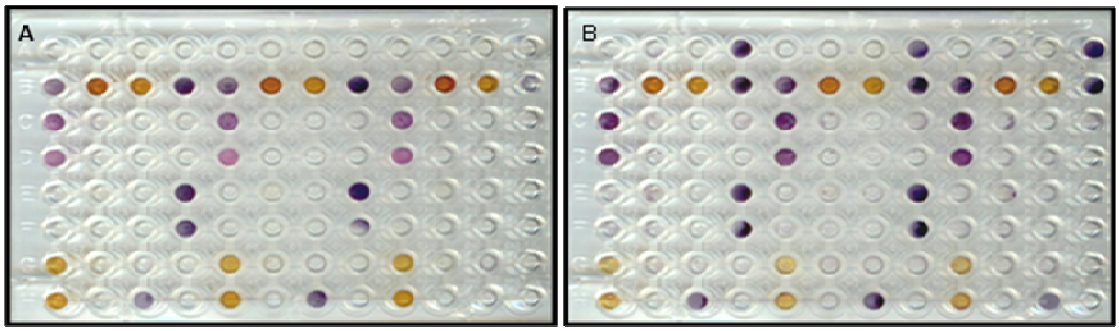


Figure 5.8: Biolog EcoPlates of Tattapani and Tapovan spring water sample. ***A:*** *Tattapani spring derived Biolog EcoPlate and* ***B:*** *Tapovan spring derived Biolog EcoPlate.*

Average metabolic rate (AMR) and community level physiological profiling (CLPP) determination:

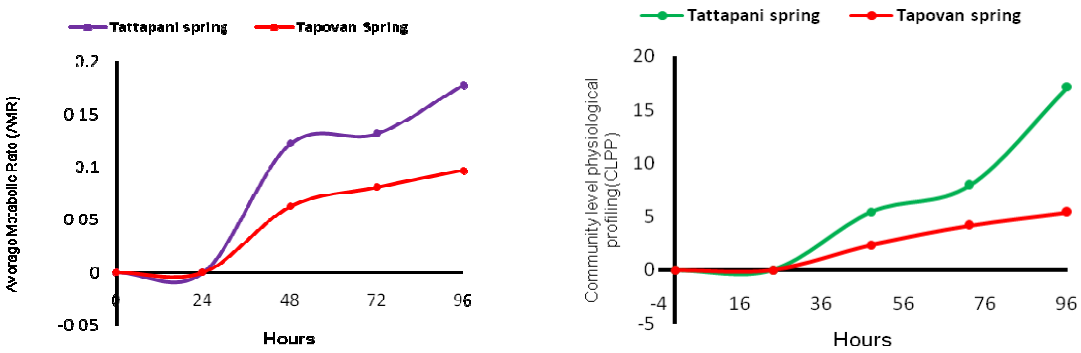


Figure 5.9: Comparison of community based metabolic profiling of Himalayan thermal springs. *Average metabolic rate (AMR) and Community level physiological profiling (CLPP) were plotted against hours.*

Table 5.4: Comparative analyses of carbon source oxidation profiles of Tattapani and Tapovan water samples using Biolog EcoPlate:

Carbon sources	Tattapani spring	Tapovan spring
Acids		
Hydroxybutryic acid	-	-
Itaconic acid	-	-
Ketobutyric acid	-	-
Mallic acid	+	+
Polymers		
Tween 40	+	+
Tween 80	+	+
Glycogen	-	-
Phenolic compounds		
2- Hydroxybenzoic acid	-	-
4- Hydroxybenzoic acid	-	-
Amine		
Phenyl ethyl amine	-	-
Putrescine	-	-
Carbohydrates		
D- Lactose	-	-
D- Cellobiose	-	-
Methyl- D- glucoside	-	-
D- Xylose	+	+
i- Erythritol	-	-
N-acetyl-D-glucosamine	-	-
D-Mannitol	-	-
D-Galactonic-lactone acid	-	-
Glucose-1-phosphate	-	-
Amino acids		
L-Asparagine	+	+
L-Arginine	-	+
L- Phenylalanine	-	-
L-serine	-	-
Glycyl-L-glutamic acid	+	+
L-Threonine	+	+
Carboxylic acids		
Pyruvic acid methyl ester	+	+
D-Glucosaminic acid	-	-
Galacturonic acid	-	-

The average metabolic rate (AMR) which is represented by the average respiration of total microbes in a community while utilizing the single C-sources in the microtiter plate [174] were calculated for both the springs and compared. Similarly, Community metabolic diversity (CMD) described as number of substrates utilized by the microbes in a community were also calculated and compared for both the springs (figure 5.9).

5.5.1. Culture isolates analysis of Tapovan hot spring:

Thermophilic isolates from Tapovan spring were analyzed microscopically. Isolates were found to be gram positive rods with variable length (figure 5.10). The strains were analyzed phylogenetically and was found that TPN 2 and 3 showed identities with *Aneurinibacillus thermoaerophilus* whereas TPN1 showed a different affiliation (figure 5.11).

Figure 5.10: Tapovan isolates (TPN 1, TPN 2, and TPN 3) visualized trough Nikon, Eclipse Ti-S compound microscope.

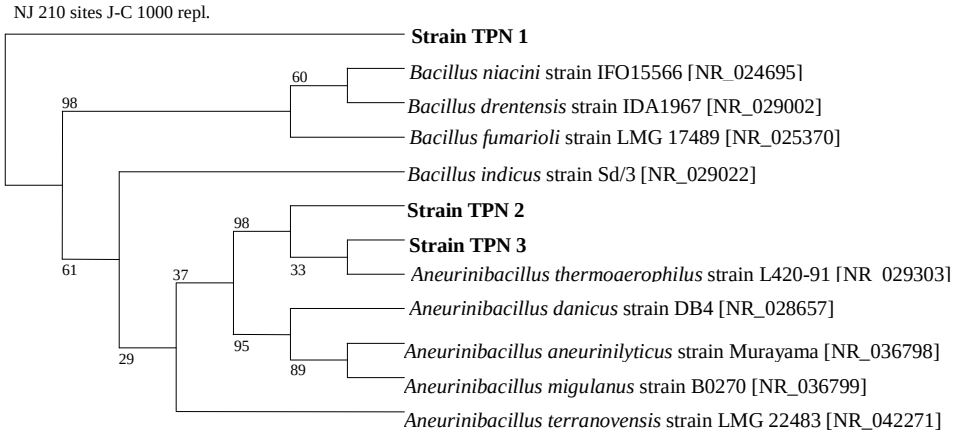


Figure 5.11: Phylogenetic tree of the cultured isolates from Tapovan with their closest relatives.

5.5.2. Fatty acid methyl ester (FAME) analysis for Tapovan isolate:

FAME analysis for Tapovan isolate TPN-1was carried out and shown in table 5.5. Which was an obligate thermophilic bacteria. Fatty acid profile of TPN-1 was compared with that of its closest bacterium *Bacillus niacin*.

Table 5.5: FAME analysis of obligate thermophile TPN-1:

S. No.	Fatty acid	TPN-1 Precent	*Bacillus niacini*
1	13:0 iso	0.91	1.58
2	14:0 iso	2.21	2.20
3	14:0	1.55	2.06
4	15:0 iso	50.20	32.32
5	15:0 anteiso	18.41	15.59
6	16:1 w7c alcohol	NIL	0.60
7	16:0 iso	3.64	1.01
8	16:1 w11c	NIL	18.90
9	16:0	5.21	7.30
10	17:1 iso w10c	NIL	3.70
11	17:0 iso	3.47	4.49
12	17:0 anteiso	5.39	1.45
13	18:1 w9c	1.22	4.13
14	18:0	0.85	3.14
15	Summed Feature 3	0.44	0.75
16	Summed Feature 4	NIL	0.76

The Tapovan derived isolates (TPN1, TPN2, and TPN3) were further characterized using Biolog GN-MicroPates containing 95 different carbon sources. The singe carbon source utilization profile was presented in table 5.6.

Table 5.6: Comparative analysis of microbial isolates from Tapovan spring for utilization of different carbon sources in Biolog GN MicroPlates.

Properties	Strain TPN1^T	Strain TPN2^T	Strain TPN3^T	Properties	Strain TPN1^T	Strain TPN2^T	Strain TPN3^T
Dextrin	-	-	-	Glycyl-L-Proline	-	-	-
D-Maltose	+	-	-	Gelatin	-	-	-
D-Trehalose	+	-	-	L-Arginine	+	+	+
D-Cellobiose	+	-	-	L-Alanine	+	+	+
Sucrose	-	-	-	L-Aspartic acid	+	+	+
Gentiobiose	-	-	-	L-Glutamic acid	+	+	+
D-Turanose	+	-	-	L-Histidine	+	+	-
Stachyose	-	-	-	L-Pyroglutamic acid	-	+	+
Positive Control	+	-	+	Lincomycin	-	-	-
pH – 6	+	-	+	L-Serine	-	-	-
pH – 5	-	-	+	Guanidine HCl	-	+	+
D-Rulfinose	-	-	-	Niaproof 4	-	-	-
α-D-Lactose	-	-	-	Pectin	+	-	-
D-Malibiose	-	-	-	D-Galacturanic acid	-	-	+
β-Methyl-D-Glucoside	-	-	-	L-Galactonic acid Lactone	-	-	-
D-Salicin	-	-	-	D-Gluconic acid	+	-	-
N-acetyl-D-Glucosamine	+	-	-	D-Glucuronamide	+	+	+
N-acetyl-β-D-Mannosamine	+	-	-	Glucuronamide	+	+	+
N-acetyl-D-Galactosamine	-	-	-	Mucic acid	-	-	-
N-acetyl-Neuraminic acid	-	-	-	Quinic acid	-	-	-
1% NaCl	+	+	+	D-Saccharic acid	-	-	-
4% NaCl	+	+	-	Vancomycin	-	-	-
8% NaCl	-	-	-	Tetracolium violet	-	-	-
α-D-Glucose	+	-	-	Tetracolium blue	-	-	-
D-Mannose	+	-	-	p Hydroxy-Phenylacetic acid	-	-	-
D-Fructose	+	-	-	Methyl Pyruvate	+	+	+
D-Galactose	-	-	-	D-Lactic acid Methyl ester	-	+	+
3-methyl glucose	-	-	-	L-Lactic acid	+	+	+
D-Fucose	+	+	-	Citric acid	-	-	-
L-Fucose	+	+	-	α-Keto Glutaric acid	-	+	+
L-Rhamnose	-	-	-	D-Malic acid	-	+	+
Inosine	+	-	-	L-Malic acid	-	+	+
1% sodium Lactate	+	+	+	Bromo Succinic acid	-	+	+
Fusidic acid	-	-	-	Nalidixic Acid	+	-	-
D-Serine	-	-	-	Lithium Chloride	+	-	-

D-Sorbitol	-	-	-	Potassium Tellurite	+	+	-
D-Mannitol	+	-	-	Tween 20	-	-	-
D-Arabitol	-	-	-	ϒ-Amino Butryric acid	-	-	-
Myo-inositol	-	-	-	α-Hydroxy Butyric acid	-	+	+
Glycerol	-	+	+	β-Hydroxy –D,L-Butyric acid	-	+	+
D-Glucose & PO4	-	-	-	α-Keto Butyric acid	-	+	-
D-Fructose & PO4	-	-	+	Propionic acid	-	-	-
D-Serine	-	-	-	Acetoacetic acid	-	-	-
D-Aspartic acid	-	-	-	Acetic acid	-	+	-
Troleandomycin	-	-	-	Formic acid	-	-	-
Rifamycin	-	-	-	Sodium Bromate	-	+	+
Minocycline	-	-	-	Sodium Butyrate	+	+	+
Aztreonam	+	+	-				

5.6. Discussion:

Bacterial diversity in Tattapani geothermal spring mat:

The presently studied, Tattapani geothermal area is likely to get submerged under Kole-hydroelectric project water reservoir. It was of prime scientific interest to study the uniqueness of the bacterial diversity of this geothermal spring before the habitat is destroyed completely. From clone libraries studied in the present work, 40% of the 16S rRNA gene clones vary from the closest phylogenetic relatives by 3 to 12% at their rRNA gene level. It is widely accepted in literature that strains showing less than 97% 16S rRNA sequence similarity to all known taxa are considered to belong to a new species [35, 55, 172]. Since these closely related species are listed as unculturable bacteria in the data base they could be designated as new species.

Proteobacteria which constitute the largest population of clone libraries generated were categorized within the classes Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, and Deltaproteobacteria. The uncultured *proteobacteria* have been largely ignored in literature [76]. Based on the level of confidence used, only 35% of the proteobacterial sequences can be assigned to known genera, inferring that many proteobacterial groups still remain to be assigned for nomenclature [55]. An interesting finding was the high frequency of *Acidobacteria*. Many of them are of uncultured type. Acidobacteria are the bacterial members frequently noticed in high temperature habitats such as hydrothermal terrestrials, deserts and marine systems [72, 76, 97] .and it is quite difficult to culture thermophilic acidobacteria in laboratory conditions. At present, research interest has increased towards the as-yet-uncultured group of phyla Acidobacteria [76,

97]. Another striking feature of the Tattapani bacterial diversity is the presence of the sequences with resemblance to Bacteroidetes. The Bacteroidetes belongs to heterotrophic bacteria and are not usually associated with geothermal springs. There have been very few reports of Bacteroidetes as major group from hot springs [80, 131]. Recently, Bacteriodetes have been reported in hot springs of Northern Patagonia constituting up to 20% of the total bacterial diversity [102]. Based on the mat composition of Patagonia, Thailand and Tattapani it is likely that Bacteroidetes prefer nutrient rich niches and a temperature range of up to 65 °C. Another feature in the Tattapani geothermal spring was presence of sequences bearing similarity to Chloroflexi. It has been suggested that the presence of autotrophic green non-sulfur bacterial species like *Chloroflexus* in such geothermal mats also encourages the growth of bacteroidetes [102].

One of the promising finding in clone library from Tattapani is that the Verrucomicrobia, which represents about 14% of the cloned sequences in the library (TPD 24,32,37,46,55,66 and 68) belonged to uncultured microbes in database. Till date, nearly all Verrucomicrobia cultured from soil have been found to be free-living and are mesophilic, facultative or obligate anaerobe, saccharolytic and oligotrophic in nature, however, relatively little information is available on uncultured verrucomicrobia in thermal spring mat samples [12]. A few Verrucomicrobia isolated from highly acidic and geothermal areas have been reported to be excellent consumers of methane and thus their application as biofilters to reduce methane emissions has been reported [46, 128]. In this work, organic rich environment and their association with other eukaryotic organisms could be accounted for the presence of Verrucomicrobia in Tattapani spring mat [12]. Another set of clones from Tattapani resembled uncultured bacteria (Accession numbers - AB426197 and FM956229 which resembled bacteria known to be growing under methanogenic conditions again pointing towards the rich biodiversity in the Tattapani geothermal springs.

The group of planctomycetes constituting 5% of the Tattapani clone library were largely represented by *Isosphaera pallida*, *Planctomyces limnophilus* and *Planctomyces brasiliensis* which are most common in brackish, marine and fresh water environments [58]. In addition, 3% of the clone sequences from Tattapani metagenome belonged to uncultured *Chloroflexi* population type identified in activated sludge and many of them showed lower value of identities with the known microbial database in National Center for Biotechnology Information (NCBI). Hence propose them as the new species under phylum *Chloroflexi*.

