Gene resources from metal polluted soils using metatranscriptomic approach

A Thesis

Submitted in fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By:

Arkadeep Mukherjee

(901300003)

Under the supervision of

Dr. M. Sudhakara Reddy

(Professor)

Department of Biotechnology
Thapar Institute of Engineering & Technology, Patiala-147004

Punjab (India)

January, 2020

CERTIFICATE

Certified that the thesis "**Gene resources from metal polluted soils using metatranscriptomic approach"** which is submitted by **Mr. Arkadeep Mukherjee,** in fulfillment of the requirement for the award of the degree of **Doctor of Philosophy** in the Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, is a record of the candidate's own independent and original research work carried out by him under my supervision and guidance. The matter embodied in this thesis has not been submitted in part or full to any other University or Institute for the award of any degree.

(Dr. M. Sudhakara Reddy)
Supervisor
Professor and HOD
Department of Biotechnology
Thapar Institute of
Engineering & Technology
Patiala- 147004, Punjab,
India

DECLARATION

I hereby declare that the work presented in the thesis entitled "**Gene resources from metal polluted soils using metatranscriptomic approach"** in the fulfillment of the requirement for the award of the Degree of **Doctor of Philosophy** in the Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, is an authentic record of my own work during the period January 2015 to December 2019, under the supervision of **Dr. M. Sudhakara Reddy**, Professor, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala. The material embodied in this thesis has not been submitted in parts or full in any other university or institute for the award of any degree in India or Abroad.

Place: Patiala

Date: 21/8/2020

Arkadeep Mukherjee

Arkadeep Mukherjee
Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala, Punjab, India

ACKNOWLEDGEMENT

"The best view comes after the hardest climb"

In pursuit of this academic endeavor, I feel that I have been exceptionally fortunate because inspiration, guidance, direction, cooperation, love and care - all came my way in abundance. I feel nostalgic when I look back into my journey and find it difficult to put into words the never ending support and encouragement of everyone throughout these years.

First and foremost, I acknowledge the grace of ***Almighty GOD*** *who paved my way through all the problems and provided me with the strength required to make this possible.*

With a profound sense of gratitude, I acknowledge my PhD Supervisor, ***Prof. M. Sudhakara Reddy****, Department of Biotechnology, TIET, Patiala for giving me this wonderful opportunity of being a part of his vibrant research team. His pertinent guidance, patronage, prudent advises motivation, enthusiasm, immense knowledge and dedication for research rendered to me during my work, without which the present endeavor would not have achieved the same status. I thank his progressive, energizing and inspirational guidance, unconditional support, admirable dedication throughout my Ph.D. programme.*

I deem it a pleasure to express my regards and gratitude to ***Prof. Moushumi Ghosh*** *(Head, Department of Biotechnology, TIET) for her support, professional guidance and valuable advice in many different ways.*

I take this opportunity to thank ***Prof. Prakash Gopalan****, Director, TIET, Patiala, for providing all the amenities and guidance for the completion of my research work. I extend my heartfelt*

word of thanks to ***Prof. Rafat Siddique****, Dean (Research and Sponsored Projects), TIET, Patiala, for providing motivating research environment during my stay in the Institute.*

It is my privilege to thank ***Prof. N. Das, Prof. Sanjai Saxena*** *Department of Biotechnology, TIET and* ***Prof. Amjad Ali,*** *School of Chemistry and Biochemistry, TIET, for their constant support during the course of my study. I am profoundly thankful to* ***Dr. M Vasundhara****, Department of Biotechnology for her valuable suggestions for the progress of my work and* ***Prof. Anil Kumar****, Coordinator, TIFAC-CORE in Agro and Industrial Biotechnology, TIET, for providing facilities, guidance and providing a brilliant environment which is very conducive for research work. I am also profoundly thankful to* ***Prof. Laurence Fraissinet-Tachet****, Department of Microbial Ecology, Claude Bernard University Lyon I, France, for her valuable suggestions for the progress of my research work. I would take this opportunity to acknowledge to all the faculty members of DBT, TIET, Patiala for their constant encouragement for my research work. I am thankful to the office and laboratory staff of Department of Biotechnology for their cooperation and help.*

I express my deep regards to my colleagues ***Mr. Sumit Joshi****,* ***Mrs. Bharti Thakur****,* ***Mrs. Shikha Khullar*** *and* ***Mrs. Tanveer Kaur*** *for their great support and valuable suggestions during need of hour. I would specially thank* ***Mr. Sumit Joshi*** *for his unconditional support throughout these years and many fruitful and refreshing discussions. Sincere thanks to my senior* ***Dr. Vinit Meshram*** *and colleagues* ***Dr. Ashutosh Sharma*** *and* ***Dr. Rachan Karmakar*** *for encouraging me in every endeavor of my life. Life at TIET has been an enjoyable experience with them. I owe special thanks to my lab mates and friends* ***Navneet, Eetika, Fatima, Anu*** *and* ***Dr. Parul*** *for inspiring discussions and support. I would also like to thank all the project students who have worked in our Laboratory for their love and support. Most of the work would have been*

incomplete without the sincere support of Lab assistants. So, I owe a word of thanks to ***Mr. Soni Singh****,* ***Mr. Ram Newal (Babbanji)****,* ***Lallanji*** *and* ***Surinder****.*

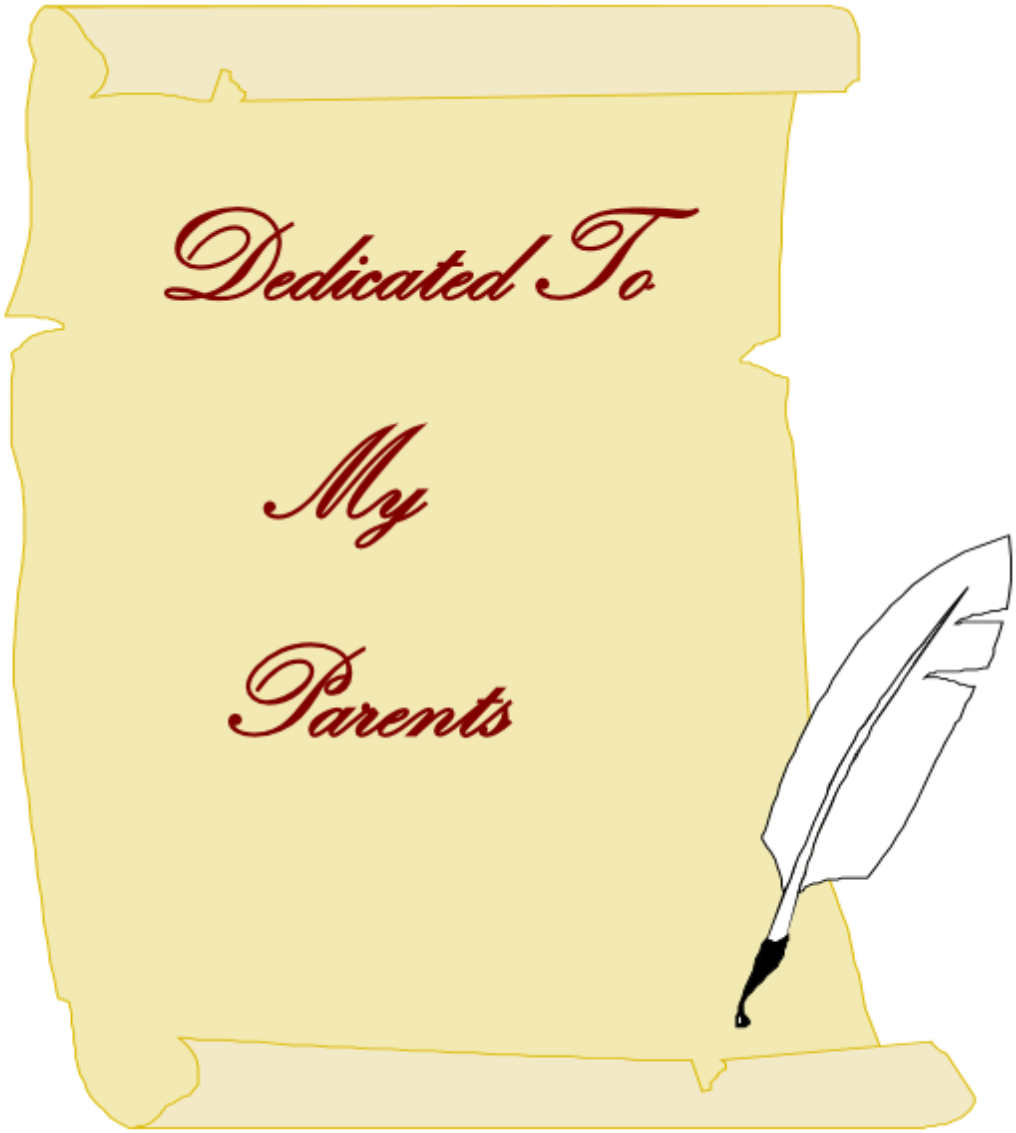
I find it difficult to express in words, my deepest sense of indebtedness towards my parents and all other family members. My family withstood all the pressure of this longish period of work with great determination and concern by my side. My indubitable gratitude is in order for the never-ending love and support of my father ***Mr. Swapan Kumar Mukhopadhyay*** *and mother* ***Mrs. Saraswati Mukherjee.*** *My father is my backbone and mother is my pillar of strength. I consider myself lucky to have such a supportive and loving family.*

Last but not the least, I wish to acknowledge all those, whose names have not figured here, but who helped me in any form during the period of my research work.

Date: 21/8/2020

Arkadeep Mukherjee

Place: Patiala **(Arkadeep Mukherjee)**


Dedicated To
My
Parents

List of Publications

Publications in Peer Reviewed Journals

- **Mukherjee A**, Yadav R, Marmeisse R, Fraissinet-Tachet L, Reddy MS (2019) Detoxification of toxic heavy metals by serine protease inhibitor isolated from polluted soil. International Biodeterioration and Biodegradation 143: 104718.
- **Mukherjee A**, Yadav R, Marmeisse R, Fraissinet-Tachet L, Reddy MS (2019). Heavy metal hypertolerant eukaryotic aldehyde dehydrogenase isolated from metal contaminated soil by metatranscriptomics approach. Biochimie 160: 183-192.
- **Mukherjee A,** Reddy MS (2020) Metatranscriptomics: An approach for retrieving novel eukaryotic genes from polluted soils and related environments. 3 Biotech (Manuscript accepted)
- **Mukherjee A**, Yadav R, Pandey AK, Marmeisse R, Fraissinet-Tachet L, Reddy MS Detoxification of metal toxicity by DHHC pamitoyl transferase-like protein isolated from contaminated soil. Ecotoxicology (Manuscript under review)

Participation in Conferences

- **Mukherjee A**, Thakur B, Reddy M.S (2016). Metatranscriptomic study of novel genes expressed in copper polluted soil and screening of environmental transcripts for potential heavy metal tolerant gene activity in ETFPP -2016 and 42nd Annual Meeting of the Mycological Society of India, Banaras Hindu University, Varanasi, India
- **Mukherjee A**, Reddy M.S (2016). Utilizing Metatranscriptomics as an unmediated approach for studying heavy metal tolerant gene activity in polluted soil in National Conference on Fungal Biotechnology Birla Institute of Scientific Research, Jaipur, India
- **Mukherjee A**, Reddy M.S (2017) The hunt for functional and novel genes expressed in heavy metal polluted land by metatranscriptomics approach in ETBWC 2017 , Biotech Research Society of India, Nagpur, India
- **Mukherjee A**, Reddy M.S (2017).Mining eukaryotic copper tolerant genes by soil metatranscriptomics in Fungal Biology: Recent Trends and Future Prospects, Mycological Society of India , Jammu, India

Table of contents

List of figures

List of tables

LIST OF ABBREVIATIONS

ORF	Open Reading Frame
pI	Isoelectric point
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
cDNA	Complementary Deoxyribonucleic Acid
BLAST	Basic local alignment search tool
SD-Ura	Synthetic defined medium without Uracil.
O.D.	Optical density
NMR	Nuclear magnetic resonance
ICP-MS	Inductively coupled plasma mass spectrometry
AAS	Atomic absorption spectroscopy
NAD	Nicotinamide adenine dinucleotide
Aa	Amino acid
MSA	Multiple sequence analysis
Wt.	Weight
OTU	Operational taxonomic unit
QIIME	Quantitative insights into microbial ecology
AMD	Acid mine drainage
NGS	Next-Generation Sequencing
PAT	Protein acyltransferase
5-FOA	5-Fluoroorotic acid
PCR	Polymerase chain reaction
ssDNA	Single stranded deoxyribonucleic acid
dsDNA	Double stranded deoxyribonucleic acid
mRNA	Messenger ribonucleic acid
rRNA	Ribosomal ribonucleic acid

CTAB	cetyl trimethylammonium bromide
LiAc	Lithium acetate
BIOM	Biological observation Matrix
RDP	Ribosomal database project
dNTP	deoxyribonucleotide triphosphate
MEGA	Molecular evolutionary genetics analysis
NCBI	National center for biotechnology information
PEG	Polyethylene glycol
Rpm	Revolutions per minute
NaOAc	Sodium acetate
DTT	Dithiothreitol
RT	Reverse transcriptase
U.V.	Ultraviolet

LIST OF SYMBOLS

%	Percentage
mg/ L	Milligram per litre
°C	Degree celcius
Psi	Pound per square inch
g	Gram
g/L	Gram per litre
µg /ml	Micrograms per milliliter
min	Minutes
Hrs	Hours
M	Molar
mM	Millimolar
µM	Micromolar
ml	Millilitre
µl	Microlitre
h	Hours
v/v	Volume by volume
w/v	Weight by volume
µm	Micrometer
mm	Millimeter
nm	Nanometer
Ng	Nanogram
mg /ml	Milligram per milliliter
mg /L	Milligram per litre
mg/kg	Milligram per kilogram
m	Meter
cm	Centimeter
L/ min	Litre per minute

ng	Nanogram
CFU/ml	Colony forming units per milliliter
mg	Milligram
mmole	Millimole
mg/dl	Milligrams per Decilitre
g/l	Gram per Litre
kHz	Kilohertz
pF/mm	Picofarad per millimeter
µF	Microfarad
M	Molar
U/L	Units per litre
mRL	meters Reduced Level
Bp	Base pair
Kb	Kilo base pair
kDa	Kilodalton
kV	Kilovolt
Ω	Omega (Resistance)

Abstract

A metatranscriptomic approach was employed to study the functional genes responsible for providing tolerance to toxic metals and also to isolate and clone them into appropriate hosts for their expression. In the first step, total RNA was extracted from metal polluted soils from India and France and cDNA was synthesized selectively from polyadenylated mRNA of eukaryotes utilizing the SMART (Switching Mechanism At 5' end of RNA) cDNA synthesis kit. cDNAs were fractionated into different sizes (A < 0.5 kb, B 0.5–1.0 kb, and C 1.0–4.0 kb) by 2-dimensional agarose gel electrophoresis. The size fractionated cDNAs were digested with SfiI restriction enzyme before ligation into yeast shuttle plasmid vector pFL61. The cDNA libraries A, B and C were initially screened for genes involved in providing tolerance to copper by functional complementation using the copper sensitive mutant *cup1*$^{\Delta}$ which is sensitive to copper at 150 μM on SD-Ura medium. More than $5x10^6$ clones were screened from library C and 155 clones were rescued from metal sensitivity. Some of the clones showing high tolerance to Cu were further sequenced and analyzed for their homology and characterized further. Apart from these, the clones rescued from contaminated soil from France were further characterized for its potential role in providing tolerance to Cu as well as other metals such as Cd, Zn and Co. Some of the genes characterized in this study are aldehyde dehydrogenase, serine protease inhibitor, DHHC palmitoyl transferase, maltose fermentation regulatory protein and chitin synthase. Functional complementation, growth response and intracellular metal accumulation studies were performed and it was found that the genes were able to confer tolerance to yeast mutants against toxic concentrations of wide range of metals like copper (150-1000 μM), cadmium (40-100 μM), zinc (10-13 mM) and cobalt (2-50 mM). The results suggested that the expression of these genes

is a part of the response of the soil eukaryotic microorganisms in relation to toxic metal stress. Characterization reveals that they are directly or indirectly involved in countering toxic metals and mitigating their harmful effects. Also, to establish a link between the organisms possessing these genes and the functional eukaryotic population in the polluted soil, molecular diversity study by Next-Generation Sequencing (NGS) was performed. Diversity of eukaryotic communities of the soil sampling sites was studied by sequencing 18S rDNA amplicons by high throughput sequencing on Illumina HiSeq platforms using suitable eukaryotic universal primers. Data analysis and taxonomic assignments revealed that phylum *Amoebozoa* and *Opisthokonta* were the major players of the metal polluted soil. This study demonstrated that metatranscriptomic approach has been particularly useful in understanding metal polluted soil ecology and can be massively fruitful to access metal tolerant species and genes whose products are important in areas of bioremediation, biosensing and other important biotechnological fields.

Chapter 1: Introduction

Soil is deemed to be the most important component of a terrestrial ecosystem. It can host a highly complex and dynamic biological environment (Bailly et al. 2007). Soils contain a tremendous diversity of microflora and microfauna present in a state of dynamic equilibrium (Roesch et al. 2007). To elucidate a soil ecosystem, one should start considering its biogeochemical composition comprising of physical (e.g. texture, density etc.), chemical (pH, available phosphorous, C:N ratio etc.) and biological (e.g. microbial biomass) properties (Dominati et al. 2010; Pointing and Belnap 2012). Maintenance and enhancement of these soil properties are the primary roles of the inhabiting microbial population that involves interplay with other members of the ecosystem. Numerous environmental parameters like carbon, nitrogen, mineral nutrients, ionic composition, moisture, temperature, pH, oxidation–reduction potential, interaction between microorganisms etc. can impact the activity and the dynamics of microbial polulation in soil (Borer et al. 2018). In turn, microbial communities strongly influence soil characteristics and are involved in important functions of the ecosystem such as decomposition of organic wastes, mobilization and geochemical cycling that are essential for sustenance of life forms. Hence, microhabitats are always in a dynamic state of existence and functioning (Tecon and Or 2017).

Studies pertaining to soil have ventured into the diversity and richness of species present in it and it has been assumed that the overall function of the soil can be maintained by conserving the microbial species richness. However, there are now increasing experimental evidences that the role of functional population is as important as that of the total microbial diversity of a soil ecosystem in maintaining soil properties (Nannipieri et al. 2014). The distribution of microbial

population in soil is discontinuous and intermittent as soil basically has less than adequate nutrients and energy sources for microbial population to survive. However, they can trap essential biological molecules such as proteins, carbohydrates nucleic acids that often become useful for microorganisms as they start to interact with soil surfaces. Often these are found as 'hotspots' or zones with enhanced biological activity like zones with accumulated organic matter around rhizospheres. This complex nature of soil makes it difficult to assess the community structure and determine microbial mediated processes. The major limitation while determining microbial mediated processes is the drawback, which is inherent within the methodology of the assay protocols. This is because all available assays in estimating the rate of metabolic activities, for example respiration and specific enzyme activities, do not have the provision of identifying that particular microbial species directly involved in the assayed process (Nannipieri et al. 2003). Thus, the most significant link is in between microbial diversity and functions occurring in soil, where novel approaches are required to study soil ecosystem. A wide range of novel approaches based on DNA and RNA study have to be incorporated in order to understand microbial ecology. This is because standard laboratory protocols for cultivation of microbes result in the isolation of less than 1% of the total microbial load present in soil at any given time (Torsvik et al. 1996). Important features of studying microbial diversity involve the complexity of interactions, number of functional processes and range of trophic levels. Hence, the success of studying microbial diversity lies in holistic approaches rather than by isolating the members of the ecosystem.

The study of soil microorganisms has been primarily focused on prokaryotes. The eukaryotes also form a diverse environment in soil and exist in a wide range of taxonomic groups (Marmeisse et al. 2017). Most of the eukaryotic microbes, being uncultivable, fail to come under the radar of researchers. Microbial eukaryotic species represent an unexplored source of

biocatalysts and metabolites (Daniel 2005). In-depth analysis of spruce and beech forest soil revealed dominance of fungi, metazoans and protists linked to enzyme activity for litter degradation for the preservation of terrestrial ecosystem fertility and making nutrients available by breaking down complex organic matter (Damon et al 2012). Major nitrogen assimilating and cell wall degrading enzymes (celluloses, pectin, lignin etc.) were reported. Also, numerous eukaryotic oligopeptide transporters which enable microorganisms to ferry small peptides generated through proteolysis in soil for further assimilation, were identified (Damon et al. 2011). Along with high-throughput sequencing technology, structure and functioning of sandy soil microflora were studied and less frequent taxa Crenarchaeota, were well represented by using rRNA-tags with the discovery of specific enzymes for ammonia oxidation and CO_2 fixation (Urich et al. 2008). The role of eukaryotes in improving soil fertility by litter degradation in forest soil has been extensively studied and different enzymatic processes for the degradation of lignocelluloses, hemicelluloses, polyphenols etc. were elucidated (Štursová et al. 2012). Eukaryotes have been associated with the important role of nutrient mobilization and maintaining soil fertility where a basidiomycete with stong acid phosphatase activity was confirmed in the organic rich layer of forest soil (Kellner et al 2011). The role of eukaryotes in heavy metal polluted soil provided a profitable field of interest for the research community and for the development of bioremediation strategies. The emergence of eukaryotic Metallothioneins (MT), a family of ubiquitous cysteine-rich proteins having the potential to chelate toxic metal ions and form metal-protein complexes, have been studied extensively in soil ecosystem (Blindauer 2013; Capdevila and Atrian 2011). However, eukaryotic MTs have been mostly studied in specific species such as vertebrates, plants, fungi, etc. Searching for a novel environmental gene from soil, capable of binding to toxic metal and conferring tolerance to soil

microorganisms to survive in metal polluted environments would therefore be a successful endeavor in the identification and bioremediation of heavy metals in soil. The studies conducted on soil involved the soil extracted mRNA converted to cDNA and was screened on the basis of their functional role of resisting toxic amounts of heavy metals (Cervantes and Gutierrez-Corona 1994). Members of the cysteine-rich proteins (CRP) were identified and those were highly similar and comparable to the already existing fungal metallothioneins (Ziller et al. 2017). Is has been well comprehended that soil microorganisms exhibit several times higher tolerance potential than plants. Several ectomycorrhizal (ECM) fungi and their association with plants have resulted in considerable lower uptake of metals in plants. The selection pressure of high metal pollution has triggered adaptive evolutionary changes in soil dwelling ECM species which include *Hebeloma sp., Pisolithus sp., Rhizopogon sp., Scleroderma sp.* and *Amanita muscaria* (Colpaert et al. 2011).

To estimate the soil diversity as well as the functionality of soil, various molecular methods have been adopted that are independent of cultivation of microbes and involve direct extraction and analysis of nucleic acids. Metagenomics is by far the fastest emerging approach for studying environmental microbial population without the need of cultivation. In this approach, soil DNA is directly extracted and sequences are analyzed with the help of reference metagenomic database. The other application of metagenomics is function based, that allows us to study novel metabolic processes and gene functions based on cloning of the metagenomic library followed by appropriate screening procedures. Once screened, the potential of the gene is analyzed for future biotechnological applications. The sequence based aspect of metagenomics allows us to predict the structure from the sequence data and further envisage its function. However, metagenomics does not give a picture of real-time activities occurring at a specific environment.

From examining the sequence data from metagenome sequencing programs, it reveals that eukaryotic sequences are outnumbered by prokaryotic ones and therefore, a metagenomic DNA library contains a very low proportion of clones of eukaryotic origin. The eukaryotes represent a major component of the ecosystem that are highly diverse in nature and are considered a rich source of biotechnological important genes. Furthermore, due to the frequent occurrence of introns in eukaryotic genes and also to the non-conservation of transcription regulatory elements between eukaryotes and prokaryotes, the genomic copy of a eukaryotic gene is unlikely to be expressed in a bacterial host (Lehembre et al. 2013).

Metatranscriptomics, as a subset of metagenomics, helps to encompass the entire microbial communities and provides valuable information of the functional potential of soil inhabiting microbes. This approach helps in depicting interplay of processes at a particular time and condition, resulting in important functions like nutrient cycling, mineralization and providing resistance towards toxic levels of pollutants (Cheeke et al. 2017). It involves the isolation and investigation of the RNA (the metatranscriptome) that is a representation of the expression profiles and gene regulation of complex microbial communities. It is based upon extraction of total environmental RNA instead of DNA, followed by different techniques to segregate the protein coding mRNA from the non-coding RNAs. This is particularly advantageous to study eukaryotic microbial communities. This involves the process of purification by affinity chromatography of the eukaryote specific polyadenylated messenger RNA from the total environmental RNA mixture. The poly-A mRNAs could be cloned in suitable expression vectors, when converted into cDNAs, (Bailly et al. 2007).

The approach of metatranscriptomics is particularly well suited to study the dynamic nature of polluted or stressed soil, as regulation of genes vastly affects dexterity of microbial communities.

Changes in soil ecosystem due to anthropogenic activities result in contamination with xenobiotic pollutants that drastically affect the soil inhabiting microorganisms. Contamination in soil may pose serious risks and hazards not only to microbes but also to the macro life forms through the food chain, direct ingestion, decrease in food quality via phytotoxicity and reduction in land usability for agricultural production causing food insecurity (Wuana and Okieimen, 2011). Out of all contaminants, toxic metal pollution has become a major problem (Abdu et al. 2017). Due to various activities like mining and smelting of metals, use of untreated wastewater to irrigate lands, use of pesticides, waste disposal etc., the soil has been contaminated with potentially toxic metals that are persistent in nature thus causing toxic effects especially on the soil microorganisms thus disturbing their normal functioning. Metals cause functional enzyme groups to be blocked, resulting in conformational modifications, denaturation and ultimately enzyme inactivation (Ochiai 2012; Qing et al. 2015). Moreover, toxic levels of metals in the soil can stimulate the formation of free radicals and reactive oxygen species, resulting in oxidative stress. This causes multiple disorders such as damage to photosynthetic pigments, ion leakage, oxidative DNA attack, redox imbalance, denature of cell structure and membrane etc. Metals can also enter into microorganisms by binding to the cell surface. Studies on metal toxicity have proved that toxic metals could be fatal to microorganisms as they are more prone to damage as compared to animals and plants growing on the same soil. As a result, this causes decline in overall microbial biomass of the soil ecosystem (Dietz et al. 1999; Jaishankar et al. 2014).

Maintaining community structure and diversity is a challenge faced by the residing microorganisms in polluted soil (Giller et al. 1998). Though, increase in toxicity or pollution has a negative impact on the diversity of community, the effect of ‘competitive exclusion’ would increase the probability of elimination of the dominant species promoting an increase in the less

diverse of usually less abundant organisms who have adapted well to the stressed soil. This is apparent from long term exposure of microbes to metal pollution (Zhang et al. 2009). To deal with the damaging effect of toxic metal pollution, microbes have adopted certain resistance or tolerance mechanisms. These mechanisms can be broadly described as cellular sequestration of metals into complexes by metal-chelating proteins, efflux of metals from cells, enzyme coupled transformations of metals and vacuolar compartmentalization. All these mechanisms have been studied in bacteria and also in eukaryotes like plants, fungi, yeast etc. and potential molecular targets have been identified. Apart from the mechanisms in direct metal detoxification, microbes have developed various other methods, where the role of several metabolically important proteins are apparent for controlling the cellular damage caused by toxic metals, leading to maintenance of overall viability (Panda et al. 2003; Gallego et al. 1996; Hall 2002).

In order to understand the relationship between ecosystem functions and microbial community structure, it is essential to assign appropriate metabolic states to the microbial groups e.g. active, growing and dormant (Blazewicz et al. 2013). Molecular surveys conducted by 'universal' and phylum specific primers to amplify small subunit of ribosomal gene (18S rDNA), represent diversity of eukaryotic microbial community. This method is based on the selective binding of the primers to the highly conserved regions of the 18S rDNA and its amplification. The variation of sequences in this gene contains phylogenetic information and can be deduced to understand taxonomic relations of the microbial species. Sequencing gene of 18S rRNA using Illumina's HiSeq® high throughput sequence analysis of cDNA as well as total soil DNA, gives us a global picture of the different eukaryotic taxonomic groups present in inhabiting soil. Metatranscriptome analysis enables us to find out the diversity of actively transcribing microbial population due to the analysis of total RNA in contrast to the metagenomic approach, where total

soil DNA is used. Amplicon sequencing is a favorable approach as far as comparative investigations in case of community response to a set of physicochemical parameters is concerned. Thus, this strategy can give important insights into the functioning of microbial communities in polluted soil.

The present study was designed to investigate the genetic resources of eukaryotes in copper and other metal polluted soils and also learn about their diversity. This allowed us to unravel novel biological molecules and their key functional roles in metal tolerance or resistance. Genes responsible for adaptation of eukaryotic life forms in metal polluted, stressful soil were characterized and novel biocatalysts were isolated from metal polluted environment, which could be implemented directly in biotechnological industries as biomarkers, for bioremediation and development of metal resistant crop plants. In this aspect, functional metatranscriptomics has been employed here to access genes of ecological importance.

Instead of culturing microorganisms, the process of metatranscriptomics employed in this study commenced from isolation of total RNA from polluted soil and synthesizing cDNA particularly from eukaryotic polyadenylated mRNA followed by preparation of sized cDNA library. cDNA library screening was carried out utilizing functional complementation of metal sensitive phenotypes of *Saccharomyces cerevisiae* mutants, which led to the isolation of several metal tolerant genes. Selected clones were tested for their specificity towards different toxic metals and the cDNA was further characterized. This enabled us to identify novel genes and compare different adaptation mechanisms in metal polluted soil.

Knowledge gap

It is difficult to isolate and cultivate soil eukaryotic microorganisms under laboratory conditions. Very few studies have included eukaryotic microorganisms possibly because they were in a

minority in the studied ecosystems or because they were physically excluded (by filtration or centrifugation on density gradients) from the biomass before DNA extraction. Hence genetic resources still largely remains untapped. There is also requirement of simultaneous information about the identity of taxa within a community and about functional processes performed.

Metagenomic study of eukaryotes faces several specific problems. The genome size of a eukaryote can be several orders of magnitude higher than the genome size of a bacterium. It is unlikely that a workable metagenomic library based on genomic DNA can capture a significant fraction of the gene content of a eukaryotic microbial community. Furthermore, the frequent presence of introns and lack of conservation of motifs in promoter sequences prevent expression of genomic copies of eukaryotic protein-coding genes.

Area of functional metatranscriptomics have not yet been substantially exploited and most of the work concentrates on the systematic sequencing of cDNAs prepared without cloning from either total RNA (mRNA+ rRNA) or total mRNA from the soil.

Hypotheses to be tested

1) A specific protocol for development and implementation for the generation of sized high quality cDNA library from very minute quantities of soil RNA.

2) A comprehensive and robust screening methodology for metal resistant genes from cDNA libraries by functional complementation assay using metal sensitive yeast mutants

3) Characterization of metal resistant genes and their role in metal tolerance and abiotic stress tolerance

4) To extensively prove that molecular diversity of the soil eukaryotes can be elucidated and studied soil through high throughput sequencing of ribosomal RNA and the data can be linked to organisms involved in metal tolerance.

Scope of the Study

Areas of functional metatranscriptomics have not yet been substantially exploited as most of the work concentrates on the systematic sequencing of cDNAs prepared without cloning from either total RNA (mRNA+ rRNA) or total mRNA from the soil. Furthermore, there are only a few studies that have included eukaryotic microorganisms probably because they were in a minority in the studied environments or because they were removed physically from the biomass during sample processing (filtration or centrifugation) prior to DNA extraction. Hence, genetic resources still largely remains untapped. Simultaneous information on taxa identity within a community and functional processes is also desirable.

The present study was aimed to isolate genes involved in providing tolerance to toxic metals and also to study the functional diversty of eukaryotic microorganisms living in the metal contaminated soil by metatranscriptomic approach. RNA isolation from the metal polluted soil was optimized and size fractionated cDNA libraries were constructed and screened for the genes involved in providing tolerance to toxic metals, by using metal sensitive mutants of *Saccharomyces cerevisiae*. Also, few cadmium tolerant clones previously obtained by Yadav et al. (2014) were characterized further with respect to different concentrations of copper, cadmium, zinc and cobalt and were also tested. Accordingly, the above mentioned study was designed and carried out with the following objectives:

1. Construction of cDNA libraries from RNA isolated from metal polluted samples
2. Identification and characterization of genes involved in metal tolerance
3. Functional diversity of eukaryotic microorganisms present in copper polluted soil

Chapter 2: Review of Literature

Soil is a complex ecosystem and hosts an incredible diversity of microbiota that is present in a state of dynamic equilibrium (Torsvik and Øvreås, 2002). These complex interactions between soil microorganisms are a result of competition and significant community based interactions. There is thus, little or no probability of a microbe to live a lonely life in such a habitat. The primary roles played by inhabiting microbes are maintaining and enhancing soil properties (Doran and Zeiss 2000; Lavelle et al. 2006). However, the properties of soil are heavily dependent on environmental factors and changes in nutrient sources, oxygen supply, humidity, temperature, etc., that can shift the ecological balance between the inhabiting microbes and cause one or many types of soil microbes to become temporarily dominant over the others. Soil microbial diversity can change drastically with the alteration of salinity, moisture, pH, redox potential and, most importantly, contamination with environmental pollutants (Zhou et al. 2002; Muller et al. 2001). Though physical properties of soil can be extensively studied, the changing dynamic nature, biodiversity and functionality can be studied by holistic approaches and not by isolating the members of the ecosystem. These techniques are expected to provide promising results when applied to soil contaminated with pollutants.

2.1 Toxicity associated with heavy metal accumulation in soil

Soil acts as a sink for accumulation of heavy metals. Unlike organic matter that is degraded rapidly and utilized by microorganisms, most heavy metals do not undergo any major degradation and can persist in nature for long periods. Heavy metals in soil are a serious threat as they can accumulate and concentrate in the food chain and pose hazards to humans and the ecosystem through direct ingestion or contact with contaminated soil, drinking of contaminated

ground water, reduction in food quality via phytotoxicity and reduction in land usability for agricultural production (Sarwar et al. 2017; Shahid et al. 2017).

Heavy metals refer to a group of chemical elements that are characterized by high density and high atomic mass. They have an important role in various chemical processes due to their ability to release electrons and take part in metabolic reactions. Hence, they are required in small quantities for the normal functioning of a cell. There are two types of heavy metals, essential (Fe, Zn, Cu, Mg, Mn, Mo and Ni) and non essential (Cd, Co, As, Pb, Hg, Cr, Ag and Se). Essential metals form micronutrients required by the plants and microbes in trace amounts for their improved metabolism, growth and development (Emamverdian et al. 2015) Deficiency of essential heavy metals may lead to stunted growth, improper metabolism and reduction in capacity of the body to remove toxins. Non essential heavy metals are known to be toxic even in small quantities. However, trace metals in excess can lead to severe form of heavy metal toxicity. Toxic amounts of heavy metals are accumulated in soil due to various anthropogenic activities like metalliferous mining and smelting of metals like copper, lead, zinc, mercury etc. Frequent irrigation of farm lands with untreated waste water from industrial effluents leads to toxic levels of cadmium, arsenic, lead, mercury etc. Also use of plant fertilizers causes deposition of harmful chemicals, rich in heavy metals like arsenic, cadmium, lead, selenium (Atafar et al. 2010; Mendes et al. 2006).

2.2 Coping with the detrimental implications of heavy metals in the environment

All life forms, in the presence of heavy metal pollution, show changes in metabolic and cellular processes in a variety of ways. Heavy metals hamper formation of proteins and their structural integrity as they attach with sulfhydryl groups, altering the active sites and disrupting protein functions (Hall, 2002). Increase in the concentration of redox active heavy metals such as Cu, Fe,

Cr and Mn, causes oxidative damage according to fenton type reactions and Haber-Weiss reactions (Jozefczak et al. 2012). This leads to production of reactive oxygen species (ROS) and O^{2-} (Superoxide free radicals) causing multiple disorders such as oxidative DNA damage, damage to photosynthetic pigments, redox imbalance, denatured cellular structure and induction of protease that may result in the activation of programmed cell death pathways (Gechev et al. 2006).

The potential of metals to cause detrimental effect on microbes depends on their bioavailability, concentration and the sensitivity of the organism (Gadd and de Rome, 1988). Some of the microbes have developed various mechanisms that allow them to tackle the problem of heavy metal contamination. Usually, they 'avoid' extracellular metals from entering into cell by reduced uptake, biosorption to cell walls and release of organic acids. However, in the second situation, microbes 'tolerate' the metal by intracellular chelation through specific ligands like metallothioneins, phytochelatins or compartmentation within vacuoles thus reducing the burden of heavy metals in the cytosol (Rajkumar et al. 2012). Figure 2.1 depicts typical cellular mechanisms of metal detoxification.

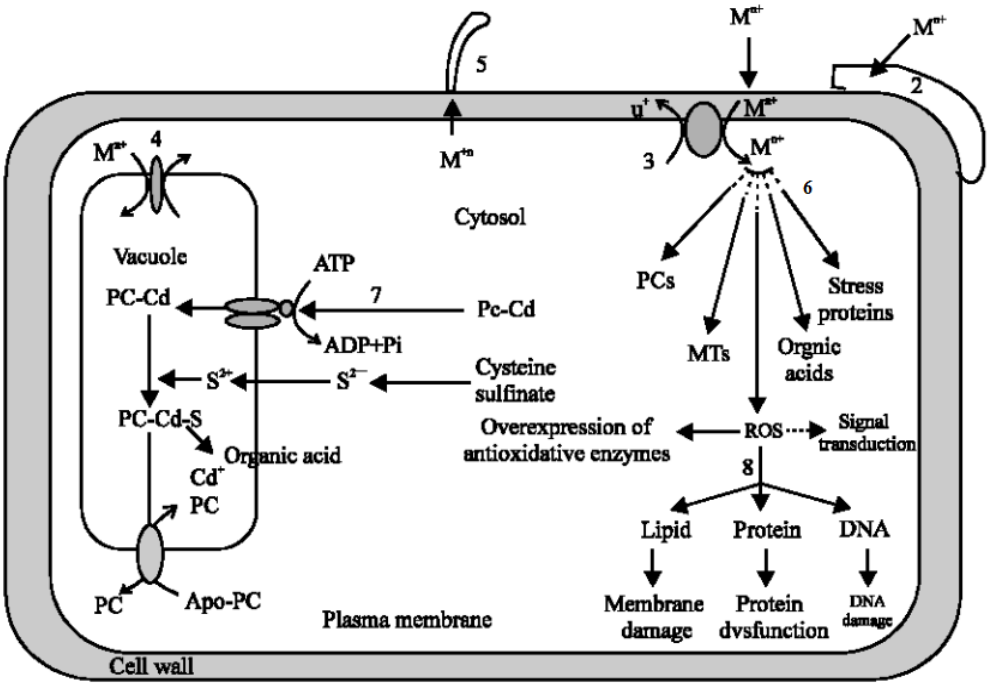


Figure 2.1: Schematic representation of typical cellular mechanisms potentially involved in metal tolerance in plants and eukaryotic microbes. 2. Ectomycorrhizal association restricts metal movement to roots. 3. Metal ions are taken up by plant roots through channel proteins and/or H^+-coupled carrier proteins. 4. From cytosol metals are transported into vacuoles, this is assisted by vacuolar proton fluxes. 5. Glandular trichomes and epidermal structures (hydropotes) sequester metals in leaves. 6. Metal ions in cytosol can be detoxified *via* these routes. 7. Mechanisms involved in Cd chelation and compartmentalization in the vacuole involving phytochelatins (PC). 8. Metal ions escaped from the complexation damage cellular macromolecules *via* the production of ROS (Adapted from Mishra and Dubey, 2006)

2.3 Extracellular mechanisms to counteract heavy metal toxicity

(i) Immobilization of heavy metals by organic acids: Extracellular chelation of metals by release of organic acids acting as ligands to form metal-ligand complex and diminish the undesired entry into the cell. Some eukaryotes and plants secrete organic acids into the rhizospheric region in

order to immobilize through precipitation or detoxify them by forming complexes (Adriaensen et al. 2005).

(ii) Heavy metal binding to extracellular components: Successful binding of metals to extracellular polysaccharides may play a role in tolerance in fungi associated with plant roots, and is considered to protect mycorrhizal roots from metal toxicity (Brown and Wilkins, 1985; Jones and Hutchinson, 1986; Marschner et al. 1998). Heavy metals are believed to bind to specific cation exchange sites in the mycelium of fungal species.

(iii) Heavy metal binding to cell wall components: The primary barrier protecting the fungal hyphae from the toxic effects of heavy metals is a polymeric cell wall containing glucan-chitin and galactosamine. Precipitation or binding to the cell wall components is also termed as biosorption and does not necessarily rely on the metabolic activity of the microbes. Specific sites for metal binding are hydroxyl, high affinity phosphate groups, free carboxyl, amino and mercapto groups. Thus biosorption of heavy metals to fungal structures may reduce the load of intracellular accumulation of heavy metals and their effect on cytoplasmic processes (Gadd and de Rome, 1988).

(iv) Heavy metal transport mechanisms: The involvement of various transport proteins capable of driving metals out from the cytoplasm to the exterior of the cells before they reach toxic levels and also allowing sequestration of metals into various intracellular compartments, have been suggested previously (Williams et al. 2000; Hall, 2002). Transport and regulation of essential elements like Cu and Zn and non essential elements like Cd have been proposed in ectomycorrhizal fungi. Cadmium transporters based on a Mg ATP driven glutathione *S*-conjugate is responsible for vacuolar sequestration. The yeast cadmium factor (*YCF1*) gene encodes a Mg ATP-energized glutathione *S*-conjugate transporter responsible for the vacuolar

sequestration of organic compounds after their *S*-conjugation with glutathione. Transport of toxic heavy metals across cell membrane have been studied in organisms where P-type heavy metal ATPases have been identified for the transport of metal ions like Cu^{2+}, Zn^{2+}, Cd^{2+}, Pb^{2+} (Ballatori, 2002).

2.4 Intracellular mechanisms to tackle heavy metal toxicity

Chelation and sequestration of metals into cellular vacuoles are important intracellular mechanisms to cope with heavy metal toxicity in the cytosol. Various metal specific proteins and peptides are responsible for metal chelation and transportation into vacuoles thus ameliorating the toxic effects of heavy metals. Many classes of peptides called metallothioneins and phytochelatins aid in the sequestration and thus are actively involved in the fate of heavy metals inside the cell (Coldsbrough, 1999).

2.4.1 Metallothioneins

Metallothioneins (MTs) are a family of cysteine-rich, low molecular weight (3500 to 14000 Da) proteins that have the capacity to bind to both physiological and xenobiotic heavy metals like Cd, Cu, Zn, Hg, Ag, Se etc. through the thiol group of cysteine residues thus regulating the concentrations of free metal ions and transport to appropriate cellular destinations. Most of the eukaryotic metallothioneins have been studied in yeasts and have been found to be highly inducible upon exposure to toxic metal stress. MTs are present in all prokaryotes and eukaryotes with affinities towards different metals. Though metallothioneins are typical of vertebrates, many MTs have been identified in plant species and various fungi as well (Bellion et al. 2007; Ramesh et al. 2009; Osobová et al. 2011). Highest induction of MTs have been observed upon treatment with Cu and Cd (Kalsotra et al. 2018) which leads to binding of metals to metallothioneins and in turn increases metal bioaccumulation in different tissues and organs (Azizi et al. 2018). The

protective roles of MTs like scavenving of Reactive Oxygen Species (ROS) and metal have been observed in some key functions of mitochondria. Some of the mitochondrial processes like utilization of oxygen, cytochrome c release, membrane permeability and electron transport have been reported to be influenced by the presence of metallothioneins (Simpkins et al. 1998; Miyayama et al. 2013; Moltó et al. 2007; Wang et al. 2001. In Arabidopsis, a type 2 metallothionein was found to interact with the outer mitochondrial membrane and help in the regulation of stress resistance against ROS generation under NaCl stress (Zhang et al. 2019). Studies pertaining to rice MT (OsMT2b) (Wong et al. 2004) and a cotton MT (GhMT3a) have proved to be effective ROS and metal scavengers (Xue et al. 2008). MTs are considered to play a critical role to regulate levels of metals in the body by efficient binding and sequestration. It reinforces the homeostatic mechanisms of Cu, Zn, and possibly Fe, and could also function as ROS absorbers in the central nervous system. Also, studies pertaining to Parkinson's disease could be linked to the rise of the principle aggregating protein α-synuclein levels which are thought to be directly proportional to the levels of heavy metal accumulation in the body (McLeary et al. 2019).

Though MTs have a common signature cysteine rich domains and regions, they exhibit little conservation of sequences between phyla and studies are more concerned towards characterization of these cysteine rich proteins (CRPs) for their metal specificities. Though MTs have been commonly studied in vertebrates, plants and fungi, a major portion of eukaryotes have been left out due to problems faced by cultivation based techniques. Interestingly, DNA/RNA based approaches have been successful in isolating and studying the environmental CRPs which majorly originated from soil dwelling eukaryotes (Ziller et al. 2017).

2.4.2 Phytochelatins

Phytochelatins (PCs) are thiol containing γ-glutamyl peptides and were first discovered in the year 1981 (Murasugi et al. 1981). These metal binding peptides are secreted mostly by higher plants, algae and fungi. They act as agents for immobilization of free and toxic heavy metal ions and form stable complexes. They are an important component of the detoxifying mechanisms and are at par with metallothioneins. It was then found in higher plants in 1985 and was named phytochelatin. By far, the most detailed characterization of the pathway of PC biosynthesis has come from studies in the fission yeast (*Schizosaccharomyces pombe*) and in *Arabidopsis*. Phytochelatins act as metal chelators (especially cadmium) and are considered vital for heavy metal detoxification in plants, fungi, algae and nematodes. Stable complexes of PCs and metals are transported towards vacuoles so that the functioning of the proteins in the cytosol is not hampered by the free metal ions. Synthesis of PCs is a result of trans-peptidation of γ-glutamyl-cysteinyl dipeptides from glutathione (GSH) via the action of constitutively expressed PC synthase, usually induced by the presence of heavy metals (Inouhe, 2005).

2.5 Diverse approaches to study microbial communities in polluted soil

To realize the functional diversity of eukaryotes and their *in situ* activities in reaction to various environmental conditions, it is essential to develop novel experimental approaches which are suitable for studying these microorganisms (Nannipieri et al. 2003). One such approach developed in recent years is metagenomics. Metagenomic approach is based on direct isolation of nucleic acids from environmental samples and has proven to be a powerful tool for comparing and exploring soil ecology (Biddle et al. 2008). Though metagenomic studies provide valuable information about the probable existence of microbial communities, it does not give a picture of the actual activities carried out by them at specific environment or changing environmental

conditions. The approach of metatranscriptomics connects community based structures and functions in single experiment, and reach beyond the community's genomic potential as assessed in DNA-based methods.

2.5.1 Molecular approaches

Several molecular methods have been developed which involves extraction of nucleic acid, directly or indirectly from soil that were able to detect species, genera, families or even higher taxonomic groups (Handelsman et al. 1998; Hirsch et al. 2010). Using techniques such as DNA re-association provides information of diversity specially the variations in GC content for detecting changes in microbial community. This technique requires large amounts of DNA and resolution suffers as several taxonomic groups may have same G+C content. Hence, it does not provide enough information about the composition and the role of microorganisms. Selected nucleic acid fingerprinting techniques with respect to heavy metals are listed in Table 2.1. The format for achieving diversity information changed significantly with the development of more complex molecular tools like the use of denaturing gradient gel electrophoresis (DGGE) (Deng et al. 2008) and eventually PCR-derived clone libraries of genes encoding the small subunit ribosomal RNA (rRNA) (Lara et al. 2007). Nucleic acids obtained from soil were subjected to different approaches of fingerprinting to assess the change in community structure with respect to environment. Comparing profiles of low molecular weight RNAs (5S rRNA, tRNA) collected from environment to their pure cultures previously obtained, provided insights into the change in structure (Liu and Stahl 2007). Amplified Ribosomal DNA Restriction Analysis (ARDRA) DNA fingerprinting technique makes use of restriction enzyme digestion of amplified 16S rRNA genes followed by gel electrophoresis (Nüsslein and Tiedje 1998). Similar technique involving terminal restriction fragment length polymorphism (T-RFLP), measures only the terminal restriction

fragment of every 16S rRNA gene thus reducing the complexity of ARDRA (Liu et al. 1997; Marsh 1999). PCR amplified intergenic spacer between 16S rRNA and 23S rRNA genes or ribosomal spacer analysis (RISA) was also used to assess the diversity of bacteria in certain soil type (Ranjard et al. 2000).The difference of length of each PCR products accounts for the heterogeneity and signifies an operational taxonomic unit (OTU). Cloning and sequencing of the automated method of ribosomal intergenic spacer analysis (ARISA) PCR products and their further basic local alignment search tool (BLAST) analysis reveals the identity of the microbes. A similar technique as that of DGGE is the Single-Strand Conformation Polymorphism (SSCP) which exploits the fact that mutated DNA bands have different electrophoretic mobility rates as that of the wild type and this is measured in a non-denaturing polyacrylamide gel electrophoresis (Widjojoatmodjo et al. 1995).

For high-throughput analysis, quantitative-PCR (qPCR) has been used extensively as a tool for identification and quantification of microorganisms and their gene expression levels. Taxa-specific quantitative amplification was performed to achieve relative abundance of a common group of soil microorganisms. The use of qPCR techniques has been helpful in the detection of single nucleotide polymorphism (SNP) and quantitative genotying. However, qPCR based analysis of soil dwelling microbes have been difficult to attain because of the poor quality of isolated RNA and DNA (Smith and Osborn, 2009).

To study the vast microbial diversity of soil, more robust techniques have to be developed which could assess the soil microbial communities. The microarray tool for simultaneous screening of microbial communities within diverse polluted environmental conditions is used to probe for the presence of specific ribotypes or phylotypes and their associated functional genes. Microarray platforms such as the GeoChip (He et al. 2007) and PhyloChip (Brodie et al. 2007) are now

providing an unparalleled opportunity to design targeted phylogenetic, marker-based and functional genes studies. However, the potential of microarrays when applied to environmental DNA is biased as the requirement of an amplification step is essential before processing.

Table 2.1: Selected nucleic acid finger-printing techniques for microbial community DNA analysis

Nucleic Acid Finger-printing techniques	**Contaminated Soil**	**References**
ARDRA	Copper	(Smit et al. 1997; Weidner et al. 1996)
T-RFLP	Hydrocarbon	(Clement et al. 1998)
RISA	Mercury, Copper	(Dell'Amico et al. 2008; Ranjard et al. 2000)
ERIC-PCR	Cadmium	(Roane and Pepper, 1999)
DGGE	Sewage, Heavy metals	(Kozdrój and van Elsas, 2000; Muyzer and Smalla, 1998, Sandaa et al. 1999)
TGGE	Zinc, Phenol	(Brim et al. 1999; Eichner et al. 1999)

2.5.2 The Metagenomic approach

Though PCR based studies are robust techniques for analyzing the metagenome, they have many drawbacks mainly due to the biasness introduced by the primers and interference of stray environmental DNA. It will eventually include genetic material not only from the live (active/inactive) sources but also from dead organisms or plant (Venter et al. 2004; Gill et al. 2006). Another major disadvantage for such techniques is the presence of PCR-inhibiting substances like humic acids, bacterial exopolysaccharides and proteins usually associated with particular soil types (Gelsomino et al. 1999; Miller et al. 1999; Opel et al. 2010).

Metagenomic based techniques rely on soil extracted DNA to be directly inserted or cloned in vector systems like plasmids or bacterial artificial chromosomes (BACs) and then propagated in

bacterial strains like *E. coli* thus creating a genomic library of all possible genetic content (Liles et al. 2003). The DNA library of the entire metagenome of soil has been estimated to approximately 10^6 Bacterial Artificial Chromosome (BAC) clones with an insert size of 100 Kb (Handelsman et al. 1998). The screening of metagenomic libraries usually depend on two broad techniques (Schloss and Handelsman 2003). Firstly, they depend on the specific enzymatic activity of the recombinant cells. This is known as activity-based screening as identifying new bioactivity does not require information about previous known sequences. Secondly, sequence driven screening (Daniel 2002, Rondon et al. 2000) which makes use of conserved DNA sequences to design hybridization probes or PCR primers, which are derived from conserved regions of already known genes or protein families, to screen metagenomic libraries for clones that contain sequences of interest. In this way, it is possible to identify only novel variants of already known functional proteins (Knietsch et al. 2003). A significant fraction of eukaryotes could not be encompassed by this method of creating gene library.

The metagenomic DNA library contain a very low proportion of clones of eukaryotic origin as prokaryotes outnumber eukaryotic organisms in soil (Lehembre et al. 2013). Also, owing to the huge size of the genome (for example genome size of 13.8 Mb of yeast *Schizosaccharomyces pombe*) (Wood et al. 2002), eukaryotic genes could not be well represented within a functional genomics library, As a result, significant number of their duties in soil still remain unconvincing (Yadav et al. 2014). Moreover, expression of eukaryotic protein-coding genes is prevented in most of the hosts (both prokaryotes and eukaryotes) mainly due to the presence of introns and lack of conservation of transcription regulatory elements in promoter sequences (Bailly et al. 2007; Takasaki et al. 2013). Metagenomic shotgun sequencing is a method employed to isolate and identify DNA from mixed microbial communities. Metagenomic sequencing has allowed

researchers to isolate functional genes from the environment and study microbial community structure and functions. This involves few major steps. DNA extraction, library construction and metagenomic sequencing which is followed by assembly of reads and downstream processing comprised of gene prediction and functional assignment (Li and Durbin 2009; Ma et al. 2016). Metagenomic DNA isolated from different sources undergoes shearing into fragments of appropriate sizes and a paired end sequencing library is constructed with adapter sequences linked to the DNA to be sequenced. Pair-end sequencing is performed on high-throughput sequencing platforms followed by downstream selection and assembly of reads. These reads are then compared to various databases for gene prediction, taxonomic analysis and function assignment leading to valuable information as to 'who is there' and 'what are they doing' (Sharpton 2014). This gives a firm idea of the functional diversity of an ecosystem. Direct sequencing of metagenomic DNA has been applied to study microbial functions in cadmium contaminated soil where few key microbial functions related to metal tolerance were found to dominate in the polluted soil (Feng et al. 2018). Vital enzymes related to cellulose and hemicelluloses enzyme were identified from grasslands (Nacke et al. 2012). Host-microbe interactions were documented with the help of metagenomic analysis which yielded valuable information on plant microbial disease mechanisms (Bulgarelli et al. 2013). It also provides insights into the ecological conditions and how unknown gene functions can have an effect on the ecology (Buttigieg et al. 2013). Researchers have a large catalogue of tools and databases to choose from concerning metagenomic DNA analysis.

2.5.3 Metatranscriptomics approach

Function based study of environmental DNA holds a role of great importance as more and more data sets are generated by cloning methods creating knowledge of putative biocatalysts, which

were previously not achievable. Discovery of such novel biocatalysts require bioinformatic tools and available databases where sequence-based metagenomic study can be deciphered. A novel protein coding gene can thus be categorized that often represents more than half of sequence similarity of the total identified ORF, to a known functional protein. With the application of this approach to identy genes from the termite hindgut, more than 600 novel putative glycosyl hydrolases were discovered (Warnecke et al. 2007). The major constraint for functional metagenomics is the availability of suitable activity assays for high-throughput screens. Some enzymatic activities may require the collective action of several proteins that could happen to be dispersed elsewhere on the host's genome. Thus, the respective activities might not be detected when screening small insert metagenomic clone libraries. Moreover, the absence of chaperons in expression hosts might fail to fold the protein and hence cannot function in the host (Warnecke and Hess 2009).

A major limitation of metagenomics is that it cannot reveal dynamic properties, such as impact of the environment on the activities of the community life. The only information obtained is the potential of the microbiome to exhibit functional properties associated with the presence of certain genes but may overlook the expression pattern of the genes in a stressful environment. Hence, the next subset of metagenomics which sheds light on the active functional profile of a microbial community is metatranscriptomics. Metatranscriptomics deal with those sets of genes which respond to a given environment at a given time to transcribe and exhibit activity (Chao-Rong and Zhang 2011; Moran 2009). Functional metatranscriptomics hence, is a powerful tool which allows characterization of genes expressed by different eukaryotic microorganisms (eg. Fungi, Protists etc.) directly in the environment. This approach has a strong potential in biotechnology to discover novel genes of interest for the bioindustry, in bioremediation and as

biomarkers. This approach allows characterization of genes implicated in adaptation to stressful conditions or involved in organic matter degradation.

Functional metatranscriptomics applied to contaminated soil involves the isolation and analysis of mRNA to help understand the regulation and expression patterns of complex microbial communities. It is based upon extraction of environmental RNA instead of DNA and on the purification by affinity chromatography of the eukaryote-specific polyadenylated mRNA from the total environmental RNA mixture. These poly(A) mRNAs are converted into cDNA, which can be cloned in suitable expression vectors such as plasmids, allow expression of the cloned genes in the eukaryotic hosts (Yadav et al. 2014). The cDNA prepared from the RNA can be directly sequenced and genetic data can be assigned to various functions (Agarwal et al. 2015) Figure 2.2 depicts methodologies for studying soil diversity.

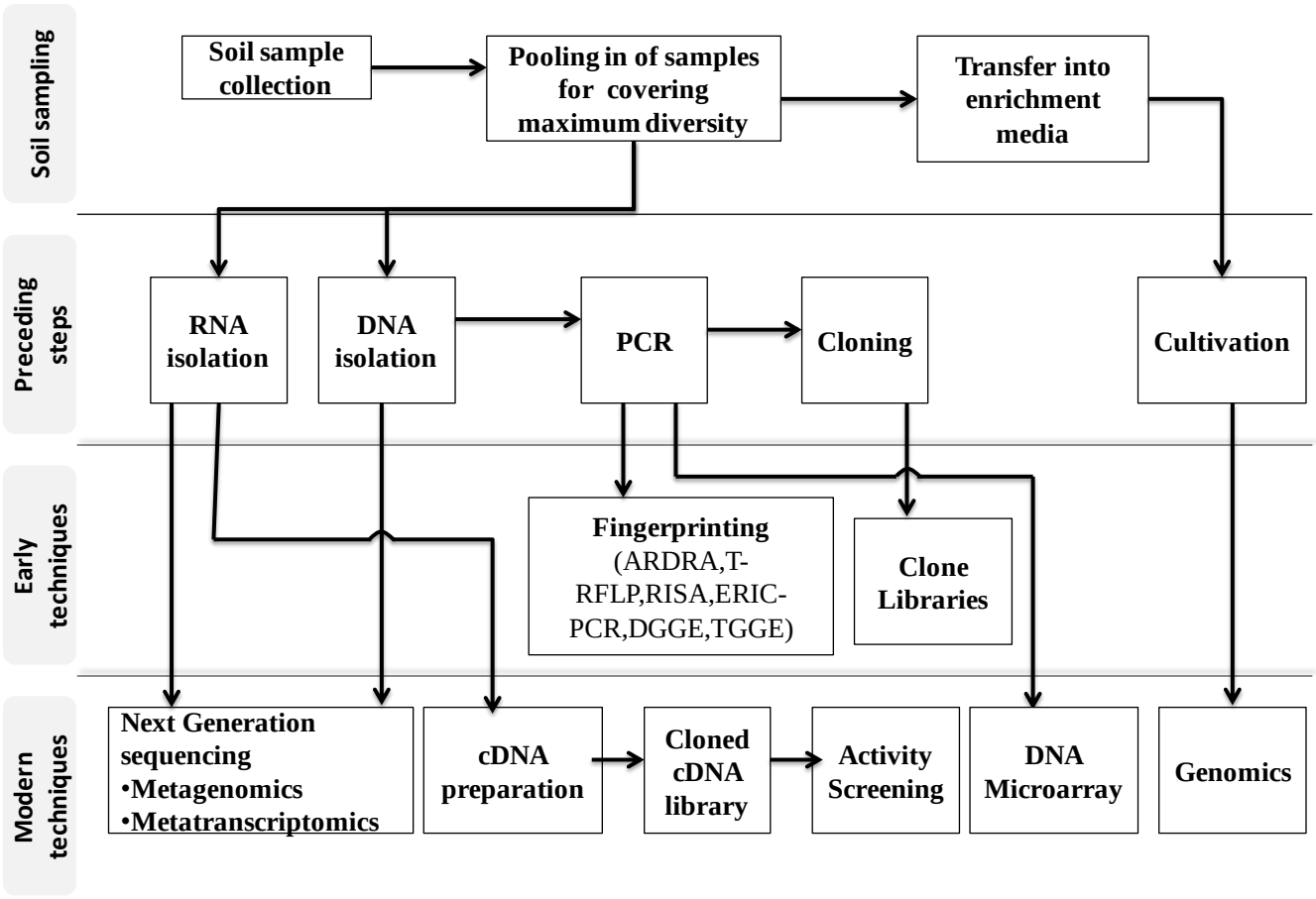


Figure 2.2: Overview of methodologies for studying soil diversity and function of microorganisms which involves both culture dependent and culture-independent techniques

2.6 Metatranscriptomics of polluted soil

2.6.1 Functional diversity of eukaryotes

Maintaining community structure and diversity is a challenge faced by the residing microorganisms in polluted soil (Giller et al. 1998). Though the diversity of community is negatively correlated to increase in stress levels, the effect of 'competitive exclusion' would enhance the removal of dominant organisms, and promote an increase in diversity of usually less abundant organisms. This is considered resilience, as the affected soil microorganisms try to regain their size, composition and function after an initial disturbance (Allison and Martiny

2008). Studies pertaining to heavy metals like cadmium and copper have indicated the presence of 'competitive exclusion' (Degens et al. 2001; Zhang et al. 2009).

Eukaryotic life forms have been previously studied in extreme environments along with prokaryotes, as they represent a significant fraction of the microbial biomass. They hold a repository of genes responsible for various activities in soil, hence it is difficult to isolate and culture them under laboratory conditions. Holistic approaches like metagenomics and metatranscriptomics have developed to identify genes and archive them as soil gene library. Information thus stored is a treasure house for novel biocatalysts and bioactive compounds, which can be coded by these genes when suitably expressed (Daniel 2005; Tringe and Rubin 2005).

Soil fungi have been under investigation for centuries (Rossman 1998), but DNA based analysis has paved the way to prove that phylogenetically, eukaryotes are as diverse as prokaryotes. A study by Fierer et al. (2007), dealt with grouping sequences by Operational Taxonomic Unit (OTU's) and revealed that archaeal or fungal OTUs appeared to be equal or exceed the number of unique bacterial OTUs in different soil types (Dessert, Grassland, Rain forest), with minimum taxonomic overlap between the soil types. In depth study of spruce and beech forest soil showed dominance of fungi, metazoans and protists linked with enzyme activities for litter degradation for maintenance of fertility of terrestrial ecosystems (Damon et al 2012).

With the advent of High-throughput or Next-Generation Sequencing (NGS), large volumes of sequence data are generated which enables efficient analysis of more complex expression profiles. The most common high-throughput sequencing platforms used in metatranscriptomic studies are the 454 Genome Sequencer FLX systems (Roche) and the HiSeq 2000 (Illumina,

Inc.).These systems make it easier to conduct a large number of individual sequencing reactions run simultaneously. Direct sequencing of DNA or cDNA with/without any cloning step (Medini et al. 2008) increases the efficiency in terms of numbers of base pairs sequenced per run and cuts down on sequencing cost. However, the new technologies produce short reads (454 pyrosequencer produced ~400 bp). Developments in sequencing technologies have improved and read lengths are comparable to traditional dye-terminator sequencing technology. Figure 2.3 shows how sequencing technologies played a major role in metagenomics and metatranscriptomics for studying the diversity, as culture based identification were slowly on the decline. This included species identification and functional annotation of genes. A clear majority in culture-independent based sequence identification and submission with the development of NGS can be seen over the years.

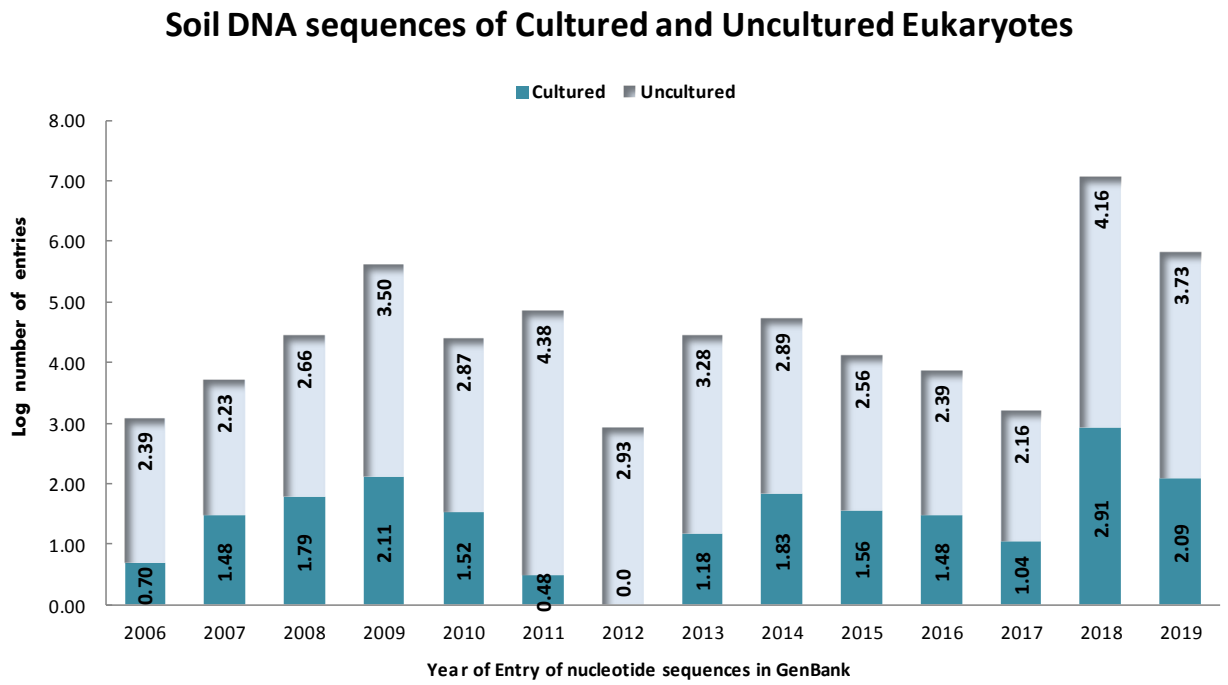


Figure 2.3: Logarithmic number of Soil DNA sequences of Cultured and Uncultured Eukaryotes submitted in GenBank as a function of year

The logarithmic number of 18S rRNA gene sequences derived from soil eukaryotes (Fungi and Protists) deposited in GenBank is shown as a function of year (data shown till December 2019). Values reported here are not absolute but an indication of the measure of number of sequences derived from soil eukaryotic microorganisms. The following search strategies were employed in Entrez. For cultured microorganisms: (18S OR SSU OR small subunit) AND (rRNA OR rDNA OR ribosomal RNA) AND (eukary*) AND (Soil) NOT (uncult* OR unidentified OR unknown). For uncultured clones: 18S OR SSU OR small subunit) AND (rRNA OR rDNA OR ribosomal RNA) AND (eukary*) AND (uncult* OR unidentified OR unknown) AND (Soil).

The first report of using 454 pyrosequencing for studying the metatranscriptome of a soil microbial community was published by Leininger et al. (2006), where they reported an abundance of archaeal ammonia oxidizers than their bacterial counterparts. A study prior to this on activated sludge and samples collected from hot springs, reported cDNA library preparation from extracted RNA exhibiting significant similarity to eukaryote mRNA-encoded protein sequences (Grant et al. 2006). Kit based RNA extraction from aquatic samples was carried out by Poretsky et al. (2005) and approximately 400 clones were analyzed after cDNA was amplified with random primers. They identified many novel genes associated to environmentally important processes such as sulfur oxidation from diverse taxonomic groups.

Pioneering work on functional diversity of eukaryotes by construction and screening of a cDNA library using poly(A) mRNA isolated from a forest soil was carried out by researchers from University of Lyon 1, France (Bailly et al 2007). A huge reservoir of genes and varied taxa were identified from sequencing over 100 cDNAs or directly from soil DNA. It was seen that fungi and especially unicellular eukaryotes (protists) dominated the microbial population with genes encoding for essential proteins involved in different biochemical processes. This was assessed by

complementation of auxotrophic mutant yeast with soil cDNA. Further, experiments were carried out by screening of environmental transcripts to reveal microbial activity in the environment and to discover novel enzymes of potential use in biotechnological applications (Kellner et al. 2011).

Studies were extended to investigate litter degradation in forest ecosystem, making nutrients available in soil by breaking down organic matter (Damon et al. 2012). Major nitrogen assimilating and cell wall degrading enzymes (celluloses, pectin, lignin etc.) were reported. Also, numerous eukaryotic oligopeptide transporters which enable microorganisms to ferry small peptides generated through proteolysis in soil for further assimilation, were identified (Damon et al. 2011). Metatranscriptomics received great impetus for studying different soil type, especially nutrient-poor soil mainly because it avoids biasness when gauging community reaction to environmental changes. Along with high-throughput sequencing technology, structure and functioning of sandy soil microflora were studied. Less frequent taxa Crenarchaeota were well represented by using rRNA-tags with the discovery of specific enzymes for ammonia oxidation and CO_2 fixation (Urich et al. 2008).

Metatranscriptomic approach was therefore a ‘hit’ for studying polluted or stressful soil. Scientists were focused in finding out ‘how’ are microorganisms thriving in harsh environmental conditions and mechanisms involved in coping up with nutrient-less or toxic environments. First study on heavy metal contaminated soil by using metatranscriptomics was conducted by Lehembre et al. (2013). A metatranscriptomic library was screened with varied diversity of sequences which was screened for heavy metal tolerance, mostly from unknown taxonomic groups. The isolated heavy metal tolerant genes were able to withstand a variety of metals even at high concentrations and were also able to accumulate metals (Ziller et al. 2017). This opened

up a huge platform for investigating the proteins and linking them with different pathways for coping up with heavy metal tolerance.

2.6.2 Novel eukaryotic gene discovery from soil employing metatranscriptomics

2.6.2.1 Methodologies to mine industrially important enzymes

Enzymes from higher eukaryotic microorganisms especially fungi have been widely used as industrial important biocatalysts ranging from applications in food industries, pharmaceutical, textile etc. (Gudynaite-Savitch and White 2016; Guerriero et al. 2016). The role of eukaryotes in improving soil fertility by litter degradation involves various enzymatic processes to degrade lignocelluloses, hemicelluloses, polyphenols etc. (Štursová et al. 2012). Litter degradation rates depend on their physicochemical properties which in turn, vary with climatic factors, plant species which form the litter and the association of saprophytic fungal microhabitats (Goldmann et al. 2015). Besides production of useful metabolites, eukaryotes are also involved in biomass treatment in biorefineries and biofuel production. Essential enzymes acting on complex organic matter like lytic polysaccharide monooxygenases cleave complex polysaccharides like cellulose and hemicelluloses resulting in oxidative cleavage. These eukaryotic microbes are extremely diverse and hence their function cannot be localized and neither they are ubiquitous in every ecosystem (Marmeisse et al. 2017). To trap the extensive enzyme resources from forest soil, the most convenient way is to create a metatranscriptomic library and perform activity screening which does not incorporate biasness towards specific species. Industrially important enzymes like acid phosphatase and laccase were identified by activity screening and semi-quantitative PCR respectively, following RNA extraction and cDNA preparation (Kellner et al. 2011; Luis et al. 2005). The only disadvantage of this process is the massive number of transformed clones which have to be screened for activity. As many as 30,000 recombinant yeast clones were

screened for potential phosphate solubilizing enzyme acid phosphatase and also an imidazoleglycerol-phosphate dehydratase by utilizing the concept of functional complementation and activity screening (Kellner et al 2011). The metatranscriptomic library was ligated to a yeast secretion vector pTEF-MF-*Sfi*I A/B which was modified with additional *SfiI* restriction sites to accommodate the library. It was used to transform *Saccharomyces cerevisiae* mutant ML20C having *pho5*− (mutation in the repressible acid phosphatase gene) and a *his3*− mutation. Positive clones complemented by plasmid-borne *his3*− gene and colony staining assays were further sequenced. Strong similarity with basidiomycete acid phosphatase from the sequence data confirmed the important role played by fungi in nutrient mobilization in the organic rich layer of forest soil.

The availability of a large number of yeast strains which are mutated for a variety of metabolic pathways, provide an opportunity to perform complementation with gene resources to screen and find out the gene of interest (Giaever and Nislow, 2014). Yeast can be genetically manipulated by plasmids having constitutive or regulated promoters for gene expression and rescuing the auxotrophs by complementation.

Along with mutant yeast strains, plasmids have also been modified to successfully clone environmental cDNA and effectively express the gene in the host. The transcription of foreign genes in mutant yeast will depend upon carefully placed compatible promoter in the plasmid upstream of the cloning site. Expression of environmental eukaryotic cDNA is efficiently cloned in pFL61 which has a multiple cloning sites between Phosphoglycerate kinase (PGK) promoter and its terminator and also bearing a 2 μm yeast origin of replication for yeast host and part of pUC19 *E. coli* vector for replication in bacterial host (Minet et al. 1992). Introduction of

restriction sites *SfiI*A and *SfiI*B (5'-GGCCNNNN//NGGCC-3') helps to accommodate and express genes from soil (Yadav et al. 2014).

An efficient technique to screen cDNA synthesized from soil-extracted polyadenylated mRNAs is by using 'Sequence capture by hybridization' (SCH) (Bragalini et al. 2014). This method of pre-selection of target genes before activity screening increases the probability of exploring novel genes by reducing the number of random environmental cDNA. It was conducted by using biotinylated degenerated RNA probes. The biotinylated probes hybridized to known and unknown sequences of members of a target gene family present in a metatranscriptomic library. They were further captured using streptavidin-coated paramagnetic beads. Tested on the glycoside hydrolase 11 gene family encoding endo-xylanases, the capture probes were able to relate >90% of the cDNA to this gene family. Sequencing of cDNA revealed many phylogenetically diverse species which were not yet included in public databases. The captured cDNAs were cloned and expressed in Uracil-auxotrophic yeast strain DSY-5 also having non-functional sugar transporter protein. Yeast cells producing endo-xylanases developed a blue halo in the presence of xylan linked to a dye confirming hydrolysis of xylan. This method used the plasmid PDR 196-SfiI-Kan modified to contain two specific *SfiI*A and *SfiI*B sites for ligation of cDNA which also contain identical sites for directional cloning and placed under the control of a yeast *PMA1* promoter to facilitate constitutive expression of cDNA. The Overview of the process is depicted in Figure 2.4.