Bacterial diversity in Tapovan geothermal spring mat:

As compared to Tattapani, Tapovan geothermal area mats are located at higher altitude, spring water has higher temperature, the soil possesses higher alkalinity and is nutrient poor. These conditions are expected to evolve unique bacterial diversity. The cloned library analysis from Tapovan confirmed that there is striking difference in the bacterial diversity from the Tattapani geothermal mats. A majority of organisms are closely related to Firmicutes. Members of this phylum possess Gram-positive cell wall, have low G+C content in genome, form endospores and are known to survive in extreme environments [118]. Amongst Firmicutes, two genera showing homology with most of the clones are represented by species of thermophilic *Anoxybacillus* and *Aneurinibacillus* (figure 5.6). Members of these two genera have been earlier reported from various ecological niches including hot springs, dairy processing plants, soil [29, 111, 116] and have potential applications related to starch and lignocellulosic biomass conversion, enzyme technology, environmental waste treatment and bioenergy production [60]. The Tapovan clone sequences showed similarity to various species of *Anoxybacillus* and *Aneurinibacillus* genera (figure 8) which indicated towards presence of several representatives of these two genera in the Tapovan thermal mats. Amongst the firmicutes, other clones which showed similarity (TPN 25, 26, 3 and 13) belonged to thermophilic *Bacillus* species isolated from Santorini volcano in Greece (JQ808139) and hot synthetic compost [41]. Clone TPN 14 and 18 showed significant similarity with *Clostridium tertium* of *Clostridiaceae* family. Close resemblance of TPN14 and 18 to *Clostridium tertium* is surprising since this genus has never been reported from thermophilic environments and warrants further in-depth investigation since *C. tertium* is known to be a pathogenic bacterium [160].

The other phylum represented in the Tapovan geothermal mats is *Proteobacteria*. Three clones (TPN 24, 27 and 32) showed significant similarity to *Pseudoxanthomonas taiwanensis*. *P. taiwanensis* has been reported from hot springs of eastern Taiwan [32] and belongs to gamma proteobacteria. One of the clone TPN8 showed similarity to beta proteobacterium *Tepidimonas* sp. albeit with low degree of certainty (figure. 8).

The water chemistry and physical parameter studies from both the springs indicate that heavy metals like Al, As, B, Mn, and Se are found significantly higher than the average standard concentrations of the metal in thermal spring environments [112]. Elemental cations like ammonium, magnesium, and anions sulphate, chloride are higher in concentration found in

Tapovan spring compared to Tattapani spring. BOD of the springs was found to be 4.1 and 4.7 respectively (Table: 4.1). It indicates that the water from Tapovan spring is having more inorganic environment than Tattapani spring. Microorganisms in these stressful environments have developed adaptations to the more inorganic and volcanic environments, hence chemolithoautotrophs are often seen [152]. Chemo(litho)autotrophic microorganisms are responsible for specific elemental transformations like S, Fe, NH_4, CH_4 oxidation, hence they have a vital role in nutrient cycling and climate regulation in the ecosystem [66]. *Pseudoxanthomonas* sp. a chemolithoautotrophic facultative anaerobic bacterium was identified in the Tapovan spring mat. Thermophilic *Anoxybacillus flavithermus* seen in the Tapovan spring is able to utilize a wide range of heavy metals like Cu, Mn, Ni, Pb, Zn [26]. Manganese is used by microorganisms as an electron acceptor in anaerobic respiration processes [119]. The ferrous ion is essential for microbes residing in anaerobic environments [119] which was also seen in the studied Tapovan spring water. Microbes growing in such environments can be used for bioremediation of metal contaminated environments. Hence it is quite important to protect and preserve microbial biodiversity and their lineages from such rare ecosystems.

Overall, it is clear that there is a striking difference in the bacterial diversity of two hot springs of Himalayan geothermal belt. A combined phylogenetic tree of the selected clones of both geothermal springs showing significant similarity to known thermophiles also segregated thermophiles from two springs taken in the study (figure. 5.7). The same can be attributed to difference in temperature, elevation from sea level, nutrient availability and pH. This study has found that Tattapani geothermal spring harbors several genera of bacteria which are known to be of industrial importance. Primary reason for this could be the method of environmental DNA isolation with a combination of electro-elution step which allowed for amplification of 16S rRNA sequences from most of the representative species.

Interestingly, 40% of the clone library candidates show less than 97% identities with the known bacterial rRNA sequences in the databases. These representatives may be categorized as new species identified from the hot spring mat as per species concept of microorganisms [35, 55, 172]. The RFLP signature analysis provided a preliminary indication about the diverse bacterial species' presence in Tattapani geothermal mats as compared to Tapovan geothermal mats. This was confirmed by the Shannon-weiner index analysis of both the hot spring mats. The diversity of two mats were studied by the rarefaction curve, which infers that the species richness is high in

Tattapani spring as compared to Tapovan spring. The microbial community of two springs have been shown to be phylogenetically distinct. Tapovan spring inhabits obligate thermophilic and facultative chemolithoautotrophic organisms which may be attributed to the poor nutrient availability and inorganic environment [87].

The difference in the bacterial community of the two geothermal springs of Himalayan geothermal belt of Indian region is remarkable. Both the sites have not been subject of in-depth studies. Unless there are concerted efforts for conservation of biodiversity of the Tattapani geothermal springs, interesting microbial species are in the danger of getting lost due to construction of Kole-hydroelectric power plant. Conservationists and environmental activists have established a clear association between hydroeletectric power plant constructions and biodiversity loss in Himalayas specifically [63]. Advancement in the molecular typing methods for biodiversity studies have provided reliable tools for quick characterization of the such unique resources so that action cum monitoring plans for conservation of such resources can be put in place. Himalayas are already home to a megadiversity center. This work stresses that the bacterial diversity of the Himalayas warrants detailed and in-depth studies. So far, a comprehensive microbial diversity study of Himalayan geothermal springs has not been taken up. This is the first attempt to fill the void in this field. The bacterial biodiversity information of both geothermal springs will serve as part of baseline data for comparing and contrasting the bacterial diversity of other geothermal springs in the Himalayan geothermal belt. Such data will be of immense help for evolution, biogeography, microbial ecology, environmental sciences, and biotechnology research. Additionally, a better understanding of such fragile ecosystems will allow for a sustainable management of such unique resources.

CHAPTER 6

RESULTS AND DISCUSSION-III

Functional metagenomic approach for characterization of Himalayan geothermal springs

6.1. Metagenomic library generation from hot spring eDNA:

6.1.1. Partial digestion of metagenomic DNA:

Metagenomic DNA isolated from Tattapani springs was partially digested using the restriction enzyme Sau3AI to obtain DNA fragment of size of range of 3 to 8 kb. Partial digestion of metagenomic DNA was optimized by digestion for different time periods. DNA digested for 10 minutes was found to be suitable for subsequent cloning as a major portion of the digested DNA was in the range of 3 to 8 kb (figure: 6.1). Gel portion containing specified partially digested DNA was excised and eDNA was eluted. The partially digested eDNA was used for functional metagenomic library construction. Size selected eDNA was then ligated to pNYL-*rygC* vector. The recombinant plasmid was electrotporated in to *Escherichia coli* DH5αZ1 host. Transformants were selected on LB agar plate containing 40μg/ml kanamicin and 20ng/ml anhydrotetracyline (ATC). Approximately 8000 clones obtained on selection plates were used for screening cellulase and xylanase enzyme coding gene fragments.

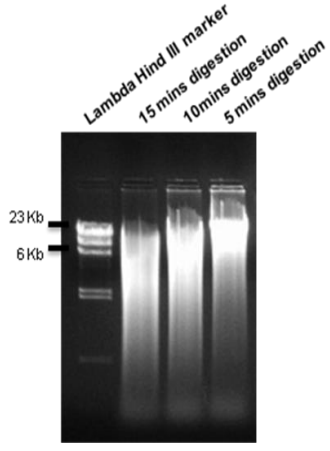


Figure 6.1. Partial digestion of metagenomic DNA. *Metagenomic DNA was partially digested using Sau3AI enzyme for 15, 10, and 5 mins respectively. The digested DNA was electrophoresed on 1% agorase gel.*

6.1.2. Cloning efficiency and clone analysis from metagenomy library:

5000 clones from each of the metagenomic library were obtained on the screening plates. They were subsequently analyzed for cloning efficiency by restriction excision on insert from the cloned plasmids.

6.1.3. Confirmation of insert in transformants of metagenomic libraries:

Clones were then analyzed by restriction digestion of plasmids from Tattapani library. Digestion of the plasmid using Kpn I and Mlu I restriction enzyme showed that there was a vector backbone corresponding to 2.2 kb. The insert of one or more fragments were visualized in the gel. The insert, which had the restriction site for Kpn I and/or Mlu I fragmented into multiple bands (figure 6.2).

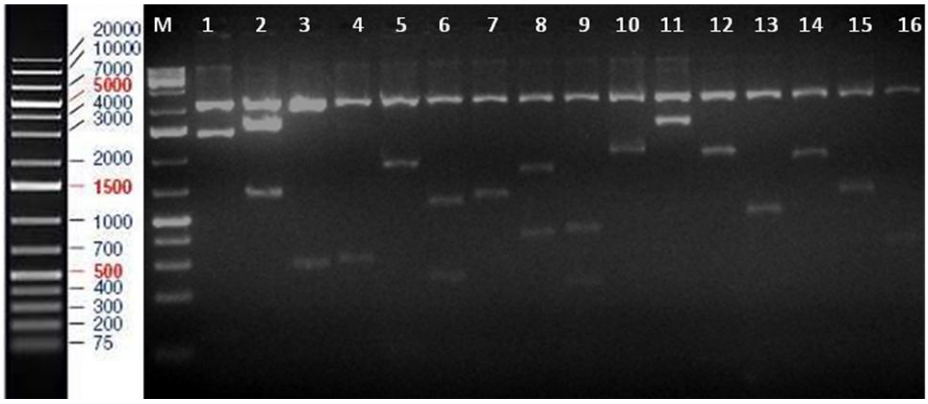


Figure 6.2. Tattapani functional metagenomic library clone analysis using Kpn I and Mlu I restriction digestion. *Lane M: Gene O Ruler 1 kb plus molecular weight marker (Fermentas, USA). Lane 1 to 16: Plasmid from Tattapani clones digested with restriction enzymes Kpn I and Mlu I.*

The same experiment was performed for Tapovan derived clone library. Similar pattern of vector backbone and insert were observed in the specified range on the gel (figure 6.3). This confirmed the successful construction of metagenomic functional libraries made from both the Himalayan eDNA.

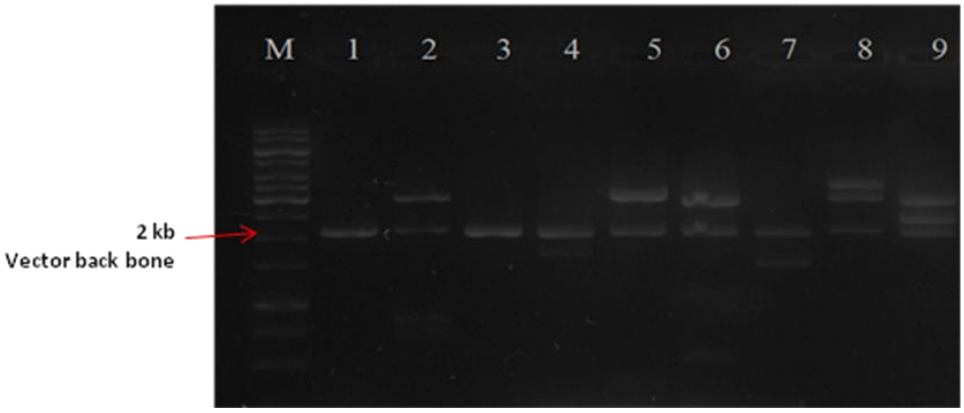


Figure 6.3. Tapovan functional metagenomic library clone analysis using Kpn I and Mlu I restriction digestion. *Lane M: Gene O Ruler 1 kb plus molecular weight marker (Fermentas, USA). Lane 1 to 9: Plasmid from Tattapani clones digested with restriction enzymes Kpn I and Mlu I.*

Screening of clones for cellulase and xylanase activity:

Around 2000 clones from each of the libraries were screened on LB kanamycin plates containing CMC and xylan separately. No cellulase or xylanase activity was observed on the agar plates containing xylan or cellulose. The reason could be difficulty in heterologous expression. Hence, to overcome this problem direct PCR based approach was adopted for functional analysis of eDNA from hot springs.

6.2. Directed PCR based metagenomic approach to study cellulase gene from hot spring eDNA:

6.2.1. Design of consensus primers for endoglucanase gene:

Degenerate primers for endoglucanase gene was constructed using the consensus sequences of cellulase hydrolase candidates of *Bacillus* sp. and from *Bacillus licheniformis* (figure 6.4).

Design of forward primer for endoglucanase (*cel5A*) gene:

Bacillus licheniformis cellulose hydrolase (*cel5H*)--- GAAACGTTCCATCTCTGTCTTCATCG-
Bacillus sp. CY1-3 CelC (*celC*) ---- GAAACGTTCCATCTCTGTCTTCATCG
Bacillus licheniformis ATCC 14580---- GAAACGTTCCATCTCTGTCTTCATCG-
Bacillus sp. NBL420 cellulase (*celA*)----- GAAACGTTCCATTTCTGTCTTCATCG
Bacillus sp. beta-1,4-endo-glucanase gene ----ACGTT--------GGTTTTATTG--
Endoglucanase forward primer GAAACGTTCCATCTCTGTCTTCAT

Construction of reverse primer for endoglucanase (*cel5A*) gene shown as reverse compement:

Bacillus licheniformis strain GXN151 (*cel9A*)	GTTTTCGGACATGAGCCCG----GTTACTA
Myxococcus stipitatus DSM 14675	
CGTGGCGAAGAGGTGCTCGCGGAAGTGCGA	
Bacillus licheniformis ATCC 14580	GTTTTCGGACATGAGCCCG----GTTACC-
Bacillus licheniformis cellulose hydrolase (*cel9A*)	GTTTTCGGACATGAGCCCG----GTTACTA
Endoglucanase reverse primer	GTTTTCGGACATGAGCCCG

Figure 6.4. Design of degenerate primers for endoglucanase gene. *Degenerate primer were constructed suing the sequence of cellulase genes of Bacillus sp. and Bacillus licheniformis.*

Purified metagenomic DNA from Tapovan spring was used for direct PCR amplification using degenerate primers of endoglucanase gene. Amplified product was electrophoresed on 1% agarose gel along with 1 kb gene O ladder (Fermentas, USA). A 1.5 Kb amplified product was vizualised on the agarose gel (figure 6.5). As the endoglucanase gene was derived from **T**apovan **M**etagenomic DNA which was identical to **Cel5A** of *Bacillus* sp., it was named as **TM-Cel5A**.

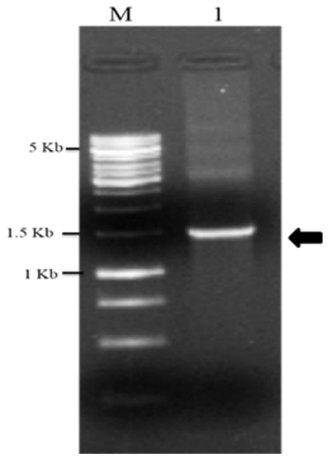


Figure 6.5. Amplification of TM-*cel5A* gene. *Lane M: represents 1 kb molecular weight marker (Fermentas,USA), Lane 1: PCR amplified product of TM-cel5A gene from Tapovan eDNA.*

6.2.2. Cloning of endoglucanase (TM-*cel5A*):

The PCR amplified and gel purified endoglucanase gene from Tapovan eDNA was ligated to pET28(a) vector and was transformed into electro-competent cells of *Escherichia coli* BL21 CodonPlus DE3. The transformants were selected on LB agar plates containing 40µg/ml kanamycin. Selected clones were further analyzed for the presence of *TM-cel5A* by restriction digestion of plasmid.