The enzymes retrieved from eukaryotic organisms have dominated the industrial processes especially where they have replaced traditional usage of chemical treatment of biomass, in textile as well as in food industries (Azizan et al. 2016). The recent advances to improve the commercial potential of industrially important enzymes have focused on the molecular

techniques such as metagenomics and metatranscriptomics for accessing the environmental genes (Niehaus et al. 2011). The practice of collecting industrially important genes from the environmental metagenomic library has been discussed in detail by Ferrer et al. (2016). Sequence based as well as functional screening of environmental metagenome to achieve a stable, 'industry-level' enzyme is a time-consuming process. This requires efficient screening and expression systems. However, the approach of metatranscriptomics has been developed into a promising and practical technique for harvesting and screening protein-coding genes from an environmental library. Forest soil has been the primary source to harvest important and novel enzyme coding genes. Glycoside hydrolase that decompose cellulosic biomass was harvested from the fungal communities of forest soil (Takasaki et al. 2013). About 40% of genes involved in metabolism of carbohydrate and amino acid out of 9,449 eukaryotic coding sequences were identified by metatranscriptomic approach. Functional metatranscriptomics was successfully applied to eukaryotes of forest soil where acid phosphatase and imidazoleglycerol-phosphate dehydratase of fungal origin was isolated and characterized using yeast mutants which could proficiently express the protein (Kellner et al. 2010). Many important eukaryotic oligopeptide transporters were discovered by functionally complementing yeast mutant defective in di/tripeptide uptake. These transporters were able to act upon a wide range of substrates and were further expressed in *Xenopus* oocytes (Damon et al. 2011). A significant number of non-fungal sources of industrially important enzymes in forest soils were explored by Damon et al. (2012), where metatranscriptomic library was probed for essential enzyme coding genes. Major plant cell wall and polymer (cellulose, hemicelluloses, pectin, lignin etc.) degrading enzymes were identified. Also, organic matter hydrolyzing enzymes like proteases, lipases, a phytase etc. were also identified. Forest soil is a host of carbohydrate active enzyme transcript sequences

(CAZymes) and essentially about 74,000 active transcripts were identified in a maple forest where soil metatranscriptome was analyzed (Hesse et al. 2015). The use of metatranscriptomics in screening polyadenylated mRNAs from soil was implemented on *endo*-xylanases encoding gene glycoside hydrolase 11 gene family. An efficient functional screening mechanism of solution hybrid selection was used screening of enzyme encoding genes (Bragalini et al. 2014).

The use of pesticides and chemical fertilizers in many agricultural lands has led to the accumulation of persistent aromatic hydrocarbon pollutants in the soil. In a study by Sharma and Sharma (2018), pesticide and chemical fertilizer polluted soil was studied and the taxonomic and functional aspect of the microbial communities was investigated by metatranscriptomic approach. Transcripts related to degradation of cypermethrin and related aromatic compounds were identified. Hence, using a single technique, both prokaryotic and eukaryotic enzymes were studied for pesticide degradation. The compound cypermethrin is degraded in a step wise process forming by products like carboxylate, benzoic acid, derivatives of aldehydes etc. It can thus be inferred that the process of removal of toxic compounds is an orchestration of a chain of functional proteins which may originate from any organisms in the soil ecosystem. Metatranscriptomic study of the wheat rhizosphere revealed various bacteria that are involved in the degradation of xenobiotics like aromatic compounds, carbazoles, naphthalene (Singh et al. 2018).

2.6.2.2 Methodologies for bioremediation of heavy metals

Though there are a limited number of published research papers, a significant amount of molecular characterization was performed with metal resistance genes obtained from the environment proving metatranscriptomics approach a lucrative study method for bioremediation

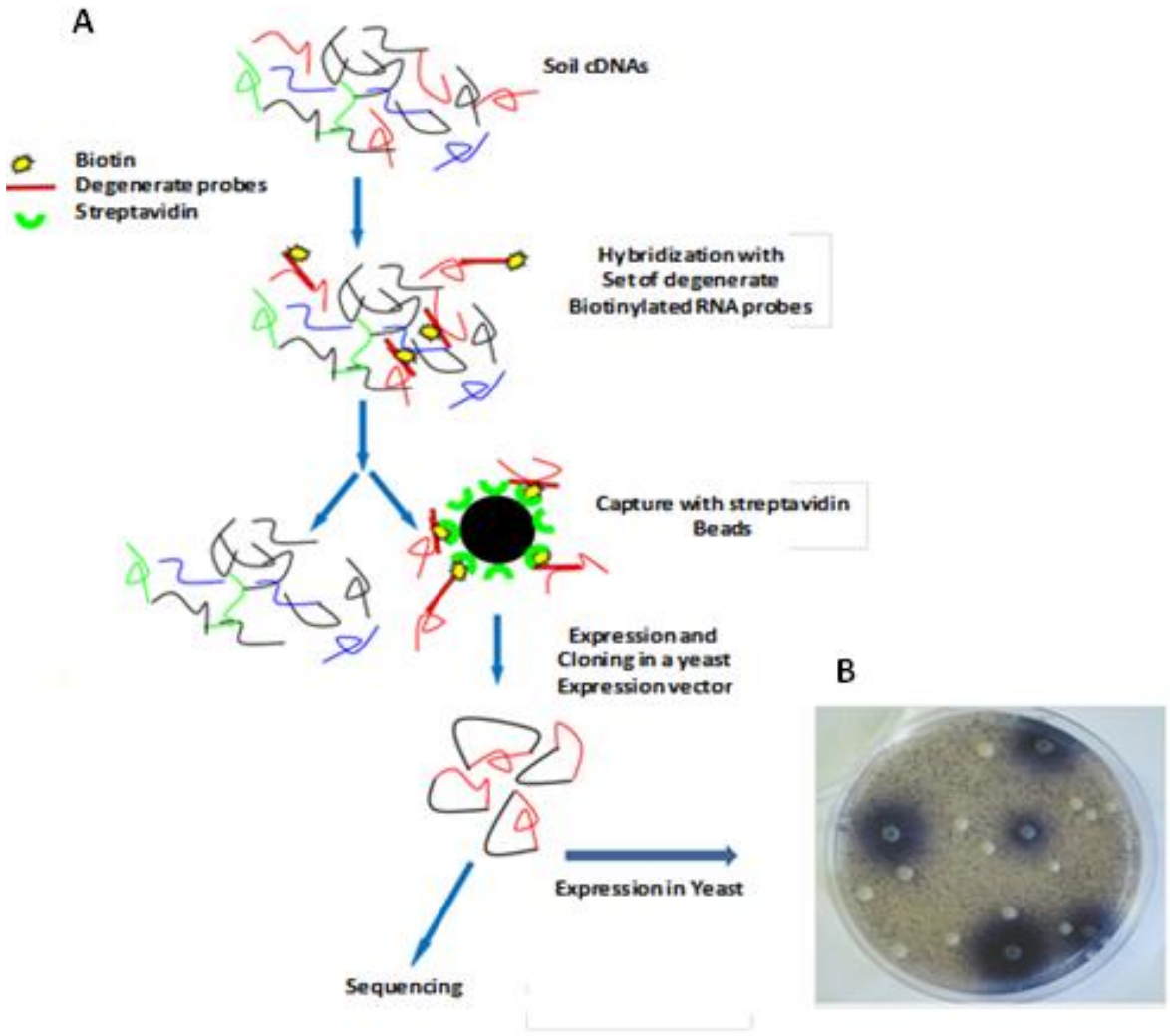


Figure 2.4: (A) Overview of the Sequence capture screening activity by degenerate probes followed by (B) Transformation of yeast DSY-5 with cloned cDNA plated on a selective medium containing AZCL-xylan (Specific substrate to endoxylanase) and without Uracil. Positive clones degrade the substrate to release dark blue dye. Modified from Bragalini et al. (2014)

of heavy metals. Metallothioneins (MT), a family of ubiquitous cysteine-rich proteins have been studied for their ability to coordinate the binding of metal ions by forming metal-thiolate complexes (Blindauer 2013; Mehra and Winge 1991). Though MTs have a defined classification system based on sequence similarity analysis (Binz and Kägi 1999), they share less overall

sequence conservation. This might be due to independent emergence of MT families throughout the course of evolution (Capdevila and Atrian 2011). Eukaryotic MTs, however, have not been covered fully as it has been studied only in specific species like vertebrates, plants, fungi etc. Thus, the search for new metal-binding proteins from environmental DNA or RNA will be a fruitful endeavor in terms of detection and bioremediation of heavy metals in soil. Soil extracted mRNA converted to cDNA was screened on the basis of their functional role of resisting toxic amounts of heavy metals (Cervantes and Gutierrez-Corona 1994; Lehembre et al. 2013). This cDNA was cloned into suitable yeast expression vectors. Metal sensitive phenotypes of yeast were then transformed and the cDNAs were screened on their ability to rescue the yeast strain by functional complementation. Members of the cysteine-rich proteins (CRPs) were identified and compared with the already existing fungal metallothioneins (Figure 2.5). The potential of these environmental genes to support the growth of sensitive yeast is much higher than existing fungal metallothioneins in media with cadmium metal stress (Lehembre et al. 2013). Further characterization of the CRPs was carried out by Ziller et al. (2017).

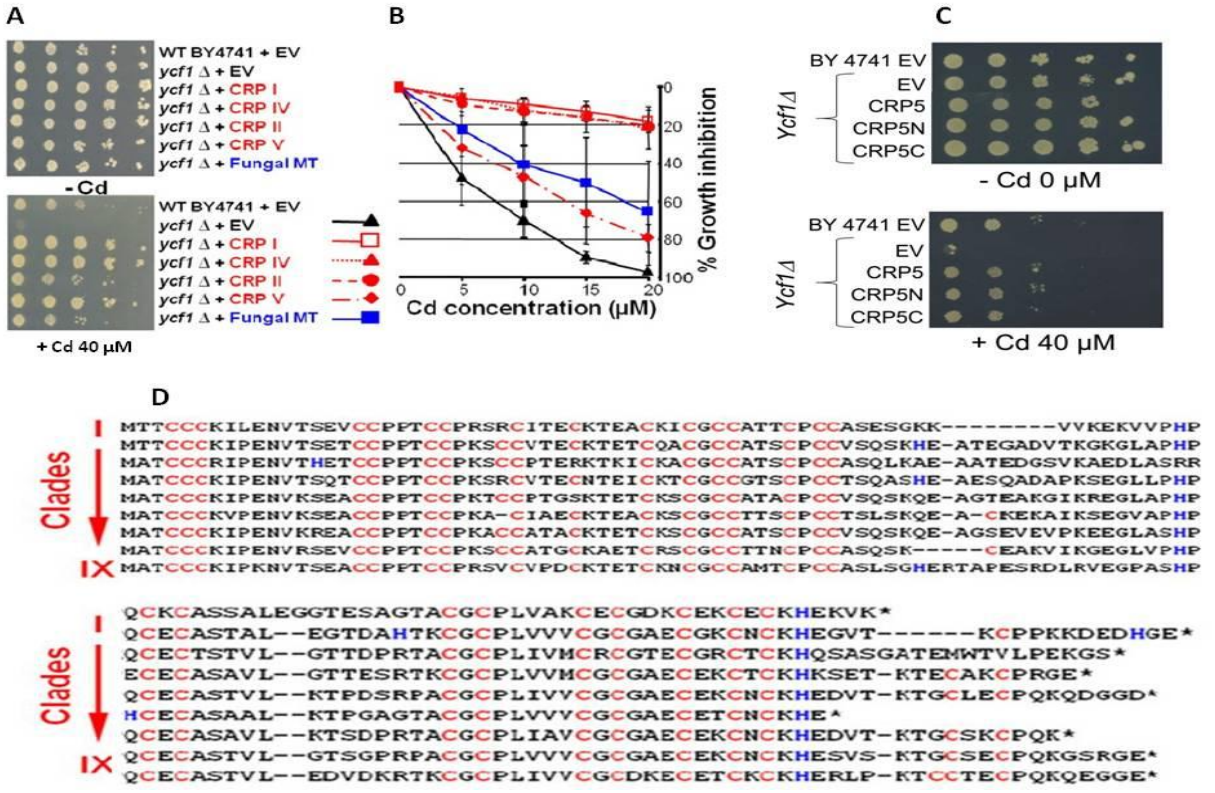


Figure 2.5: (A) Functional complementation of the CRPs in Cadmium sensitive yeast phenotypes *ycf1Δ* compared to the fungal metallothionein gene. (B) Inhibition of growth (%) in liquid broth spiked with cadmium. (C) CRP 1-5 expression in Zinc sensitive phenotypes *Zrc1Δ* in presence and absence of Zinc metal exhibited multi-metal resistance potential. (D) Nine cysteine-rich proteins (CRP2) were identified from soils of 12 different sites with their amino acids alignments showing conserved cysteine motifs. Adapted from Lehembre et al. (2013) and Ziller et al. (2017)

Amino acid features were analyzed for these CRPs and their metal- binding abilities towards Cd, Zn and Cu metal ions. Previously isolated soil metatranscriptomes were sub-cloned and expressed as a fusion protein. These novel proteins or 'environmental metallothioneins' had the capability to chelate Cd(II), Zn(II) and Cu(I). Yeast mutants (*ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$) derived from the wild type strain BY4741 (MATa *his3*$^{\Delta}$*1 leu2*$^{\Delta}$*0 met15*$^{\Delta}$*0 ura3*$^{\Delta}$*0*) were used is this study thereby

enabling efficient screening by functional complementation. The *YCF1* gene codes for an ABC transporter that represents Cd tolerance by transporting Cd conjugates into vacuoles. *ZRC1* gene encodes a transporter that sequesters Zn into vacuoles. Copper sensitive yeast mutant DTY4 have a deleted copper-binding gene *CUP1* hence making it susceptible to copper in growth media (Lehembre et al. 2013).

The potential of genes identified by functional metatranscriptomics can be extended to be used as biomarkers whose level of expression can be used to access the extent of toxicity of heavy metals in crop fields. A recent study on yeast-based biosensors for detection of environmental pollutants provides an explanation as to why eukaryotic models are now being utilized for this purpose. Yeast-based systems have an edge over bacteria based and enzyme based biosensor systems as bacteria has low relevance for representing eukaryotic community and enzyme based biosensors have lower relevance for the cell or the whole organisms (Jarque et al. 2016). Already existing yeast based heavy metal detection are depicted in the Table 2.2.

Ubiquitin domain containing zinc finger protein was isolated and characterized by Thakur et al. (2018) which was screened in *S. cerevisiae* mutants and was able to impart tolerance towards different toxic metals. Similar study revealed a novel von Willebrand factor type D of vitellogenin protein from polluted soil environment which depicted higher expression levels in presence of toxic metals and was able to confer tolerance towards toxic concentrations of metals (Thakur et al. 2019). Vitellogenin proteins are known for their antioxidant properties. From these studies it can be concluded that the technique to isolate enzyme coding genes from perturbed soil could serve as important biomarkers and genes suitable for bioremediation of environmental soil pollution.

Table 2.2: Yeast bio-detection systems of heavy metals in environment

Detected Metal(s)	Receptor/ Gene	Detection	Level of Detection (Range mM)	Ligand	References
Copper	CUP1	Amperometry	0.11–500	Copper	(Lehmann et al. 2000)
		Colorimetry	1		(Vopálenská et al. 2015)
		Fluorescence	0.5		
		Luminescence	0.5		(Shetty et al. 2004)
					(Roda et al. 2011)
Cadmium	AtPCS1	Fluorescence	0.2	Cadmium	(Matsuura et al. 2013)
Heavy metals	cAMP-PKA	Fluorescence	12–45		(Radhika et al. 2005)

Abbreviations: AtPCS1, phytochelatin synthase from *Arabidopsis thaliana*; cAMP-PKA, Cyclic adenosine monophosphate-Protein kinase A (Adapted from Jarque et al. 2016)

Nonspecific and specific biosensors can detect toxic heavy metals. For non-specific biosensors, the sensing factor is based on the cAMP/PKA (cAMP-dependent protein kinase) signaling pathway which mediates gene expression in response to common toxic metals in yeast. These biosensors have been shown to recognize As^{3+}, Fe^{2+}, Pb^{2+} and Cd^{2+} as general toxic metals (Radhika et al. 2005). A representation of a general biosensing detection system with yeasts is depicted in Figure 2.6.

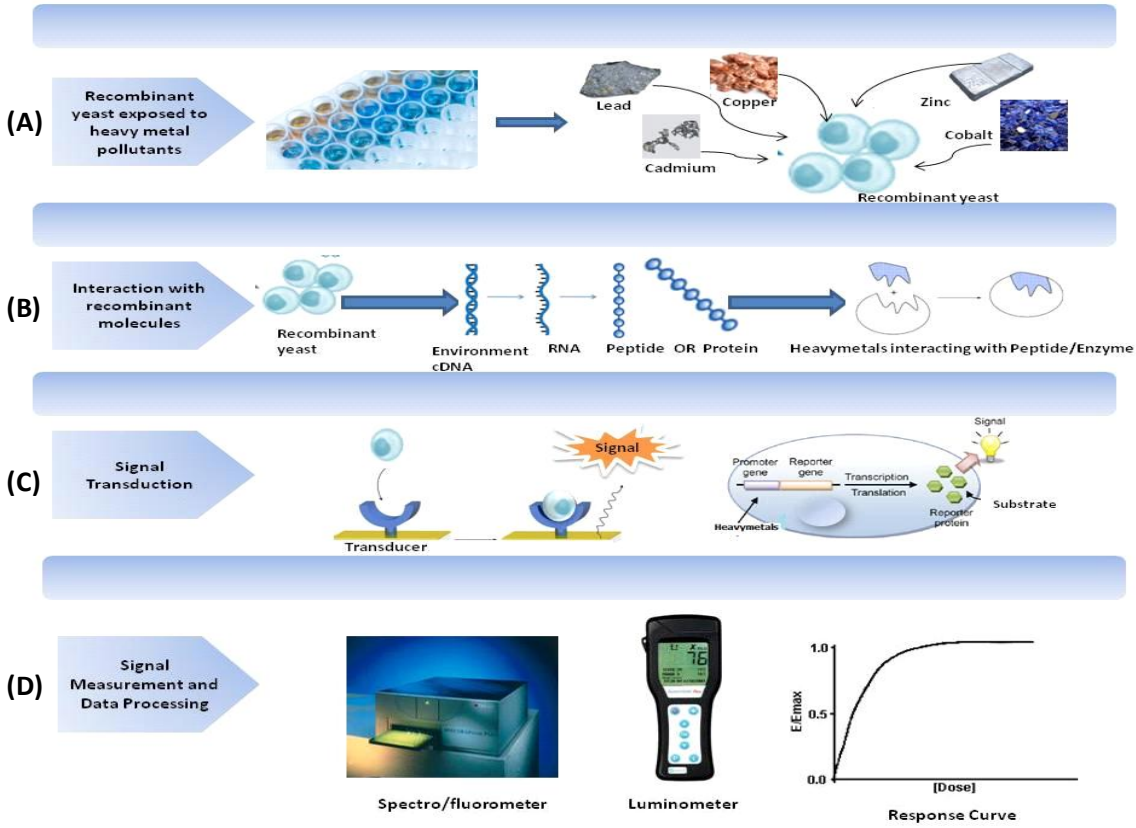


Figure 2.6: Yeast Biosensing Detection Principle. (A) Preparation of environmental samples and its exposure to yeast-based sensing system (B) Recombinant product- environmental metal resistant gene reacts with heavymetals inducing activity of (C) a reported gene or generation of electrical signal and (D) The signal associated with the response might be amplified and processed into valuable quantifiable data

Critically designed experiments which capture microbial response reveal newer, underrepresented species, performing environmental activities like nutrient mobilization, breakdown of organic matter/hydrocarbons. Studies by Lehembre et al. (2013) and Ziller et al. (2017) have captured different roles of eukaryotes involved in metal chelation and homeostasis. This will greatly accelerate the tapping of natural resources for industrially important biocatalysts and also help to solve problems of soil fertility and pollution. Table 2.3 summarizes all published work on soil metatranscriptomics.

Table 2.3: Published metatranscriptomics studies on eukaryotic soil microorganisms

Soil Sample	RNA extraction method	mRNA enrichment	cDNA preparation	Sequencing/Screening	Size of dataset	Significant find	Reference
Forest Soil	Guanidine isothiocyanate, bead-beating,	Poly(A) tail mRNA captured by affinity capture	Poly-dT primer (SMART cDNA kit,	Sanger sequencing	119 clone sequences	Eukaryotic diversity, novel enzymes	(Bailly et al. 2007)
Sandy, Nutrient-poor soil	CTAB, P/C/I Lysing matrix E tube (Q-Biogene) PEG 8000	None	Random primer	Pyrosequencing (Direct sequencing)	314,041	CO_2 fixation and ammonia oidizers enzymes	(Urich et al. 2008)
Sugar maple forest soil	Bead beating, , 8-hydroxyquinoline, RNA/ DNA Midi kit (Qiagen)	None	Selective amplification of eukaryotic mRNA by(SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	10 transformed yeast sequenced	Acid phosphatase, Imidazoleglycerol-phosphate dehydratase	(Kellner et al. 2011)
Spruce forest	Guanidine isothiocyanate, bead-beating	Affinity capture of polyadenylated mRNA on poly-dTcoated paramagnetic beads	Poly-dT primer (SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	7.7×10^5 screened transformants, 25 putative genes	Novel fungal oligopeptide transporter	(Damon et al. 2011)
Beech and Spruce forest	Guanidine isothiocyanate, bead-beating	Poly-dT beads for separating eukaryotic polyadenylated RNA	SMART cDNA Library Construction Kit (Clontech)	Sanger sequencing	10,000 each	Eukaryotic Diversity, novel enzymes for organic matter degradation	(Damon et al. 2012)
Soil near metal smelter	Guanidine isothiocyanate, bead-beating	Poly(A) tail mRNA captured by affinity capture	Poly-dT primer (SMART cDNA kit,	Sanger sequencing (direct sequencing of cDNA)	30k cDNA for each soil sample	Metal tolerant genes	(Lehembre et al. 2013)

Soil near metal smelter	Guanidine isothiocyanate, bead-beating	Poly(A) tail mRNA captured by affinity capture	Poly-dT primer (SMART cDNA kit,	Characterization of Cysteine-rich proteins from Cadmium tolerant clones	Fve cysteine-rich proteins isolated from soil	Novel metal-binding metallothionein families	(Lehembre et al. 2013)
Agro forestry contaminated soil	RNA isolation Kit (Mo Bio labs)	None	Selective amplification of eukaryotic mRNA by(SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	0.5 kb- 1.0 kb size library	Ubiquitin fusion protein/Heavy metal tolerance gene	Thakur et al. 2018
Agro forestry contaminated soil	RNA isolation Kit (Mo Bio labs)	None	Selective amplification of eukaryotic mRNA by(SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	1kb – 4 kb size library	VWD like protein/multi-metal torlerant	Thakur et al. 2019

2.7 Metatranscriptomics in microbiome research

Being a holistic approach of study, metatranscriptomics has been employed in microbiome research majorly to characterize the microbiome of human gut and classifying the active population of the gut microbiota. This will help us to gain insights into the host gene expression profile and also to study the gene expression of microbial communities in a given environment where culture-dependant techniques cannot assess their true functional roles e.g. gut, oral cavity, respiratory tract etc. From such data, metabolic pathways in the microbial communities can thus be identified and can be associated to particular environment. This can actively be associated with health, diseases and metabolism (Bashiardes et al. 2016). Studies on the rumen microbiome of buffaloes, lead to the understanding of the role of microbial carbohydrate assimilating enzymes present in the buffalo gut with the abundance of polysaccharide degrading bacteria (Patel et al. 2014). Moreover, the colonization of the bacteria in the rumen of cattle throughout its lifetime was studied using next-generation sequencing technology which has lead to useful information of the digestive capabilities of the ruminant (Koringa et al. 2019).

The rumen gut is particularly interesting from a study point as it holds a repertoire of carbohydrate assimilating active enzymes, which can degrade plant matter. It has been reported that breakdown of cellulose involves the action of three different types of cellulolytic enzymes endo-β-1, 4-glucanase, glycosyl hydrolase family, cellobiohydrolase and β-glucosidase (Terry et al. 2019). Several studies have employed the approach of metatranscriptomics in studying the eukaryotic population of rumen and their significant contribution towards plant fiber digestion. The presence of anaerobic fungi of phylum Neocallimastigomycota in the rumen has been associated with degradation of plant cell wall material where CAZyme families have been active in hemicelluloses digestion (Gruninger et al. 2018). High-throughput RNA sequencing technique

for estimation of active population applied to rumen gut has found a total of 21.8% of the small subunit RNA belonging to the eukaryotes (Comtet-Marre et al. 2017). The protozoan population was also found to be an active member of the gut microflora and about 3-6% total non-rRNA reads were from protozoan species. However, Piromyces and Neocallimastix accounted for the largest fungal cellulases identified from the rumen gut (Dai et al. 2015). Söllinger et al. (2018) correlated the activity of the transcripts along with the metabolites of rumen and cellulase transcripts originated from two eukaryotic groups, Neocallimastigaceae and Ciliophora.

Metatranscriptomic approach to study the rumen gut suggested the predominance of exoglucanases, which degrade crystalline cellulose. Many glysoside hydrolases, which were previously not identified by metagenomic studies, were confirmed to be active in the metatranscriptomic study (Dai et al. 2015). Major cellulose binding modules were also found in the eukaryotic transcriptomes along with polysaccharide utilization loci which aid in the binding, transport and degradation of polymeric glycan structures (Seshadri et al. 2018). A detailed review by Terry et al. (2019) on the metagenomic and metatranscriptomic studies further describes the rumen gut microflora.

2.8 Practical hindrances in metatranscriptomics

Every method for studying natural habitats seems to have their own benefits and short comings. The first limitation of metatranscriptomic study which has already been dealt upon is the instability of RNA molecules. Unlike DNA, RNA is prone to degradation by various sources once they are isolated, average half-lives being seconds to minutes (Iost and Dreyfus 1995). There can be variations in RNA stability depending on the type of soil, microbial species and nutrients present (Redon et al. 2005). In order to save the RNA in the collected sample, liquid nitrogen must be used to freeze the samples post-collection or to shift them to RNA protective

solutions (*e.g.* RNAlater™ Invitrogen ™, USA). Isolation of RNA from soil is incredibly challenging as it involves inefficient cell lysis due to the tough matrix soil provides. Downstream processing becomes difficult due to co-precipitation of complex molecules which inhibit PCR reactions. Nucleic acids are easily adsorbed on soil particles which makes extraction very difficult, especially with low pH buffers, which specifically separate RNA from total nucleic acid. A detailed evaluation of challenges faced by scientists practicing metatranscriptomics as a tool to understand soil ecosystems has already been presented by Warnecke and Hess (2009) and Carvalhais et al. (2012).

A major step to bypass the mRNA enrichment stage and selectively amplify soil eukaryotic mRNA was first used by Bailley et al. (2007) where the technique of Switching Mechanism At the 5' end of RNA Template (SMART ™) was used to synthesize cDNA. Traditional method of synthesizing cDNA has many shortcomings. They produce truncated cDNA due to premature termination of reverse transcription and as a result most of the 5' terminal sequence information is lost, resulting in underrepresentation of 5' end in cDNA library (Botero et al. 2005; Gubler and Hoffman 1983; MD'Alessio and Gerard 1988). Moreover, use of ligation adapters for cloning cDNA generated undesirable byproducts (chimera).

SMART cDNA synthesis starts with as low as 2 nanogram of total RNA. A modified oligo(dT) primer containing specific restriction sites primes the first-strand synthesis reaction (Zhu et al. 2001). When the Reverse transcriptase reaches the 5' end of the mRNA, the terminal transferase activity of the enzyme adds few additional nucleotides to the 3' end of the cDNA. The oligonucleotide then base-pairs with the non-template nucleotide stretch, creating an extended template. The reverse transcriptase then switches templates and continues replicating to the end of the oligonucleotide. Overview of the process of metatranscriptomics is depicted in Figure 2.7**.**

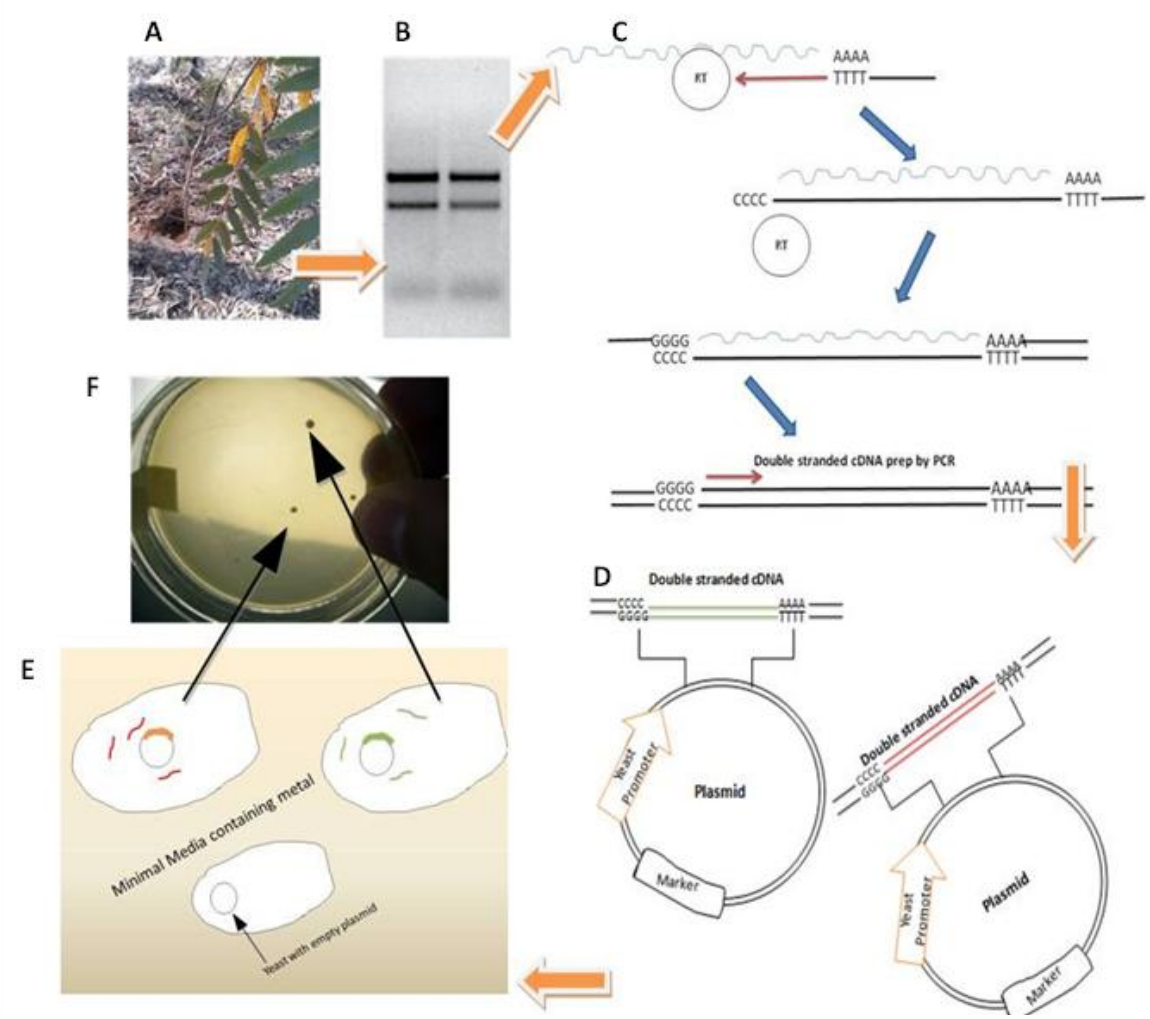


Figure 2.7: Overview of metatranscriptomics for mining heavy metal tolerant genes from Soil. [A] and [B] Soil collection and total RNA isolation [C] cDNA synthesis by SMART ^{TM}, Reverse Transcriptase (RT) converts soil mRNA to cDNA [D] Double stranded cDNA is cloned in plasmids having yeast promoter and suitable gene marker for selection in yeast and *E. coli* [E] Mutant/Auxotrophic yeast transformed with recombinant plasmids are able to survive metal stress due to the expression of plasmid-bourne cDNA; control having empty plasmid cannot survive the stress [F] Surviving yeast cells are selected on minimal media containing specific metal ions

The designed primer serves well as a priming site for the overall exponential amplification of the synthesized cDNA in contrast to incomplete cDNA or genomic DNA which may otherwise be contaminating the cDNA library. However the only downside of this method is the presence of truncated RNAs in poor quality RNA starting material which will eventually be amplified, and will contaminate the final cDNA library (Chenchik et al. 1998).

In this study, a cDNA library was synthesized from all the functional genes of eukaryotic microorganisms residing in metal polluted soil from the copper mining site of India. This library was screened for isolating the genes having potential to confer metal tolerance to yeast cells that are sensitive to metals. Several genes were identified and isolated from the cDNA library. Most of the genes and their protein products are well known for their cellular and metabolic functions. However, apart from their usual duties, it was clear that these genes also contribute to normalize the toxic effects of metals by providing tolerance to the microorganisms directly or indirectly as a stress response. Few of the genes isolated from the French soil are aldehyde dehydrogenase, serine protease inhibitor DHHC palmitoyl transferase, maltose fermentation regulatory protein and chitin synthase were further characterized in this study with respect to tolerance towards toxic concentration of metals.

2.9 Aldehyde dehydrogenase

Aldehyde dehydrogenase (ALDH) (EC 1.2.1.3) is an enzyme superfamily that takes part in detoxifying aldehydes by catalyzing the oxidation of various endogenous and exogenous aldehydes to respective carboxylic acids (Kirch et al. 2004). These important aldehydes are metabolized by the superfamily of $NAD(P)^+$ dependant aldehyde dehydrogenase

The following chemical reaction is catalyzed by ALDH

$$RCHO + NAD^+ + H_2O \rightarrow RCOOH + NADH + H^+$$

It is a NAD(P)$^+$-dependent reaction, where the aldehyde enters the active site of the enzyme through a channel from its surface. The active site consists of a Rossman fold, and the reaction is based on the interactions between the cofactor and the Rossman fold (Liu et al. 1997).

On the carbonyl carbon of the aldehydes, the initiation of the reaction occurs via a nucleophilic attack followed by the reduction of NAD(P)$^+$ to NAD(P)H and a release of a hydrogen ion. Consequently, the active site of the enzyme undergoes an isomorphic change, which removes the NAD(P)H for a molecule of water to access the substrate. The glutamate in the active site then allows the water molecule to make a nucleophilic attack on the carbonyl carbon, removing the sulfur (Figure 2.8).

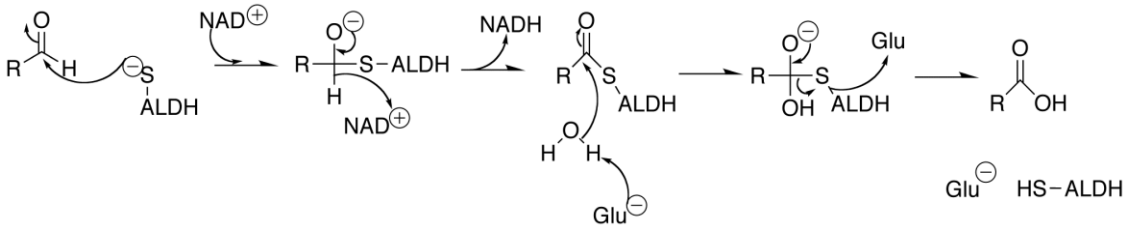


Figure 2.8: Mechanism of action of aldehyde dehydrogenase. Adapted from Liu et al. (1997)

The first purified ALDH was obtained in the year 1949 and the amino acid structure was determined in the year 1984. Nowadays more than 30 different ALDHs are known to exist. The first structural information was put forth by Liu et al. (1997) derived with the help of X-ray crystallography. The enzyme is functional as a dimer with two identical 452 amino acid residues. Each subunit has a catalytic domain, an NAD(P)$^+$ binding domain and an arm like bridging domain. These three structures can be identified structurally as the ALDH glutamic acid signature sequence MELGGNA, the Rossman fold which is the GXGXXG coenzyme binding site and the catalytic thiol (Kirch et al. 2004). In the eukaryotic system, ALDH of *Arabidopsis* is

completely sequenced within the plant genome that enables the classification of different families of ALDH and examines the phylogenetic and molecular relationship of all the ALDH genes.

In this study we focus on the ALDHs that are associated with environmental stress response. *Arabidopsis* has a gene ALDH7B4, which codes for a turgor pressure responsive gene which is activated when there is an environmental stress. This is also called as a stress response protein. Aldehydes are extensive in nature occurring in common things such as plant material and the environment. However, they are also generated in the living organisms due to the physiologically derived intermediates which arise during metabolism of compounds. Mostly, aldehydes are generated while metabolizing various xenobiotic compounds (Vasiliou et al. 2004). Environmental factors such as vehicular emission, smoke, pesticides, food additives etc. contributes to the generation of aldehydes. Also an important factor of aldehyde generation is lipid peroxidation of cellular phospholipids which induces the formation of about 200 reactive aldehyde species (Marchitti et al. 2008; Conklin et al. 2006; Ishitani et al. 1977). Besides a few essential roles, aldehydes by far are cytotoxic and carcinogenic. Aldehyde dehydrogenases (ALDHs) are enzymes which can effectively metabolize these molecules. Living organisms are always in contact with reactive oxygen species (ROS), which cause tremendous oxidative stress in living beings.

ALDHs are known to reduce oxidative stress particularly those that are caused by aldehydes. The induction of Aldehydes are found to be highest in prokaryotic and eukaryotic systems when exposed to biotic and abiotic shock (e.g. high temperature, low moisture, salinity, ultraviolet rays, pesticides and heavy metals). Mechanical trauma and injury are also responsible for its induction. A large number of ALDHs are expressed in organisms like yeasts and *C. elegans* in

response to a variety of oxidative stress. In plants, ALDHs are found throughout the cell in the cytosol, mitochondria, peroxisomes, endoplasmic reticulum, leucoplasts, and chloroplasts, and their expression patterns are tissue-specific (Li et al. 2000; Missihoun, 2011).

Plants have few ALDHs that seem to express constitutively, while others require an array of biotic and abiotic stresses to be induced. Most plant ALDH gene expression has a highly species-specific response to stress. Studies in experimental plant models investigated the defensive behavior of ectopic ALDH against stress and also their increased vulnerability to stress as a result of ALDH deficiency. Most of the plant based ALDH are highly responsive towards dehydration and salinity. A water imbalance in plant cells and tissues causes an osmotic stress and as a result there is a simultaneous increase in ROS generation. The upregulation of ALDH has multiple functions too. ALDH7 and ALDH10 are involved in the production of large number of osmolytes including betaine and dimethylsulfoniopropionate for normalizing osmoregulation and stress (Rodrigues et al. 2006).

The plant ALDHs helps to maintain the overall cell functioning and normal metabolic activities in conditions where there are no stressors by synthesis and catabolizing biomolecules such as carbohydrate lipids and amino acids. But the versatility of ALDH in promoting stress tolerance is being studied for the generation of crops which are resistant towards environments having extreme temperature, water scarcity, heavy metal pollution and nutrient deficiency.

The cellular targets for ROS species are DNA, RNA, proteins and lipids. Majority of the damage is caused due to highly reactive hydroxyl radicals generated from H_2O_2, which happens by the Fenton reactions and require divalent metal ions like Fe and Cu and NADH as a source of reducing equivalent (Figure 2.9).

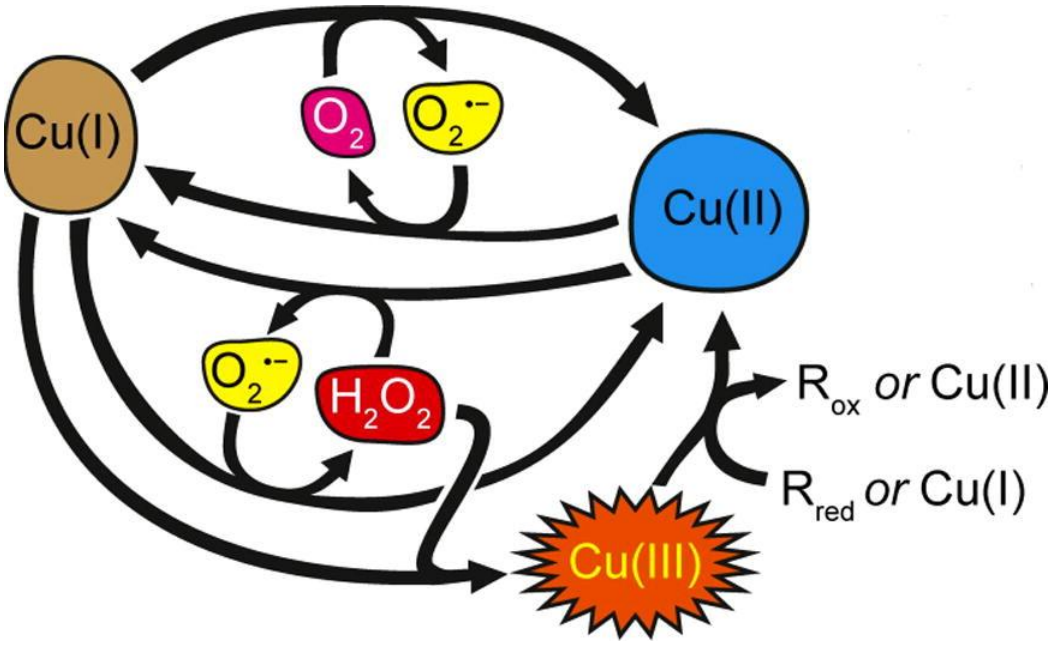


Figure 2.9: Generation of highly reactive hydroxyl radical in presence of Copper. Adapted from Pham et al. (2013).

The toxicity of copper lies in the fact that its existence in a free hydrated form (Cu^{2+}) can be potentially toxic to organism and plants. They can alter the membrane permeability and affect chromatin structure, inhibit various enzyme activities and hinder in protein synthesis. The ability to take part in generation of hydroxyl radicals ($HO^{\cdot}$) and other ROS as a result of reaction between H_2O_2 and copper in a Fenton-like reaction, directly increases oxidative stress and causes neurodegenerative diseases like Alzheimer's and Parkinson's disease. Together with aldehydes which include a majority of the products of lipid peroxidation, contribute to the pathophysiological effects alongside oxidative stress. Increased ALDH activity thus indicates a defense strategy in the detoxification of aldehydes and increased tolerance towards abiotic stresses (Pham et al. 2013).

2.10 Serine Protease inhibitors (Serpins)

Protease Inhibitors are a group of enzymes that are widely present in plants, animals and insects. They target serine proteases as these proteases contain a catalytic triad in their active site having

a nucleophilic serine residue. They initially form a complex with the target protease and irreversibly inhibit the proteolytic activity mainly through conformational change and thus disrupting the active site. Protease inhibitors are grouped into cysteine, serine, aspartic and metallo protease inhibitors. Out of these, most abundant are the serine protease inhibitors that are responsive towards environmental stresses like hypersalinity, radiation, osmotic stress and heavy metals (Rehman et al. 2017).

Stucturally, most serpins are identical comprising of 7-9 α-helices and 3 β-sheets. The primary feature of Serpin is the reactive center loop (RCL) that binds to the target serine protease protein near the C terminus of the sequence. Serpin contain a scissile bond in the RCL and once cleaved, a major conformational change is triggered. This unconventional conformational change is exhibited on cysteine and serine proteases that catalyze the cleavage of peptide bonds. The RCL is still bound to the protease when the protease hydrolyzes the serpin. This makes the serpin change its state of energy. It transitions from a stressed to a relaxed or low energy state (S to R transition). Consequently, the bound protease is pulled from top to bottom of the serpin, distorting the protease's active sites. The distorted protease remains covalently bonded to the serpin for days (Figure 2.10). Serpins are classified as irreversible or suicidal inhibitors as every time the protease is inactivated, it causes the serpin to lose its activity too hence becoming a ‘one time use’ protein (Huntington et al. 2000).

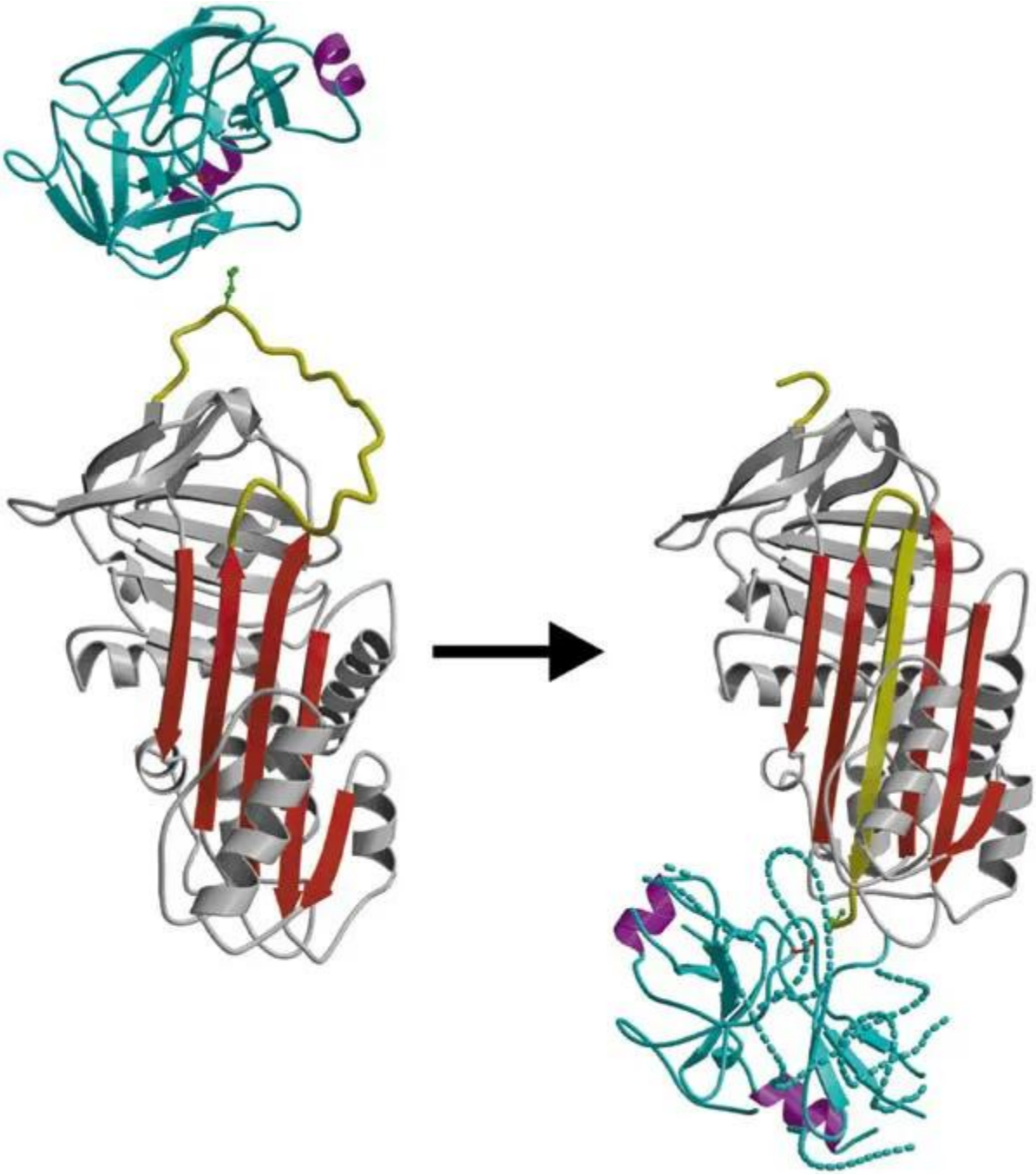


Figure 2.10: The mechanism of serine protease inhibition by serpin. Trypsin aligns with the serpin and complex is formed by shifting of residues in the RCL of the serpin cleaved by the action of trypsin. The complex formed disordered structure and remains attached to the serpin for extended perios of time. Adapted from Huntington et al. (2000)

When the structural integrity of the protease is lost, it is susceptible to attack by other protease and eventually degrades.

In nature serpins are distributed widely and is present in almost all eukaryotes and also in some viruses. However, fungal serpins are rare and was found in the anaerobic fungus *Piromyces sp.* Most pronounced occurrence of serpins are in plants. It was believed that they were absent from bacteria was but studies have proved this wrong (Irving et al. 2002). The occurrence of serpins is both intracellular as well as extracellular; however, in most multicellular organisms they are more of an extracellular protein. Mostly they have been located within the cytoplasm; however, reports suggest other possible sites of occurrence such as the nuclear compartment in case of some intracellular serpins (Stein and Carrell, 1995).

About 16 serpin groups have been identified in human which are termed from serpin A to P and they include 29 inhibitory and 7 non-inhibitory serpins. Certain mammalian serpins were identified having no orthology with the human counterpart. In insects, the *Drosophila melanogaster* has about 32 genes encoding for serpins. The nematode *C. elegans*, which has been sequenced and analysed, revealed eight sequences for serpin-like proteins. Plant serpins have been extensively studied and were the first member of the superfamily that was identified. Also, viral serpins have been identified, e.g. serpin CrmA found in the genome of cowpox virus that aims to neutralize host caspases to prevent cell death (Zhou et al. 1997). In eukaryotes, the serpins have both inhibitory and non-inhibitory roles. Non-inhibitory serpins include functions like binding to hormones and transport of the hormone thyroxine (Pemberton et al. 1988). Plant serpins are the most extensively studied group of serpins. They also form a distinct group in the evolution of serpins, which is considered evolutionary divergence from a common ancestry of animals and plants (Cohen et al. 2019). Two protease inhibitors studied extensively in plants are Bowman-Birk and the Kunitz trypsin inhibitors. The Bowman-Birk inhibitors (BBIs) are found

abundantly in di and mono-cotyledonous plants. These BBIs are found to be stable to heat and acidic pH due to the presence of a number of disulphide bonds. During the evolutionary process, the number of cysteine residues has remained to be conserved for the BBIs. The Kunitz trypsin inhibitor was first isolated in the year 1945 by Kunitz which was a soyabean trypsin inhibitor having a molecular mass of 18-22 kDa, which inhibits both trypsin and chymotrypsin (Kunitz, 1945).

The role of a serpin is prominent in the environmental responses that it exhibits (Huang et al. 2007). Adverse environmental stresses like mechanical wounding and microbial infection can significantly enhance the level of protease inhibitors. Oxidative stress (H_2O_2) and other abiotic stresses can lead to the induction of protease inhibitors, majority of them belonging to the Kunitz family of serpins. Similar study identified that these PIs are up-regulated under high salt and also in low moisture conditions (Gosti et al. 1995).

Several studies were also conducted on plants like Potato and *Brassica napus* under water deficient environments (Downing et al. 1992). Protease inhibitor motif was found to be induced by drought and temperature stresses (Satoh et al. 2001). Salt stressed tomatoes showed signs of strong induction of serpin (Dombrowski, 2003). A rice chymotrypsin inhibitor was also found to be induced under dehydration and ABA treatments (Huang et al. 2007). Abiotic stress in terms of exposure to heavy metals also induced the protease inhibitor gene in Populus roots and when this gene was isolated and cloned, the expression of the gene resulted in the accumulation of heavy metals. All these reports clearly indicate that the protease inhibitors are also involved in abiotic stress, although their exact role in the tolerance of abiotic stress is still to be identified.

2.11 DHHC Palmitoyl transferase

Palmitoyltransferase protein assists in the reversible post-translational addition of palmitic acid groups on cysteine residues with the help of thioester linkage. This process is termed as S-palmitoylation. Almost 10 % of the protein encoded by a human genome is acylated at cysteine residues with fatty acids. Majority of the proteins are modified by DHHC Palmitoyl transferase activity. This post translational modification increases the hydrophobicity of the proteins and increases its affinity towards the membrane thus providing anchorage (Martin et al. 2012).

Protein palmitoylation was first described in the early 1980s where several proteins have been identified as substrates of palmitoylation. To study palmitoylation, the classical methodology was incorporation of ^{3}H palmitate after they were extracted from cells by immunoprecipitation. Proteins like G protein-coupled receptors, Ras proteins, neurotransmitter receptors, endothelial nitric oxide synthase etc. were identified using this approach (O'Brien and Zatz, 1984; Buss and Sefton, 1986; Robinson and Busconi, 1995). Palmitoylation's primary role is to provide a hydrophobic membrane for many proteins that lack a transmembrane domain, so that membrane anchorage is possible. As a consequence of membrane attachment, the proteins are directed to a particular membrane compartment e.g. many peripherial proteins which are palmitoylated at the Golgi travel towards plasma membrane (Rocks et al. 2010). This method of protein sorting is not limited to proteins without anchorage, but it also modifies and directs the transmembrane proteins towards intracellular locations (Greaves et al. 2009). Apart from trafficking of proteins, palmitoylation regulates protein stability as well as protein–protein interactions. Significant work was done regarding *S. cerevisiae* where effect on Ras function due to palmitoylation was studied (Lobo et al. 2002). Also yeast palmitoyl transferase Akr1 was proved to be essential for the palmitoylation of yeast casein kinase. It was revealed that palmitoylation activity of both the Ras

proteins and Akr1 exactingly contained an intact DHHC motif (Mitchell et al. 2010). Further studies in yeast have discovered that for the activity of the palmitoyl transferases, a cysteine-rich group is essential to the DHHC protein family along with protein cofactors. However, it was also revealed that other DHHC proteins can function independently. Studies have also revealed the localization of the proteins having distinct intracellular membranes (Politis et al. 2005).

The process of protein palmitoylation is one of most vital post-translational modifications which is also reversible and affects a variety of proteins by controlling the cellular trafficking and membrane localization functions. It is a two-step process which transfers the palmitoyl group to their substrates. This process begins with the autoacylation of the enzyme and in the second stage; it transfers the attached palmitoyl-residue to the target protein (Figure 2.11).

Though palmitates are the most common group of fatty acid attached to cysteines, the process of palmitoylation can attach other acyl groups with different chain length and degree of unsaturation to cysteines groups in a similar manner (Liang et al. 2001). The S-palmitoylation process is normally catalyzed by the polytopic transmembrane protein class known as protein acyl transferases (PATs), consisting of a zinc-finger and DHHC (aspartate–histidine–histidine–cysteine) domain. There are over 23 members of Zinc finger DHHC family of genes in mammals (Korycka et al. 2012).

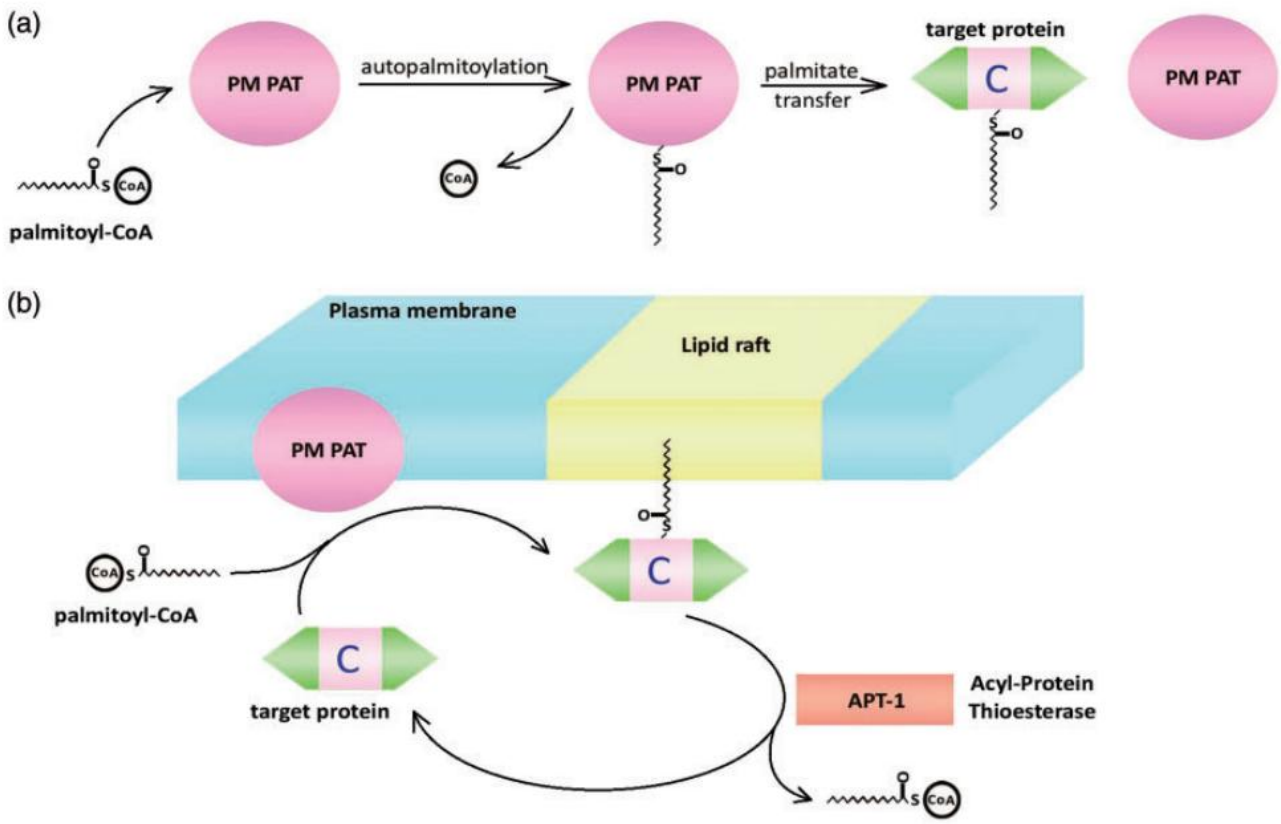


Figure 2.11: Mechanism of palmitoylation. In the first step, the palmitoyl-CoA reacts with the enzyme which leads to the formation of a PAT intermediate, releasing the CoA. In the second step, transfer of palmitate to the target protein takes place (b) the cycle of palmitoylation and depalmitoylation. This enables the target protein to be localized to various cellular compartments like the lipid layer of membranes and the removal of the palmitoyl group by the action of acyl-protein thioesterase, relocates the protein back to the cytosol. Adapted from Tabaczar et al. (2017).

The most common structural characteristics of a DHHC palmitoyl transferase are the transmembrane domains which have N- and C- termini in the cytosol (Figure 2.12)

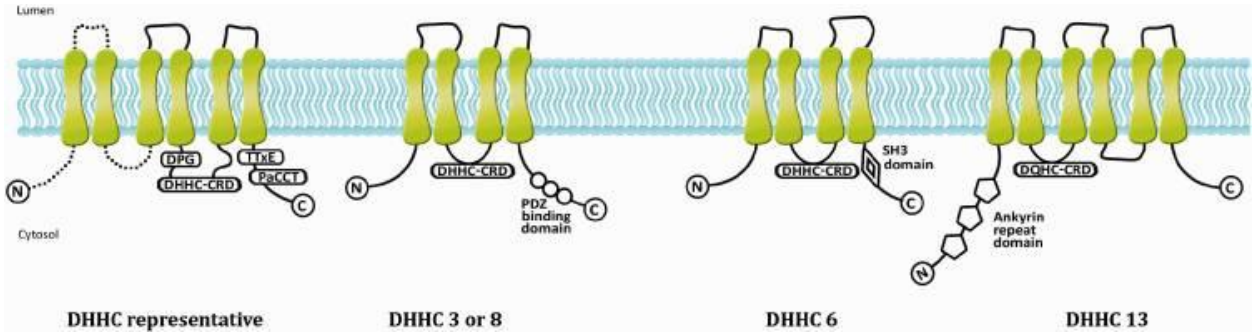


Figure 2.12: The predicted structure of DHHC acyltransferases. They usually have four transmembrane domains (six rarely). The conserved, catalytic DHHC motif, where the protein becomes autopalmitoylated, is located in a cytosolic region. The signature region of DHHC contains a cysteine rich domain (CRD-domain). DHHC proteins contain a conserved DPG motif facing the cytosol. Moreover, most DHHC enzymes have a C-terminal signature domain consisting of threonine-threonine-x-glutamate (TTxE; x stands for any amino acid residue) and palmitoyltransferase conserved C-terminus (PaCCT) motifs that are predicted to localize to the cytosol. DHHC13 is the only DHHC protein with DQHC motif. Adapted from Tabaczar et al. (2017).

Usually, the DHHC signature is embedded in a larger cysteine-rich domain (CRD-domain) which is a region of approximately fifty amino acids that depicts many conserved amino acids with a canonical $CX_2CX_9HCX_2CX_2CX_4DHHCX_5CX_4NX_3FX_4$ domain. The process of autoplamitoylation occurs at the cysteine residue of the DHHC motif and forms the most crucial functional part of the protein involved in activity of several PATs. Studies related to mutation of this domain have concluded the loss of enzyme function in both *in vitro* and *in vivo* conditions (Valdez-Taubas and Pelham, 2005; Roth et al. 2002; Smotrys et al. 2005). Studies pertaining to localization of DHHC proteins have indicated Golgi and the endoplaspic reticulum as the primary sites. Similar studies have proved that the DHH superfamily of proteins has different

specificity towards substrates and act as various important enzymes e.g. phosphoesterases, exonucleases, phosphatases etc. This protein superfamily plays significant roles in replication, recombination, repair, nucleic acid metabolism, stress tolerance and homeostasis maintenance (Srivastav et al. 2018). The DHHC protein contains a zinc finger domain which has a conserved pattern of cysteine and histidine residues. The primary role of zinc fingers is the binding of the metal Zn which provides a structural stability to the protein, and any alterations in the genetic information of the protein could lead to the loss of protein functions. There are also several reports which emphasize the importance of a selective pressure for the maintenance of zinc finger in the protein (Gottlieb and Linder, 2017). The most studied properties of the zinc fingers genes are the responses towards external metals. A number of reports have shown an increase in the expression of the zinc finger with the increase in external metal concentration like copper and cadmium, in which a zinc finger transcriptional factor played important role in tolerance through glutathione-dependant pathway (Chen et al. 2016).

2.12 Maltose fermentation regulatory protein

Saccharomyces cerevisiae are the most important yeast species which are industrially important in brewing and baking. The ability of the yeast to utilize maltose is fundamental to its activity in fermentation. The maltose inducible expression of the MAL gene results in the utilization of maltose. The MAL loci are present in the telomeric region of various chromosomes. These five loci are MAL1(VII), MAL2(III), MAL3(II), MAL4(XI) and MAL6(VIII). Each MAL locus has a set of genes which are responsible for the metabolism of maltose. However, all the MAL loci show incredible homology between the sequences as well as at the functional level. Each of these gene loci comprises of a three gene complex containing MALR (regulatory or activator), MALT (maltose transport or permease) and MALS (structural or maltase). The MAL activator gene is a

DNA-binding transcription activator essential for the induction of the structural genes (Hu et al. 1999; Needleman, 1991). In a report by Charron et al. (1989), the MAL63 gene, which is one of the nearly identical genes referred to as MAL13, MAL23, MAL33, MAL43 or MAL63 depending upon the locus position. The structure of the predicted Mal63p protein is a sequence of length 470 amino-acids. The most important domain of this protein is the presence of a cysteine-rich amino-terminal region which share homology with the zinc-cluster DNA-binding proteins having cysteine-rich domains previously characterized in yeast activator proteins like Gal4p (Kim and Michels 1988). All these genes are transcribed under the regulation of a common promoter. The individual genes related to Maltose fermentation have been studied with mutation analysis such as the N-terminal region of the cysteine-rich region of Mal63p, clearly demonstrated the importance of the domain for its regulatory function. The function of this domain was predicted, as it demonstrated homology with Gal4p containing DNA-binding zinc-cluster region of which formed a bi-nucleate metal-ion binding cluster (Vallee et al. 1991). In yeast *S. cerevisiae* the region responsible for DNA binding was also a conserved region, and the region adjacent to the C-terminal domain was suggested to contain the dimerization domain of the proteins necessary for the binding of the DNA. The C-terminal domain (284-470) of the activator protein was identified as a region whose absence did not affect the constitutive expression of MAL63 gene (Hu et al. 1999). However in the MAL gene, various constitutive mutations have been observed throughout the C-terminal i.e. from residue 300 to the end, which did not produce a constitutive activator. This suggests that the C-terminal had a positive effect on MAL-activator, which is needed to negate the inhibitory effect of the some of the residues adjacent to the DNA binding domain.

One of the major families of proteins that carry out transcriptional regulation in eukaryotes is the Zinc-binding proteins. They display variable secondary structures and are functionally diverse. The zinc finger consists of α-helix and a pair of anti-parallel β strands. Zn(II)2Cys6 DNA binding domains, like those of other DNA binding domain classes, interact with similar DNA binding sites (Schjerling and Holmberg, 1996) which have revealed sequence and well as structural similarities of the DNA binding sites recognized by Zn(II)2Cys6 proteins. The major classes of zinc binding proteins along with their domain structures are shown in Figure 2.13.

Nearly the majority of DNA-binding zinc proteins have multiple zinc atoms. Consequently, structural data gives us the measured inter atomic distances between the Zinc atoms which could be of significant importance. A sugar regulatory protein like GAL4, contains six cysteine residues that forms a coordination site for two zinc atoms thus forming a zinc cluster. There are also two zinc atoms in the glucocorticoid receptor, but these form two separate sites. For every site, there are four different cysteine ligands that bind to one zinc with an interatomic distance of 13 Å between zinc ions (Figure 2.13). In TFIIIA, such inter-zinc ionic distances are not known because a three-dimensional structure has yet not been documented.

The most important characteristics of DNA transcriptional control, DNA packaging, DNA repair etc. is the binding of regulatory proteins to the DNA. When these structural characteristics were analyzed, it revealed distinct domains which are able to interact with DNA. The six-cysteine (Zn(II)2Cys6 or C6 zinc) binuclear cluster DNA binding domain was first characterized in the yeast *Saccharomyces cerevisiae* GAL4 protein. The arrangement of the six cysteine residues which are arranged in a manner of $CX_2CX_6CX_6CX_2CX_6C$, are able to coordinate with 2 zinc(II) ions to form a cloverleaf-shaped structure (Figure 2.14) (Gardner et al. 1991). Functional domains of the zinc cluster protein are depicted in Figure 2.15.

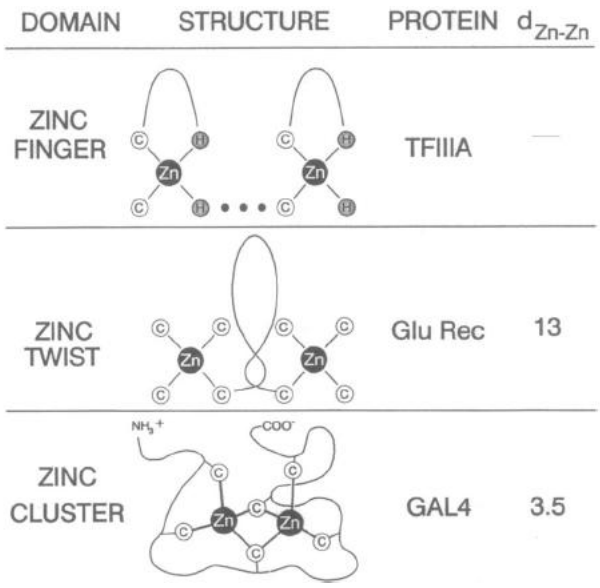


Figure 2.13: Different types of zinc ion binding sites present in DNA binding domains of transcriptional factors. The zinc-zinc interatomic distance, d_{Zn-Zn} is given in Å. Adapted from Vallee et al. (1991)

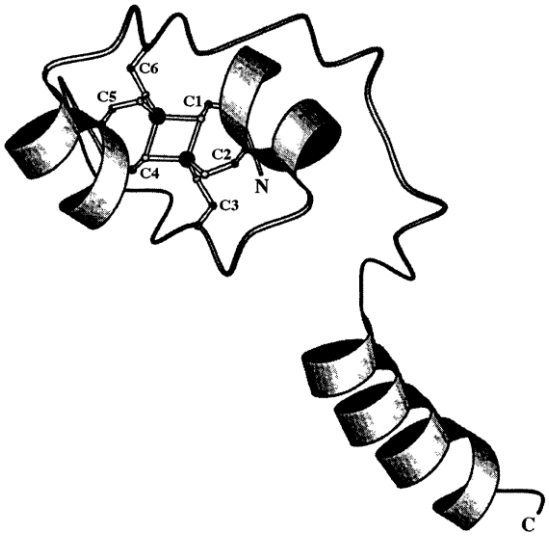


Figure 2.14: The structure of MAL regulatory protein. The Zn(II)2Cys6 cluster is marked on the left with the Zn(II) ions represented by large solid spheres at the center of the structure. The amino (N) and carbon (C) terminal ends are denoted. Adapted from Todd and Andrianopoulos (1997)

Figure 2.15: Functional domains of Zinc cluster protein. Zinc cluster proteins can be divided into three functional domains: the DNA binding domain, the regulatory domain with middle homology region, and the acidic region. The DNA binding region is compartmentalized into three subregions: the zinc finger, the linker, and the dimerization domain. Adapted from MacPherson et al. (2006).

A significant percentage of zinc cluster proteins found in *Saccharomyces cerevisiae* are believed to have a role in the cell's response to stress, as the yeast cells definitely rely on this group of regulators to manage environmental stress. More often than not, the zinc finger proteins are targets for toxic metals. Even at very low concentrations, these metals such as cadmium, nickel, cobalt etc. interfere and disrupt various cellular processes primary being DNA transcription and repair. To reveal possible interactions between toxic metals with zinc-binding protein domains, several studies have been conducted. The structural and functional changes due to displacement of zinc by cadmium or cobalt have been stuied. Similarly, due to incomplete coordination of toxic metal ions, the formation of mixed complexes has also been studied (Hartwig, 2001).

A major feature of zinc binding proteins is the role of zinc ion, which does not take part in interacting with the DNA directly but is instead necessary for the right folding of the protein. This enables DNA binding and other protein-protein interactions. Circular dichroism studies indicated the metal dependant folding of zinc finger transcriptional factor IIIA (Frankel et al. 1987). In the absence of the metal ion, the protein tends to lose its function. However, studies have reported that the extent of binding of the zinc finger domain with metal ions such as Co^{2+}, Cd^{2+}, Fe^{2+} and Ni^{2+} were many times lower than the binding affinity with Zn^{2+}. Nevertheless, the

pH of the environment plays a major factor in determining the strength of bond (Krizek and Berg, 1992). With respect to Co^{2+}, the dissociation constants were significantly lower for Zn ions and the affinity for Cd noticeably increased by five times with the of cysteine residues. Thus it results in preferential binding of Cd over Zn (Krizek et al. 1993). Several studies pertaining to the effects of metal replacement on zinc finger transcriptional factors have been studied e.g. TFIIIA, transcription factor Sp1, metal-response element-binding transcription factor 1 (MTF-1), as well as Tramtrack transcription factor (TTK) (Makowski and Sunderman, 1992). Transcriptional factor Sp1 containing the C-terminus zinc finger which binds to the GC rich promoter binding sites, was found to weaken in the absence of Zn but was recovered and function was restored in the presence of Hg, Pb, Cd, Co (Cook et al. 1999).

2.13 Chitin synthase

The enzyme chitin synthase (EC 2.4.1.16) belongs to the family glycosyltransferases. The substrates for this enzyme to catalyze are Uridine Diphosphate-N-acetyl-D-glucosamine and [1,4-(N-acetyl-beta-D-glucosaminyl)]n, to form long chain polymer of N-acetylglucosamine units in β-(1→4)-linkage (also called as chitin) and Uridine Diphosphate.

The core polysaccharide constituent of the cell wall which encapsulates the fungal cell is chitin. Two major functions of chitin are providing strength and rigidity to the cells by forming an interaction between the fungal cell and their host. For the yeast *S. cerevisiae*, there are three chitin synthase genes, Chs1, Chs2, and Chs3. Though the Chs proteins originate in the rough endoplasmic reticulum, they operate mostly in the plasma membrane and are an important component of the cell wall. Other major components of the cell wall include glucan and mannan, which are synthesized continuously during the cell cycle and are distributed uniformly over the cell surface. On the other hand, the synthesis of chitin occurs only during certain points of the

cell cycle and is distributed asymmetrically in the cell wall (Cabib et al. 1982). Chitin is normally deposited between the neck, in between the mother and the emerging bud during the formation of the division septum. The process of chitin deposition during emerging of the bud and formation of primary septum is governed by the actions of Chs2 and Chs3.

The synthesis of chitin synthase is facilitated by trypsin and other serine protease inhibitors in various fungal and insect systems. This indicates that chitin synthase is basically synthesized as a zymogen which requires activation. However, the presence of high concentrations of N-acetylglucosamine stimulates all the three yeast chitin synthase up to a few folds (Sburlati and Cabib, 1986). When the Chs1 is activated by proteolysis, it shows the maximum *in vitro* activity of chitin synthases assayed in membranes from wild-type cells (Orlean, 1987). These assayed levels of Chs1 activity from membranes were found to be same in both exponentially growing and yeast cells in the stationary phase of growth (Orlean 1987), also the levels of Chs1 remain unchanged throught the process of the cell division (Ziman et al. 1996). However the activity of Chs1 increased *in vitro* and became detectable only after trypsin activation. Chitin synthase I and II are together responsible for only less than 10% of the cellular chitin. It has been observed that in specific harsh environmental conditions like in acidic media, chitin synthase I is necessary for normal budding. The requirement of chitin synthase II in yeast cells is for normal morphology, septation, and cell separation. On the other hand, chitin synthase III is accountable for the synthesis of 90% of the cellular chitin. A representation of basic structure of the fungal cell wall is depicted in Figure 2.16.