6.2.3. Confirmation of endoglucanase gene from transformant of Tapovan:

Selected transformant on the selection plate was taken for confirmation of endoglucanase gene. Plasmid was isolated from the transformant and was digested with Bam HI and Xho I restriction enzymes. The digested product was electrophoresed on 1% agarose gel along with 1 kb molecular weight marker (Fermentas USA).Vector backbone of 5.3 kb and insert of 1.5 kb were obtained on agarose gel (figure 6.6).

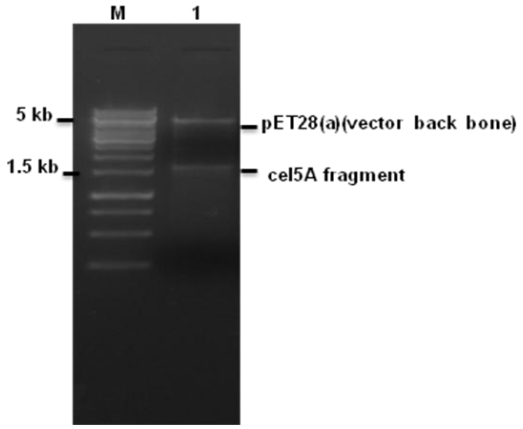


Figure 6.6. TM-*cel5A* clone analysis. *M: 1 kb molecular weight marker (Fermentas USA).Lane 1: Digestion and conformation of TM-cel5A gene insert in clone. TM-Cel5A gene fragment was obtained by digesting with Bam HI and Hind III enzymes.*

6.2.4. Analysis of cellulolytic activity of clones from Tapovan eDNA:

Selected clone harboring pET28(a)-TM-*cel5A* plasmid was used for cellulolytic activity screening on the agar plate. LB agar plate containg 1% CMC and kanamycin (40 µg/ml) was used for screening of activity of the clone. *Escherichia coli* cell carrying pET28(a)-TM-*cel5A* were grown overnight on the screening plate containing Carboxymethyl cellulose (CMC). This screening plate was flooded with congo red dye for 15 minutes followed by 1M NaCl for 10 minutes. A zone of hydrolysis was visualized around the positive colony. Similar experiment was carried out using *Escherichia coli* BL21- CodonPlus cells harboring pET28(a) plasmid without the endoglucanase gene, which showed no zone of hydrolysis (figure 6.7).

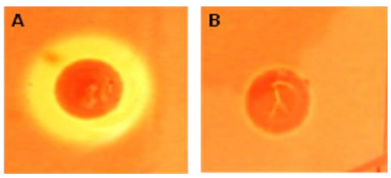


Figure 6.7. Activity of positive clone on LB CMC plates. A: *Escherichia coli BL21CodonPlus cell carrying pET28(a)-TM-cel5A derived from Tapovan eDNA, screened on LB plates containing 1% CMC stained with congo red* ***B:*** *Escherichia coli BL21- CodonPlus cell harboring pET28(a) plasmid without the endoglucanase gene as negative control.*

Endoglucanases, which cleave at the amorphous region of CMC lead to the generation of reducing and non reducing ends for cellobiohydrolases. Agar plates containing CMC with its amorphous structure facilitate the detection of endoglucanases when stained with congo red. The result showed a clear and distinct zone of hydrolysis around the positive clones on the agar plate containing CMC when compared with the non producer clone (Figure 6.7). 47 randomly selected clones when screened on the CMC containing agar plates showed difference in their expression pattern. The clone with largest zone of hydrolysis was taken for further expression and characterization of endoglucanase (Figure 6.8).

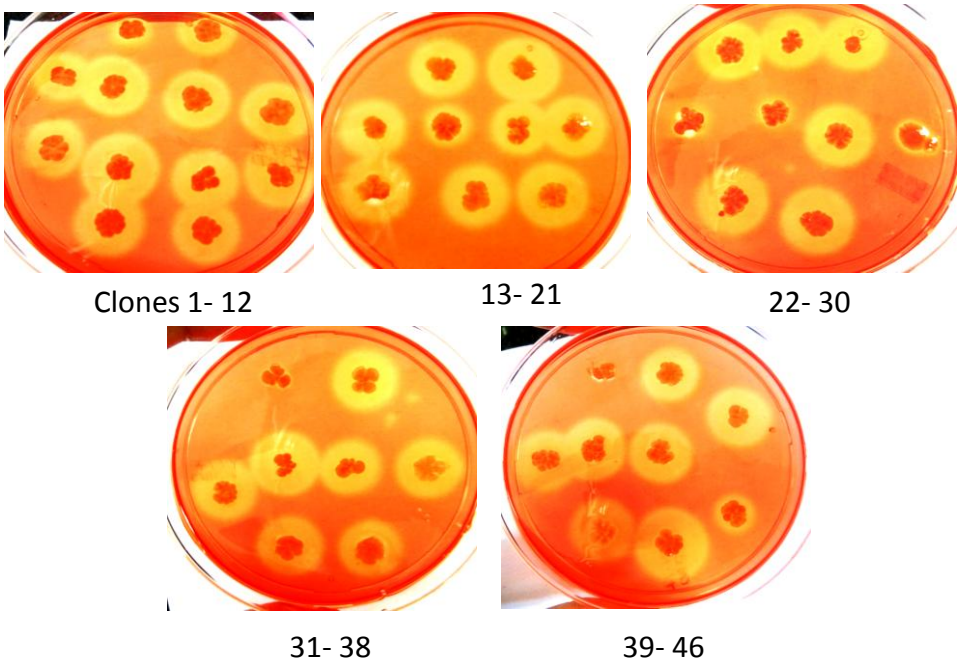


Figure 6.8. Difference in zone of hydrolysis pattern of different clones derived from Tapovan on LB- CMC plates. *Zones of hydrolysis of 47 clones on LB-CMC agar plate stained with Congo red.*

6.2.5. Sequence based analysis of TM-Cel5A:

The TM-Cel5A gene from the positive clone was sequenced. BLAST result of the cloned TM-*cel5A* gene sequence, showed similarity to endoglucanase gene from *Bacillus* sp. and endoglucanase gene of *Bacillus licheniformis.*

The amino acid sequence of TM-Cel5A protein share about 97% identity at amino acid level with endoglucanase gene of *Bacillus* sp. and *Bacillus licheniformis* (ACY72379). A phylogenetic tree was constructed using TM-Cel5A amino acid sequence and its closest relatives, with poisson distribution at 1000 bootstrap replicator. Phylogenetic tree of TM-Cel5A amino acid sequence showed that TM-Cel5A is closely identical to endoglucanase cel5A protein from *Bacillus* sp., *Bacillus licheniformis.* To validate this observation, amino acid multiple sequence alignment of TM-Cel5A and other endoglucanases from *Bacillus mojavensi*, *Bacillus amyloliquefaciens*, *Bacillus pumilus*, *Paenibacillus polymyxa*, *Geobacillus* sp, *Streptomyces* sp, *Hypocrea rufa*, *Caldivirga maquilingensi* and an endo-1,4-beta-glucanase from an uncultured bacterium were carried out (figure 6.9).

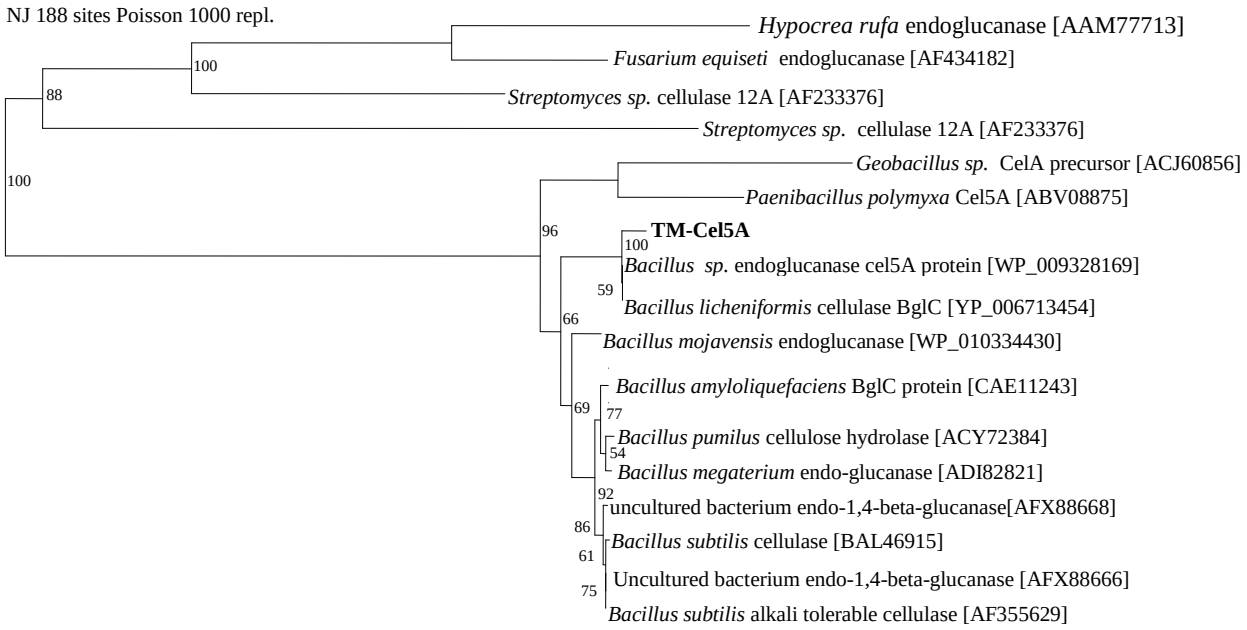


Figure 6.9. Phylogenetic classification of TM- Cel5A through amino acid sequence analyses. *Amino acid sequences from TM-Cel5A and other endoglucanases analyzed by poisson distribution. Bootstrap replicator of 1000 was taken for the phylogenetic tree construction. GeneBank accession numbers are given in the parentheses. Phylogenetic relation shows that TM-Cel5A is closely related to endoglucanase gene of Bacillus sp. and Bacillus licheniformis.*

Multiple sequence alignment of TM-Cel5A with its closest relatives was carried out using CLUSTALW at European Bioinformatics Institute (EBI). The multiple amino acid sequence alignment of TM-Cel5A protein with other identical endoglucanase proteins from *Bacillus* sp., *Bacillus licheniformis, Bacillus amyloliquefacien, Bacillus subtilis, Bacillus mojavensis* and an uncultured bacterium showed that there was a significant difference in the multiple alignment profile. Though the TM-Cel5A protein showed highest identity with *Bacillus* sp. and *Bacillus licheniformis* protein it also showed differences at amino acid level indicating TM-Cel5A to be a different protein from the one from *Bacillus licheniformis*. Three stretches of amino acid sequences of TM-Cel5A were found different even within the closest relatives (figure 6.10).

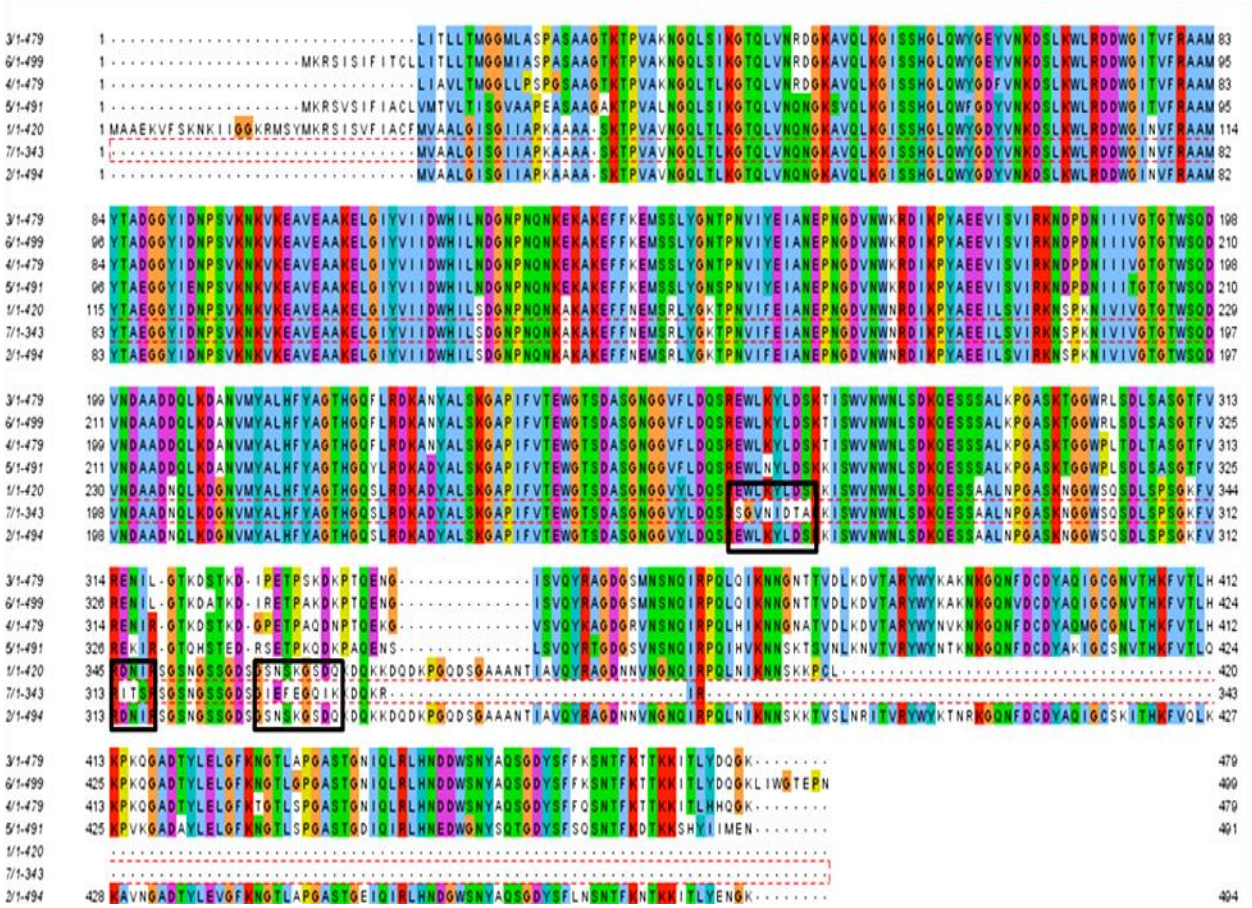


Figure 6.10. Sequence comparison of TM-Cel5A. *Amino acid multiple sequence alignment of 1: Bacillus sp. endoglucanase cel5A protein (WP_009328169) 2: Bacillus licheniformis cellulase BglC (YP_006713454) 3: Uncultured bacterium endo-1,4-beta-glucanase (AFX88668) 4: Bacillus amyloliquefaciens BglC protein (CAE11243) 5: Bacillus mojavensis endoglucanase (WP_010334430) 6: Bacillus subtilis cellulase (ADH93702) and 7: TM-Cel5A endoglucanase.*

The boxes indicate the stretches of amino acids which differ from its closest relative of Bacillus sp. and Bacillus licheniformis.

6.2.6. Expression and purification of recombinant endoglucanase in *Escherichia coli* DE3 CodonPlus cell:

Selected clone harboring pET28(a)-TM-*cel5A* plasmid was inoculated in LB broth containing 40µg/ml kanamycin and cultured overnight. Subsequent culture was done with 1% inoculum of the overnight culture. Induction of the subcultured cells were done at 0.6 OD with 0.5, mM IPTG at 37 ^{0}C for 5 hours at 200 rpm. Cell supernatant of the induced and uninduced cells was collected. The cell supernatant were analyzed for beta-endoglucanase (TM-Cel5A) protein on 12% polyacrylamide gel (figure 6.11). 52 kDa protein was found to be over expressed as compared to the uninduced cell supernatant. The culture supernatant was subjected to ammonium precipitation at 80% saturation. Precipitated pellet was collected centrifugation at 14000 rpm for 15 minutes. Collected protein pellet was dissolved in sodium phosphate buffer (50 mM pH-8.0). The crude protein was subjected to dialysis for overnight at 4°C in sodium phosphate buffer (50 mM pH-8.0).