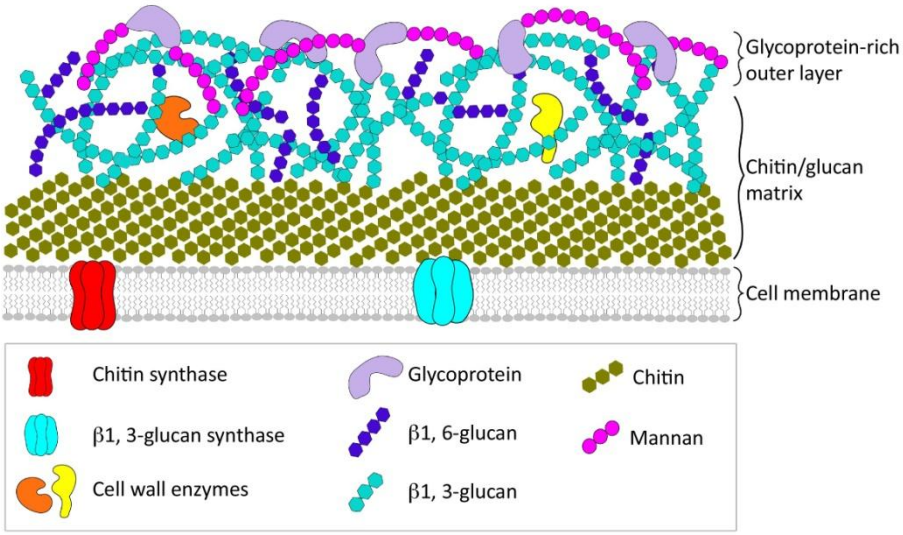


Figure 2.16: Various enzymes involved in the formation of the chitin-glucan rich matrix. Adapted from Geoghegan et al. (2017)

Although chitin synthesis is not entirely dependant upon Chitin synthase 1 entirely, reports have indicatd a network of proteins that interact with the genes CHS1 and CHS3. A total number of 57 genes were identified which were divided into two categories. Few of these genes from the first catergory, when mutated, affected the integritiy of the cells as they impacted β-1,3-glucan synthesis. This indicated the essentiality of Chs1 in providing a buffer for the cell wall when faced with changes or mutations, affecting the robustness of the cell. In the other category, the genes of the CHS 1 interaction network, involved in budding and recycling of endocytic proteins, could directly influence the functioning of Chs2 protein. This implies that Chs1 also acts as a defence against the inadequacies occurring to Chs2. However, Chs1 and Chs3 network genes are quite distinct from each other. Mutations in CHS1 itself or any of the related network genes do not have any effect on the Chs3-dependent chitin stress response. Therefore Chs1 and Chs3 have different functions and hence do not compliment each other (Lesage et al. 2005).

Each chitin synthase thus has a different role when it comes to cell wall and septum synthesis. The basic reaction it catalyzes is the same which is the transfer of N-acetylglucosamine (GlcNAc) from UDP- GlcNAc to form a growing chitin chain. It is however, difficult to measure one enzyme activity alone in presence of the others. However, the availability of mutant strains lacking one or more than one chitin synthase activity can be helpful in measuring the functions. For the functioning of chitin synthase, the requirement of divalent cations is essential. In a report by Sburlati and Cabib (1986), the differences between the activities of Chs1 and Chs2 were identified. The divalent cation requirement for both the enzymes to function was different. It was found that Co^{2+} highly stimulates the activity of Chs2 while the activity of Chs1 was inhibited. Also, different pH optimum of the two enzymes was recorded, 6.5 for Chs1 and 7.5-8.0 for Chs2. The specificity of the different chitn synthase towards different metals was further demonstrated by Choi and Cabib (1994). The stimulation of Chs1 activity by Mg^{2+} was evident; however, the activity was inhibited by the presence of Co^{2+} and Ni^{2+}. The activity of Chs2 was enhanced by Co^{2+} but not much with Mg^{2+}. Both the enzymes; however, were inhibited in the presence of Ni^{2+}.

Expression or functioning of chitin synthase under stressed conditions have been studied in fungal systems especially those which are pathogenically important. The polymorphic *Wangiella dermatitidis* was studied with respect to the chitin synthase genes due to its invasive strategy during infection (Harris and Szaniszlo, 1986). The isolated chitin synthase gene was able to highly express under stress conditions like changes in temperature, pH and also it was found that the total activity was stimulated with trypsin treatment (Wang and Szaniszlo, 2000). The presence of metals in the environment is by far, the most critical environmental stressor. Contaminating metals such as Zn, Cu, Mn, Fe etc. have been toxic for fungi as well. The

response of fungi towards excess of metal in the surroundings comes from both cellular and molecular defense mechanisms, which leads to various levels of tolerance behaviors. This has lead scientists to study the different cell wall biogenesis mechanisms in presence of toxic metals and the variation in the level of chitin synthase transcripts occurring in different fungi during the life cycle and morphogenetic events. The tolerance to metals comes from the ability of fungi to form complexes with the cell wall which is considered as a key detoxification mechanism. The fungi *Rhizopus stolonifer* and *Cunninghamella blakesleeana* that were exposed to elevated concentrations of copper, also acquired an elevated tolerance towards other toxic concentrations of cadmium, cobalt, nickel, and lead. In this study, the mycelia was able to remove twice as much copper from solution which suggested that the tolerance to copper resulted from cell wall-metal complexation or by other intracellular detoxification mechanism (Garcia-Toledo et al. 1985). The cell wall-metal association may be due to a variety of factors like ion exchange, cell wall absorption, metal complexation, precipitation and crystallization, as fungal cell walls contain potential sites involved in metal sequestration e.g. due to the presence of carboxyl, hydroxyl, amine, phosphate and sulfhydryl groups (Strandberg et al. 1981). Studies have proved that chitin is responsible for adsorption of metals, and also different derivatives of chitin are also responsible of chelation of metals to chitin. Various derivatives of chitins for e.g. chitin phosphate (P-chitin) or DA-chitin phosphate (P-DA-chitin) can be prepared very conveniently which showed high Cu^{2+} binding ability (Nishi et al. 1987). In yeast *Saccharornyces cerevisiae*, cell wall binding of metals was studied in greater details. The cell wall was isolated using chemical and enzymatic processes and subsequently they were allowed to react with heavy metal cation solutions. Accumulation of metals on cell wall material was determined. It was observed that the Cu^{2+} ion was accumulated to a greater extent compared to Co^{2+} ions. This study reported

the determination of the major components of the yeast cell wall that are involved in the binding of metal cations (Brady et al. 1994).There are reports on the studies on the chitin-metal interactions with purified chitin or chitin derivatives. However, study of chitin-metal complexes ^{13}C-NMR, X-ray diffraction and IR studies on cell wall of *Neurospora crassa* to find out the chitin-cadmium complexation when grown under conditions of Cd-toxicity, was reported by Bhanoori and Venkateswerlu (2000) where a novel chitin-cadmium binding model was proposed and the studies revealed that the conformation of chitin was not altered even after binding of cadmium.

Studies on an ascomycete which are able of establish ericoid mycorrhizas, was used to examine the effect of zinc ions on the cellular mechanisms of fungal growth. Though a significant reduction of the fungal biomass was observed when treated with zinc; an apparent visual alteration of cell wall structure was observed as the metal stimulated apical swellings and hyphal branching. Further examination proved that localization of chitin resulted in the inner parts of the hyphal walls of the fungus. In another study by Lanfranco et al. (2004) which proved that exposure to toxic amounts of zinc resulted in the accumulation of Chs 4 mRNA, which was higher than the accumulation of Chs 2 mRNA. Though the content of chitin in metal-treated mycelia is considerably higher and it increases with increase in metal concentration of a Class IV CHS, it remains to be seen whether other genes of the isoforms of the enzyme are affected when exposed to metals. The effect of toxic concentrations of copper on cell wall composition was first studied on *Fusarium oxysporum*, where the strain was grown up to a concentration of 600 mg/L. It was reported that the content of copper in the cell wall of the mycelia was 1.5 times higher in presence of 600 mg/L than that of the cells obtained from mycelium grown at 200 mg Cu/L. The

chelation of metals to cell wall was attributed to the fact that chitin and chitosan are strong metal chelators.

The strong adsorbing capability of chitn and chitin-derived compounds were studied by (Labidi et al. 2016). The work focused on the comparison of different forms of copper adsorption on chitin-based adsorbents, *i.e.* chitin (CH), chitosan (CS) and chitosan- ethylenediaminetetra-acetic acid (CS-EDTA), which could be used for removal of toxic copper concentrations from water. Removal of metals by adsorption of toxic metals such as copper, chromium and arsenic from waste materials utilizing chitin or chitin derived compounds were studied as effective remediation strategies (Kartal and Imamura, 2005).

To sum up, the approach of metatranscriptomics to study the eukaryotic microbial population is a dependable and robust tool to discover novel functions and possible roles of eukaryotic microorganisms concerning soil ecosystem. The survival strategies in heavy metal polluted soil have encouraged scientists to develop clear-cut holistic approaches to identify and discover new and underrepresented, and yet, important microbes and functions. To understand the complex and dynamic nature of soil microbial community function and environmental impact due to metal contamination, it is essential to focus on the expressed subset of total genes and not work with the entire metagenome (Wasserman et al. 2008). Here metatranscriptomics helps us to understand the responses of microbial population with environmental modification or perturbation, which will reveal the services performed pertaining to nutrient mobilization, breakdown of complex matter, imparting tolerance or resistance in response to heavy metals in environment etc. Sequence based discovery of participating organisms and their functional information from the RNA reveals a lot more than what isolation or cultivation based techniques would have revealed.

The genes and their products isolated through metatranscriptomics approach is a result of overexpression which is triggered due to metal stressed condition (Mitchell et al. 2011). This environmental condition increased the activity of a particular community which would otherwise be dormant under ordinary conditions. The genes thus obtained, contribute to the huge repository of information in the form of sequences, which is expanding due to high-throughput sequencing facilities that everybody has access to. As the soil biota is well represented, more and more taxonomic annotations are assigned to previously unknown species, especially eukaryotes.

The eukaryotic genes recovered from polluted soil through a metatranscriptomic approach thus can be used to study unknown taxonomic groups and their contribution towards metal tolerance. This will have great biotechnological importance such as in biomarkers or organism based bio-sensors. It will also help us manipulate the metal resistance level in plants which would be useful for remediation and revegetation of polluted soil (Khan 2005). Though these are lucrative options, they do not come hurdle-free. Finding suitable host for expression of environmental genes other than yeast can be a problem but there are ongoing studies carefully designed to express these genes in other living systems for example Damon et al. (2011) successfully introduced an environment oligopeptide transporter into *Xenopus* oocytes which could successfully transport dipeptides. Once they can be suitably expressed in different hosts, the novel genes and its products in the form of biocatalysts would thus open up numerous profitable paths in the area of biotechnology.

Chapter 3: Materials and Methods

3.1 Biological materials, chemicals and growth media

Biological materials used in this study include yeast strains, bacterial strains and plasmids. All mutant yeasts and their parent type were obtained from Euroscarf (http://www.euroscarf.de/) except *pbi2*$^{\Delta}$ and *akr1*$^{\Delta}$ which were kindly donated by Dr. A.K Pandey (National Agri-Food Biotechnology Institute, Mohali). Details of yeast stains used are provided in Table 3.1. Details of primers used in this study are provided in Table 3.2. Standard media components were obtained from Takara (Japan), Clontech (USA), Sigma Aldrich (St. Louis, MO, USA) and HiMedia Laboratories (India). For growth and propagation of all yeast strains, YPD medium was used (Appendix I). Synthetic defined media (SD) without Uracil supplemented with 2% glucose was used for yeast screening experiments. SD-Ura medium is suitable for complementation of auxotrophic yeast mutants lacking a functional Ura3 gene. Luria broth was used for bacterial growth. The strains DH5α and DH10β were used for transformation. Electrocomp™ *E. coli* (Thermofisher, USA) was used for bacterial transformations. Super Optimal broth with Catabolite repression (SOC) medium was used for cultivation of transformed bacterial cells. A modified plasmid DNA pFL61 with SfiIA and SfiIB restriction sites were used throughout this study (Minet et al. 1992; Yadav et al. 2014). All media solutions were sterilized by autoclaving at 121 °C and 15 psi for 20 minutes prior to use. Protease K and lysozyme were purchased from Sigma Aldrich (St. Louis, MO, USA). Hydrogen peroxide (H_2O_2) was purchased from Merk Millipore (USA).

Table 3.1: Yeast strains used in this study

Yeast Strain	Genotype	Sensitivity	Reference
BY4741	*MATa his3$^{\Delta}$1 leu2$^{\Delta}$0 met15$^{\Delta}$0, ura3$^{\Delta}$0*	Wild type	Brachmann et al. 1998
BY4742	*MATα his3$^{\Delta}$1 leu2$^{\Delta}$0 lys2$^{\Delta}$0 ura3$^{\Delta}$0*	Wild type	Brachmann et al. 1998
ycf1$^{\Delta}$	BY4741; *MATa; ura3$^{\Delta}$0; leu2$^{\Delta}$0; his3$^{\Delta}$1; met15$^{\Delta}$0; YDR135c::kanMX4*	Cadmium	Li et al. 1997
cot1$^{\Delta}$	BY4741; *MATa; ura3$^{\Delta}$0; leu2$^{\Delta}$0; his3$^{\Delta}$1; met15$^{\Delta}$0; YOR316c::kanMX4*	Cobalt	Conklin et al. 1992
cup1$^{\Delta}$ (DTY4)	DTY3 with *cup1::URA3*	Copper	Longo et al. 1996
cup1^{S} (DTY3)	*MATα leu2-3, 112 his3$^{\Delta}$1 trp1-1 ura3-50 gal1 CUP1^{S}*	Slightly sensitive	Welch et al. 1989
zrc1$^{\Delta}$	BY4741; *MATa; ura3$^{\Delta}$0; leu2$^{\Delta}$0; his3$^{\Delta}$1; met15$^{\Delta}$0; YMR243c::kanMX4*	Zinc	Li and Kaplan, 2001
sod2$^{\Delta}$	BY4742; MATα; *ura3$^{\Delta}$0*; *leu2$^{\Delta}$0*; *his3$^{\Delta}$1*; *lys2$^{\Delta}$0*; *YHR008c::kanMX4*	Oxidative Stress	Van Raamsdonk and Hekimi, 2009
akr1$^{\Delta}$	BY4741; *MATa; ura3$^{\Delta}$0; leu2$^{\Delta}$0; his3$^{\Delta}$1; met15$^{\Delta}$0; YDR264c::kanMX4*	Heavy metals	Roth et al. 2002
pbi2$^{\Delta}$	BY4741;*YNL015w*:*kanMX*	Serine Proteases	Schu et al. 1991

Table 3.2: Primers used in this study

Primer	Sequence	Use
pFL61 NF pFL61 NR	5'-CAGATCATCAAGGAAGTAATTATCTAC-3' 5'-CAGAAAAGCAGGCTGGGAAGC-3'	pFL61 vector specific amplification
3' CDS adapters	5’AAGCAGTGGTATCAACGCAGAGTGGCCGAGGCGGCC(T)4G(T)6C (T)13VN -3’	Preparation of cDNA
5' end PlugOligo adapters	5’-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGG-3’	
PCR primer M1	5’-AAGCAGTGGTATCAACGCAGAGT-3’	Amplifying cDNA

EF1F	5’- GTCGTYGTYATYGGHCAYGT-3’	Validation of size fractionation
EF1R	5’- TGYTCNCGRGTYTGNCCRTCYTT-3’	
TAReukF	5’- CCAGCASCYGCGGTAATTCC-3’	Eukaryotic degenerate primers for V4 region
TAReukR	5’- ACTTTCGTTCTTGATYRA-3’	

3.2 Collection of soil samples

Malanjkhand copper deposit (22°0’54’’N and 80°43’20’’E) is an open cast copper mine situated in Malanjkhand at a distance of 90 Km North East of Balaghat in Madhya Pradesh. It is India’s largest base metal copper open pit mine. Exploitation of copper deposits is presently carried out by M/s Hindustan Copper Limited. The zone of mineralization hosts scanty and patchy vegetation. Rhizopheric soil (1 foot from plant base) was collected from 15 different locations all throughout the mining area including tailing ponds. Soil samples were also collected from the agro-forest soil populated by Poplar trees in Pierrelaye (PL) (49°1′45”N, 2°10′32″E) near Paris, France, contaminated with toxic metals. This agricultural site was contaminated due to repeated usage of untreated waste water for irrigation purpose and contains elevated levels of metas like Cadmium, Copper, Zinc etc. This site is now populated with Poplar trees (Foulon et al. 2016). Composite soil was prepared to cover a wide range of soil microbial activity by combining equal volumes from each sample. They were then sieved (2 mm mesh) to remove debris and plant material, and equal volumes of each sample was combined to form composite samples. Soil samples were stored in disinfected plastic bags and preserved at -70 °C until RNA extraction. Physicochemical properties such pH, organic carbon (Walkley, 1947), total Phosphorous (Kitson and Mellon, 1944) available Phosphorous (Olsen, 1954) and total nitrogen (Piper, 1966) of soil samples were analyzed. To estimate the metal content of the soil, samples were digested with

nitric acid-perchloric acid in a ratio of 2:1 (v/v) (Idera et al. 2014). The contents of metals were analyzed by ICP-MS (Element XR, ThermoFisher Scientific, Germany).

3.3 Physicochemical properties of metal polluted soil

Physicochemical properties of soil sample were analyzed by various methods.

3.3.1 Total organic carbon (Walkley, 1947)

Organic carbon was determined by the following method:

Reagents

1) Potassium Dichromate Normal Solution - 49.035 g of potassium dichromate was dissolved in distilled water and volume was made upto one liter

2) $FeSO_4$ solution, 0.5 N solution: About 140 g of $FeSO_4$ was dissolved in 0.5 N H_2SO_4 to make 1 liter of solution. For 0.5 N H_2SO_4 , 14 ml of concentrated sulphuric acid was added to water to make 1 liter of solution.

3) Sulphuric Acid, Concentrated.

4) Orthophosphoric Acid, 85 %

5) Indicator solution was prepared by measuring 0.25 g of sodium diphenylamine-sulphonate and dissolving in 100 ml of distilled water.

Protocol

Standardization of ferrous sulphate solution:

Ten millilitres of potassium dichromate solution was added to 20 ml of concentrated sulphuric acid. Two hundred millilitres of distilled water, 10 ml of phosphoric acid and 1 ml of the mixed indicator was added to the mixture. The mixture was shaken vigorously. Now, ferrous sulphate solution was added to the mixture slowly from a burette until the solution's colour changed from blue to green. Further, 0.5 ml of potassium dichromate solution was added to change the colour

back to blue. Now, dropwise addition of ferrous sulphate solution with continued swirling changed the colour of the solution from blue to green again. The total volume of ferrous sulphate solution used (X) was noted

For soil analysis:

1) The soil sample was taken and oven-dried to remove all the moisture. The weight of the sample was recorded as W_1. The sample was then sieved through a 10 mm sieve and the sieved sample weight was recorded as W_2. Further, the samples were powdered so that it passes through the 425-micron Sieve. Final sample was separated and used for the test.

2) A measured 5 g of soil sample was placed in a flask.

3) To the sample, 10 ml of potassium dichromate solution and 20 ml of concentrated sulphuric acid was added from a measuring cylinder. The mixture was thoroughly mixed by swirling for about one minute and the mixture was allowed to stand for 30 min to allow oxidation of the organic matter to proceed.

4) To the mixture, 200 ml of distilled water was added along with 10 ml of orthophosphoric acid and 1 ml of the mixed indicator. The mixture was shaken vigorously.

5) The mixture was titrated against Ferrous sulphate solution added added in 0.5 ml increments, with constant mixing, until the color of the solution changed from blue to green.

6) An additional 0.5 ml of potassium dichromate solution was added to change the color from green to blue

7) Now ferrous sulphate solution was added drop-wise and the change of color from blue to green was noted immediatelt after the addition of a single drop.

Calculations

The total volume (V ml) of potassium dichromate required to oxidize organic matter in the soil is given by the following formula: V= 10.5 (1 - r/X)

Where

r = Total volume of ferrous sulphate used

X= Total volume of ferrous sulphate used in standardized test

The percentage of organic matter present in the sample was calculated accordingly:

$$\text{Organic matter, percent by weight} = \frac{0.67\ W2V}{W1W3}$$

Where

W_2 = weight of soil sample passing 10mm Seive

V = total volume of potassium dichromate used to oxidize the organic matter

W_1 = weight of soil sample taken before seiving

W_3 = weight of soil sample used for the experiment

3.3.2 Total nitrogen (Kjeldahl method) (Piper, 1966)

The principle of the determining ammonia nirtrogen in organic matter sample is as follows: In the presence of a catalyst, the soil sample is digested with concentrated sulfuric acid this converts the organic nitrogen into ammonium sulphate which releases ammonia by distillation with concentrated alkali solution. The ammonia formed in this way is absorbed by sulphuric acid and the excess acid is titrated with standard alkaline solution.

<u>Reagents</u>

1) Potassium Sulphate or Anhydrous Sodium Sulphate

2) Mixed catalyst (Copper Sulphate)

3) Concentrated H_2SO_4

4) Sodium hydroxide solution (450 g of NaOH pellets in 1000 ml of water)

5) Standard H_2SO_4 solution 0.5 N

6) Standard NaOH solution 0.25 N

7) Alkaline Sodium Sulphide Solution- 20 g of sodium sulphide in water, diluted to 50 ml and 600 ml of sodium hydroxide solution added to it.

8) Methyl Red Indicator solution

Protocol

1) The soil sample was finely crushed and 1 gm of the sample was added to the Kjeldahl flask along with the 10 g of catalyst and 30 ml of H_2SO_4.

2) The contents of the flask is heated to boiling and is boiled until the solution becomes clear. Then the contents were further boiled gently for 2 hours.

3) Contents were cooled and transferred to the round bottom flask of the distillation unit

4) H_2SO_4 solution was added to the beaker and few drops of methyl red indicator was added to it

5) In the separating funnel sodium hydroxide solution was added and mixed with the contents of the flask to make it alkaline. Distillation was performed till one third of the volume of solution in the flask

6) Titration of the excess H_2SO_4 in the beaker was done against standard sodium hydroxide solution

7) Blank determination of the samples were conducted in presence of all chemicals but in the absence of the soil sample

Calculation:

$$\text{Nitrogen percentage by weight} = \frac{1.4\,(V2-V1)N}{W}$$

Where

V1= Volume of the standard sodium hydroxide solution used to neutralize the excess acid in case of sample (in ml)

V2= Volume of the standard sodium hydroxide solution used to neutralize the excess acid in case of blank (in ml)

N = Normality of the standard sodium hydroxide solution

W = Weight of the soil sample in g

3.3.3 Total phosphorous (Kitson and Mellon, 1944)

Total phosphorus was determined spectrophotometrically with acid digested soil samples.

Acid digestion of soil samples

The soil samples were digested with a combination of nitric acid and perchloric acid (Idera et al. 2014). The protocol is as follows:

1) 1 gm of soil sample was measured and placed it on a digestion tube. To the sample, 10 ml of conc. nitric acid (HNO_3) was added and the contents of the tube were digested for 1 hr at 145°C in acid proof digestion having fume exhaust.

2) The reaction mixture was allowd to cool and 10 ml of conc. nitric acid and 5 ml of perchloric acid ($HClO_4$) was added and heated at about 100 °C for 1 hr and then the temperature was raised to 200°C.

3) The digestion procedure was continued until the contents became colorless and white matter remained in the digestion tube.

4) After removing from the heating mantle the solution was sufficiently cooled and 50 ml of distilled water was added to the tube. Contents were filtered through a Whatman filter paper number 42.

5) Elemental analysis of different metals were performed ICP-MS (Element XR, ThermoFisher Scientific, Germany).

Reagents

1) Concentrated nitric acid.

2) Perchloric acid HClO4 60%.

3) 0.5N HCl (41 ml concentrated HCl in one liter).

4) 5N H_2SO_4 (139 mL concentrated H_2SO_4 in one liter)

6) p-nitrophenol

7) Solution A: 12 g of ammonium molybdate in 250 ml of distilled water. In another flask, 0.2908 g of antimony potassium tartrate in 100 ml of water. these two solutions were added 10 1000 ml of 5N H_2SO_4. The volume was made upto 2 liters with distilled water and stored in a dark pyrex bottle.

8) Solution B: 1.056 g of ascorbic acid was dissolved in 200 ml of solution A and mixed thoroughly. This solution was freshly prepared.

Colorimetric determination of Phosphorous

1) The digested samples were transferred to 50 ml volumetric flasks and 35 ml of distilled water was added to it.

2) To the samples 10 ml of Reagent B was added and volume was made upto 50 ml with distilled distilled water. The color was allowed to develop and the intensities were read at 690 or 880 nm.

3) The standard curve was prepared with P standard (KH_2PO_4) with concentration of 0-1 µg/ml

Calculations

$$\text{Total \%P} = \frac{\text{Concentration of P in final solution (µg per ml)} \times 10}{\text{Sample weight (mg)} \times \text{aliqout (ml)}}$$

3.3.4 Available Phosphorous (Olsen, 1954)

Available Phosphorus in Soil was determined by Dickman and Bray's Method

Reagents

1) Ammonium Fluoride: 37 g of NH_4F was dissolved in distilled water and the solution was diluted to 1 liter. This solution was stored in polyethylene bottle.

2) 0.5 N HCl: 20.2 mL Con. HCl in 500 ml of distilled water

3) Extracting Solution: 15 ml of 1N NH_4F and 25 mL 0.5 N HCl was mixed to 450 ml of distilled water.

4) Dickman and Bray's Reagent: 15 g of ammonium molybdate $(NH_4)_6Mo_7O_{24}.4H_2O$) dissolved in 300 ml distilled water. It was warmed to 60 °C filtered and cooled to room temperature. To that add 34.3 ml of concentrated HCl and volume made upto 1 lit.

5) Stannous Chloride ($SnCl_2$. 2 H_2O) Solution: 2 g $SnCl_2$. 2 H_2O crystals was dissolved in in 8.3 ml Con. HCl and diluted to 100 ml

6) Preparation of Standard Phosphorous Solution:

0.439 g of Potassium Dihydrogen Phosphate (KH_2PO_4) was mixed with 25 ml of conc. H_2SO_4. To make a 2ppm stock, 2 ml from the mixture was added to a beaker and volume was made upto 10 ml

7) Extraction of Phosphorous:

To 5 g of the soil sample, 50 ml of extraction solution was added. Contents were shaken for 5 mins and filtered.

8) Estimation of available P:

To 5 ml of soil extract, 5 ml of Dickman and Bray's Reagent was added and to the mixture 7.5 ml of Boric acid was added. Finally, 1 ml of $SnCl_2$ was added and the solution was mixed thoroughly.

9) For preparing standard curve:

To 5 ml of each Standard Phosphorous Solution, 5 ml of Dickman and Bray's Reagent was added and to the mixture 7.5 ml of Boric acid was added. Finally, 1 ml of $SnCl_2$ was added and the solution was mixed thoroughly. Blank was prepared by replacing 5 ml of Standard Phosphorous Solution with 5 ml distilled water.

10) The intensity of the blue color developed was measured at 690 nm

Calculation:

$$\text{ppm of Phosphorous in soil} = \text{ppm of P from graph} \times \frac{\text{Total volume of extraction solution (ml)}}{\text{Weight of soil}}$$

3.4 Extraction of total RNA from soil

Soil samples were pulverized into a fine powder using mortar and pestle in presence of liquid nitrogen. This method of freeze grinding with the help of liquid nitrogen ensured maximum release of nucleic acids from the microbial population due to efficient rupturing of the cell wall. Total RNA from soil was isolated using the RNA PowerSoil® Total RNA Isolation Kit (MOBIO Laboratories, Carlsbad, CA) (Damon et al. 2011). Total RNA isolation kit is designed to improve the RNA yield and also remove the humic substances, fulvic acids and other PCR inhibitory substances. Briefly, bead beating of the finely ground soil in presence of solutions containing disrupting agents like SDS led to proper dispersal of cells followed by complete cell lysis. Phenol: Chloroform: isoamyl alcohol solution separated the total nucleic acids in the sample and was precipitated by isopropanol. Finally, the nucleic acid was captured on a column and eluted by a buffer which preferentially elutes RNA thus leaving out the DNA. After the final precipitation step, the RNA pellet was resuspended in nuclease-free water and stored at -20 °C.

3.5 Analysis of extracted RNA

Integrity of RNA was tested by electrophoresis and purity was determined by nanodrop spectrophotometry (Thermo Fisher). Agarose gels (0.7%- 1 % w/v) were prepared in 0.5X TBE buffer (pH 8.0) and about 100 ng of total RNA was loaded on the gel (Appendix I) using a 6X loading buffer (Appendix I). Prior to casting, the agarose gels were stained with Ethidium bromide (EtBr) (0.5 µg/ml). The migrated nucleic acids were visualized on a U.V. transilluminator (312 nm). All electrophoresis devices were washed with detergent and rinsed in autoclaved double distilled water to eliminate ribonuclease contamination prior to use.

3.6 Construction of cDNA from eukaryotic mRNA

Construction of cDNAs from eukaryotic mRNA was performed with the help of Mint-2 cDNA synthesis kit (Evrogen, Moscow) which facilitates full length enriched double stranded cDNA from total RNA. The kit uses specific adapters and features of Mint reverse transcriptase (RT) to construct cDNA. First strand cDNA synthesis starts from the 3'-end adapter containing an oligo(dT) sequence that anneals to poly(A) stretches of RNA.First strand synthesis is started by Mint RT and as it reaches the 5'-end of the mRNA, addition of several non-template deoxycytidines at the 3'-end of the newly synthesized first-strand cDNA takes place. This stretch of oligo(dC) is complementary to a olido dG sequence located in the 3' sequence of the DNA adapter called PlugOligo. The RT carries on with reverse transcription to the end of the PlugOligo adapter thus forming single stranded cDNA, incorporating the sequence into the 5'-end of cDNA. At the final step, the cDNA is amplified by PCR leading to the formation of double stranded cDNA which is flanked by PlugOligo and CDS adapter sequences (Appendix I). A schematic representation of the process is depicted in Figure 3.1.

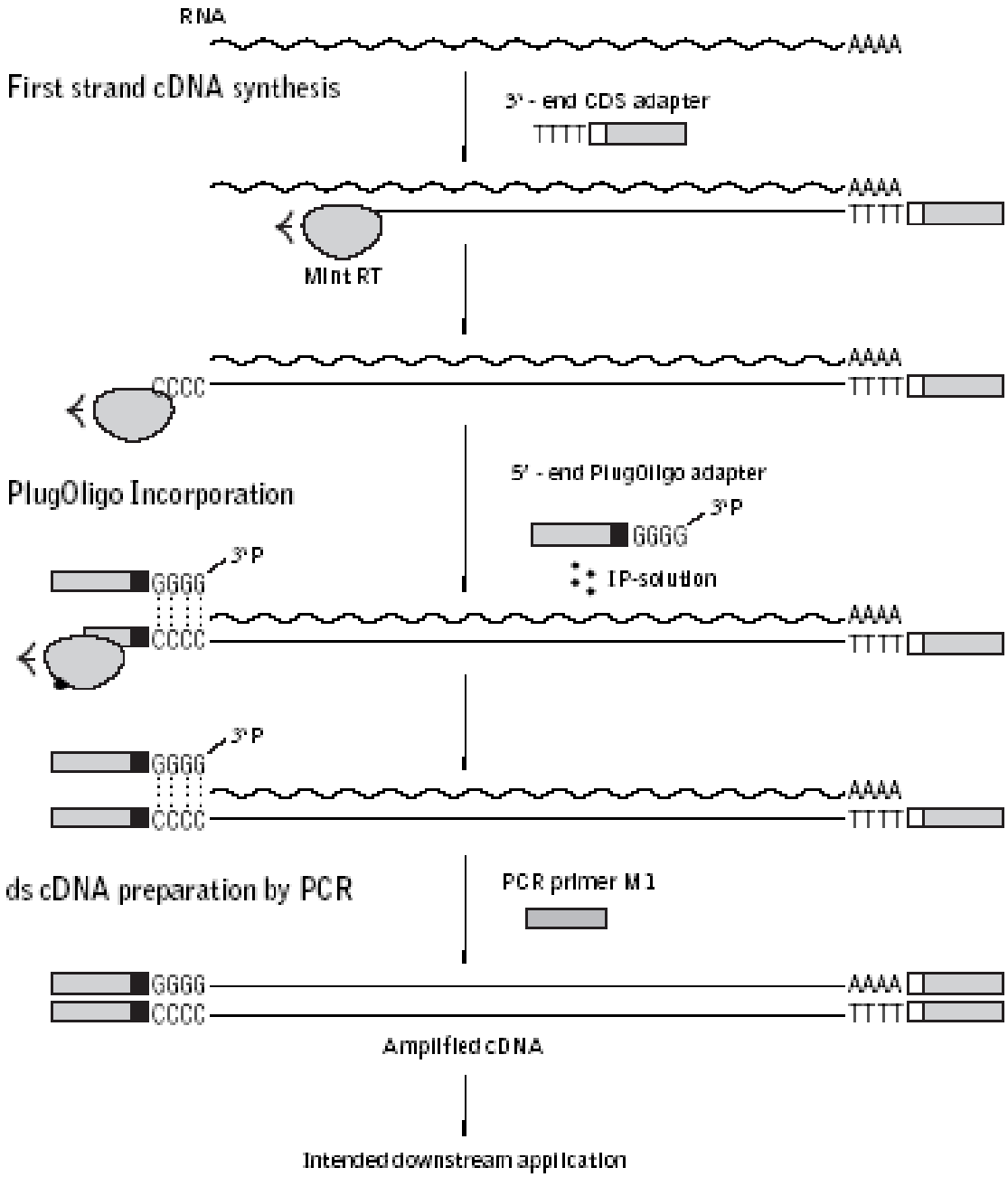


Figure 3.1: Schematic representation of the cDNA synthesis from eukaryotic PolyA mRNA by Mint2 cDNA synthesis kit (Adapted from Mint2 cDNA synthesis kit manual)

The protocol used in this analysis is based on the manufacturer's instructions. Brief description of the protocol is listed below.

1) 250 ng of the total RNA extracted from soil was briefly heated to 65 °C before use to prevent aggregation

2) For each reaction the following components were combined:

- 3 µl of RNA (750 ng)
- 1 µl of CDS adapter (10 µM)
- 1 µl of PlugOligo Adapter (10 µM)

  ---

  5 µl Total volume

3) The mixture was incubated in thermal cycler at 70 °C for 2min.

4) For the reverse transcription step:

- 2 µl 5X First-strand buffer
- 1 µl Dithiothreitol (20 mM)
- 1 µl dNTP (10 mM each)
- 1 µl Mint reverse transcriptase

  ---

  5 µl Total volume

5) The components were gently mixed and 5 µl of the RT mix was added to the mixture from step 2

6) The tubes were incubated at 42°C for 30 min and 5 µl of IP- solution was added to the reaction tube. The reaction mixture was incubated for 1.5 h at 42°C.

7) The first strand synthesis reaction was terminated by keeping on ice.

8) Double stranded cDNA was synthesized from the first strand using the following PCR reaction:

- 80 µl RNase-free water
- 10 µl 10X Advantage 2 PCR buffer
- 2 µl dNTP mix (10 mM each)
- 4 µl PCR primer-M1 (10 µM)
- 2 µl First-strand cDNA
- 2 µl 50X Advantage2 polymerase mix

  ---

  100 µl Total Volume

9) The tubes were gently mixed and contents were spun down in a microcentrifuge.

10) 24 µl of the reaction mix was aliquoted into 3 tubes and were labeled S1, S2 and S3.

11) PCR was performed in a thermal cycler with the following program:

- Initial denaturation 1 cycle at 95°C of 1 min
- Cycling 'X' cycles of
  - ❖ 95°C 15 sec
  - ❖ 66°C 20 sec
  - ❖ 72°C 3 min

  Where X is 5 cycles for S1, 18 cycles for S2 and 21 cycles for S3

12) When cycling was complete, the cDNA was visualized in a 1.2 % agarose gel along with 1 Kb DNA molecular marker.

13) Optimum number of PCR amplification cycles was determined in this way and both under cycling and over cycling was avoided.

14) The double stranded cDNA was purified using a QIAQIAquick PCR Purification Kit (Qiagen, Germany).

15) Elution of double stranded cDNA was done with with 50 µl of sterile RNase-free water.

3.7 Fractionation of constructed cDNA based on size

To focus on full-length genes of specific size ranges from the entire double stranded cDNA prepared from total soil RNA, size fractionation of cDNAs was performed by two-dimensional agarose gel electrophoresis as described by Yadav et al. (2014).

Prior to fractionation of the cDNA, the DNA molecular marker was used to optimize the conditions of electrophoresis so as to capture the entire size range of the cDNA during two-dimensional gel electrophoresis. Agarose gels of 0.7 % were prepared in 40 ml 0.5X TBE buffer and volume of tank buffer (0.5 X TBE), DNA molecular marker loading volume, length of gel lane to be cut and electrophoresis running conditions were optimized.