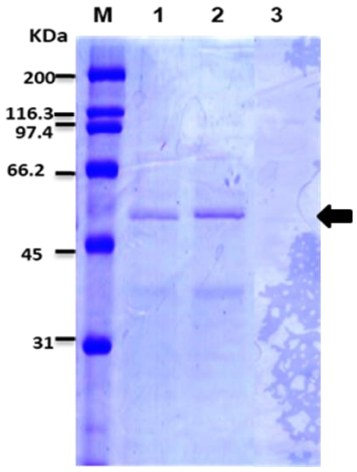


Figure 6.11. TM-Cel5A protein expression in *Escherichia coli* BL21-CodonPlus host. *M: Broad range protein marker Biorad (31-200 kDa) 1: Supernatant of cells induced with 0.5 mM IPTG. 2: Supernatant of cells induced with 0.5 mM IPTG after purification and concentration 3: Uninduced cell supernatant.*

6.2.7. Biochemical characterization of TM-Cel5A:

Determination of cellulolytic activity by Dinitrosalicillic acid (DNS) method:

The partially purified TM-Cel5A protein was used for determination of activity using liquid based biochemical assay against carboxy methyl cellulose (CMC). Equal volume (50µl each) of TM-Cel5A protein along with 1% CMC solution was incubated at 60°C for 20 minutes. 150µl of Dinitrosalicillic acid (DNS) was added to the reaction mixture. The reaction mixture was incubated at boiling water bath for 5 minutes. The reaction tube was cooled immediately. The reaction was stopped by addition of 50 µl 40% sodium potassium tartarate. The generated reduced sugar was determined by measuring OD 540 nm. The same assay was performed with negative control which was the above mentioned reaction mixture without the enzyme (figure 6.12).

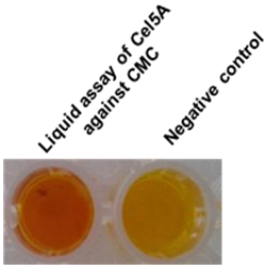


Figure 6.12. Liquid assay of TM-Cel5A against CMC. *Activity of TM-Cel5A was determined against carboxymethyl cellulose (CMC) and compared with negative control containing no enzyme.*

6.2.7.1. Thermal stability of TM-Cel5A:

Thermal stability of TM-Cel5A protein was determined by measuring its endoglucanase activity against CMC at temperature ranges from 20ºC to 90 º C temperatures range. The reaction mixture containing TM-Cel5A protein and 1% (w/v) CMC in 50 mM sodium phosphate buffer pH 8.0 was incubated at temperature ranges from 20ºC to 90ºC with 10ºC increment for 20 minutes. Generation of reduced sugar was determined using DNS method.

Maximum enzyme activity was found to be at about 60ºC. There was 50% decrease in activity when the reaction was carried out at 20ºC. Similarly, there was upto 40% decrease in the enzyme activity when the reaction was carried out at 90ºC. Thermal stability was determined at temperature vary from 60°C to 80°C by incubating the reaction mixture for 120 minutes. The

enzyme was found to be stable at 60ºC for 1hour and its activity decreased by 12% after 2 hours of incubation. Similarly, the enzyme showed a decrease in activity at higher temperatures such as of 70ºC, 80ºC and 90ºC (figure 6.13).

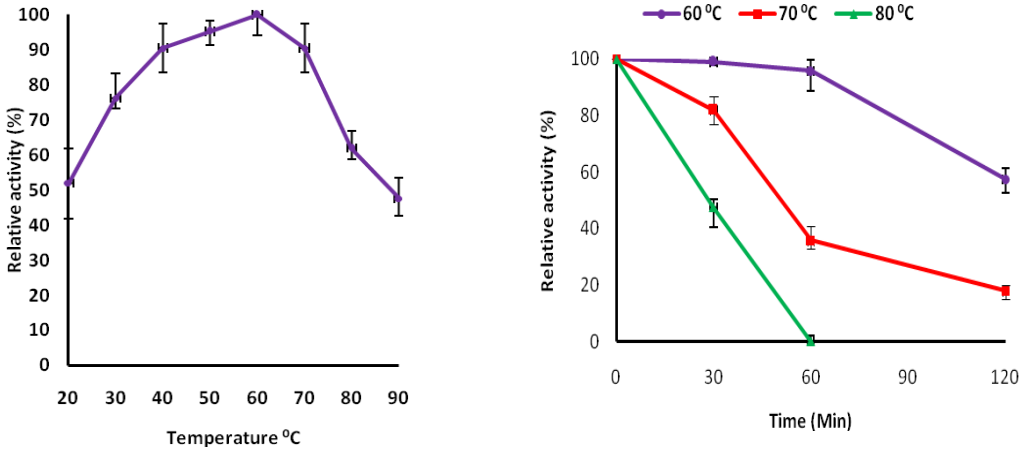


Figure 6.13.Thermal stability of TM-Cel5A protein . ***A:*** *Relative percentage enzyme activity of TM-Cel5A at different temperature ranges.* ***B:*** *Thermal stability of TM-Cel5A protein showing highest stability at 60ºC.*

6.2.7.2. Determination of optimum pH for TM-Cel5A activity:

Enzymatic activity of TM-Cel5A endoglucanase was determined by measuring its activity against CMC within the pH range of 4 to 9. Different buffers used for determination of activity of TM-Cel5A protein were 50 mM sodium acetate buffer for pH 4 to 6, 50 mM Sodium phosphate buffer for pH 6 to 8 and 50 mM Tris buffer for pH 8 and 9. TM-Cel5A protein showed maximum activity at pH 8.0. The pH stability of TM-Cel5A was determined by incubating TM-Cel5A protein at different pH ranging from 6.0 to 9.0 at 4ºC for 16 hours. The pH stability was estimated using CMC substrate. The reduced sugar generated was measured using DNS. The enzyme was found to be stable at pH 8 and pH 7 for 2 hours. Similarly, enzyme showed decrease in activity at pH 6.0 and 9.0 (figure 6.14).

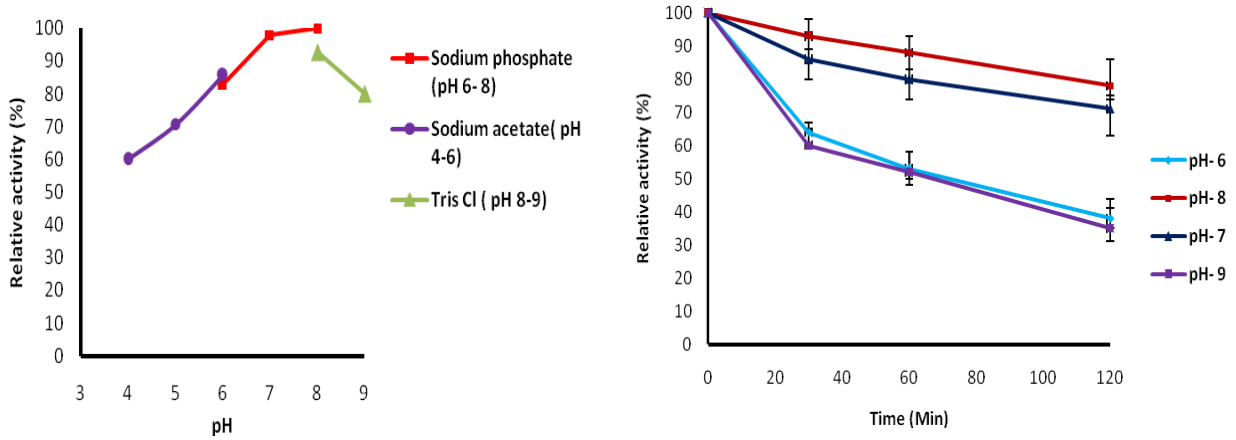


Figure 6.14. Optimum pH determination for TM-Cel5A protein: ***A:*** *Relative percentage enzyme activity of TM-Cel5A at different pH ranges* ***B:*** *pH stability of TM-Cel5A protein at different pH ranges(6.0 to 9.0).*

Bovine serum albumin (BSA) standard curve for protein estimation:

A standard curve was plotted using bovine serum albumin (BSA). Protein estimation was done using bicinchoninic (BCA) protein assay kit (Thermo scientific USA). BSA in the range of 0.5 to 2 mg/ml was taken for the assay. 25µl of standard (BSA) protein was taken in microtiter plate separately. 200 µl of BCA working reagent (WR) was added to both the wells. The content was mixed thoroughly. The plate was incubated at 37°C for 30 mins. Absorbance was taken at A_{562}nm. BSA standard curve was plotted for different dilutions taking absorbance versus protein concentrations (figure 6.15). Linear equation was obtained from the standard curve. Similarly, TM-Cel5A protein was processed to obtain the OD at A_{562}nm. Concentration of TM-Cel5A protein was determined using the BSA standard curve derived linear equation. TM-Cel5A protein concentration was determined to be 0.8 mg/ml.

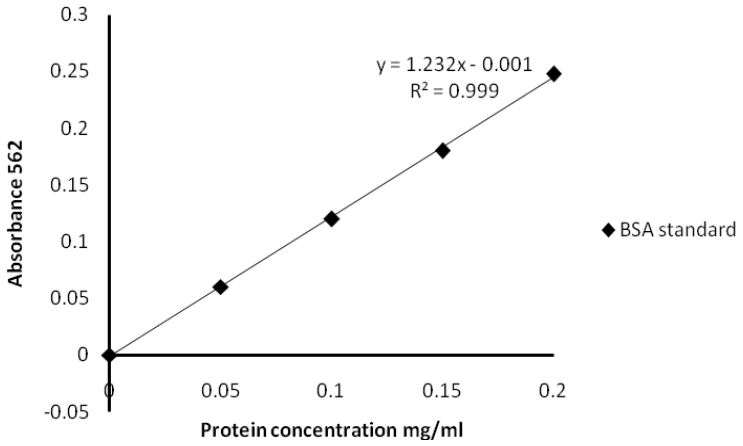


Figure 6.15. BSA standard curve determination by bicinchoninic (BCA) method.

Glucose standard curve determined by dinitrosalicillic acid (DNS) method:

Glucose (Himedia, India) in the range of 0.5 to 2.5 mg/ml concentration was taken in different test tubes. They were treated with 1.5 ml DNS (sigma, USA) solution. The reaction mixture was kept on boiling water for 5 minutes. Tubes were cooled down under running tape water. Reaction was terminated by addition of 100 µl sodium potassium tartarate. The color development was measured spectrophotometrically at A_{540} nm. A standard curve was plotted for different dilutions taking absorbance versus glucose concentrations (figure 6.16). Linear equation was obtained from the standard curve. The equation was used for determination of enzyme activity of TM-Cel5A protein.

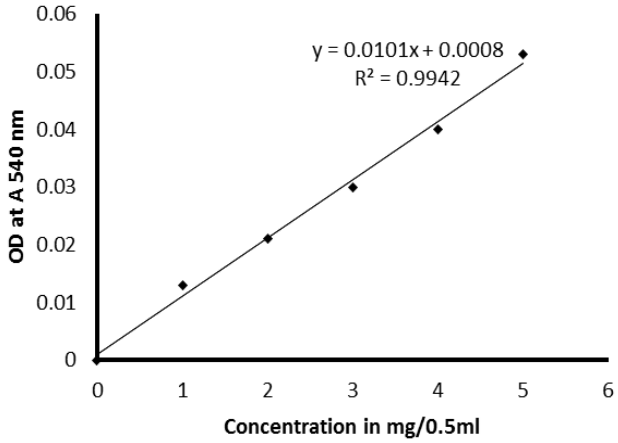


Figure 6.16. Glucose standard curve determination by DNS method.

6.2.7.3. Substrate specificity analysis of TM-Cel5A enzyme:

A variety of substrates were treated with TM-Cel5A enzyme for determination of substrate specificity. Substrates were incubated with TM-Cel5A protein at 60°C at pH 8.0 for 20 minutes. Generation of reduced sugar due to TM-Cel5A activity was determined by DNS method. Substrate specificity assay depicted that TM-Cel5A was able to utilize a wide range of substrates (figure 6.17). Apart from crystalline CMC it has the ability to hydrolyze other forms of natural polysaccharides like Beta-D-glucan. Specific activity of TM-Cel5A was found to be 100 for CMC, 130.82 for β-D-glucan, 9.17 for laminarin,64 for salicin, 323.67 for *p*-Nitrophenyl-β-D-Glucopyranoside (pNPG), 84.54 for *p*-Nitrophenyl-β-D-xylopyranoside (pNPX), 35 for Filter paper (Whatman No.1) and nil against avicel, richwood xylan (Table 6.1).

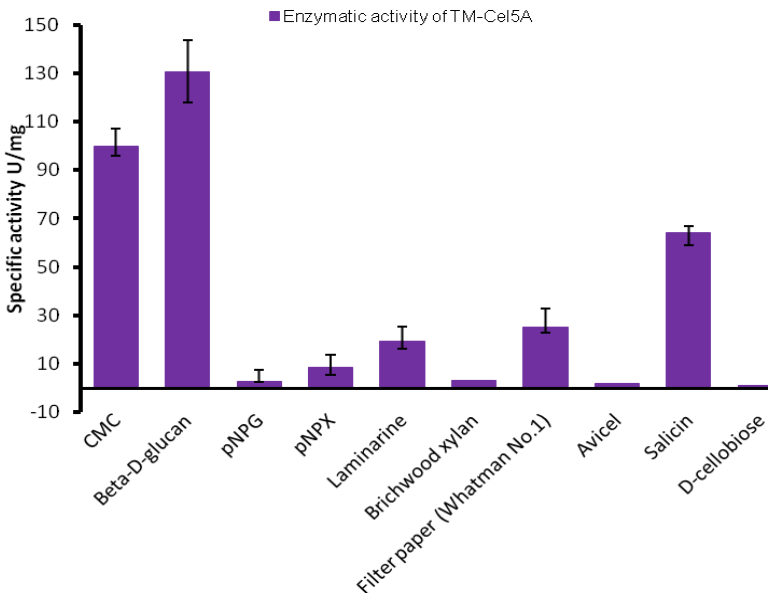


Figure 6.17. Spcific enzymatic activity of TM-Cel5A against various substrates. *Substrate specificity of TM-Cel5A against against CMC, β- D-glucan, p-Nitrophenyl-β-D-Glucopyranoside (pNPG), p-Nitrophenyl-β-D-xylopyranoside (pNPX), laminarine, xylan, filter paper,avicel,D-cellobiose, salicine are shown with standard error.*

Table 6.1. Substrate specificity of TM-Cel5A against various polysaccharide substrates.

Substrates	Specific activity (U/mg)
CMC	100
Beta-D-glucan	130.82
pNPG	2.67
pNPX	8.54
Laminarine	19.17
Brichwood xylan	3
Filter paper (Whatman No.1)	35
Avicel	2
D-cellobiose	1
Salicin	64

6.3. Cloning of L-asparaginase-II from Tapovan derived isolate:

6.3.1. Production of L-asparaginase in Tapovan derived isolate:

The Tapovan spring derived isolates, enriched through Biolog EcoPlate were cultured in minimal medium containing L-asparagine as the sole carbon source (figure 6.18). Confirmation of L-asparaginase production was done on Modified Czapek Dox medium containing phenol red. The change of color from yellow to red in Czapek Dox agar plates provided us with the preliminary indication for L-asparaginase production. Ammonia production due to L-asparaginase activity was detected by treatment with nessler's reagent. Change in color from light yellow to deep yellow in nessler's reagent confirmed the production of L-asparaginase (figure 6.18).