For size fractionation, two identical but separate 0.7 % agarose gels were cast and both cDNA and 20 µl of pre-stained DNA molecular marker (1 Kb) were electrophoresed under identical running conditions (6 $V.cm^{-1}$, 90 mins). Following electrophoresis, the gel lanes containing ds cDNAs and DNA marker were excised (2 cm from top and 2 cm from bottom) rotated at 90° and placed into two separate but identical gel casting trays. Forty ml of 1.4% of low melting point agarose (Himedia, Mumbai) was used in both trays to cast the gels and electrophoresis was carried out at 2.6 V.cm−1 for 5 h. Ethidium bromide solution (0.5 µg/ml) was used to stain the gel containing the DNA marker. The unstained gel containing cDNAs was then superimposed over the stained gel which was visualized using a transilluminator. Three gel slices corresponding to the cDNA size fractions A, 0.1–0.5 Kb; B, 0.5–1 Kb and C, 1–4 Kb, were excised from the unstained gel. cDNAs were extracted from each gel slice by using GeneJET Gel Extraction Kit (Thermo Scientific, USA). A schematic representation of the process of size fractionation is shown in Figure 3.2.

Each fraction was amplified by PCR using the M1 primer with PCR program as described above, which was used to create the dsDNA, but with higher number of cycles. The largest fraction C (1-4 Kb) was amplified by 21 cycles, fraction B (0.5-1 Kb) amplified by 18 cycles and fraction A (0.1-0.5 Kb) amplified by 16 cycles.

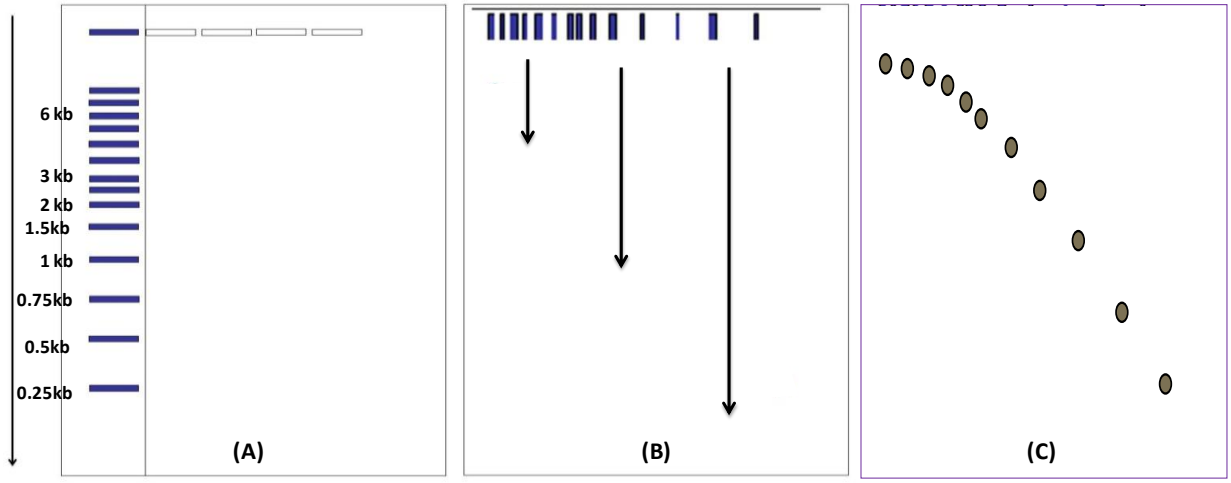


Figure 3.2: Schematic representation of the process of size fractionation. (A) Electrophoresis of DNA size marker (B) Gel lane was cut and rotated 90° and electrophoresed (C) representation of the DNA size fractions

3.8 Validation of size fractions by PCR

Validation of each size fraction was carried out by PCR amplification to check for cross contamination. This was done by amplifying each fraction with a set of primers of a known conserved gene of different length. For the largest fraction C (1-4 Kb), primers for amplifying elongation factor (EF1α) were used. The PCR products were visualized on a 0.7 % agarose gel.

Reaction mixture for performing 1 reaction was setup as follows:

Components	1x premix	Final Conc.
• 10 x PCR buffer	5 µl	1X
• $MgCl_2$ (25 mM)	2 µl	1mM
• dNTPs (2 mM each)	5 µl	each 0.2 mM
• NF (forward primer, 10 µM)	1 µl	0.2 µM
• NR (reverse primer, 10 µM)	1 µl	0.2 µM
• Template DNA	3 µl	~ 1µg
• Taq polymerase	0.25 µl	1.25 U/50 µl
• Deionized H_2O	7.75 µl	-

Total volume	50 µl

PCR program used for this study was as follows:

1 cycle of Initial denaturation 95 °C of 1 min

35 cycles of

- ❖ Denaturation 95 °C for 30 secs
- ❖ Annealing 55 °C for 1 min
- ❖ Extension 72 °C for 1 min

1 cycle of Final Extension 72 °C for 10 mins

3.9 Construction of cDNA library

The cDNA fractions thus obtained, were further processed before library construction of each fraction. The incorporation of SfiI (A and B) sites in the PlugOligo and CDS adapters flanking the cDNA helps directional cloning into vectors. SfiI sites are extremely rare in eukaryotic DNA hence all the cDNA remain intact after digestion.

3.9.1 Restriction digestion of cDNA

All amplified cDNA size fractions flanked by adapters harboring SfiI sites were digested with SfiI enzyme (Thermo Scientific, USA) to form asymmetric SfiI A and SfiI B sites.

For each fractions of cDNA, the following reagents were combined in a labeled PCR tube:

- 44 µl of Amplified cDNA fraction
- 5 µl of 10X SfiI buffer (10 µM) (Thermo
- 1 µl of SfiI restriction endonuclease (10-20 U)

  ---

  50 µl total volume

The tubes were gently mixed and incubated for 3 hrs at 50 °C. Following digestion, the contents of the tubes were purified by QIAquick PCR purification kit (Qiagen, Germany) according to manufacturer's protocol.

3.9.2 Restriction digestion of plasmid pFL61 modified with SfiI sites

Plasmid DNA pFL61 modified with SfiI A and SfiI B sites was used. This digestion procedure enabled efficient directional cloning of cDNA fragment downstream of a constitutively expressed PGK (3-phosphoglycerate kinase) promoter in pFL61 plasmid.

The plasmid digestion protocol used was as follows:

In a 1.5 ml centrifuge tube, the following reagents were combined:

- Nuclease free water 16 µl
- 10X Buffer G 2 µl
- Plasmid DNA(1µg) 1 µl
- SfiI enzyme (10 U/µl) 1 µl

The contents of the tube were mixed gently and spun down for few seconds

The tube was incubated at 50 °C for 16 hrs

3.9.3 Purification of restriction digested plasmid DNA

Restriction digestion was carried out in 5 reaction tubes. After completion, the contents of all the tubes were pooled and purified according to the following protocol:

1) About equal volume of phenol was added to the tube and was mixed by vortexing at slow speed.

2) The upper aqueous layer was transferred to a fresh 1.5 ml centrifuge tube.

3) 100 µl of chloroform was added to each tube and contents were mixed by slow vortexing followed by centrifugation at 10,000 g for 10 mins at room temperature.

4) The upper aqueous layer was transferred into a fresh tube and twice the volume (200 µl) of ice cold 95% ethanol, 1/10 volume (9.5 µl) of 3 M NaOAc (pH 4.5) were added and precipitated at -20°C for 2 hrs.

5) DNA was pelleted by centrifuging at 14,000 g for 20 mins at 4°C.

6) The supernatant was discarded and 100 µl of ice cold 80% ethanol was used to wash the pellet by centrifugation at 14,000 g for 10 minutes at 4°C.

7) The supernatant was discarded and pellet was air dried.

8) The dried pellets were reconstituted in 20 µl of nuclease free water.

9) One microlitre of purified digested DNA was run on a 0.6% (w/v) agarose gel to determine the approximate amount of DNA after purification.

3.9.4 Ligation of cDNA to modified pFL61 plasmid vector

cDNA and pFL61 plasmid vector both digested with SfiI enzyme were ligated with the help of T4 DNA ligase which catalyzes the combination of two strands of DNA between the 5´-phosphate and the 3´-hydroxyl groups of adjacent nucleotides.

Three separate ligation reaction mixtures were designed for each size fractions. Tubes were labeled as 1:1, 1:3 and 3:1 based on the vector:insert ratio used for ligation reactions.

The amount of insert to be used for ligation was determined by the formula

$$\frac{\text{ng of vector} \times \text{insert size in Kb}}{\text{vector size in Kb}} \times \text{molar ratio of } \frac{\text{Insert}}{\text{Vector}} = \text{ng of insert}$$

For a typical reaction mixture with 1:3 vector:insert for the size fraction C (~ 1Kb), the components were combined in a centrifuge tube as follows:

- Vector DNA(5.47 Kb) 100 ng
- Insert DNA (1 Kb) 54.84 ng
- Ligase 10X buffer 1 µl
- T4 DNA ligase 1 µl

Nuclease free water was added to make up to 10 µl.

The reaction mixture was incubated at 15°C for 16 hrs.

This ligation mixture was used for transformation of electrocompetent *E. coli* cells through electroporation.

3.9.5 Transformation of electrocompetent *E. coli* cells

Electroporation provides a method of high efficiency genetic transformation of bacterial cells with efficiency as high as 10^9 to 10^{10} transformants per µg of DNA. Cloned cDNA in pFL61 plasmid is transformed in electrocompetent DH10β cells using a high voltage electrical pulse which is applied to a sample having high resistance. Gene Pulser® II Electroporation system (Bio- Rad, USA) was used in this study.

3.9.5.1 Preparation of electrocompetent *E. coli* cells

1) 50 ml of Luria Broth was inoculated with 500 µl of fresh overnight grown DH10β *E. coil* cells.

2) Cells were allowed to grow at 37°C at 200 rpm to OD_{600} of approximately 0.5-0.7 (early log-mid-log phase).

3) Further growth was arrested by keeping the cells on ice for 20 mins.

4) Cells were harvested by centrifugation at 4000 g for 15 mins at 4°C.

5) The cells were then resuspended in 50 ml of ice cold 10 % glycerol in a chilled falcon tube. The cells were centrifuged at 4000 g for 15 mins at 4°C.

6) The supernatant was carefully poured off and the cells were resuspended in 25 ml of ice cold 10 % glycerol. The cells were again centrifuged at 4000 g for 15 mins at 4°C.

7) Following removal of supernatant, cells were again resuspended in 2 ml of ice cold 10 % glycerol. This process was repeated and finally the cells were harvested by centrifugation at 4000 g for 15 mins at 4°C.

8) In a final volume of 500 µl the cells were resuspended in ice cold 10 % glycerol. At this stage, aliquots of electrocompetant cells were stored at -70°C.

3.9.5.2 Electroporation

1) Cells were thawed on ice. A 1.5 ml microfuge tube and 0.2 cm electroporation cuvette were placed on ice.

2) In the ice cold microfuge tube, 40 µl of the cell suspension was mixed with 1-2 µl of the recombinant plasmid DNA. The mixture was incubated on ice for 1 min.

3) The micropulser was set to the following values:

Voltage = 2 KV, Resistance = 200 Ω, Capacitance = 25 µF

4) The mixture of cell suspension and DNA was transferred to a cold electroporation cuvette. Contents of the cuvette were tapped to make it settle to the bottom.

5) The cuvette was then secured in the shocking chamber by placing it in between the electrical contacts. An electrical pulse was passed.

6) The cuvette was quickly removed and immediately 1 ml of SOC medium (Appendix I) was added and cells were resuspended in the media.

7) The time constant recorded ranged from 4.6 to 5 milliseconds.

8) The cell suspension was transferred to a sterile falcon tube and incubated at 37°C for 2 hrs at 200 rpm.

9) The cells were then spread on LA plates containing 100 µg/ml of Ampicillin.

10) Plates were incubated at 37°C overnight.

3.9.6 Validation of cDNA library by colony PCR

The colonies growing on LA+ Ampicillin media represented the cDNA library. However validation of inserts and check for correct sizes was essential before proceeding any further.

Colony PCR is a rapid method to check the insert size of the recombinant plasmids and was employed here.

1) For each size library, 20 PCR tubes were set up each having 5 µl of sterile distilled water.

2) A fresh sterile toothpick was taken and was touched upon a small colony and immediately dipped into the sterile water in the PCR tube then streaked on a Luria agar plate containing Ampicillin (100 µg/ml)

3) For positive and negative control plasmid pFL61 and distilled water was used respectively as templates

4) The PCR tubes were placed in a Thermal cycler and were heated to 95°C for 10 mins.

5) PCR reaction mix was set up accordingly:

Components	**1x premix**	**20 x premix**	**Final Conc.**
• 10 x PCR buffer	5 µl	100 µl	1X
• $MgCl_2$ (25 mM)	2 µl	40 µl	1mM
• dNTPs (2 mM each)	5 µl	100 µl	each 0.2 mM
• NF (forward primer, 10 µM)	1 µl	20 µl	0.2 µM
• NR (reverse primer, 10 µM)	1 µl	20 µl	0.2 µM
• Template DNA	3 µl	3 µl each	~ 1µg
• Taq polymerase	0.25 µl	5 µl	1.25 U/50 µl
• Deionized H_2O	7.75 µl	155 µl	-
Total volume	50 µl	443 µl	

PCR program used was as follows:

1 cycle of Initial denaturation 95 °C of 30 sec

35 cycles of

- ❖ Denaturation 95 °C for 30 secs

- Annealing 55 °C for 30 secs
- Extension 68 °C for 1 min

1 cycle of Final Extension 72 °C for 10 mins

6) From every PCR tube, 5 µl was loaded on a 0.8 % agarose gel alongside a 1 Kb DNA marker.

3.9.7 Pooling of *E. coli* library cells and plasmid extraction

Transformed *E. coli* cells from all agar plates were gently scraped from the surface and pooled to constitute the size fraction cDNA libraries. The library cells were resuspended in LB broth and later used for plasmid extraction.

ZymoPURE™ II Plasmid Maxiprep Kit was used for the extraction of plasmid DNA from the library cells according to the protocol mentioned by the manufacturer. Almost 150 ml of library cells were used to extract plasmid DNA in this method and eluted in a volume of 1 ml of elution buffer. The pooled plasmids contain the cDNA which was used to transform mutant yeast cells and functionally complement them to screen the libraries of important genes.

3.10 Transformation of metal sensitive mutant yeast with constructed library

Yeast transformation protocol followed the standard Lithium Acetate (LiAc) method (Gietz et al. 1995). This method involves three main steps: making the yeast cells competent for transformation, transformation with plasmid DNA and plating on selective medium to screen the transformants. They are usually selected on the basis of auxotrophic markers hence the appropriate synthetic dropout medium is used for screening transformants. All mutants yeast cells used in this study and wild type yeast cells were transformed according to the following protocol.

3.10.1 Preparation of competent yeast cells

1) One colony of yeast was used to inoculate 20 ml of YPD (Appendix I). The flask was incubated at 30°C, 200 rpm.

2) This culture was used to inoculate 40 ml of YPD in a fresh flask and was incubated at 30°C, 200 rpm until the OD reached 2.

3) The culture was transferred to a falcon tube and centrifuged for 5 mins at 700 g at room temperature.

4) Supernatant was discarded and the pellets were resuspended in 25 ml of sterile distilled water and centrifuged for 5 mins at 700 g at room temperature.

5) The supernatant was discarded and yeast cells were suspended in 1 ml of sterile distilled water.

6) The cell suspension was transferred into a 1.5 ml microfuge tube and centrifuged at 700 g at room temperature. Supernatant was discarded and cells were once again resuspended in 1 ml of sterile distilled water.

7) 100 µl aliquots of the competent yeast cells were distributed in separate microfuge tubes.

3.10.2 Yeast transformation

1) Tubes containing 100 µl of competent cells were centrifuged and supernatant was discarded.

2) Meanwhile, Salmon sperm DNA (Promega, USA) was placed in 100 °C for 5 mins and then placed on ice immediately.

3) To each tube the following components were added and mixed gently:

- PEG 3350 50% w/v = 240 µl
- LiAc (1M) = 36 µl
- Salmon sperm DNA (10 mg/ml) = 10 µl
- Plasmid DNA = 1 µl
- Sterile distilled water = 74 µl

4) The tubes were incubated at 42°C in a water bath for 1 hr.

5) Following heat shock, the tubes were centrifuged at 700 g for 1 min and supernatant was discarded.

6) YPD was added to the tubes and cells were resuspended. The tubes were incubated for 1 hr at 30°C.

7) Following incubation, tubes were centrifuged and supernatant was discarded.

8) Cells were washed once in sterile distilled water, supernatant was discarded and cells were resuspended in 1 ml of sterile distilled water.

9) For library screening of metal tolerant genes, transformed cells were plated on Synthetic defined (SD –Ura) (Appendix I) amended with 150 μM of copper and also on SD-Ura without metal (control). The cfu/ml of the control plate was used to calculate the number of metal tolerant colonies screened for every transformation procedure. The process was repeated until 5 $x10^6$ colonies per size library were screened for metal tolerance.

3.11 Functional complementation by drop assay

All yeast mutants that were able to grow on SD-Ura medium amended with copper were selected and were further tested by drop-test on the same media supplemented with copper.

1) Transformed *cup1*$^{\Delta}$ cells were grown in SD-Ura broth overnight and cells were harvested by centrifugation.

2) Cells pellets were washed and reconstituted with sterile distilled water to achieve an OD_{600} 1

3) Serial dilutions of 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} were prepared in sterile distilled water and 5 μl from every dilution was spotted on SD-Ura agar plates containing 150 μM of copper. They were also spotted on 300 μM and 500 μM copper plates.

4) Clones which were growing up to all dilution levels were considered highly resistant, those which were growing up to the first four dilution levels were considered moderately resistant while those growing only up to three dilution levels were grouped into low resistant categories

To make sure the resistance showed by the clones was due to the recombinant plasmid, all clones were first grown in SD+Ura+5-FOA media.

3.12 Yeast Plasmid isolation

All metal resistant yeast clones were streaked on SD-Ura plates and also preserved as glycerol stocks. To sequence the cDNA inserts from each yeast clone, plasmid DNA was isolated by employing the following protocol.

1) Metal resistant yeast strains were grown in 10 ml SD-Ura broth at 30°C overnight.

2) Cells were harvested by centrifugation at 14,000 rpm for 2 mins.

3) Pellets were resuspended in 400 µl of yeast breaking buffer (Appendix I) with 0.3 g of glass beads.

4) To the mixture, 0.2 ml of phenol: chloroform mixture was added and was vortexed vigorously for 2 mins.

5) Contents were centrifuged at 14,000 rpm for 5 mins and upper aqueous layer was transferred to a fresh microfuge tube.

6) DNA was precipitated by adding 1/10 volume of 3 M NaOAc and 2.5 vol of ethanol.

7) The DNA pellets were centrifuged at 14,000 rpm for 15 mins followed by washing of pellets by adding 0.5 ml of 70 % ethanol.

8) Lastly, DNA was resuspended in 35 µl of sterile distilled water.

3.13 Bacterial transformation of *E. coli* DH5α by calcium chloride

The plasmids derived from yeast cells were used to transform *E. coli* cells to increase the copy number which would be suitable for sequencing. The protocol followed for bacterial transformation is described as follows:

3.13.1 Preparation of competent bacterial cells

1) Stock *E. coli* DH5α cells were streaked on Luria Agar (LA) plate and incubated at 37°C overnight.

2) Single bacterial colony was picked and 20 ml of Luria Broth (LB) was inoculated. The media was kept for incubation at 37°C overnight.

3) 100 µl from the overnight grown saturated culture was used to inoculate 50 ml of fresh Luria broth. The flask was incubated at 37°C at 180 rpm till the OD_{600} reached 0.5-0.8.

4) Cells were transferred to a pre-chilled 50 ml oakridge tube and kept on ice for 30 mins and then the tube was centrifuged at 8000 rpm for 10 mins at 4°C.

5) The supernatant was discarded and 10 ml of sterile, ice cold solution of 0.1M $CaCl_2$ was added. Cells were incubated on ice for 15 mins.

6) Following incubation, the tubes were again centrifuged and supernatant was discarded. Again, 1 ml of 0.1M $CaCl_2$ was added and the tubes were incubated on ice overnight to make competent cells. Following incubation, 100 µl aliquots of cells were prepared in pre-chilled 1.5 ml microfuge tubes and stored for transformation procedure.

3.13.2 Bacterial transformation

1) To 100 µl of competent cells, 5 µl of plasmid DNA was added and mixed gently. The mixture was kept on ice for 30 mins to allow binding of the plasmid to the cells surface.

2) Cells were removed from ice and heat shock treatment was provided by keeping the tube in a water bath set at 42°C for exactly 2 mins. Immediately the tubes were transferred into ice and kept for 2-3 mins.

3) 1 ml of LB + Ampicillin (50 µg/ml) was added to the tube and incubated at 37°C for 1 hr for expression of the plasmid DNA.

4) 100 µl of the transformed cells were plated on LB + Ampicillin (100 µg/ml) media for selection of antibiotic resistant transformants.

5) The plates were incubated at 37°C for 16-20 hours and checked for appearance of bacterial colonies. All bacterial colonies were streaked and plasmids were isolated from them.

3.13.3 Isolation of plasmid DNA from bacteria

1) The recombinant colonies were picked up from the plates with the help of sterile toothpicks and inoculated in 2 ml of LB + Ampicillin (100 µg/ml) and incubated overnight at 37°C.

2) About 1.5 ml of cell suspension was transferred to a microfuge tube and cells were harvested by centrifugation at 8000 rpm for 10 mins at room temperature.

3) To the cell pellet, 100 µl of Solution 1 (AppendixI) was added and the cells were resuspended in the solution by vortexing.

4) Solution 2 (Appendix I) was freshly prepared and 200 µl was added to the cells. The contents of the tube were mixed by inverting the tube 10-12 times rapidly and was kept on ice for 2 mins.

5) A total of 150 µl of Solution 3 (Appendix I) was added and the tubes were inverted slowly 10-12 times and kept in ice for 10-15 minutes.

6) 400 µl of phenol:chloroform:isoamylalcohol (25:24:1) mixture was added to the tube and contents were mixed by inversion for 2 mins and centrifuged at 12,000 g for 10 minutes.

7) The upper aqueous layer was transferred to a fresh tube and 800 µl of 95% ethanol was added.

8) After proper mixing, the tube was kept at room temperature for 5 minutes and centrifuged at 12,000 g for 5 minutes to precipitate plasmid DNA.

9) The supernatant was discarded and the pellet was washed with 400 µl of 70% ethanol and centrifuged 12,000 g for 5 minutes.

10) The supernatant was removed and the pellet was air-dried and dissolved in 50 µl of sterile nuclease-free water.

3.13.4 PCR validation of the plasmid DNA

All inserts of the plasmids were validated before sequencing by PCR. The PCR program to validate

Components	**1x premix**	**20 x premix**	**Final Conc.**
• 10 x PCR buffer	5 µl	100 µl	1X
• $MgCl_2$ (25 mM)	2 µl	40 µl	1mM
• dNTPs (2 mM each)	5 µl	100 µl	each 0.2 mM
• NF (forward primer, 10 µM)	1 µl	20 µl	0.2 µM
• NR (reverse primer, 10 µM)	1 µl	20 µl	0.2 µM
• Template DNA	3 µl	3 µl each	~ 1µg
• Taq polymerase	0.25 µl	5 µl	1.25 U/50 µl
• Deionized H_2O	7.75 µl	155 µl	-
Total volume	50 µl	443 µl	

PCR program used was as follows:

1 cycle of Initial denaturation 95 °C of 30 sec

35 cycles of

- ❖ Denaturation 95 °C for 30 secs
- ❖ Annealing 55 °C for 30 secs
- ❖ Extension 68 °C for 1 min

1 cycle of Final Extension 72 °C for 10 mins

3.14 Sequencing of cDNA and analysis

The cDNA inserts in the plasmids were sequenced at DNA sequencing facility, Department of Biochemistry, South campus, Delhi university, New Delhi, India using the Sanger method (Sanger et al. 1977) with an Applied Biosystems automatic sequencer. The sequences of cDNA were cleaned by removing vector sequences and queried against non-redundant protein database (BLASTX) at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to study homology to known sequences in the GenBank database. Few of the cDNA sequences were selected for further characterization and were submitted to NCBI database.

3.15 Characterization of metal tolerant genes

Out of the sequenced cDNA, few of the clones rescued from contaminated soil from Pierrelaye, France, were characterized further. These clones were rescued by screening of the metatranscriptomic library for cadmium tolerance. The potential of the cDNA in providing tolerance against different concentrations of metals such as Copper, Cadmium, Zinc and Cobalt were further studied.

Apart from these, ARK3, a Maltose fermentation regulatory protein MAL33 and ARK1 which is a chitin synthase, demonstrating high tolerance to copper was characterized to understand the biological mechanism of metal tolerance.

3.16 Multi-metal tolerance potential of the isolated genes

To investigate the potential of the genes responsible for metal tolerance and the way they respond to a number of toxic metals even at higher concentrations, drop test and growth kinetic study was carried out.

3.16.1 Drop assay

1) Yeast mutant strains *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ that are susceptible towards Copper, Cadmium, Zinc and Cobalt, respectively were transformed by the recombinant plasmid pFL61 carrying the isolated cDNA according to the LiAc method.

2) All transformants carrying the plasmid-borne cDNA were selected on respective SD-Ura agar plates amended with metals; *cup1*$^{\Delta}$ on 150 μM Cu, *ycf1*$^{\Delta}$ on 40 μM Cd, *zrc1*$^{\Delta}$ on 10mM Zn, and *cot1*$^{\Delta}$ on 2mM Co.

3) Each of the transformants were grown in separate SD-Ura media at 30°C at 200 rpm.

4) Cells pellets were adjusted to OD_{600} 1 with sterile distilled water.

5) Serial dilutions of 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} were prepared in sterile distilled water and 5 μl from every dilution was spotted on SD-Ura agar plates amended with 150 μM of Cu, 40 μM Cd, 10mM of Zn and 2mM Co.

3.16.2 Growth assay

1) Growth response of all the transformants was carried out in SD- Ura broth supplemented with metals to study the growth response in presence of varying metal concentrations.

2) For growth assay, 20 ml of SD-Ura broth was inoculated with mid-log precultures of *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with the cDNA and all the transformants were allowed to grow till initial OD_{600} reached 0.02.

3) Cells were allowed to grow at 30°C at 220 rpm for 3 hrs. Then, varying concentration of Cu (0 - 1000 mM), Cd (0 - 100 mM), Zn (0 - 13 mM) and Co (0 - 50 mM) were supplemented separately to the respective yeast mutant cells carrying the metal tolerant gene and the growth was measured till 40 h at an interval of 3 h.

4) For comparative study, wild type strain BY4741 transformed with empty pFL61vector was used. Also all yeast mutants transformed with empty vector were used.

3.16.3 Metal accumulation studies

1) For metal accumulation studies, cells were grown and 20 ml of SD-Ura broth was inoculated with mid-log precultures of all the transformants and they were allowed to grow till the yeast cultures reached an OD_{600} of 0.02.

2) Cells were allowed to grow at 30°C at 220 rpm for 3 hrs. Then, varying concentration of Cu (0 - 1000 mM), Cd (0 - 100 mM), Zn (0 - 13 mM) and Co (0 - 50 mM) were supplemented separately to the respective yeast mutant cells carrying the cDNAs. All the yeast transformants were allowed to grow till 40 h.

3) Cells were harvested by centrifuging at 700 g or 5 mins at room temperature and washed thrice with 10 mM of EDTA solution and dried at 80°C for 24 hrs.

Nitric acid and perchloric acid digestion of yeast cells

1) 1 mg of dry yeast cell biomass was mixed with 10 ml concentrated nitric acid (HNO_3) in a 250 ml digestion tube.

2) The mixture was boiled gently for 30 – 40 mins to oxidize the organic matter. It was allowed to cool to room temperature.

3) After the digestion, tube and its contents were cooled to room temperature, then 5 ml of 70% perchloric acid ($HClO_4$) was added and the mixture was again boiled until dense white fumes became visible.

4) After cooling the mixture to room temperature, 50 ml of distilled water was added and mixture was filtered through Whatman No. 42 filter paper. The digested sample was transferred into a 50 ml falcon tube.

5) Metal was estimated using ICP-MS (Element XR, Thermo Fisher Scientific,Germany).

3.16.4 Sequence characterization of the genes responsible for metal tolerance

BLASTX sequence tool was used to search for sequence homology of the cDNA obtained from the metal tolerant clones (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The sequences were aligned with a number of homologous sequences to find out conserved domains. The tools and databases of CBS Prediction Servers (http://www.cbs.dtu.dk/services/) were used for glycosylation (NetOGlyc 4.0 Server) and phosphorylation sites prediction (NetPhos 3.1 Server). For multiple sequence alignment, the protein sequences of the genes were first aligned with homologous sequences using the default setting of the Clustal Omega multiple sequence alignment program (http://www.ebi.ac.uk/Tools/msa/clustalo). For phylogenetic analysis, protein sequences from various taxa were aligned using Clustal W program in MEGA version 7.0 (Kumar et al. 2016) and neighbor-joining method was employed to construct phyllogenetic tree with a bootstrap resampling value of 1000.

3.17 Mechanism of metal tolerance

The genes expressed in contaminated soil as a result of metal pollution were characterized further based on their respective gene functions. This was performed in order to establish a link between metal tolerance and the biological activity of the genes. PLCc38 (Aldehyde dehydrogenase) was tested for its ability to tolerate oxidative stress which is a probable consequence of metal stress. PLCg49 (serine protease inhibitor) was tested for its role in improving the viability of the yeast transformants against protease activity. Lastly, the cDNA PLCg39 (DHHC palmitoyl transferase) was also tested for its role in conferring metal tolerance.

3.17.1 Oxidative stress tolerance potential of Aldehyde dehydrogenase

1) To confirm the role of Aldehyde dehydrogenase as a stress response protein, the cDNA PLCc38 was transformed in oxidative stress sensitive mutant *sod2*$^{\Delta}$ using the standard LiAc method as described in previous section. The transformant *sod2*$^{\Delta}$/PLCc38 was selected on SD-Ura plates supplemented with 1mM H_2O_2.

2) Functional complementation by drop assay was performed according to the previously described method on SD-Ura agar plates containing 0 mM, 1 mM, 2mM and 3mM of H_2O_2

3) Growth response in SD-Ura broth supplemented with 0 mM, 1 mM, 2mM and 3mM of H_2O_2 was performed as described in previous section. Wild type BY4742 and *sod2*$^{\Delta}$ transformed with empty pFL61 served as positive and negative controls, respectively. Also, BY4742 was transformed with PLCc38 to understand the synergistic effect of PLCc38 and the wild type BY4742 in response to H_2O_2 stress. Growth response was recorded till 30 h at an interval of 5 h.

3.17.2 Tolerance potential of Serpin in response to protease treatment

1) The role of serpin in providing tolerance to the yeast transformant to lytic serine protease enzyme was validated, the cDNA PLCg49 was transformed in protease sensitive yeast mutant *pbi2*$^{\Delta}$ using the standard LiAc method as described in previous section.

2) Also, PLCg49 was transformed in wild type BY4741 to understand the combined effect of PLCg49 and the wild type BY4741 in response to protease treatment.

3) Transformed yeast cells were grown in SD-Ura medium till optical density (OD_{600}) reached 1.0. Cells were harvested by centrifugation at 700×g for 5 min at room temperature, washed thrice with 0.1M Potassium Phosphate buffer (pH 7) and re-suspended in the same buffer

4) Varying concentrations of trypsin (0‑500 µg/ml) and proteinase K (0‑30 mg/ml) were prepared in 0.1M Potassium Phosphate buffer (pH 7) (Appendix I).

5) Equal volumes of yeast cell cultures and proteolytic enzymes were mixed and OD_{660} was measured immediately

6) The mixture was incubated at 30 °C for 2.5 h and OD_{660} was measured again.

7) Yeast cultures mixed with equal volumes of buffer instead of proteolytic enzymes, served as control.

8) The percent of yeast cell lysis was calculated by using the following formula:

$$\left(\frac{\text{Absorbance660 at 2.5 hrs for assay without enzyme}}{\text{Absorbance660 at 0 hr for assay without enzyme}}\right) - \left(\frac{\text{Absorbance660 at 2.5 hrs for assay with enzyme}}{\text{Absorbance660 at 0 h for assay with enzyme}}\right) \times 100$$

3.17.3 Role of DHHC in metal tolerance

1) To verify the role of DHHC palmitoyl transferase as a stress response protein, the cDNA PLCg39 was transformed in DHHC palmitoyl transferase mutant *akr1*$^{\Delta}$ following the standard LiAc method as described in previous section. The transformant was screened on SD-Ura agar plates supplemented with 150 μM of Cu.

2) Drop assay was performed in presence of various concentrations of metals according to the methodology already described.

3) Also, to confirm the role of DHHC palmitoyl transferase in metal tolerance, the wild type strain BY4741 was transformed with PLCg39 and was compared to BY4741 transformed with empty vector pFL61. Growth assay was performed according to the method described in previous section.

To sum up the process of metatranscriptomics used in this study, an overview of the process is presented in Figure 3.3

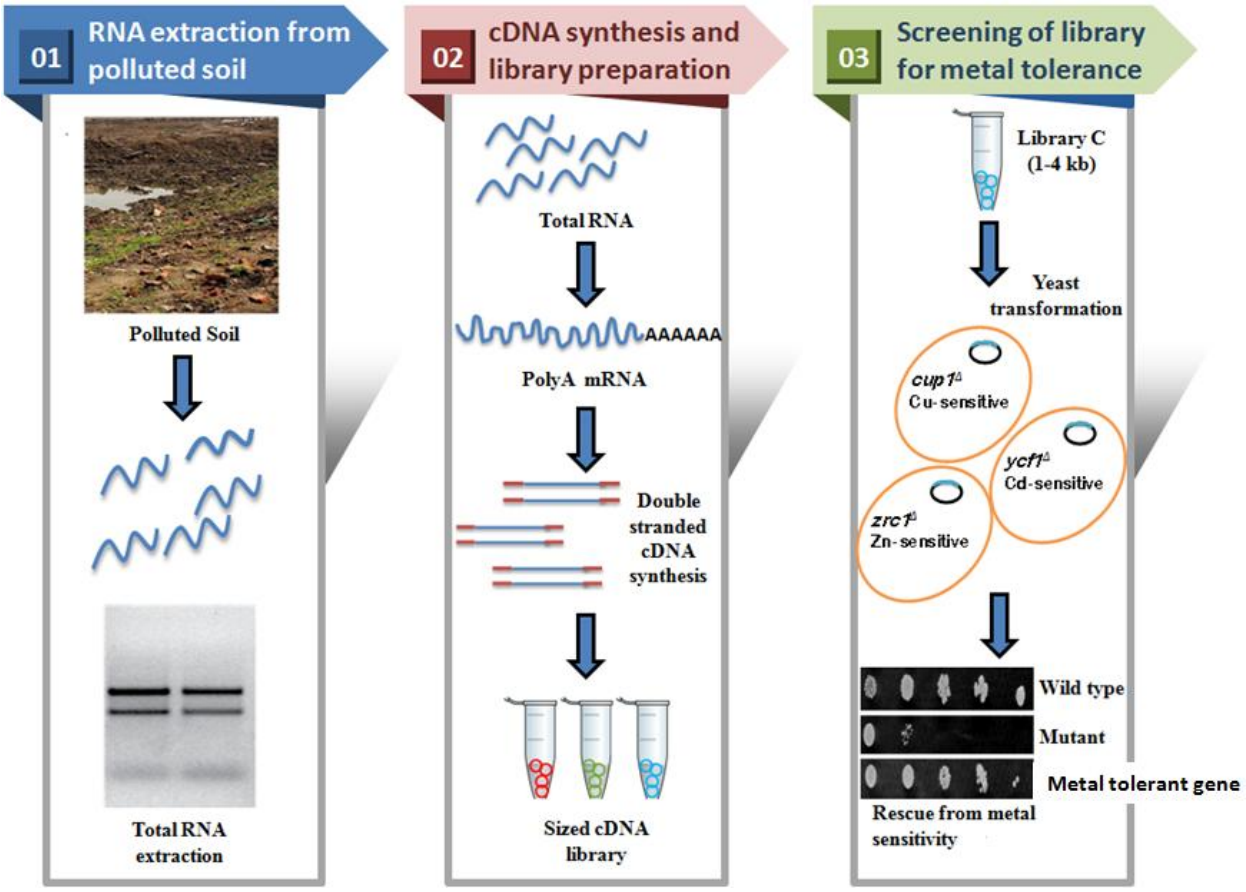


Figure 3.3 Overview of the steps involved in metatranscriptomic lbrary approach employed in this study

3.18 Functional diversity of eukaryotic microorganisms in Copper polluted soil

Diversity of eukaryotic communities of the soil sampling sites was studied by sequencing 18S rDNA amplicons by high throughput sequencing on Illumina HiSeq platforms using suitable eukaryotic universal primers.

3.18.1 Extraction of total RNA from soil

For functional diversity study, total RNA was extracted according to the procedure already described previously.

3.18.2 First strand cDNA synthesis

cDNA synthesis was performed from total RNA by using PrimeScript first strand cDNA synthesis kit (Takara, Japan), according to the manufaturer's instruction.

1) In a PCR tube the following components were added:

Reagents	**Volume**
Oligo dT primers (50 μM)	1 μl
dNTP Mixture (10 mM each)	1 μl
Template RNA	1 μg
RNase free dH_2O	7 μl
Total	10 μl

2) Contents were incubated at 65°C for 5 mins and immediately cooled on ice

3) In a PCR tube the following components were added:

Reagents	**Volume**
Template RNA Primer Mixture	10 μl
5X PrimeScript Buffer	4 μl
RNase Inhibitor (40 U/μl)	0.5 μl (20 units)
PrimeScript RTase (200 U/μl)	1 μl (200 units)
RNase Free dH_2O	4.5 μl

4) The contents were gently mixed

5) The reaction mixture was incubated at 42°C for 60 mins.

6) To inactivate the enzyme mixture, it was incubated at 72°C for 15 mins and cooled on ice.

3.18.3 Isolation of genomic DNA from soil

Soil DNA was extracted using Nucleospin® Soil Genomic DNA extraction kit (Macherey Nagel, Germany). Five hundred milligrams of composite soil sample was used as the starting material and DNA was extracted according to the protocol supplied by the manufacturer. The method involved mechanically homogenizing the soil sample in presence of lysis buffers provided in the kit and precipitating the contaminants by precipitating agents. The contents were centrifuged at 11,000 g for 1 min. The conditions for binding of the DNA to the membrane were set by buffers in the kit and it was filtered through the membrane. The silica membrane was washed with wash buffers and was dried before eluting the DNA with elution buffer.

3.19 Molecular Diversity Study by Next-Generation Sequencing (NGS)

The active and functional eukaryotic population diversity was studied by employing Next Generation Sequencing. cDNA thus constructed was validated by PCR using Eukaryotic Degenerate Primers targeting the hypervariable V4 region of Eukaryotic 18S rDNA (Russo et al. 2016; Stoeck et al. 2010) (Figure 3.4).

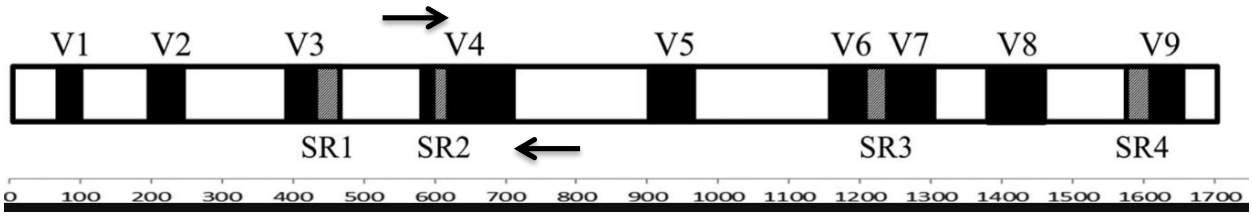


Figure 3.4: Variable and signature regions of the 18S rDNA

3.19.1 PCR amplification of V4 region with degenerate primers

1) cDNA and genomic DNA were validated by PCR using Eukaryotic Degenerate Primers TAReukF and TAReukR targeting the hypervariable V4 region of Eukaryotic 18S rDNA.

3) PCR reaction mix:

Components	1x premix	Final Conc.
• 10 x PCR buffer	5 µl	1X
• $MgCl_2$ (25 mM)	2 µl	1mM
• dNTPs (2 mM each)	5 µl	each 0.2 mM
• TAReukF (forward primer, 10 µM)	1 µl	0.2 µM
• TAReukR (reverse primer, 10 µM)	1 µl	0.2 µM
• Template DNA	1 µl	~ 1µg
• Taq polymerase	0.25 µl	1.25 U/50 µl
• Deionized H_2O	7.75 µl	-
Total volume	50 µl	

PCR program used for this study was as follows:

1 cycle of Initial denaturation 95 °C of 5 mins

35 cycles of

- ❖ Denaturation 94 °C for 30 secs
- ❖ Annealing 54 °C for 40 secs
- ❖ Extension 72 °C for 1 min

1 cycle of Final Extension 72 °C for 10 mins

3.19.2 Metagenomic High-throughput sequencing for understanding eukaryotic diversity

1) The amplified PCR product was purified using QIAquick PCR purification kit Qiagen (Germany) according to the manufacturer's protocol.

2) Purified PCR product was pooled and required concentration of DNA (~5 µg) was used for amplicon sequencing. High-throughput metagenomic sequencing was performed in AgriGenome Labs Pvt Ltd, Cochin, India.

3) Illumina HiSeq (Rapid) platform was chosen as the platform for sequencing to create a library of 2X250 Paried End reads.

4) Read quality checking – A quality report was generated based on:

a. Base quality score distribution

b. Sequence quality score distribution

c. Average base content per read

d. GC distribution in the reads

5) Demultiplexing of samples – The fastq files were de-multiplexed into separate samples based on the barcode assigned to it. Statistics in terms of total reads belonging to each sample was generated

6) Building of consensus sequence – For each sample, highly similar reads were clustered together and a consensus sequence was generated. The consensus sequence represents single eukaryotes in the sample. The number of reads used for building the consensus was noted. This was helpful in quantifying the presence of eukaryotic species.

7) Taxonomy assignment – Each consensus sequence was assigned a taxonomy based on similarity of the consensus sequence with existing eukaryotic species. Quantification of each phylum was performed.

8) Eukaryotic diversity analysis within each sample was performed to build shannon index for diversity and rarefaction plots.

An overview of the NGS metagenomic workflow is presented in Figure 3.5

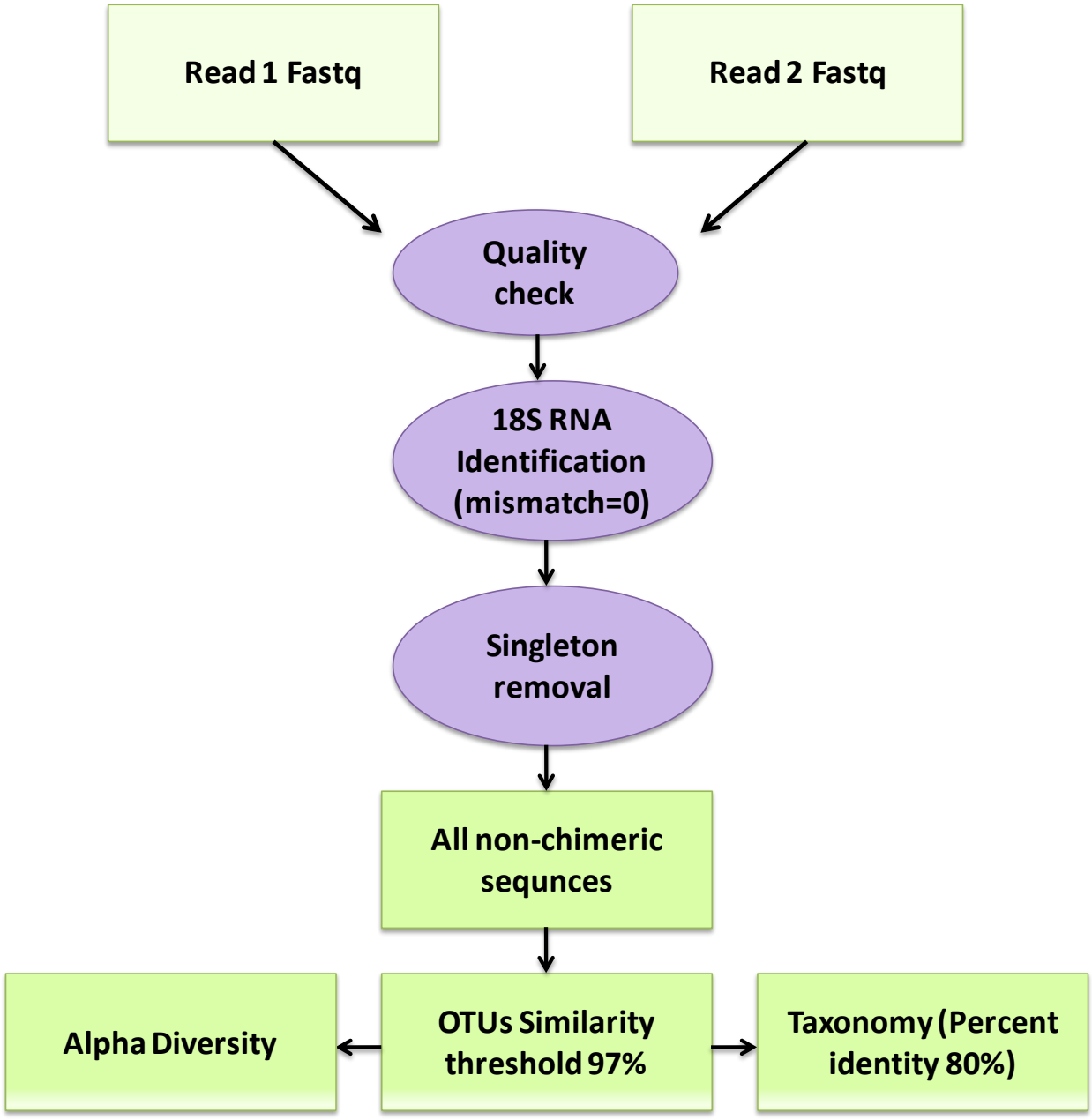


Figure 3.5: Overview of the metagenomic workflow

3.19.3 Sequence analysis workflow in QIIME

1) Input data after sequencing:

The workflow starts with importing the sequence raw reads from the Illumina HiSeq sequencing platform consisting of FASTQ files which includes the header file, quality and the sequence of the reads. The raw reads were both from sample cDNA and genomic DNA. The files were paired end reads and the raw data were obtained as cDNA-MJK_R1.fastq and cDNA-MJK_R2.fastq for sample cDNA and Genomic-MJK_R1.fastq and Genomic-MJK_R2.fastq for the sample Genomic DNA.

2) Quality check of data:

The quality of the sequences was analyzed with the FASTQC tool. The imported fastq files of the samples were checked for the sequencing quality. FASTQC tool includes the parameters like, base sequence, sequence quality and base content. A good quality sequence has a Phred score above 30. The Phred score of the datasets were 37.76. The theoretical and observed GC count was found to be equal and it was checked for over representation in the data sets. When all these parameters were checked, the sequences were allowed for further processing.

3) QIIME workflow:

The workflow begins with pre-processing of the reads by combining various filtering processes which is briefly demultiplexing samples from the reads, primer removal, clean up of sequences and quality filtering followed by OTU clustering, assigning taxonomies and alpha diversity analysis. QIIME combines the pre-processing steps of demultiplexing, primer removal, and quality-filtering in a single script, **split_libraries.py**. This generated a fatsa format file in the same folder as seq.fna which contains all the demultiplexed sequences after quality check.

4) OTU picking and clustering:

Next in the pipeline is clustering of the preprocessed sequences into OTUs, which represent cluster of organism defined by core phenotypic similarity that consists of taxonomic data. Sequences were clustered on the basis of similarity which was at the default value of 97%. UCLUST is the default clustering method which uses a *de novo* OTU picking method. The script in QIIME for picking OTUs and clustering is **pick_otus.py**. The default mechanism of Uclust or Usearch is they use the most abundant sequences in each OTUs and use it as the reference sequence. Output of running this script was a rep_set.fna file. The file contains the information about the sequence ID and samples from which it came from.

4) Assignment of taxonomy:

Taxonomy was assigned to the 18S rRNA gene sequences with reference to the database using the ribosomal database project (RDP) classifier against the SILVA 18S database. For this the script **assign_taxonomy.py** script was used. It created a new directory containing the taxonomy results as a text file named rep_set_tax_assignments.txt file containing an entry of each representative sequences which were according to the set threshold value. This file provides information of the organisms and their taxonomic values from the BLAST as well as the RDP classifier.

5) Generation of OTU table

The OTU table is generated by the script **make_otu_table.py**. This created an OTU table file called otu_table.biom which is a BIOM file (Biological Observation Matrix) which is suitable for any downstream analysis.

6) Alpha diversity study

For calculating alpha diversity, that is the mean species diversity in a particular site or environment, the QIIME script **alpha_diversity.py** was run to calculate alpha diversity of

samples. This command computes the alpha diversity metrics for all of the rarefied OTU tables and places the results in a new directory called alpha_rare. Chao1, Shannon Index and observed species were calculated for the number of sequences.

3.20 Statistical Analysis

All experiments were performed in triplicates. One way analysis of variance (ANOVA) was performed to test the significant differences among the treatments and the means were compared with Tukey's test at significance level of $P < 0.05$. All statistical analyses were performed using Graph Pad Prism 5.1 software.

Chapter 4: Results and Discussion

4.1 Collection of soil samples

The Indian soil was collected from Malanjkhand copper deposit with coordinates 22°0’54’’N and 80°43’20’’E. It is India's largest open-pit base metal copper mine located 90 km north-east of Balaghat in Madhya Pradesh (Figure 4.1). The copper project of Malanjkhand was founded in 1982 by M/s Hindustan Copper Limited, which set up the copper mining project. The copper deposits in the open pit mine in Malanjkhand is presently in the exploitation phase.

Although vegetation was scanty and patchy, soil samples were collected from the rhizospheres of to microbial diversity. Soil samples were collected from 15 locations in the core mining area and 4 samples were collected from the copper mine waste dumps as highlighted in (Figure 4.2). All samples were sieved to remove debris and plant materials. To capture most of the microbial diversity, the soil samples were pooled and stored in sterile zip bags and transported on dry ice to the laboratory where they were stored at -70 °C. The soil samples collected from the rhizospheric region are shown in Figure 4.3.

The district of Balaghat is broadly covered by three types of soil, sandy loamy and lateritic. The soil was found to be acidic with a mean pH of 5.87. This is due to the impact of acid mine drainage, which occurs as a result of rocks containing sulphide being converted to sulphuric acid when exposed to the atmosphere. The acid here dissolves heavy metals including lead, zinc, copper, arsenic, selenium, mercury and cadmium occurring in waste rocks and tailings, into ground and surface water thus directly affecting the soil quality and contaminating it with toxic metals. By far, the mining activity at Malanjkhand Copper project has produced huge piles of

waste rock which remain rich in heavy metals like iron, sulphur, lead, cadmium, zinc, manganese etc. after the copper ore has been extracted. These metals find its way into the surroundings through percolation and drainage thus affecting surrounding soil and water bodies (Pandey et al. 2007). The physico-chemical properties of the soil are listed in Table 4.1.

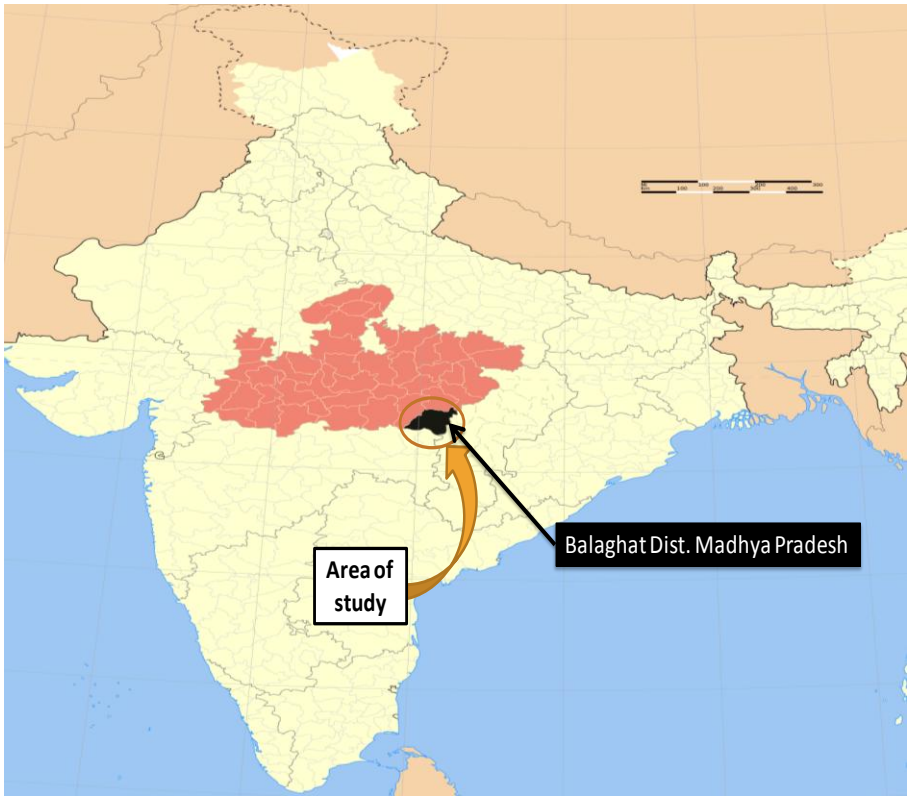


Figure 4.1: Location of Balaghat District, Madhya Pradesh

Figure 4.2: Sampling areas are shown in red

Figure 4.3: Sampling of rhizospheric soil

Table 4.1: Physico-chemical properties of soil sample collected from Malanjkhand copper deposit

Parameters	**Content**
pH	5.87
Total Organic Carbon	0.87 %
Total nitrogen	0.12 %
Available Phosphorous	4.01 mg kg^{-1}
Total Phosphorous	65.12 mg kg^{-1}
Metal content	
Copper	990 mg kg^{-1}
Cadmium	0.45 mg kg^{-1}
Cobalt	4.5 mg kg^{-1}
Zinc	79.2 mg kg^{-1}
Lead	56.5 mg kg^{-1}
Chromium	64.35 mg kg^{-1}

French soil samples were collected from Pierrelaye's agro-forest land (PL) (49°1′ 45″ N, 2°10′ 32″ E) near Paris, France, which is contaminated with heavy metals. This land was initially dedicated to vegetable plantation. However, the problem of metal pollution arose when the soils of Pierrelaye were heavily irrigated with untreated wastewater until the year 2002 (Foulon et al. 2016). As a consequence, the concentration of cadmium, zinc and copper were ten times higher than non-irrigated soil. However, after 2007 this has been converted to a Poplar plantation site. The physico-chemical properties of soil samples are listed in Table 4.2. Size fractionated cDNA library was constructed from the total RNA isolated from this soil and the library was further screened to isolate cadmium tolerant clones (Yadav et al. 2014). In the present study, the clones bearing the cDNA that were rescued from contaminated soil from France, were further characterized for its potential role in providing tolerance to Copper as well as other metals such as Cadmium, Zinc and Cobalt.

Table 4.2: Physico-chemical properties of soil sample collected from Pierrelaye, France

Parameters	Agro-forest soil, Pierrelaye
pH	7.15
Organic carbon	1.6 %
Total Nitrogen	0.12 %
Available Phosphorous	14.2 mg kg^{-1}
Total Phosphorous	291 mg kg^{-1}
Metal Content	
Cadmium	2.5 mg kg^{-1}
Copper	64 mg kg^{-1}
Zinc	385 mg kg^{-1}

4.2 Isolation of total RNA from soil

The composite soil samples were freeze ground with liquid nitrogen to improve better release of nucleic acid from the inhabiting microbial population. The soil sample, when stored in a -80 °C freezer, reduces cellular enzymatic activities and preserves the nucleic acid and many proteins. However, this method is unsuitable for preserving RNA in the sample because even at -80°C there is a considerable amount of water activity which causes nuclease activity, ultimately degrading RNA. Rapid freezing ceases all biological activity when combined with grinding of samples with mortar and pestle helps with the extraction of nucleic acids. One of the most traditional ways for harvesting nucleic acids from samples is the use of mortar and pestle combined with liquid nitrogen. This technique of cryogenic grinding of samples like soil and tough plant tissues is effective as the samples become fragile at a temperature of -196 °C and turn to fine dust provided, they remain in the chilled state. Extraction of total RNA from soil was performed using the RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories, Carlsbad, CA). Previously, several methods have been used to isolate RNA for metatranscriptomic analysis. Bailey et al. (2007) used a method using guanidine isothiocyanate and bead beating which was done in presence of denaturing agents. The use of guanidine isothiocyanate has been

referred to as ‘single step’ RNA extraction method. RNA is selectively separated from DNA by using an acidic solution consisting of guanidinium thiocyanate, sodium acetate, phenol and chloroform. Total RNA remains in the upper aqueous phase under acidic conditions, whereas most DNA and proteins remain in the interphase or lower organic phase. The RNA thus obtained was precipitated overnight in 4 M LiAc. However the method was modified and the amount of SDS and guanidine isothiocyanate was varied in another study by Damon et al. (2011) as soil DNA was purified additionally by the RNA isolation kit. In other studies, extraction of RNA was performed by the addition of 0.5 ml of hexadecyltrimethylammonium bromide (CTAB) extraction buffer and the RNA was extracted from the aqueous layer with Polyethyene glycol (Griffiths et al. 2000; Urich et al. 2008).

In the present study, using the Total RNA Isolation Kit, the yield was 50 ng of total RNA per gram of soil. The RNA from soil was repeatedly extracted and pooled to constitute approximately 500 ng of total RNA which was then used for cDNA synthesis. Purity of the RNA was checked using the nanodrop spectrophotometer and integrity was verified by gel electrophoresis. Two distinct bands were observed with band intensity of 2:1 for 28S RNA when compared to 18S RNA (Figure 4.4). The Absorbance of A_{260}/A_{280} value of the RNA was 2.04. Nucleotides like dsDNA, ssDNA and RNA all absorb at a wavelength of 260 nm therefore all of them will contribute towards the total absorbance of the sample. However, to assess the purity of specifically DNA or RNA, measurement of a ratio of 260/280 is employed. Generally, a ratio of ~1.8 is generally accepted as “pure” for DNA; a ratio of ~2.0 is generally accepted as “pure” for RNA. The absorbance of 260/230 is the secondary measure of nucleic acid purity. A lower value of the ratio would indicate the presence of contaminants which absorb at 230 nm.

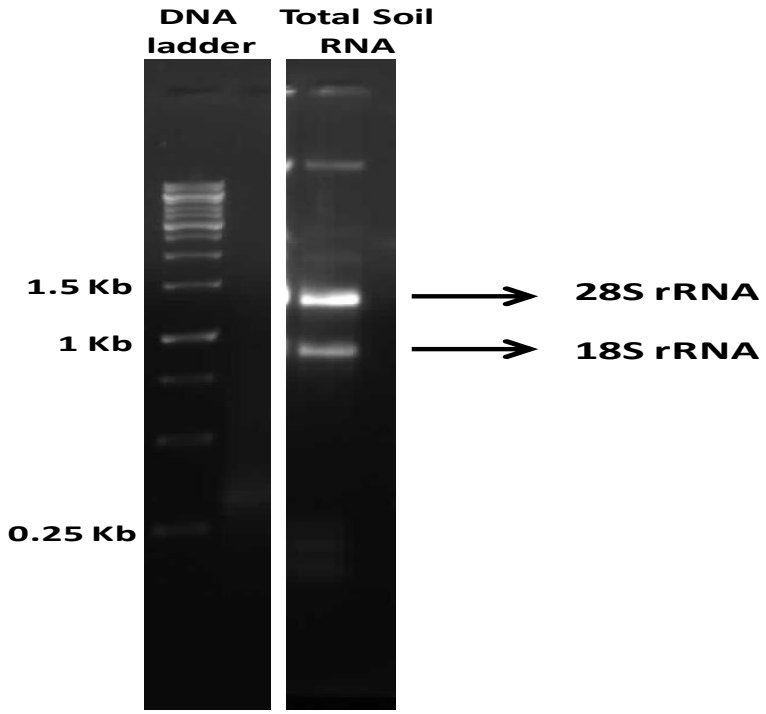


Figure 4.4: Extracted total RNA from soil using RNA PowerSoil® Total RNA Isolation Kit (MOBIO Laboratories, Carlsbad, CA) visualized on a 0.8 % non-denaturing agarose gel

4.3 cDNA synthesis from total RNA extracted from soil

Eukaryotic mRNA was specifically converted to cDNA using the Mint-2 cDNA synthesis kit. It is of utmost importance to have a full length cDNA library when working in the field of metatranscriptomics. However, conventional cDNA synthesis technique fails to build a full length cDNA from eukaryotic RNA which is primarily due to premature termination of reverse transcription of mRNA particularly longer than 2 Kb and also due to loss of 5'terminal sequence (Alessio and Gerard, 1988; Kazuo and Sumio, 1994).

Several studies have focused on the cap structure of mRNA to build full length cDNA (Edery et al. 1995; Seishi et al. 1994). However, they lead to degradation of the mRNA and involve complicated manipulation of the cap structure during every step of cDNA preparation. Another major disadvantage is with the low efficiency of enzymatic steps with this method and thus

requires larger amounts of RNA (20-100 μg). However, RNA from soil can be quite a limiting factor as it is largely dependent on the activities of the microbial population inhabiting a particular soil type.

The method used in this study ensured synthesis of full length clones by employing the reverse transcriptase function of Moloney Murine Leukemia virus (MMLV) reverse transcriptase and also its template switching mechanism. In this method, eukaryotic mRNA was selectively converted to cDNA amongst the huge pool of bacterial messenger RNA and ribosomal RNA. In earlier studies, the requirement of purification of poly 'A' mRNA was necessary by affinity chromatography before cDNA conversion (Damon et al. 2012; Lehembre et al. 2013). However, this method was able to reduce this additional step by using special adapters binding specifically to the 3'poly 'A' tail of the eukaryotic mRNA directly. Also the adapters contain SfiI sites which are very rare in eukaryotic DNA hence they provide stable sites for directional cloning (Zhu et al. 2001).

When cDNA was visualized on agarose gel, it appeared like a smear with indistinct amplification bands. The number of cycles used for the amplification of the double stranded DNA varied. Five, 18 and 21 cycles were used to prepare the double stranded cDNA. Five cycles produced a very weak band; however, 18 and 21 cycles produced the desirable result (Figure 4.5).

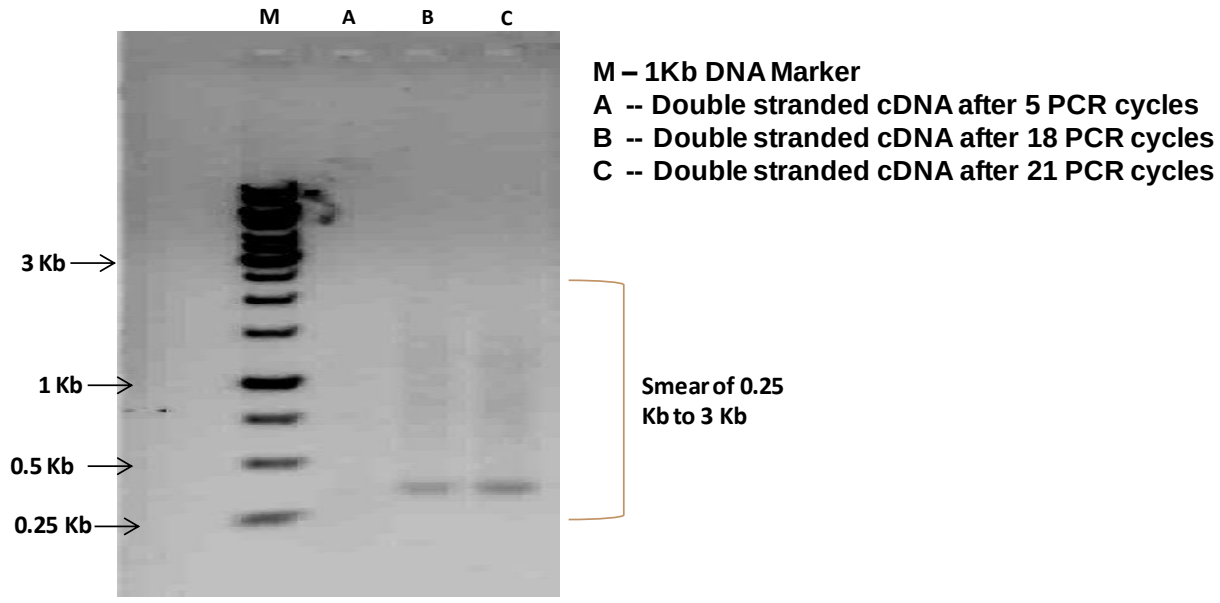


Figure 4.5: Synthesis of double stranded cDNA from total RNA by switching template cDNA synthesis technique

4.4 Generation of size fractions from fully enriched cDNA

Amplified cDNAs were size fractionated in parallel with DNA molecular markers on 2-D gel electrophoresis. Following electrophoresis the gel fractions were excised and purified. The size fractions were A, 0.1–0.5 Kb; B, 0.5–1 Kb and C, 1–4 Kb (Figure 4.6). Amplification of the total enriched cDNA prior cloning could lead to PCR biasness against the larger cDNA as smaller fragments of cDNA are preferentially amplified that might outnumber the larger full length cDNAs (Figure 4.7). This could create a misrepresentation of functional genes in the library. In this study, a size fractionation step ensured each size fractions to be amplified separately by varying number of PCR cycles before cloning and constructing cDNA library. However, the number of PCR cycles was kept low in order to reduce errors introduced by PCR polymerase. Using this method, larger genes could be efficiently represented along with the smaller ones (Wellenreuther et al. 2004).

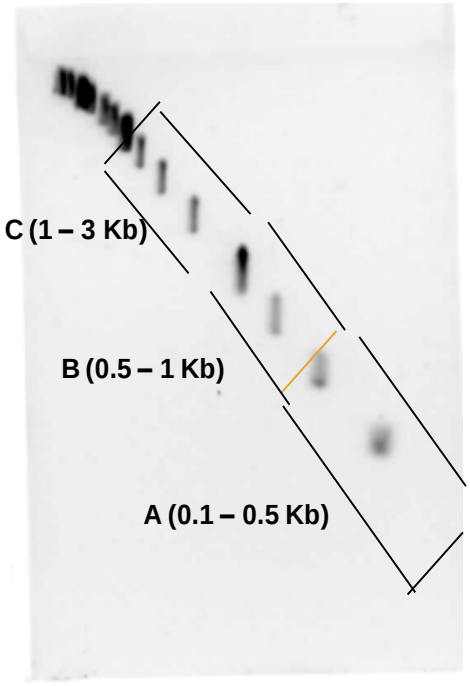


Figure 4.6: cDNA size fractionation by 2-D agarose gel electrophoresis. Three different size fractions were excised from the gel

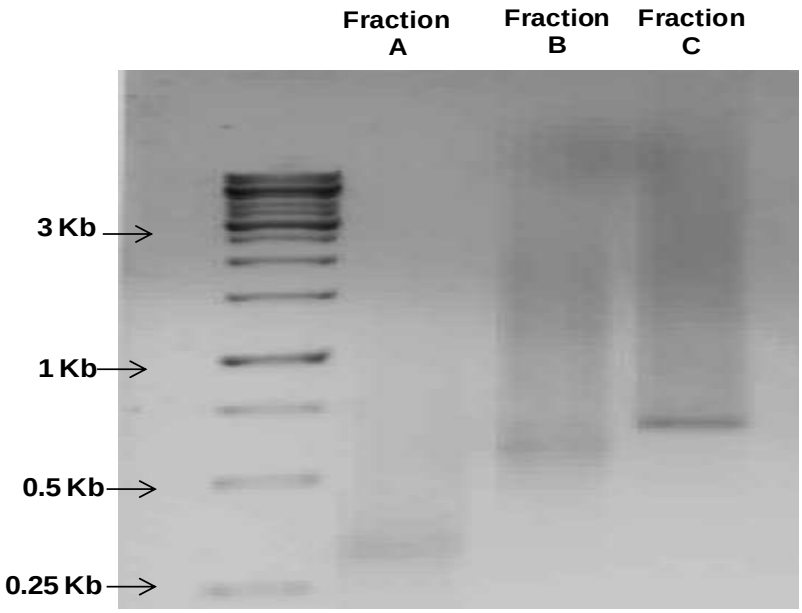


Figure 4.7: Amplified cDNA fragments visualized on 0.8 % agarose gel

The cDNA fractions were tested for cross-contamination of sized genes by amplifying each fraction with elongation factor (EF1α). Amplification was positive for only fraction C (1-4 Kb)

as amplified gene product of size 0.4 Kb of EF1α (1.4 Kb gene) was observed. This indicates a perfect correlation between the sized fraction and detection of PCR product (Figure 4.8).

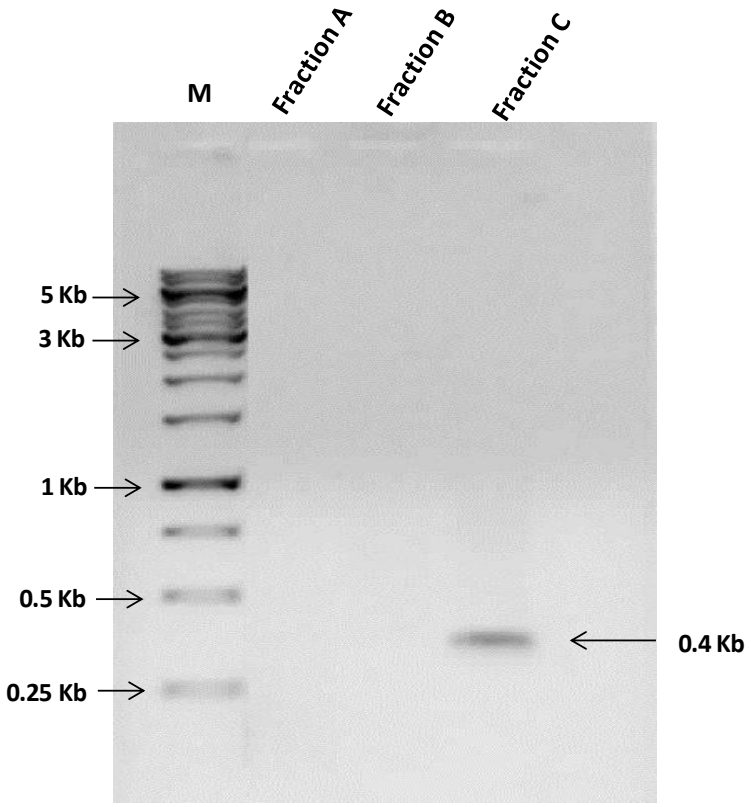


Figure 4.8: Validation of size fractions by PCR amplification by EF1α primers. Amplification of the gene product of size 0.4 Kb was only observed in fraction C (1-4 Kb)

4.5 Construction of cDNA library

The cDNA fractions were further processed and digested with SfiI restriction enzyme before ligation into yeast shuttle plasmid vector pFL61 (Figure 4.9) modified with the addition of SfiI restriction enzyme recognition sites (Figure 4.10). This is done to successfully clone a foreign gene and express in a yeast host. The sites are placed under the control of Phosphoglycerate kinase (PGK) promoter and its terminator from *S. cerevisiae*.

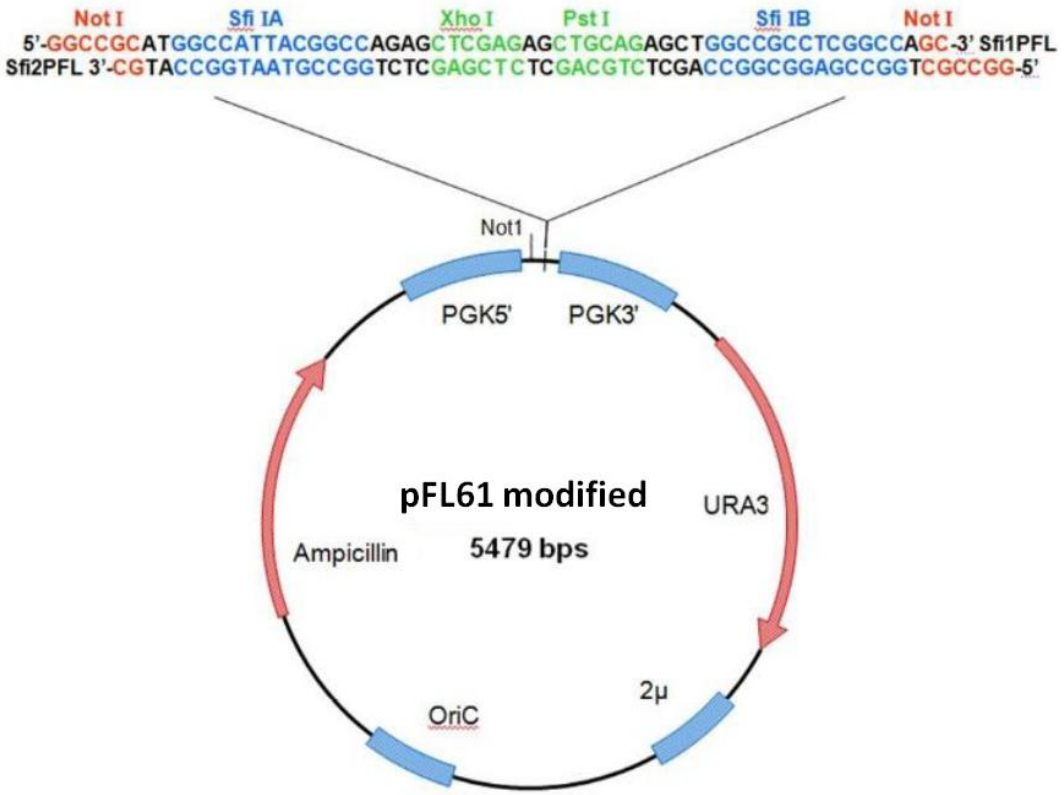


Figure 4.9: Restriction map of shuttle plasmid pFL61 modified by SfiI A and SfiI B sites within the yeast PGK promoter and terminator. OriC: bacterial origin of replication for propagation in *E.coli* . 2µ: 2µ yeast origin of replication for *S. cerevisiae*. Amplicilin: Selection marker for maintenance of plasmids in *E. coli*; URA3: Nutritional selection marker for maintenance of plasmid in yeast *S. cerevisiae*. PGK5' and PGK3' promoter and terminator sequences derived from *S. cerevisiae* phosphoglycerate kinase

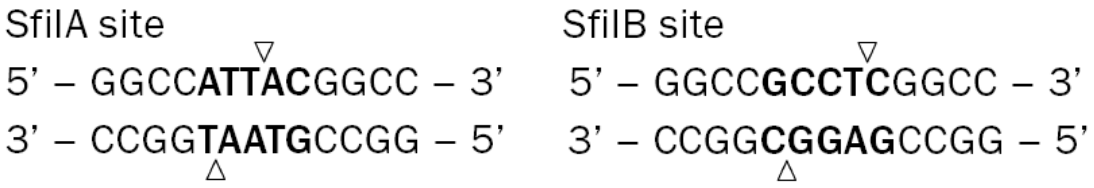


Figure 4.10: SfiI recognition sites SfiIA in the 5'-end PlugOligo adapters and SfiIB in the 3'-end CDS adapter

Following purification of digested cDNA and plasmid DNA, the ligation into vector was carried out using T4 DNA ligase and transformed into electrocompetent *E. coli* cells through electroporation. The ligation ratio of 1:3 vector:insert exhibited the best result and was used to carry on further ligation reactions. All libraries were screened for more than $1x10^6$ independent colonies on Luria agar supplemented with ampicillin (100 μg/ml). Colonies representing each library were pooled and plasmids were extracted from them.