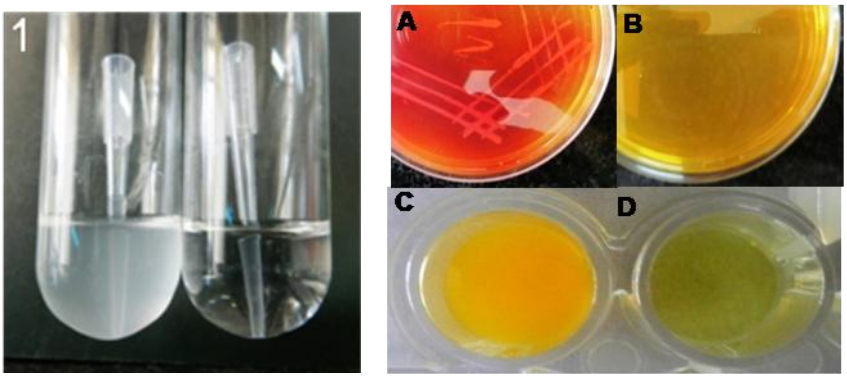


Figure 6.18. Detection of L-asparaginase producing strain in Modified Czapek Dox medium. ***[1]*** *Growth of L-asparaginase producer in minimal medium with 2% L-asparagine as the sole carbon source.* ***[A]*** *L-aspararaginase producer on modified Czapek Dox agar plate with L-asparagine as the only carbon source and containing phenol red at pH of 5.5.* ***[B]*** *Negative control of modified Czapek Dox agar plate with L-asparagine as the only carbon source and containing phenol red at pH of 5.5 in the absence of bacterial isolate.* ***[C]*** *Liquid assay for ammonia production from cell supernatant of due to L-asparaginase activity. [**D**] Negative control of liquid assay for ammonia production.*

6.3.2. Cloning of L-asparaginase-II gene from bacterial isolate of Tapovan spring:

The isolated bacterium from Tapovan spring water with L-asparaginase activity was identified by 16S rRNA gene sequencing and was found to be a close homolog of *Ralstonia* sp.

A 1 kb gene fragment was amplified from the genomic DNA of the isolated bacteria (figure 6.19-A). The cloned L-asparaginase-II gene from pET28(a) vector was digested using restriction enzymes Bam HI and Xho I. The recombinant plasmid upon digestion a 1 kb L-asparaginase- II gene fragment was obtained along with vector the backbone (figure 6.19-B).

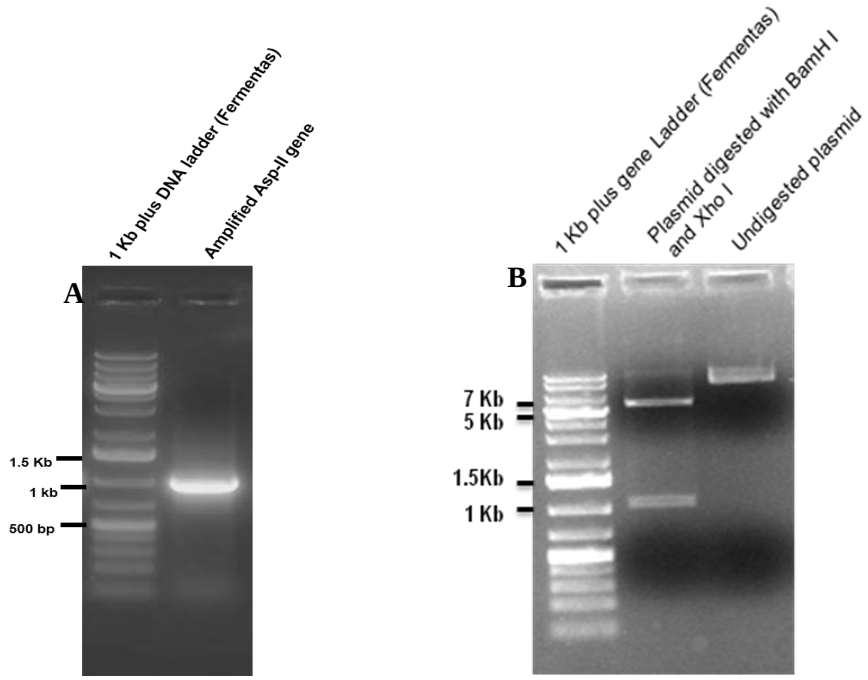


Figure 6.19: Amplification and confirmation of the clone with L-Asparaginase-II gene. ***A:*** *L- asparaginase- II gene amplified product in 1% agarose gel along with 1 kb plus DNA ladder (Fermentas USA).* ***B:*** *pET28(a)-asparaginase-II plasmid from clone was digested with Bam HI and Xho I, a 1 kb fragment of L-asparaginase-II gene is visualized on the gel. Compared with undigested pET28(a)-asparaginase-II plasmid.*

6.3.3. Expression of L-asparaginase type –II in *Escherichia coli* BL21 CodonPlus cell:

Selected clone harboring pET28(a)-asparaginase-II plasmid was inoculated in LB broth containing 40µg/ml kanamycin and cultured overnight. Subsequent culture was done with 1% inoculum of the overnight culture. Induction of the subcultured cells were done at 0.6 OD with 0.4 TO 0.6 mM IPTG at 37 ^{0}C for 5 hours at 200 rpm. Cell pellet of the induced and uninduced cells was collected and lysed. The cell lyset were analyzed for L-asparaginase-II protein on 12% polyacrylamide gel (figure 6.20).

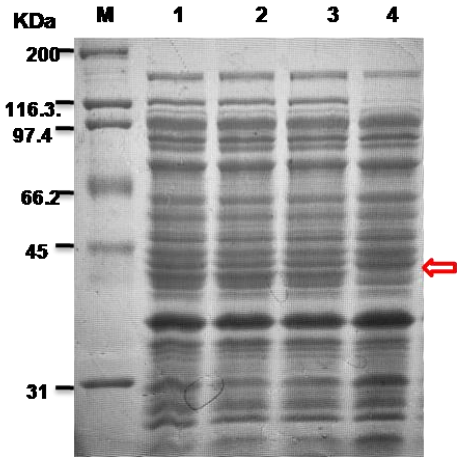


Figure 6.20. Expression of L-asparaginase type –II protein. ***M:*** *Broad range protein marker Biorad.* ***1:*** *0.4 mM IPTG induced cell lysate of L-asparaginase clone.* ***2:*** *0.5 mM IPTG induced cell lysate* ***3:*** *0.6 mM IPTG induced cell lysate* ***4:*** *uninduced cell lysate.*

6.3.4. Phylogenetic relation of cloned L-apapsraginase-II:

Cloned L-asparaginase-II gene was sequenced and from the sequence information a phylogenetic tree was constructed with the closest protein homologs of L-asparaginase. The phylogenetic tree indicated that the cloned gene was identical to L-asparaginase type- II protein of *Ralstonia* sp. It showed significant difference from asparaginase protein of *Ralstonia solanacearum, Ralstonia syzygii*, and *Burkholderia* sp (figure 6.21).

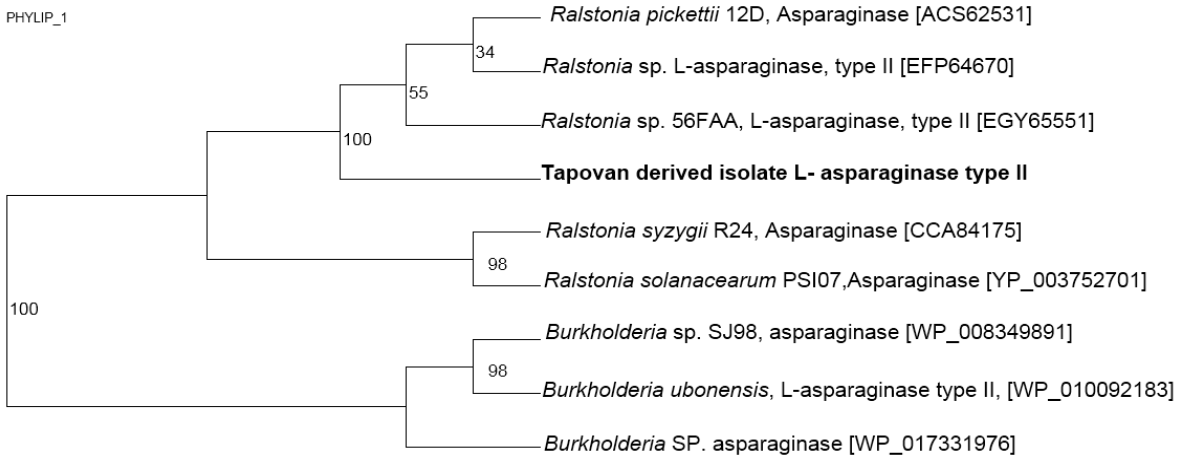


Figure 6.21. Phylogetic tree of L- asparaginase from Tapaovan derived isolate. *Phylogenetic tree of cloned L-asparaginase-II showed identity with Ralstonis sp., Ralstonia picketti, and was phylogenetically distinct from Burkholderia sp.*

6.4. Discussion:

In the present study, the cloning and characterization of Cel5A endoglucanase from metagenomic DNA preparation of Tapovan hot spring mat is reported. The greatest challenge of the study includes heterologous expression of metagenomic DNA derived from a thermophilic environment. The use of *Escherichia coli* BL-21 CodonPlus cells with extra copies of t-RNA for minimizing the codon biasness made the study relatively less complicated [142]. A 1.5 kb fragment corresponding to endoglucanase gene was cloned in an expression vector. The overexpressed TM-Cel5A protein had a 17 bp signal sequence which allowed it to secreted out of the cell into the culture medium. Amino acid sequence comparison of TM-Cel5A endoglucanase with endoglucanase of *Bacillus* sp. and *Bacillus licheniformis* showed its significance difference. It can be inferred from the amino acid study of TM-Cel5A that it contains a catalytic domain for glycosyl hydrolase family 5 (GH5).

Use of highly sensitive Congo red based agar plate assay made the screening feasible for the detection of beta endo-glucanase active clones [161]. Congo red is an indicator dye, which binds strongly to β-1,4-linked D-glucopyranosyl units of cellulose fiber [178]. Hydrolysis pattern of CMC around the positive clones was readily visualized by using CMC-congo red indicator plates. TM-Cel5A showed maximum activity at temperature 60 °C and pH 8.0. The enzyme was active at a wide temperature range. TM-Cel5A showed 60% active at pH 4.0, indicating its wide range of pH stability. TM-Cel5A protein had the ability to hydrolyze a wide variety of substrates. TM-Cel5A was able to hydrolyze beta D- glucan (beta 1,4 linkage) and CMC (beta 1,4 linkage) more efficiently. TM-Cel5A protein had negligible activity against insoluble avicel (commonly acted by exo-glucanase). The use of glucan (Beta- D- glucan) by TM-Cel5A indicates its strong cellulolytic activity as it is a storage form of natural cellulose. Beta- D- glucan polysaccharide are found in plant, baker's yeast, certain fungi and mushrooms [163]. Due to extraordinary characteristics of TM-5A, it could be a useful enzyme for industrial applications. The enzyme could also be used to obtain fermentable form of sugars and bioethanol. To the best of our knowledge this is the first report of cloning and characterization of beta endoglucanase gene from Himalayan geothermal spring metagenomic DNA.

In the second part of the study, L-asparaginase protein was cloned from a strain of *Ralstonia* sp. This is the first report of cloning and expression of L-asparaginase gene from *Ralstonia* sp. Amino acid sequence analysis indicated that the cloned gene was identical to L-asparaginase type

II. Many members of *Ralstonia* sp. like *Ralstonia solanacearum*, *Ralstonia eutropha*, *Ralstonia pickettii* are known for harboring the gene responsible asparaginase activity. Only sequence based information of asparaginase gene from different species of *Ralstonia* sp. is available in literature. Detailed study is yet to be done.

CHAPTER 7

RESULTS AND DISCUSSION-IV

Real-time PCR approach to study genes responsible for transporters and drug resistance in eDNA isolated from Tapovan

7.1. Real-time PCR based detection (C_T value calculation) of genes responsible for efflux of heavy metals and antibiotic resistance in Tapovan samples:

Real-time PCR based detection for transporters and antibiotic resistance gene was carried out for Tapovan metagenomic DNA. As studied from the Inductively Coupled Plasma Mass Spectroscopy (ICPMS) in chapter-4, the heavy metals were beyond their normal concentration in Tapovan spring water. Tapovan spring also has high load of sulphur ion with poor nutrient availability. Microorganisms in this stressful environment are expected to have developed adaptation mechanisms for survival. CusCFBA family transporters are known for their ability to efflux out heavy metals like copper and silver [96]. Similarly, AcrB a member of Resistance Nodulation cell Division (RND) family transporters remains overexpressed in nutrient deficient conditions [100]. These transporters often play the dual role for efflux of other amphiphilic substances including antibiotics. Therefore, a real-time PCR based approach was adopted for demonstration of transporter genes and antibiotic resistance genes in Tapovan environmental DNA (eDNA). In contrast, Tattapani spring is rich in nutrients and heavy metals are also comparatively low in concentration. Hence Tapovan eDNA was studied for the real-time based transporter gene detection.

List of the primers used for real-time PCR of the study are given in the table 3.2. Real-time PCR was carried out with the purified eDNA of Tapovan spring. Real-time PCR based threshold value (C_T) calculated for efflux pumps genes *acrB*, *acrD* and *cusA* were found to be 19.1, 18.09 and 17.78 respectively. Similarly, the threshold value for glycopeptides, tetracycline and amino glycosides antibiotic resistance gene were found to be 21.21, 21.09 and 21.12 respectively. The relatively low C_T values indicate their significance amount in Tapovan eDNA. Genes responsible for macrolides and beta-lactam antibiotics were not found in the spring DNA (Table 6.1).

Table 7.1. C_T (Threshold value) detected through real-time PCR against various antibiotics and transporter genes.

Transporter and antibiotic resistance gene analyzed through qPCR	CT (Threshold concentration detected)
Beta-lactam antibiotic	NA
Macrolides	NA
Multidrug efflux pump (*acrB*)	**19.1**
cusA	**17.78**
acrD	**18.09**
Glycopeptide antibiotics	21.21
Tetracycline	21.09
Aminoglycosides	21.12
Plants (cpDNA)	24.29
Human (mtDNA)	27.74
E. coli plasmid pET23a-GFP	28.06

A graph was plotted taking C_T values of transporters, antibiotics resistance specific gene with number of PCR cycles. The C_T value were determined as number of cycle at which the fluorescence level crosed minimum limit of 30. The threshold value was found low for metal transporter *cusA* and antibiotics resistance of *acrD* and *acrB*, indicating their abundance in Tapovan metagenomic DNA (figure 7.1).

7.2. Assessment of purity of Tapovan eDNA:

While studying the transporters and antibiotic resistance genes in metagenomic DNA of Tapovan spring, it was utmost necessary to clarify if these genes have originated from anthropogenic or modern days laboratory sources. Hence, the purity of DNA was assessed using plant (cpDNA), Human (mtDNA) and *E. coli* plasmid pET23a-GFP specific primers in real-time PCR. Chloroplast DNA (cpDNA) specific primer were used for assessment of plant originated contaminations, Human mitochondrial DNA (mtDNA) for anthropogenic allied contaminations, *E. coli* plasmid pET23a-GFP for modern days laboratory contaminations. C_T values for plant (cpDNA), Human (mtDNA) and *E. coli* plasmid pET23a-GFP were found to be 24.29, 27.74 and 28.06 respectively (Table 6.1).

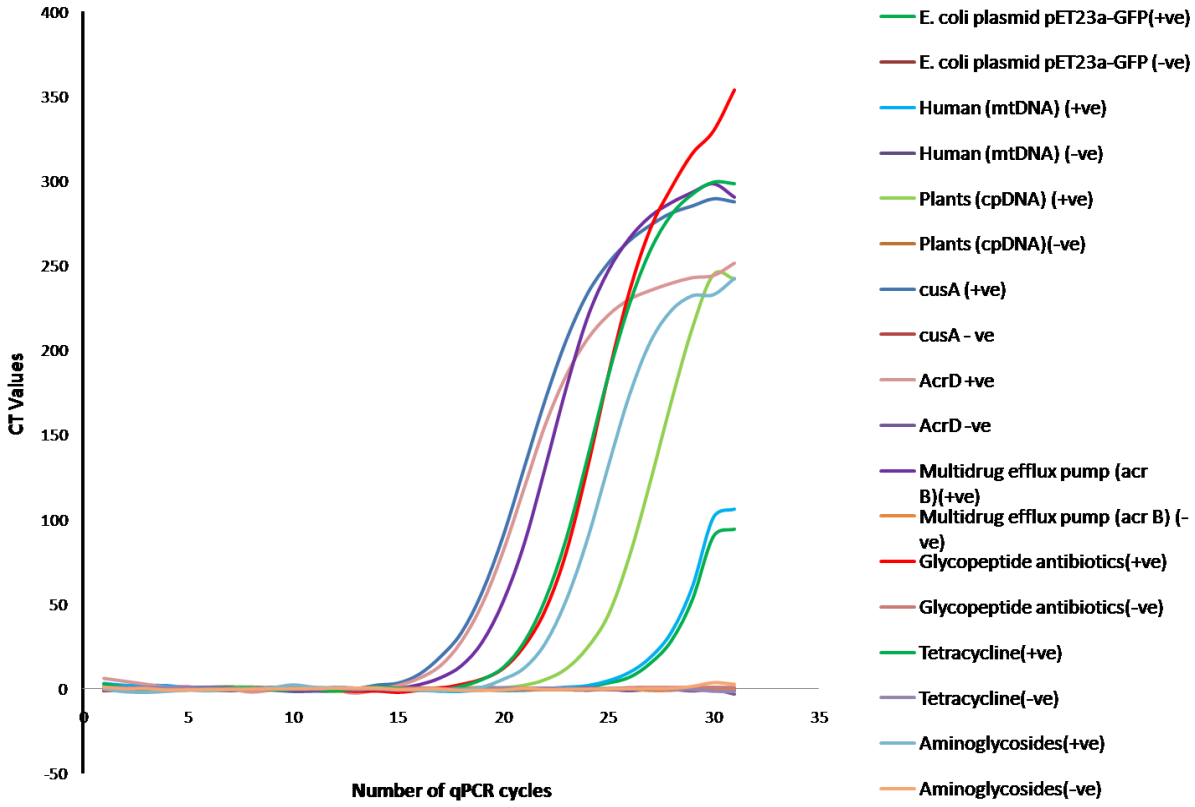


Figure 7.1. Determination of threshold value (C_T) for transporter and antibiotic resistance gene in Tapovan DNA through metagenomic approach. *C_T values were determined with positive and negative controls for transporters and antibiotic resistance genes.*

The C_T values indicate that there may be negligible contamination with plant originated DNA but it was free from anthropogenic and modern laboratory contamination. Again the origin of transporter and antibiotic resistance genes can be easily ruled out from plant sources.

7.3. Cloning and sequencing of *acrB* gene from Tapovan eDNA:

Purified environmental DNA (eDNA) from Tapovan spring was used for PCR amplification using acrB primer set. A 160 bp gene fragment corresponding to *acrB* gene was visualized on the gel (figure 7.2). *acrB* amplified gene was gel purified and ligated into pTZ57R/T vector and cloned using *Escherichia coli* DH5α host. Nucleotide sequence for cloned *acrB* gene was determined.

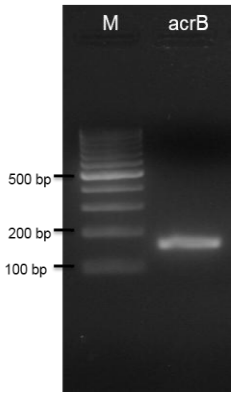


Figure 7.2. Amplification of *acrB* gene. *Lane M: 100 bp molecular weight marker and amplified acrB gene of 160 bp was electrophoresed in 2% agarose gel*.

7.4. Phylogenetic affiliation of cloned acrB gene:

Cloned acrB gene sequence derived from Tapovan DNA was used to complete phylogenetic relationship.

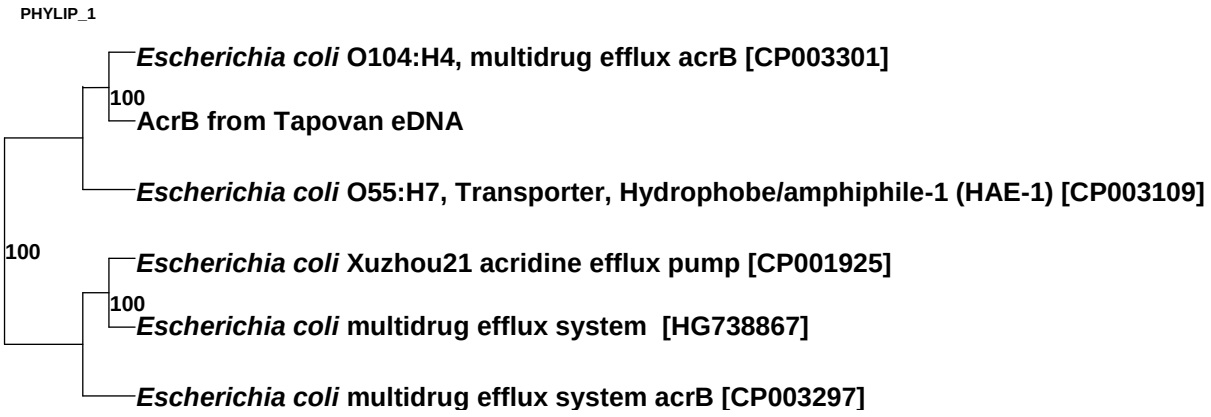


Figure 7.3. Phylogenetic relation of acrB. Phylogenetic tree showing relation of cloned *acrB* with closest relatives.

Phylogenetic relationship was obtained with the closest relative *acrB* gene of *Escherichia coli* O104:H4 and *Escherichia coli* O55:H7, transporter, Hydrophobe/amphiphile-1, *Escherichia coli* Xuzhou21 acridine efflux pump, *Escherichia coli* multidrug efflux system, *Escherichia coli* multidrug efflux system It was found that the gene showed identity with multidrug efflux protein acrB and with hydrophobe/amphiphile-1 (HAE-1) of *Escherichia coli* (figure 7.3).

7.5. Discussion:

As studied by ICPMS Tapovan spring showed significant amount of heavy metals like Al, As, Cu, Fe, B and Se. Extremely high temperature and high load of sulphur ions along with nutrient deprive condition creates a stressful environment in the spring site. Microbes residing in such adverse environmental conditions should have developed the adoptive mechanisms for survival. Membrane transporters responsible for the efflux of such heavy metals and toxic compounds from cell cytoplasm to outer environment play a major role in defending from unfavorable environmental conditions. In the present study, primers responsible for RND family of transporters and certain antibiotic resistance gene were used for detection of corresponding genes in Tapovan eDNA. CusA is such a transporter responsible for efflux of Copper and silver in microorganisms. CusA is a member of heavy metal efflux (HME) system [96]. AcrB is another transporter for transport of hydrophobe/amphiphile substances like acriflavin dye, detergents and multiple antibiotics. AcrB consist of hydrophobe/amphiphile efflux-1 (HAE-1) family of transporter. HAE-1 and HME system are the member of Resistance Nodulation cell Division (RND) super family of transporters involve in efflux of heavy metals $Zn^{2+,}$ Cu^{2+} [157]. *acrB* gene remains overexpressed in general stress like nutrient deficient condition [100]. Hence the presence of *acrB* gene in Himalayan thermal spring (Tapovan) is supported. The validity of the study is again supported from the purity assessment report of Tapovan eDNA.

The present study is a report of presence of efflux pump genes in extreme environments. While ruling out the contamination of these genes from anthropogenic, plant or modern research laboratory sources.

CONCLUSIONS

In the present study, the major focus was on microbial analysis through culture independent approach in high temperature environments, but the general idea illustrated here is applicable to other habitats too. Two major geothermal springs of the Indian Himalayan range were studied. The thermal springs with rare and low microbial load provided a challenge in isolation and purification of metagenomic DNA. The manual metagenomic DNA isolation and their purification using methods like electro-elution resulted in high quality and higher concentration of metagenomic DNA. Presently described environmental DNA purification methodology can be useful to obtain a superior quality genomic DNA from extreme environments supporting rare microbial populations.

In the first section of the study, analysis of microbial diversity in the hot sulphur springs at Tattapani and Tapovan geothermal area of Himalayas were carried out. So far, a comprehensive microbial diversity study of Himalayan geothermal springs has not been taken up. This is the first attempt for looking into the unculturable bacterial diversity of Himalayan geothermal springs by applying a metagenomic approach. The bacterial biodiversity information of the geothermal springs will serve as part of baseline data for comparing and contrasting the bacterial diversity of other geothermal springs in the Himalayan geothermal belt. Such data will be helpful for study of microbial ecology and evolution, biogeography, environmental sciences and biotechnology research. An understanding of such fragile ecosystems will also help the authorities to chalk out conservation programs for sustainable management of such unique resources. Coupling of sequence data with biological community function will help in better understanding of microbes from such ecosystem. Therefore, the community based functional analysis of microbes were carried out using Biolog EcoPlates. Such culture based sole carbon source utilization profile has helped in finding the correlation between diverse microbes present in the community and their function [173].

Alleviating the problems of heterologous expression and proper screening of metagenomic library, direct PCR based method was employed to assess functional potentiality of environmental DNA. The demonstrated thermostable beta endo-glucanase (TM-Cel5A) enzyme from high altitude Indian Himalayan geothermal spring DNA is a valuable enzyme showing significant difference in terms of sequence and functional analysis. It has the ability to hydrolyze crystalline cellulose and is active against glucan (a storage form of natural cellulose). Apart from its thermal

stability the enzyme is found to be stable over a wide pH range. TM-Cel5A no doubt is a promising enzyme for economical conversion of lignocellulolytic substances. It can be used for generation of fermentable sugars and bioethanol. The direct approach of such functional screening from metagenomic DNA can be employed for discovery of lignocellulolytic and other hydrolytic enzymes. The presently discussed functional metagenomic approach of direct PCR based screening will improve the strategies for quicker discovery of novel enzymes and secondary metabolites from yet uncultured microbes. The identified L-asparaginase enzyme from a bacterial isolate of Tapovan spring was identified as type-II group. The type-II asparaginase is commonly used in treatment of acute lymphoblastic leukemia (ALL). This is the first report of L-asparaginase from *Ralstonia* sp.

In the preceding section of our study, the real-time PCR based analysis of Himalayan thermal spring DNA revealed the presence of RND family efflux pumps. The most probable reason for the presence of AcrB, CusA transporters in hot spring environment is efflux of heavy metals and poor nutrient availability. The studied efflux pumps plays the dual role in efflux of certain of antibiotics hence presence of drug resistance gene is also supported. The contamination assessment assay identified that the eDNA was free from anthropogenic, and modern days laboratory contaminations.

Conclusively, the use of culture independent along with culture dependent approach for the study of microorganisms from these Himalayan thermal springs provided an illuminating knowledge about their community structure and function. These facts can be applied to microorganisms of other geothermal springs in Himalayas as well as across the globe. A huge number of enzymes, secondary metabolites with extraordinary properties from the microbial resource of Himalayan thermal springs can be expected. The microbial resource from these unique rare ecosystems need to be protected and require great attention for their conservation.

Future prospective:

Tattapani spring shows the presence of uncultured Verrucomicrobia, which are known to have bioremediation properties. Other thermophilic organisms identified in Tattapani spring could be the source of many enzymes and secondary metabolites. Similarly, Thermophilic *Aneurinibacillus* sp., *Bacillus thermoaerophilus, Anoxybacillus* sp. identified in the Tapovan mat are known to be source of a variety of valuable secondary metabolites and enzymes [111,116]. Therefore, detailed

functional studies of microorganisms from this high altitude hot springs is required for discovery of enzymes and secondary metabolites with novel properties. *Pseudoxanthomonas* sp. identified in Tapovan spring is a chemolithoautotrophic facultative anaerobic bacterium whos role in elemental transformations of S, Fe, NH_4, CH_4 can also be studied. This study could help in understanding the vital role in nutrient cycling and climate regulation in geothermal ecosystems. The endoglucanase identified from Tapovan eDNA can be optimized for high scale production. Enzymatic activity of endoglucanase can also be exploited for degradation of various natural substrates like rice straw, filter paper and molasses. The enzyme could also be used to obtain fermentable form of sugars and bioethanol. L-asparaginase enzyme demonstrated in the study can be optimized for over production. It can be studied for inhibition of acrylamide formation in starchy food processing. L-asparaginase can also be studied for inhibition of leukemic cell development in animal cell lines and in-vivo systems. The transporter identified in the metagenomic DNA isolated from Himalayan hot water spring can be further characterized to understand the relation between their origin and function.

REFERENCES

1. Aakvik, T., Degnes, K. F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Völker, U., Ellingsen, T. E. and Valla, S. 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(2):149-158 (2009).
2. Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L. and Polz, M. F. Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430(6999):551-554 (2004).
3. Aditiawati, P., Yohandini, H. and Madayanti, F. Microbial diversity of acidic hot spring (kawah hujan B) in geothermal field of kamojang area, west Java-Indonesia. *The open microbiology journal* 3:58 (2009).
4. Aires, J. R. and Nikaido, H. Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of *Escherichia coli*. *Journal of Bacteriology* 187(6):1923-1929 (2005).
5. Amann, R. I., Ludwig, W. and Schleifer, K.-H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews* 59(1):143-169 (1995).
6. Anderson-Carpenter, L. L., McLachlan, J. S., Jackson, S. T., Kuch, M., Lumibao, C. Y. and Poinar, H. N. Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology* 11(1):30 (2011).
7. Ashworth, L. A. and MacLennan, A. P. Comparison of the L-asparaginases from *Escherichia coli* and *Erwinia carotovora* as immunosuppressants. *Cancer research* 34(6):1353-1359 (1974).
8. Banik, J. J. and Brady, S. F. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proceedings of the National Academy of Sciences* 105(45):17273-17277 (2008).

9. Banik, J. J., Craig, J. W., Calle, P. Y. and Brady, S. F. Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. *Journal of the American Chemical Society* 132(44):15661-15670 (2010).
10. Bastida, F., Moreno, J., Nicolas, C., Hernandez, T. and Garcia, C. Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *European Journal of Soil Science* 60(6):845-859 (2009).
11. Beales, N. Adaptation of microorganisms to cold temperatures, weak acid preservatives, low pH, and osmotic stress: a review. *Comprehensive Reviews in Food Science and Food Safety* 3(1):1-20 (2004).
12. Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., Knight, R. and Fierer, N. The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biology and Biochemistry* 43(7):1450-1455 (2011).
13. Berthelet, M., Whyte, L. G. and Greer, C. W. Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. *FEMS Microbiology Letters* 138(1):17-22 (1996).
14. Bhat, M. Cellulases and related enzymes in biotechnology. *Biotechnology advances* 18(5):355-383 (2000).
15. Bhat, M. and Bhat, S. Cellulose degrading enzymes and their potential industrial applications. *Biotechnology advances* 15(3):583-620 (1997).
16. Bodelier, P. L. Toward understanding, managing, and protecting microbial ecosystems. *Frontiers in microbiology* 2 (2011).
17. BOEGL, Z. C.-E. K.-P. Impact of L-asparaginase on acrylamide content in potato products. *Journal of Food and Nutrition Research* 45(4):141-146 (2006).
18. Bohannan, B. J. and Hughes, J. New approaches to analyzing microbial biodiversity data. *Current Opinion in Microbiology* 6(3):282-287 (2003).
19. Brady, S. F. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nature protocols* 2(5):1297-1305 (2007).
20. Brady, S. F., Chao, C. J., Handelsman, J. and Clardy, J. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Organic letters* 3(13):1981-1984 (2001).

21. Brady, S. F. and Clardy, J. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *Journal of the American Chemical Society* 122(51):12903-12904 (2000).
22. Braker, G., Fesefeldt, A. and Witzel, K.-P. Development of PCR primer systems for amplification of nitrite reductase genes (nirK and nirS) to detect denitrifying bacteria in environmental samples. *Applied and environmental microbiology* 64(10):3769-3775 (1998).
23. Braker, G., Zhou, J., Wu, L., Devol, A. H. and Tiedje, J. M. Nitrite Reductase Genes (nirK andnirS) as Functional Markers To Investigate Diversity of Denitrifying Bacteria in Pacific Northwest Marine Sediment Communities. *Applied and environmental microbiology* 66(5):2096-2104 (2000).
24. Breznak, J. A. Intestinal microbiota of termites and other xylophagous insects. *Annual Reviews in Microbiology* 36(1):323-323 (1982).
25. Bui, D. M., Gregan, J., Jarosch, E., Ragnini, A. and Schweyen, R. J. The bacterial magnesium transporter CorA can functionally substitute for its putative homologue Mrs2p in the yeast inner mitochondrial membrane. *Journal of Biological Chemistry* 274(29):20438-20443 (1999).
26. Burnett, P.-G. G., Handley, K., Peak, D. and Daughney, C. J. Divalent metal adsorption by the thermophile *Anoxybacillus flavithermus* in single and multi-metal systems. *Chemical Geology* 244(3):493-506 (2007).
27. Campbell, H., Mashburn, L., Boyse, E. and Old, L. Two L-Asparaginases from *Escherichia coli* B. Their Separation, Purification, and Antitumor Activity. *Biochemistry* 6(3):721-730 (1967).
28. CAPIZZI, R. L., Bertino, J., Skeel, R., Creasey, W., Zanes, R., Olayon, C., Peterson, R. and Handschumacher, R. L-asparaginase: clinical, biochemical, pharmacological, and immunological studies. *Annals of internal medicine* 74(6):893-901 (1971).
29. Caspers, M. P., Boekhorst, J., Abee, T., Siezen, R. J. and Kort, R. Complete genome sequence of *Anoxybacillus flavithermus* TNO-09.006, a thermophilic sporeformer associated with a dairy-processing environment. *Genome announcements* 1(1) (2013).
30. Chauhan, N. S., Ranjan, R., Purohit, H. J., Kalia, V. C. and Sharma, R. Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library. *FEMS microbiology ecology* 67(1):130-139 (2009).

31. Chen, K. and Pachter, L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS computational biology* 1(2):e24 (2005).
32. Chen, M.-Y., Tsay, S.-S., Chen, K.-Y., Shi, Y.-C., Lin, Y.-T. and Lin, G.-H. *Pseudoxanthomonas taiwanensis* sp. nov., a novel thermophilic, N2O-producing species isolated from hot springs. *International journal of systematic and evolutionary microbiology* 52(6):2155-2161 (2002).
33. Chistoserdova, L. Recent progress and new challenges in metagenomics for biotechnology. *Biotechnology letters* 32(10):1351-1359 (2010).
34. Chung, E. J., Lim, H. K., Kim, J.-C., Choi, G. J., Park, E. J., Lee, M. H., Chung, Y. R. and Lee, S.-W. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Applied and environmental microbiology* 74(3):723-730 (2008).
35. Cohan, F. M. What are bacterial species? *Annual Reviews in Microbiology* 56(1):457-487 (2002).
36. Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R. and Wittwer, P. Metagenomic gene discovery: past, present and future. *Trends in biotechnology* 23(6):321-329 (2005).
37. Craig, J. W., Chang, F.-Y., Kim, J. H., Obiajulu, S. C. and Brady, S. F. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and environmental microbiology* 76(5):1633-1641 (2010).
38. D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W., Schwarz, C., Froese, D., Zazula, G., Calmels, F. and Debruyne, R. Antibiotic resistance is ancient. *Nature* 477(7365):457-461 (2011).
39. Daniel, R. The metagenomics of soil. *Nature Reviews Microbiology* 3(6):470-478 (2005).
40. Debbab, A., Aly, A. H., Lin, W. H. and Proksch, P. Bioactive compounds from marine bacteria and fungi. *Microbial Biotechnology* 3(5):544-563 (2010).
41. Dees, P. M. and Ghiorse, W. C. Microbial diversity in hot synthetic compost as revealed by PCR-amplified rRNA sequences from cultivated isolates and extracted DNA. *FEMS microbiology ecology* 35(2):207-216 (2001).
42. Degrange, V. and Bardin, R. Detection and counting of Nitrobacter populations in soil by PCR. *Applied and environmental microbiology* 61(6):2093-2098 (1995).

43. DeSantis, T. Z., Brodie, E. L., Moberg, J. P., Zubieta, I. X., Piceno, Y. M. and Andersen, G. L. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microbial ecology* 53(3):371-383 (2007).

44. Doi, R. Cellulases of mesophilic microorganisms. *Annals of the New York Academy of Sciences* 1125(1):267 (2008).

45. Doolittle, W. F. and Zhaxybayeva, O. Metagenomics and the units of biological organization. *BioScience* 60(2):102-112 (2010).

46. Dunfield, P. F., Yuryev, A., Senin, P., Smirnova, A. V., Stott, M. B., Hou, S., Ly, B., Saw, J. H., Zhou, Z. and Ren, Y. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* 450(7171):879-882 (2007).

47. Ekkers, D. M., Cretoiu, M. S., Kielak, A. M. and van Elsas, J. D. The great screen anomaly—a new frontier in product discovery through functional metagenomics. *Applied microbiology and biotechnology* 93(3):1005-1020 (2012).

48. Elshahed, M. S., Senko, J. M., Najar, F. Z., Kenton, S. M., Roe, B. A., Dewers, T. A., Spear, J. R. and Krumholz, L. R. Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. *Applied and environmental microbiology* 69(9):5609-5621 (2003).

49. Ferrer, M., Golyshina, O., Beloqui, A. and Golyshin, P. N. Mining enzymes from extreme environments. *Current Opinion in Microbiology* 10(3):207-214 (2007).

50. Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A., Strompl, C., Elborough, K., Jarvis, G. and Neef, A. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology* 7(12):1996-2010 (2005).

51. Fidler, I. J., Montgomery, P. C. and Cesarini, J. P. Modification of surface topography of lymphocytes by L-asparaginase. *Cancer research* 33(12):3176-3180 (1973).

52. Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y., Mendieta, K. S., Jones, K. L., Ocampo, A., Rudd, K. E. and Storz, G. Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Molecular microbiology* 70(5):1076-1093 (2008).

53. Gabor, E. M., De Vries, E. J. and Janssen, D. B. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology* 6(9):948-958 (2004).

54. Gadd, G. M. and Griffiths, A. J. Microorganisms and heavy metal toxicity. *Microbial ecology* 4(4):303-317 (1977).
55. Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P. and Thompson, F. L. Re-evaluating prokaryotic species. *Nature Reviews Microbiology* 3(9):733-739 (2005).
56. Gilkes, N., Henrissat, B., Kilburn, D., Miller, R. and Warren, R. Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiological reviews* 55(2):303-315 (1991).
57. Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., Clardy, J., Goodman, R. M. and Handelsman, J. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and environmental microbiology* 68(9):4301-4306 (2002).
58. Giovannoni, S., Schabtach, E. and Castenholz, R. *Isosphaera pallida*, gen. and comb. nov., a gliding, budding eubacterium from hot springs. *Archives of microbiology* 147(3):276-284 (1987).
59. Godiska, R., Patterson, M., Schoenfeld, T. and Mead, D. A. Beyond pUC: Vectors for cloning unstable DNA. *DNA sequencing: optimizing the process and analysis* 1:55-76 (2005).
60. Goh, K. M., Kahar, U. M., Chai, Y. Y., Chong, C. S., Chai, K. P., Ranjani, V., Illias, R. M. and Chan, K.-G. Recent discoveries and applications of *Anoxybacillus*. *Applied microbiology and biotechnology* 97(4):1475-1488 (2013).
61. Gomez, E., Ferreras, L. and Toresani, S. Soil bacterial functional diversity as influenced by organic amendment application. *Bioresource Technology* 97(13):1484-1489 (2006).
62. Gonzalez-Toril, E., Llobet-Brossa, E., Casamayor, E., Amann, R. and Amils, R. Microbial ecology of an extreme acidic environment, the Tinto River. *Applied and environmental microbiology* 69(8):4853-4865 (2003).
63. Grumbine, R. E. and Pandit, M. K. Threats from India's Himalaya Dams. *Science* 339(6115):36-37 (2013).
64. Gupta, N., Vohra, R. and Hoondal, G. A thermostable extracellular xylanase from alkalophilic *Bacillus* sp. NG-27. *Biotechnology letters* 14(11):1045-1046 (1992).
65. Gupta, R., Govil, T., Capalash, N. and Sharma, P. Characterization of a Glycoside Hydrolase Family 1 β-Galactosidase from Hot Spring Metagenome with

Transglycosylation Activity. *Applied biochemistry and biotechnology* 168(6):1681-1693 (2012).

66. Haaijer, S. C. M. Microbial mediators of the sulfur, nitrogen, and iron cycles in freshwater ecosystems: [Sl: sn] (2007).
67. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* 68(4):669-685 (2004).
68. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. and Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology* 5(10):R245-R249 (1998).
69. Hårdeman, F. and Sjöling, S. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS microbiology ecology* 59(2):524-534 (2007).
70. Hochstein, M. and Regenauer-Lieb, K. Heat generation associated with collision of two plates: The Himalayan geothermal belt. *Journal of Volcanology and Geothermal Research* 83(1):75-92 (1998).
71. Hong, S.-H., Bunge, J., Jeon, S.-O. and Epstein, S. S. Predicting microbial species richness. *Proceedings of the National Academy of Sciences of the United States of America* 103(1):117-122 (2006).
72. Hugenholtz, P., Pitulle, C., Hershberger, K. L. and Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *Journal of Bacteriology* 180(2):366-376 (1998).
73. Ingram, L., Gomez, P., Lai, X., Moniruzzaman, M., Wood, B., Yomano, L. and York, S. Metabolic engineering of bacteria for ethanol production. *Biotechnology and bioengineering* 58(2-3):204-214 (1998).
74. Innis, M. A., Myambo, K. B., Gelfand, D. H. and Brow, M. DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proceedings of the National Academy of Sciences* 85(24):9436-9440 (1988).
75. Iwai, S., Chai, B., Sul, W. J., Cole, J. R., Hashsham, S. A. and Tiedje, J. M. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *The ISME journal* 4(2):279-285 (2009).

76. Janssen, P. H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and environmental microbiology* 72(3):1719-1728 (2006).
77. Johnston, C. G. and Aust, S. D. Detection of Phanerochaete chrysosporium in soil by PCR and restriction enzyme analysis. *Applied and environmental microbiology* 60(7):2350-2354 (1994).
78. Joice, H. P., Mauricio, L. S., Jorge, L. N., Rodrigo, Z. N., Junio, C., Patricia, K. A., Zaira, B. H., Andreia, N. M., Juliana, H. S. and Maria, L. N. Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochemical Journal* 441(1):95-104 (2012).
79. Kang, S., Park, Y., Lee, J., Hong, S. and Kim, S. Production of cellulases and hemicellulases by *Aspergillus niger* KK2 from lignocellulosic biomass. *Bioresource Technology* 91(2):153-156 (2004).
80. Kanokratana, P., Chanapan, S., Pootanakit, K. and Eurwilaichitr, L. Diversity and abundance of Bacteria and Archaea in the Bor Khlueng hot spring in Thailand. *Journal of basic microbiology* 44(6):430-444 (2004).
81. Kapardar, R. K., Ranjan, R., Grover, A., Puri, M. and Sharma, R. Identification and characterization of genes conferring salt tolerance to *Escherichia coli* from pond water metagenome. *Bioresource Technology* 101(11):3917-3924 (2010).
82. Kauffmann, I. M., Schmitt, J. and Schmid, R. D. DNA isolation from soil samples for cloning in different hosts. *Applied microbiology and biotechnology* 64(5):665-670 (2004).
83. Keiblinger, K. M., Wilhartitz, I. C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., Riedel, K. and Zechmeister-Boltenstern, S. Soil metaproteomics–Comparative evaluation of protein extraction protocols. *Soil Biology and Biochemistry* 54:14-24 (2012).
84. Khalaf, M. Effect of the fractionation and immobilization on the sorption properties of humic acid: Universitätsbibliothek; (2003).
85. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. and Daniel, R. 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(3):1408-1416 (2003).
86. Kuhad, R. C., Gupta, R. and Singh, A. Microbial cellulases and their industrial applications. *Enzyme research* 2011 (2011).

87. Kumar, B., Trivedi, P., Kumar Mishra, A., Pandey, A. and Palni, L. M. S. Microbial diversity of soil from two hot springs in Uttaranchal Himalaya. *Microbiological research* 159(2):141-146 (2004).

88. Kumar, C. G., Malik, R. and Tiwari, M. Novel enzyme-based detergents: an Indian perspective. *Curr Sci* 75(12):1312-1318 (1998).

89. Lakay, F., Botha, A. and Prior, B. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *Journal of applied microbiology* 102(1):265-273 (2007).

90. Lee, S.-Y., Bollinger, J., Bezdicek, D. and Ogram, A. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. *Applied and environmental microbiology* 62(10):3787-3793 (1996).

91. Leelavathi, S., Gupta, N., Maiti, S., Ghosh, A. and Reddy, V. S. Overproduction of an alkali-and thermo-stable xylanase in tobacco chloroplasts and efficient recovery of the enzyme. *Molecular Breeding* 11(1):59-67 (2003).

92. Lefevre, F., Robe, P., Jarrin, C., Ginolhac, A., Zago, C., Auriol, D., Vogel, T. M., Simonet, P. and Nalin, R. Drugs from hidden bugs: their discovery via untapped resources. *Research in microbiology* 159(3):153-161 (2008).

93. Lewin, A., Wentzel, A. and Valla, S. Metagenomics of microbial life in extreme temperature environments. *Current opinion in biotechnology* (2012).

94. Lichtfouse, E., Chenu, C., Baudin, F., Leblond, C., Da Silva, M., Behar, F., Derenne, S., Largeau, C., Wehrung, P. and Albrecht, P. A novel pathway of soil organic matter formation by selective preservation of resistant straight-chain biopolymers: chemical and isotope evidence. *Organic Geochemistry* 28(6):411-415 (1998).

95. Lim, H. K., Chung, E. J., Kim, J.-C., Choi, G. J., Jang, K. S., Chung, Y. R., Cho, K. Y. and Lee, S.-W. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Applied and environmental microbiology* 71(12):7768-7777 (2005).

96. Long, F., Su, C.-C., Zimmermann, M. T., Boyken, S. E., Rajashankar, K. R., Jernigan, R. L. and Edward, W. Y. Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature* 467(7314):484-488 (2010).

97. López-García, P., Duperron, S., Philippot, P., Foriel, J., Susini, J. and Moreira, D. Bacterial diversity in hydrothermal sediment and epsilonproteobacterial dominance in

experimental microcolonizers at the Mid-Atlantic Ridge. *Environmental Microbiology* 5(10):961-976 (2003).

98. Lorenz, P. and Eck, J. Metagenomics and industrial applications. *Nature Reviews Microbiology* 3(6):510-516 (2005).

99. Louis, P. and Flint, H. J. Development of a semiquantitative degenerate real-time PCR-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples. *Applied and environmental microbiology* 73(6):2009-2012 (2007).

100. Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H. and Hearst, J. E. Genes acrA and acrB encode a stress-induced efflux system of *Escherichia coli*. *Molecular microbiology* 16(1):45-55 (1995).

101. Mackay, I. M. Real-time PCR in the microbiology laboratory. *Clinical Microbiology and Infection* 10(3):190-212 (2004).

102. Mackenzie, R., Pedrós-Alió, C. and Díez, B. Bacterial composition of microbial mats in hot springs in Northern Patagonia: variations with seasons and temperature. *Extremophiles* 17(1):123-136 (2013).

103. MacNeil, I., Tiong, C., Minor, C., August, P., Grossman, T., Loiacono, K., Lynch, B., Phillips, T., Narula, S. and Sundaramoorthi, R. Expression and isolation of antimicrobial small molecules from soil DNA libraries. *Journal of molecular microbiology and biotechnology* 3(2):301-308 (2001).

104. Magnani, D. and Solioz, M. How bacteria handle copper. In: *Molecular microbiology of heavy metals*ed.eds), p.pp. 259-285. Springer, (2007).

105. Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J. and Wade, W. G. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Applied and environmental microbiology* 64(2):795-799 (1998).

106. Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends in genetics* 24(3):133-141 (2008).

107. Mardis, E. R. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:387-402 (2008).

108. Margolles-Clark, E., Ihnen, M. and Penttilä, M. Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources. *Journal of biotechnology* 57(1):167-179 (1997).

109. Marteinsson, V. T., Hauksdóttir, S., Hobel, C. F., Kristmannsdóttir, H., Hreggvidsson, G. O. and Kristjánsson, J. K. Phylogenetic diversity analysis of subterranean hot springs in Iceland. *Applied and environmental microbiology* 67(9):4242-4248 (2001).

110. Martin, A. P. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Applied and environmental microbiology* 68(8):3673-3682 (2002).

111. Masomian, M., Rahman, R. N. Z. R. A., Salleh, A. B. and Basri, M. A unique thermostable and organic solvent tolerant lipase from newly isolated *Aneurinibacillus thermoaerophilus* strain HZ: physical factor studies. *World Journal of Microbiology and Biotechnology* 26(9):1693-1701 (2010).

112. Mathur, J., Bizzoco, R. W., Ellis, D. G., Lipson, D. A., Poole, A. W., Levine, R. and Kelley, S. T. Effects of abiotic factors on the phylogenetic diversity of bacterial communities in acidic thermal springs. *Applied and environmental microbiology* 73(8):2612-2623 (2007).

113. McRoberts, N., Sinclair, W., McPherson, A., Franke, A., Saharan, R., Malik, R., Singh, S. and Marshall, G. An assessment of genetic diversity within and between populations of Phalaris minor using ISSR markers. *Weed Research* 45(6):431-439 (2005).

114. Mok, W. W., Navani, N. K., Barker, C., Sawchyn, B. L., Gu, J., Pathania, R., Zhu, R. D., Brown, E. D. and Li, Y. Identification of a toxic peptide through bidirectional expression of small RNAs. *Chembiochem* 10(2):238-241 (2009).

115. Moré, M. I., Herrick, J. B., Silva, M. C., Ghiorse, W. C. and Madsen, E. L. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Applied and environmental microbiology* 60(5):1572-1580 (1994).

116. Namsaraev, Z., Babasanova, O., Dunaevsky, Y., Akimov, V., Barkhutova, D., Gorlenko, V. and Namsaraev, B. *Anoxybacillus mongoliensis* sp. nov., a novel thermophilic proteinase producing bacterium isolated from alkaline hot spring, Central Mongolia. *Microbiology* 79(4):491-499 (2010).

117. Nawani, N. and Kaur, J. Purification, characterization and thermostability of lipase from a thermophilic *Bacillus* sp. J33. *Molecular and cellular biochemistry* 206(1-2):91-96 (2000).
118. Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J. and Setlow, P. Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and Molecular Biology Reviews* 64(3):548-572 (2000).
119. Nies, D. H. Microbial heavy-metal resistance. *Applied microbiology and biotechnology* 51(6):730-750 (1999).
120. Nies, D. H. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS microbiology reviews* 27(2-3):313-339 (2003).
121. Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S. and Pritchard, J. K. Sequencing and analysis of Neanderthal genomic DNA. *Science* 314(5802):1113-1118 (2006).
122. Oettgen, H. F., Stephenson, P. A., Schwartz, M. K., Leeper, R. D., Tallal, L., Tan, C. C., Clarkson, B. D., Golbey, R. B., Krakoff, I. H. and Karnofsky, D. A. Toxicity of E. coli L-asparaginase in man. *Cancer* 25(2):253-278 (1970).
123. Oren, A. The dying Dead Sea: The microbiology of an increasingly extreme environment. *Lakes & Reservoirs: Research & Management* 15(3):215-222 (2010).
124. Paul, E. A. Soil microbiology, ecology and biochemistry: Academic press (2006).
125. Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A. and Versalovic, J. Metagenomic pyrosequencing and microbial identification. *Clinical chemistry* 55(5):856-866 (2009).
126. Pettit, R. K. Soil DNA libraries for anticancer drug discovery. *Cancer chemotherapy and pharmacology* 54(1):1-6 (2004).
127. Pidiyar, V. J., Jangid, K., Patole, M. S. and Shouche, Y. S. Studies on cultured and uncultured microbiota of wild Culex quinquefasciatus mosquito midgut based on 16S ribosomal RNA gene analysis. *The American journal of tropical medicine and hygiene* 70(6):597-603 (2004).
128. Pol, A., Heijmans, K., Harhangi, H., Tedesco, D., Jetten, M. and den Camp, H. 630 2007. Methanotrophy below pH1 by a new Verrucomicrobia species. *Nature* 450:874-631.
129. Poole, K. Multidrug resistance in Gram-negative bacteria. *Current Opinion in Microbiology* 4(5):500-508 (2001).

130. Porter, T. M. and Brian Golding, G. Are similarity-or phylogeny-based methods more appropriate for classifying internal transcribed spacer (ITS) metagenomic amplicons? *New Phytologist* 192(3):775-782 (2011).
131. Portillo, M., Sririn, V., Kanoksilapatham, W. and Gonzalez, J. Differential microbial communities in hot spring mats from Western Thailand. *Extremophiles* 13(2):321-331 (2009).
132. Purdy, K., Embley, T., Takii, S. and Nedwell, D. Rapid Extraction of DNA and rRNA from Sediments by a Novel Hydroxyapatite Spin-Column Method. *Applied and environmental microbiology* 62(10):3905-3907 (1996).
133. Raha, S., Roy, S., Dey, S. and Chakrabarty, S. Purification and properties of an L-asparaginase from *Cylindrocarpon obtusisporum* MB-10. *Biochemistry international* 21(6):987 (1990).
134. Ranjan, R., Grover, A., Kapardar, R. K. and Sharma, R. Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochemical and biophysical research communications* 335(1):57-65 (2005).
135. Reese, E. T. and Mandels, M. Stability of the cellulase of *Trichoderma reesei* under use conditions. *Biotechnology and bioengineering* 22(2):323-335 (1980).
136. Reysenbach, A.-L., Wickham, G. S. and Pace, N. R. Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Applied and environmental microbiology* 60(6):2113-2119 (1994).
137. Riesenfeld, C. S., Schloss, P. D. and Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38:525-552 (2004).
138. Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A. and Minor, C. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and environmental microbiology* 66(6):2541-2547 (2000).
139. Roose-Amsaleg, C., Garnier-Sillam, E. and Harry, M. Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology* 18(1):47-60 (2001).
140. Rosenberg, E. Y., Ma, D. and Nikaido, H. AcrD of *Escherichia coli* is an aminoglycoside efflux pump. *Journal of Bacteriology* 182(6):1754-1756 (2000).

141. Rothschild, L. J. and Mancinelli, R. L. Life in extreme environments. *Nature* 409(6823):1092-1101 (2001).
142. Sakanyan, V., Snapyan, M., Ghochikyan, A. and Lecocq, F.m. Method of RNA and protein synthesis. EP Patent 1,412,504, (2008).
143. Satyanarayana, T., Raghukumar, C. and Shivaji, S. Extremophilic microbes: Diversity and perspectives. *Current science* 89(1):78-90p (2005).
144. Savitri, A. N. and Azmi, W. Microbial L-asparaginase: A potent antitumour enzyme. *Indian journal of Biotechnology* 2:184-194 (2003).
145. Schleifer, K. H. Classification of Bacteria and Archaea: Past, present and future. *Systematic and applied microbiology* 32(8):533-542 (2009).
146. Schloss, P. D. and Handelsman, J. Biotechnological prospects from metagenomics. *Current opinion in biotechnology* 14(3):303-310 (2003).
147. Schloss, P. D. and Handelsman, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and environmental microbiology* 71(3):1501-1506 (2005).
148. Schloss, P. D. and Handelsman, J. Toward a census of bacteria in soil. *PLoS computational biology* 2(7):e92 (2006).
149. Schwab, H. Principles of genetic engineering for *Escherichia coli*. *Biotechnology Set, Second Edition*:372-425 (1993).
150. Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S. and Huttenhower, C. Computational meta'omics for microbial community studies. *Molecular systems biology* 9(1) (2013).
151. Senesi, N. and Loffredo, E. Soil humic substances. *Biopolymers Online* (2001).
152. Shock, E. L., McCollom, T. and Schulte, M. D. Geochemical constraints on chemolithoautotrophic reactions in hydrothermal systems. *Origins of Life and Evolution of the Biosphere* 25(1-3):141-159 (1995).
153. Simon, C. and Daniel, R. Metagenomic analyses: past and future trends. *Applied and environmental microbiology* 77(4):1153-1161 (2011).
154. Singh, P., Raghukumar, C., Verma, P. and Shouche, Y. Fungal community analysis in the deep-sea sediments of the Central Indian Basin by culture-independent approach. *Microbial ecology* 61(3):507-517 (2011).

155. Smith, C. J. and Osborn, A. M. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS microbiology ecology* 67(1):6-20 (2009).
156. Solioz, M. and Vulpe, C. CPx-type ATPases: a class of P-type ATPases that pump heavy metals. *Trends in biochemical sciences* 21(7):237-241 (1996).
157. Su, C.-C., Yang, F., Long, F., Reyon, D., Routh, M. D., Kuo, D. W., Mokhtari, A. K., Van Ornam, J. D., Rabe, K. L. and Hoy, J. A. Crystal Structure of the Membrane Fusion Protein CusB from *Escherichia coli*. *Journal of molecular biology* 393(2):342-355 (2009).
158. Sukumaran, R. K., Singhania, R. R. and Pandey, A. Microbial cellulases-Production, applications and challenges. *Journal of Scientific and Industrial Research* 64(11):832 (2005).
159. Tahlan, K., Park, H. U. and Jensen, S. E. Three unlinked gene clusters are involved in clavam metabolite biosynthesis in Streptomyces clavuligerus. *Canadian journal of microbiology* 50(10):803-810 (2004).
160. Tappe, D., Dirks, J., Müller, R., Brederlau, J., Abele-Horn, M., Suerbaum, S. and Kurzai, O. Fatal *Clostridium tertium* septicemia in a nonneutropenic patient. *Journal of Infection* 50(1):76-80 (2005).
161. Teather, R. M. and Wood, P. J. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and environmental microbiology* 43(4):777-780 (1982).
162. Tebbe, C. C. and Vahjen, W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and environmental microbiology* 59(8):2657-2665 (1993).
163. Teeri, T. T. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in biotechnology* 15(5):160-167 (1997).
164. Tirawongsaroj, P., Sriprang, R., Harnpicharnchai, P., Thongaram, T., Champreda, V., Tanapongpipat, S., Pootanakit, K. and Eurwilaichitr, L. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *Journal of biotechnology* 133(1):42-49 (2008).
165. Tomme, P., Warren, R. and Gilkes, N. Cellulose hydrolysis by bacteria and fungi. *Advances in microbial physiology* 37:1-81 (1995).

166. Tringe, S. G. and Rubin, E. M. Metagenomics: DNA sequencing of environmental samples. *Nature reviews genetics* 6(11):805-814 (2005).

167. Tsai, Y.-L. and Olson, B. H. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and environmental microbiology* 58(7):2292-2295 (1992).

168. Varrot, A., Schülein, M. and Davies, G. J. Insights into ligand-induced conformational change in Cel5A from *Bacillus agaradhaerens* revealed by a catalytically active crystal form. *Journal of molecular biology* 297(3):819-828 (2000).

169. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E. and Nelson, W. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304(5667):66-74 (2004).

170. Verma, N., Kumar, K., Kaur, G. and Anand, S. L-asparaginase: a promising chemotherapeutic agent. *Critical Reviews in Biotechnology* 27(1):45-62 (2007).

171. Volossiouk, T., Robb, E. J. and Nazar, R. N. Direct DNA extraction for PCR-mediated assays of soil organisms. *Applied and environmental microbiology* 61(11):3972-3976 (1995).

172. Ward, D. M. A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1(3):271-277 (1998).

173. Weber, K. P., Grove, J. A., Gehder, M., Anderson, W. A. and Legge, R. L. Data transformations in the analysis of community-level substrate utilization data from microplates. *Journal of microbiological methods* 69(3):461-469 (2007).

174. Weber, K. P. and Legge, R. L. One-dimensional metric for tracking bacterial community divergence using sole carbon source utilization patterns. *Journal of microbiological methods* 79(1):55-61 (2009).

175. Weller, R., Bateson, M. M., Heimbuch, B. K., Kopczynski, E. D. and Ward, D. M. Uncultivated cyanobacteria, Chloroflexus-like inhabitants, and spirochete-like inhabitants of a hot spring microbial mat. *Applied and environmental microbiology* 58(12):3964-3969 (1992).

176. Wilson, I. G. Inhibition and facilitation of nucleic acid amplification. *Applied and environmental microbiology* 63(10):3741 (1997).

177. Winsley, T., van Dorst, J., Brown, M. and Ferrari, B. Capturing greater 16S rRNA gene sequence diversity within the domain Bacteria. *Applied and environmental microbiology* 78(16):5938-5941 (2012).

178. Wiwat, C. and Sillapee, J. Prospecting for Cellulase Enzymes Based on Sequences and Functional Screening from Metagenomic Libraries.

179. Wu, L., Wang, H., Zhang, Z., Lin, R., Zhang, Z. and Lin, W. Comparative metaproteomic analysis on consecutively Rehmannia glutinosa-monocultured rhizosphere soil. *PloS one* 6(5):e20611 (2011).

180. Wu, X., Jörnvall, H., Berndt, K. D. and Oppermann, U. Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochemical and biophysical research communications* 313(1):89-96 (2004).

181. Yeh, Y.-F., Chang, S. C.-y., Kuo, H.-W., Tong, C.-G., Yu, S.-M. and Ho, T.-H. D. A metagenomic approach for the identification and cloning of an endoglucanase from rice straw compost. *Gene* (2012).

182. Yun, J., Kang, S., Park, S., Yoon, H., Kim, M.-J., Heu, S. and Ryu, S. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and environmental microbiology* 70(12):7229-7235 (2004).

183. Zhou, J., Bruns, M. A. and Tiedje, J. M. DNA recovery from soils of diverse composition. *Applied and environmental microbiology* 62(2):316-322 (1996).