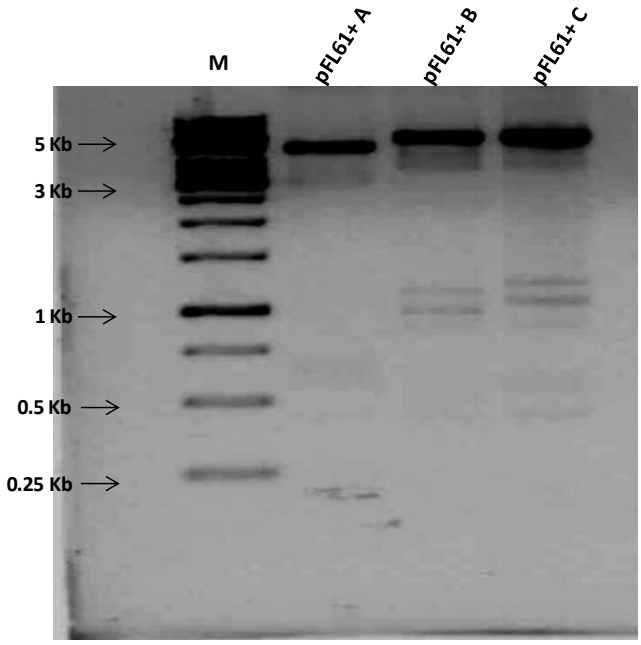


Figure 4.11: Soil cDNA size fractions after cloning into pFL61 (modified). For each sized library the size range of cDNA inserts is similar before and after cloning. M is the DNA molecular marker and size fractions pFL61+A, pFL61+B and pFL61+C correspond to the three size libraries A, B and C cloned into the plasmid pFL61

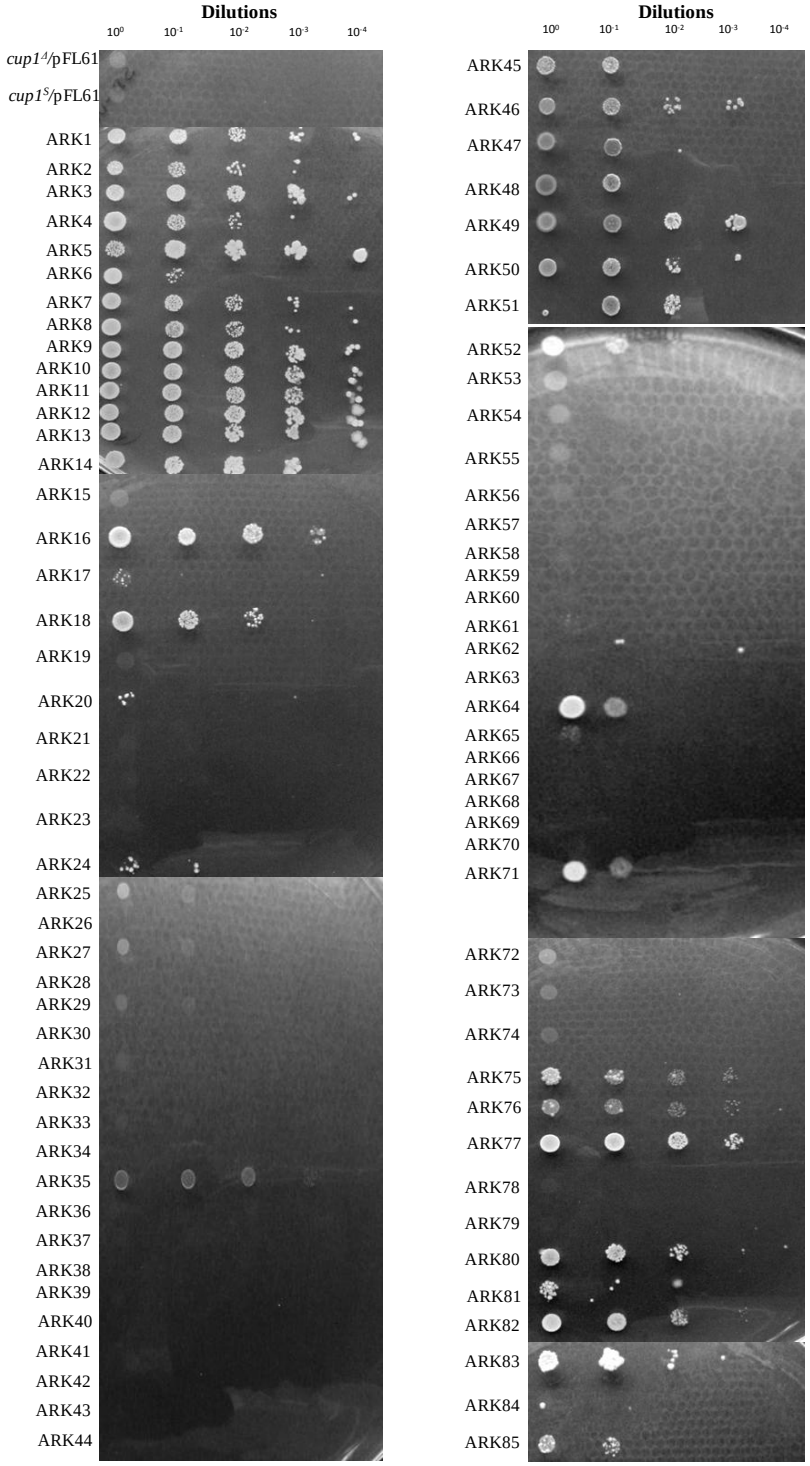
Quality control of the cDNA inserts for each library was performed by amplifying with NF and NR primers and visualizing on 0.8 % agarose gel. The plasmids were clearly visible and every amplified product appeared as a blot, the size of which corresponded to the original cDNA size fractions (Figure 4.11).

With this method of fractionation and cloning, larger gene transcripts (> 1Kb) could be efficiently converted into cDNA and cloned into plasmids for screening. In a study by Wellenreuther et al. (2004), a correlation between percentage full length cDNA and percentage full length ORF was estimated in a human cell line library which was size fractionated and cloned. It was observed that even with the increase in insert size, the full-length content of the cDNA or ORF did not decrease significantly as compared to the conventionally created cDNA library.

4.6 Screening of cDNA libraries by yeast complementation assay

The cDNA libraries were initially screened for copper tolerance by functional complementation using the copper mutant *cup1*$^{\Delta}$ which is sensitive to copper. The cDNA was used to rescue the sensitive yeast strain in presence of 150 μM of Cu selected on SD-Ura. Only the plasmid bearing cDNA with URA3 gene was able to grow on the Synthetically Defined medium without uracil in the control plates and plasmid carrying cDNA, having a potential to tolerate heavy metal will be able to grow both in control medium and in the presence of external metal. The pooled library cells of the size library C (1-4 Kb) was used to transform the mutant yeasts *cup1*$^{\Delta}$ to check for Cu tolerance. Many clones were screened where the total number of clones screened was 5 times the number of bacterial clones in the library. Approximately 160 Cu tolerant yeast transformants were chosen out of which 155 were tested on Cu amended media by drop assay (Figure 4.12). They were categorized into highly, moderately and low tolerant groups based on the appearance of growth in drop assay according to increasing dilutions (Table 4.3). For library B (0.5-1 Kb) a total of 1 x 10^6 clones were screened and 17 clones were identified which exhibited high tolerance and 5 and 3 clones were categorized in moderately tolerant and low tolerant groups respectively. A total of 1.2 x 10^6 clones were screened for library A (0.1-0.5 Kb) where a total of

52 Cu tolerant clones were identified. 23 clones were categorized into highly tolerant, 15 and 10 into moderately tolerant and low tolerant groups.



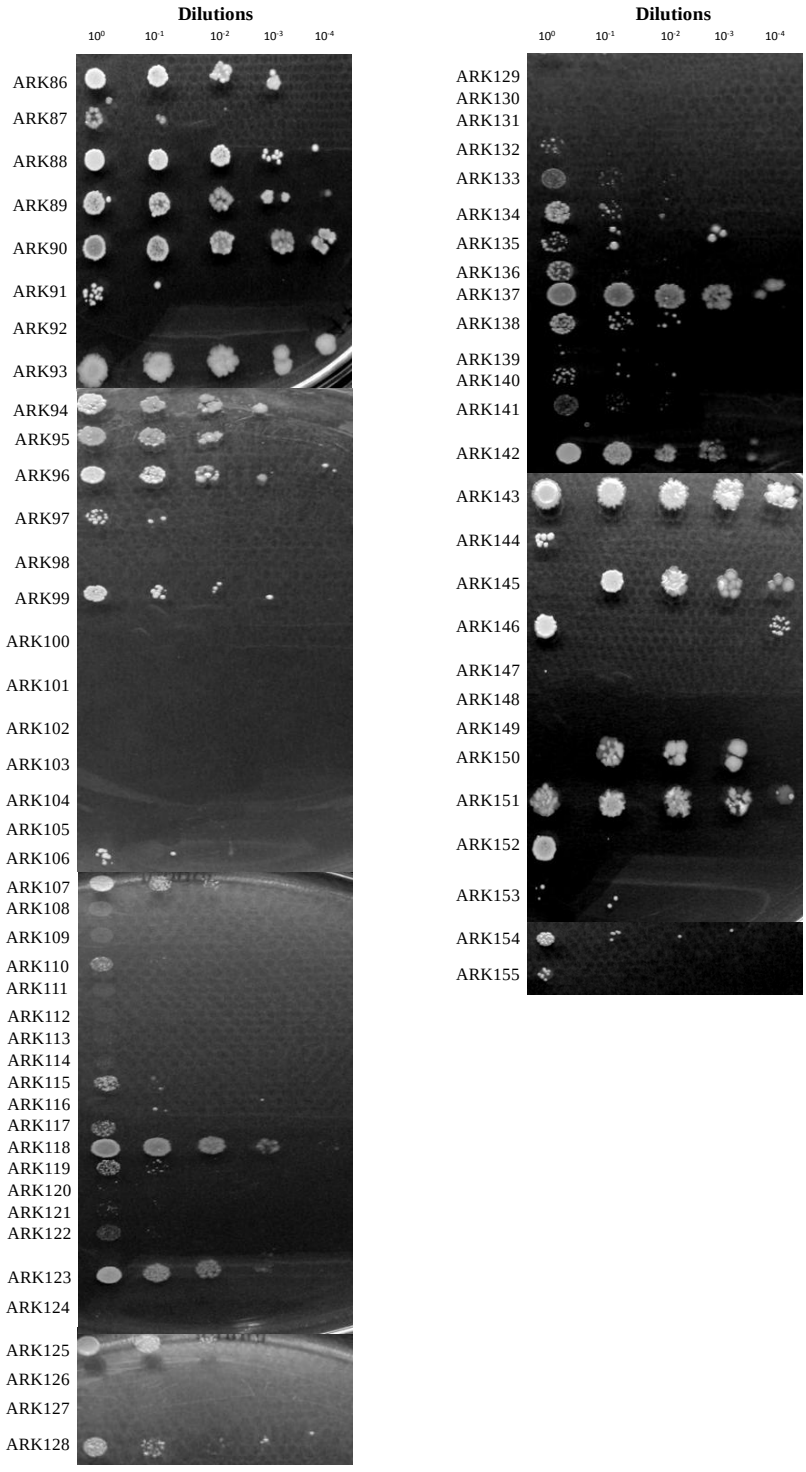


Figure 4.12: Drop test of transformants on SD-Ura supplemented with 150 μM of Copper

Table 4.3: Analysis of cDNA clones from library C for tolerance towards 150 µM of Copper

Clone name	Cu (150 µM)				
	Dilutions				
	10^{0}	10^{-1}	10^{-2}	10^{-3}	10^{-4}
ARK1	5	5	5	5	5
ARK2	5	5	5	5	0
ARK3	5	5	5	5	5
ARK4	5	5	5	5	0
ARK5	5	5	5	5	0
ARK6	5	4	0	0	0
ARK7	5	5	5	5	5
ARK8	5	5	5	5	5
ARK9	5	5	4	3	2
ARK10	5	4	3	2	0
ARK11	5	5	5	5	5
ARK12	5	5	5	5	5
ARK13	5	5	5	5	5
ARK14	5	5	5	5	4
ARK15	1	0	0	0	0
ARK16	5	5	5	3	0
ARK17	1	0	0	0	0
ARK18	5	5	4	0	0
ARK19	0	0	0	0	0
ARK20	4	1	0	0	0
ARK21	0	0	0	0	0
ARK22	0	0	0	0	0
ARK23	0	0	0	0	0
ARK24	3	2	0	0	0
ARK25	1	0	0	0	0
ARK26	0	0	0	0	0
ARK27	1	0	0	0	0
ARK28	0	0	0	0	0
ARK29	1	0	0	0	0
ARK30	0	0	0	0	0
ARK31	0	0	0	0	0
ARK32	0	0	0	0	0
ARK33	0	0	0	0	0
ARK34	0	0	0	0	0

ARK35	4	4	3	2	0
ARK36	0	0	0	0	0
ARK37	0	0	0	0	0
ARK38	0	0	0	0	0
ARK39	0	0	0	0	0
ARK40	0	0	0	0	0
ARK41	0	0	0	0	0
ARK42	0	0	0	0	0
ARK43	0	0	0	0	0
ARK44	0	0	0	0	0
ARK45	5	5	0	0	0
ARK46	5	5	4	4	0
ARK47	5	5	1	0	0
ARK48	5	5	0	0	0
ARK49	5	5	5	5	0
ARK50	5	5	5	2	0
ARK51	4	5	5	0	0
ARK52	5	4	1	0	0
ARK53	1	0	0	0	0
ARK54	1	0	0	0	0
ARK55	0	0	0	0	0
ARK56	0	0	0	0	0
ARK57	0	0	0	0	0
ARK58	0	0	0	0	0
ARK59	0	0	0	0	0
ARK60	0	0	0	0	0
ARK61	0	0	0	0	0
ARK62	0	0	0	0	0
ARK63	0	0	0	0	0
ARK64	5	5	2	0	0
ARK65	3	0	0	0	0
ARK66	0	0	0	0	0
ARK67	0	0	0	0	0
ARK68	0	0	0	0	0
ARK69	0	0	0	0	0
ARK70	0	0	0	0	0
ARK71	5	4	0	0	0
ARK72	1	0	0	0	0

ARK73	1	0	0	0	0
ARK74	1	0	0	0	0
ARK75	5	4	3	2	0
ARK76	4	3	2	1	0
ARK77	5	5	5	4	1
ARK78	0	0	0	0	0
ARK79	0	0	0	0	0
ARK80	5	5	5	4	1
ARK81	5	4	3	0	0
ARK82	5	5	4	0	0
ARK83	5	5	4	0	0
ARK84	2	0	0	0	0
ARK85	4	3	0	0	0
ARK86	5	5	5	5	5
ARK87	5	4	0	0	0
ARK88	5	5	5	5	5
ARK89	5	5	5	5	5
ARK90	5	5	5	5	5
ARK91	3	2	0	0	0
ARK92	0	0	0	0	0
ARK93	5	5	5	5	5
ARK94	5	5	5	5	0
ARK95	5	4	4	0	0
ARK96	5	5	5	5	5
ARK97	3	2	0	0	0
ARK98	0	0	0	0	0
ARK99	5	4	3	2	0
ARK100	0	0	0	0	0
ARK101	0	0	0	0	0
ARK102	0	0	0	0	0
ARK103	0	0	0	0	0
ARK104	0	0	0	0	0
ARK105	0	0	0	0	0
ARK106	4	2	0	0	0
ARK107	5	4	3	0	0
ARK108	1	0	0	0	0
ARK109	1	0	0	0	0
ARK110	1	0	0	0	0

ARK111	0	0	0	0	0
ARK112	0	0	0	0	0
ARK113	0	0	0	0	0
ARK114	0	0	0	0	0
ARK115	3	1	0	0	0
ARK116	0	0	0	0	0
ARK117	1	0	0	0	0
ARK118	5	5	5	5	4
ARK119	4	3	1	0	0
ARK120	1	0	0	0	0
ARK121	1	0	0	0	0
ARK122	4	2	0	0	0
ARK123	5	4	4	3	0
ARK124	0	0	0	0	0
ARK125	5	5	4	0	0
ARK126	0	0	0	0	0
ARK127	0	0	0	0	0
ARK128	5	4	3	2	0
ARK129	0	0	0	0	0
ARK130	0	0	0	0	0
ARK131	0	0	0	0	0
ARK132	3	0	0	0	0
ARK133	4	2	0	0	0
ARK134	5	4	2	0	0
ARK135	4	3	0	0	0
ARK136	4	1	0	0	0
ARK137	5	5	5	5	5
ARK138	5	4	4	0	0
ARK139	2	1	0	0	0
ARK140	4	3	2	0	0
ARK141	4	3	2	0	0
ARK142	5	5	5	5	2
ARK143	5	5	5	5	5
ARK144	3	0	0	0	0
ARK145	0	4	4	4	2
ARK146	5	0	0	0	3
ARK147	0	0	0	0	0
ARK148	0	0	0	0	0

ARK149	0	0	0	0	0
ARK150	2	4	4	4	0
ARK151	5	5	5	5	5
ARK152	5	1	0	0	0
ARK153	3	2	0	0	0
ARK154	5	4	3	1	1
ARK155	3	2	0	0	0

Colour code	**Level of tolerance**
	Highly tolerant clones
	Moderately tolerant clones
	Low tolerant clones

Analysis of cDNA clones from library C categorized as highly, moderately and low Cu tolerant based on their apprearance of growth in different dilutions. Highest value ‘5’ being assigned to clones growing in all the dilutions and subsequently no growth is assigned a value ‘0’

Further, plasmids from metal tolerant clones were treated with 5Fluoroorotic Acid (5-FOA) for the counter-selection of yeast (Figure 4.13). The use of 5-Fluoroorotic Acid (5-FOA) as a genetic screening method is a common practice which is based on the counterselection of yeast. Plasmid curing, plasmid shuffling, two-hybrid screens etc. are few of the methods that can employ the use of 5-FOA. 5-FOA is nontoxic to all yeast strains, however, 5-FOA is converted to the toxic form 5-flurouracil in strains expressing the functional URA3 gene that encodes for orotine-5-monophosphate decarboxylase that is responsible in the synthesis of uracil.

Yeast strains that are auxotrophic for uracil were tested for growth in presence of 5-FOA. BY4741 and DTY4 both Ura- were able to grow in 5-FOA plates; however, transformants carrying plasmids or plasmid-borne cDNAwere unable to grow in 5-FOA as 5-FOA is converted to a toxic form (5-flurouracil) in strains expressing the URA3 gene which codes for orotine 5’-

monophosphate decarboxylase that is involved in the synthesis of uracil (Boeke et al. 1987). Table 4.4 depicts the summary of the screening of the libraries.

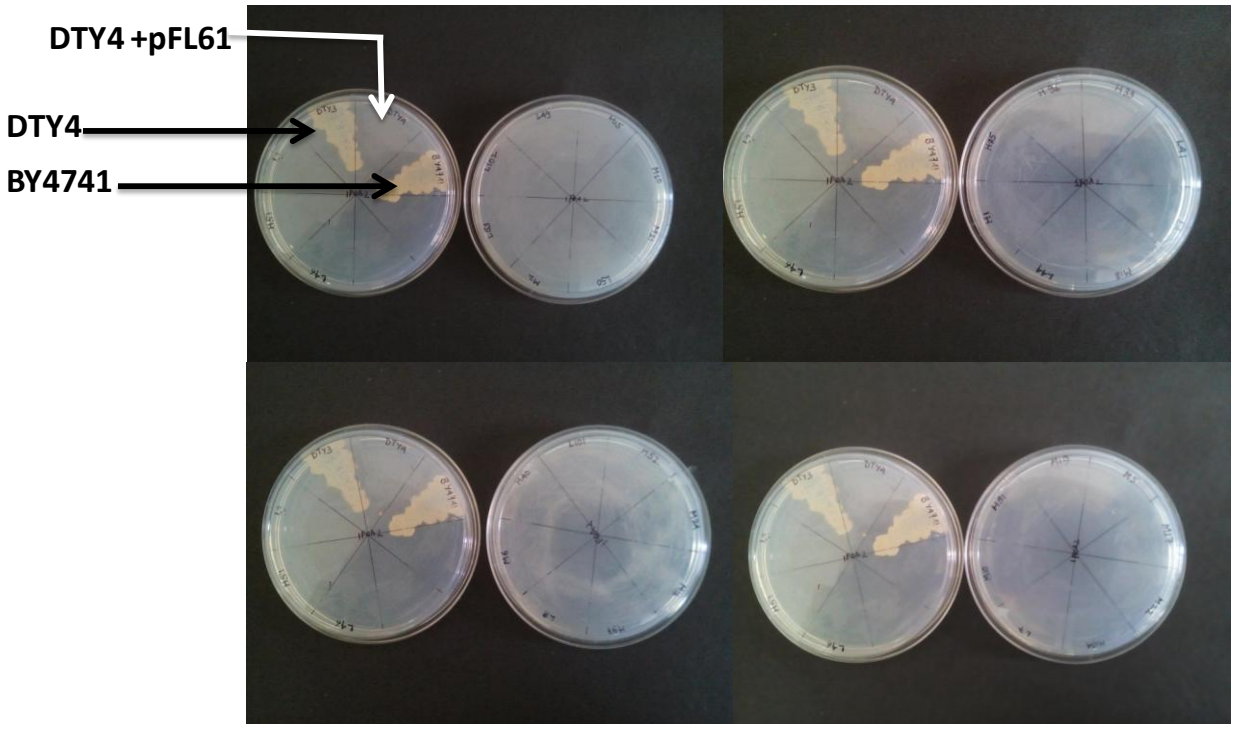


Figure 4.13: 5-FOA treatment. Parent type BY4741 and DTY4 was able to grow in SD+Ura+5-FOA agar plates since they are Ura$^-$ strains. DTY4 carrying the pFL61 plasmid did not grow because of the presence of Ura gene. Other transformants unable to grow on the agar medium were true plasmid-bearing clones and were selected for further work

Table 4.4: Summary of screening of cDNA libraries

Summary		Library C (1-4 Kb)	Library B (0.5-1 Kb)	Library A (0.1-0.5 Kb)
Total number of clones in library		$>10^6$	$>10^6$	$>10^6$
Total number of yeast transformants screened		$5x10^6$	$1x10^6$	$1.2x10^6$
Total number of Cu tolerant clones selected		160	57	57
Total number of tolerant clones tested by drop assay		155	57	52
Conclusion after drop assay	Highly tolerant clones	77	45	23
	Moderately tolerant clones	20	12	15
	Low tolerant clones	58	10	10
Final number of selected clones after FOA treatment and drop assay	Highly tolerant clones	77	17	23
	Moderately tolerant clones	20	5	15
	Low tolerant clones	58	3	10

4.7 PCR amplification of yeast plasmids

Following the drop assay, PCR was performed to ensure successful transformation of cDNA insert in mutant yeast by amplifying the insert by NF and NR primers and visualizing the amplified band for appropriate amplicon size (Figures 4.14). Few clones showing positive amplifications were selected and were further characterized.

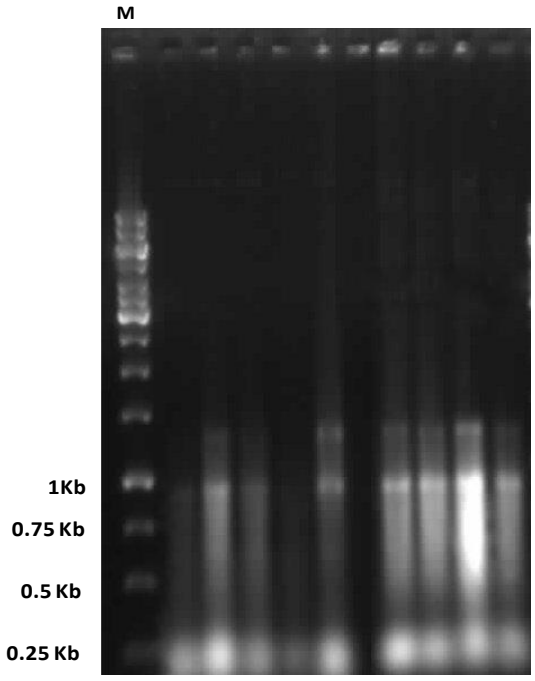


Figure 4.14: PCR amplification of the cDNA inserts in metal resistant yeast clones. M is the DNA molecular marker

To focus on gene size of around 1 Kb and understanding the role of larger proteins with respect to metal tolerance, plasmid DNA from some of the clones that showed high tolerance towards copper were further sequenced and analyzed for homology using bioinformatic tools. They were further characterized on the basis of its potential to tolerate high concentrations of copper. The genes isolated and sequenced from metal polluted soils are listed below in Table 4.5.

Table 4.5: Identified metal tolerant genes from polluted soils

Clone ID	Length (bp)	Description	E-Value	Organism	% Identity	% Query coverage
PLCc38	1667	Aldehyde dehydrogenase	190e-152	*Stylonychia lemnae*	62.7	92
PLCg49	1233	Serine protease inhibitor	5e-68	*Hypsibius dujardini*	37.94	77
PLCg39	1355	DHHC palmitoyl transferase	8e-60	*Acanthamoeba castellanii*	58.97	34
ARK1	849	Chitin synthase	1e-123	*Saccharomyces cerevisiae*	99.37	96
AKR3	904	Maltose fermentation regulatory protein MAL33	6e-86	*Saccharomyces cerevisiae*	98.47	43
ARKM1	1281	Hypothetical protein	2e-91	*Neospora caninum*	100	32
ARK10	1185	Hypothetical protein	2e-68	*Neospora caninum*	100	34

4.8 Characterization of genes involved in metal tolerance

Out of the genes involved in metal tolerance identified and isolated from the metal contaminated soil, hypothetical proteins were omitted from characterization. Aldehyde dehydrogenase, serine protease inhibitor, DHHC palmitoyl transferase, maltose fermentation regulatory protein MAL33 and chitin synthase were characterized with respect to metal tolerance.

For most of the genes, bio-accumulation and tolerance conferred by the genes to toxic concentrations of metals have been studied here. The exposure of eukaryotic microorganisms to various levels of toxic metal pollutants resulted in differential expression of genes and accumulation of transcripts encoding various functional proteins. The functions of these proteins are towards mitigating the effect of toxic metals encountered by the organisms by not strictly

confining to direct metal tolerance strategies as discussed earlier but also to the consequences arising due to the presence of external metal ions on cellular functioning. Hence the functions of the transcripts described in this study cannot be solely held responsible for metal tolerance alone. Along with metal tolerance/avoidance, there are several other stress related factors which are dealt with by the 'stress response' proteins. A concerted effort of the proteins which participate in several processes such as metal binding, transporting and saving, oxidative-reduction balance maintaining, stress response, signal transduction, etc, where significant up-regulation of expression of the genes has been observed confirmed by proteome analysis. Binding of toxic metals with metallothionein-like proteins and storing toxic metals in metal-rich granules in insoluble forms were the important strategies for detoxification. Studies by Luo et al. (2014) have revealed that most of the stress and immunity responsive proteins, such as heat shock proteins (HSP), extracellular superoxide dismutase (ECSOD) and cavortin, and the cellular redox reaction relative proteins such as 20G-Fe (II) oxygenase family oxidoreductase, aldehyde dehydrogenase and retinal dehydrogenase 2, were upregulated in the gills of oysters exposed to long term heavy metal contaminated environments.

Various transcription profiling data have shown a clear indication of the induction of genes as a result of exposure to Ag^{+}, Al^{3+}, As^{3+}, Cd^{2+}, Co^{2+}, Hg^{2+}, Mn^{2+}, Ni^{2+}, V^{3+}, and Zn^{2+}, following short-term exposure. It was noticed that induction of central elements of the oxidative stress response by the majority of investigated metals is the basic detoxification process against short-term metal exposure. General detoxification mechanisms also comprised the induction of genes coding for chaperones and those for chelation of metal ions via siderophores and amino acids.

In plants the turgor-responsive genes increased rapidly when turgor pressure dropped due to dehydration where an army of transcripts responsible for oxidative stress management became

abundant. This clearly indicates that a single gene cannot be assigned a solitary function in case of an event like soil metal toxicity. There are several factors and genes involved which work in unison to perform the task of detoxification. Hence, expression of a stress response gene cannot be directly implicated in metal accumulation but works together in a cascading effect which includes signal transduction, metal binding, transporting and cellular detoxification.

4.9 Aldehyde dehydrogenase (ALDH)

One of the clones from library C of Cd contaminated soil from France, PLCc38 showed high tolerance to Cd was analyzed. Sequence study indicated that PLCc38 demonstrated homology with aldehyde dehydrogenase and shared common ancestry from protista kingdom, phylum Ciliophora.

4.9.1 Bioinformatic analysis of PLCc38

Sequencing of the cDNA by NF and NR primers revealed the size of cDNA PLCc38 to be 1667 bp bearing an ORF of 720 bp nucleotides which encoded a protein of 239 amino acids. BLAST analysis revealed 62.7 % homology with aldehyde dehydrogenase (ALDH) family 7 member of *Stylonychia lemnae* phylum Ciliophora (Figure 4.15 and Figure 4.16). The predicted molecular mass and pI were 26.4 kDa and 6.9 respectively. Highly conserved regions were found by aligning the deduced sequence of PLCc38 with other closely related ALDH amino acids from same phylum and also from few plants (Figure 4.17).

Sequences producing significant alignments Download Manage Columns Show 100

select all *0 sequences selected* GenPept Graphics

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☐	aldehyde dehydrogenase [uncultured eukaryote]	491	491	45%	1e-170	100.00%	QBQ70680.1
☐	turgor pressure sensor [Stylonychia lemnae]	332	625	92%	1e-151	62.70%	CDW90834.1
☐	Turgor pressure sensor [Oxytricha trifallax]	326	607	92%	8e-147	62.30%	EJY81836.1
☐	mitotic spindle assembly checkpoint protein [Plasmopara halstedii]	301	595	91%	1e-146	55.34%	XP_024577216.1
☐	hypothetical protein FGO68_gene13629 [Halteria grandinella]	322	601	92%	1e-145	59.46%	TNV79540.1
☐	aldehyde dehydrogenase family 7 member A1 [Pythium insidiosum]	312	584	91%	1e-139	58.50%	GAX96345.1
☐	PREDICTED: aldehyde dehydrogenase family 7 member A1 [Brassica oleracea var. oleracea]	295	583	91%	2e-138	55.16%	XP_013630745.1
☐	turgor pressure sensor [Sterkiella nova]	310	580	92%	3e-138	57.31%	AED87003.1
☐	turgor [Sterkiella histriomuscorum]	310	578	92%	6e-138	57.31%	AAZ23912.1
☐	PREDICTED: aldehyde dehydrogenase family 7 member A1 [Raphanus sativus]	292	578	91%	8e-138	54.58%	XP_018451019.1

Figure 4.15: Sequence homology search for cDNA PLCc38 using BLASTx through NCBI

```
gcgg ggg atc ggt gaa ggt gat gtt tac aat agc atc aac cct cac aat gga gag gtg atc
  R   G   I   G   E   G   D   V   Y   N   S   I   N   P   H   N   G   E   V   I
gcc acc atc cgc tac gga acc aaa acc aac tac gaa gaa gct gta cag tcc atg caa cga
 A   T   I   R   Y   G   T   K   T   N   Y   E   E   A   V   Q   S   M   Q   R
gga aag gct gag tgg atg acc acc cct gct cca aac aga gga gaa atc gtc aga cag atc
 G   K   A   E   W   M   T   T   P   A   P   N   R   G   E   I   V   R   Q   I
ggt gac gcc ttc aga aaa cac aag gac gca ttg gga aga ctc atc tcg ctt gaa atg gga
 G   D   A   F   R   K   H   K   D   A   L   G   R   L   I   S   L   E   M   G
aag atc ttg agt gaa ggc ctt gga gag gtc caa gaa ttc att gac gtc tgc gac atg gca
 K   I   L   S   E   G   L   G   E   V   Q   E   F   I   D   V   C   D   M   A
gtc ggc atg agc aga act atc gat gga aag atc ttg aac tcc gaa aga cca aac cac ttc
 V   G   M   S   R   T   I   D   G   K   I   L   N   S   E   R   P   N   H   F
atg atg gaa caa tgg aat cct ctt ggt ctc atc ggt gtc atc act gcc ttc aac ttc cct
 M   M   E   Q   W   N   P   L   G   L   I   G   V   I   T   A   F   N   F   P
tgc gcc gtc ttc ggc tgg aac ttt gct atc gct gct atc tgc ggt gac ttg act ctt tgg
 C   A   V   F   G   W   N   F   A   I   A   A   I   C   G   D   L   T   L   W
aaa ggc tct gaa agc act tca ttg gtg aca ctc gcc act gca aaa atc gtt gct gaa gtc
 K   G   S   E   S   T   S   L   V   T   L   A   T   A   K   I   V   A   E   V
ctt gag aga aac aag tgc cca gaa ggt gtc atg acc gtc gtt tct ggt gaa gga aag aca
 L   E   R   N   K   C   P   E   G   V   M   T   V   V   S   G   E   G   K   T
att gga gag ctt atc acc aat gac cca aga atg gaa ctc gtc tca ttc act gga tca acc
 I   G   E   L   I   T   N   D   P   R   M   E   L   V   S   F   T   G   S   T
ttt gtg gga aga aga gtc agc gag atc gtc cac aga aga ttc gga aga act gtc ctc gaa
 F   V   G   R   R   V   S   E   I   V   H   R   R   F   G   R   T   V   L   E
ctc gga gga nac aat gct att att gtt tgc gag gac gga gat aca tat ggc ctt aag agc
 L   G   G   X   N   A   I   I   V   C   E   D   G   D   T   Y   G   L   K   S
tgt tac ctt cgc tgc tgt cgg aac ctg cgg aca gag atg cac aac cgt tag aag act gat
 C   Y   L   R   C   C   R   N   L   R   T   E   M   H   N   R   -   K   T   D
cgt cca cga gaa gca cta tga cgg ttt cgt caa gca gtt gct tga agt cta ccc aac tat
 R   P   R   E   A   L   -   R   F   R   Q   A   V   A   -   S   L   P   N   Y
caa gat cgg caa ccc aat gga aag aag gta ctc tct gcg gtc cac ttc aca cca aga acg
 Q   D   R   Q   P   N   G   K   K   V   L   S   A   V   H   F   T   P   R   T
```

Figure 4.16: Translation of cDNA PLCc38 encoding for a protein which showed homology with aldehyde dehydrogenase

```
         10         20         30         40         50         60
MQRGKAEWMTTPAPNRGEIVRQIGDAFRKHKDALGRLISLEMGKILSEGLGEVQEFIDVC

         70         80         90        100        110        120
DMAVGMSRTIDGKILNSERPNHFMMEQWNPLGLIGVITAFNFPCAVFGWNFAIAAICGDL

        130        140        150        160        170        180
TLWKGSESTSLVTLATAKIVAEVLERNKCPEGVMTVVSGEGKTIGELITNDPRMELVSFT

        190        200        210        220        230       239
GSTFVGRRVSEIVHRRFGRTVLELGGNNAIIVCEDGDTYGLKSCYLRCCRNLRTEMHNR
```

Figure 4.17: Amino acid sequence of PLCc38. PR box is highlighted in green, NAD^+ Binding sites are highlighted in blue and pink. Glycine-rich region is underlined. Glutamic acid active site is in yellow. Catalytic thiol is represented in red.

ALDH mainly acts as an aldehyde metabolizing enzyme and metabolizes aldehydes into its carboxylic acid. This is achieved by reducing NAD^+, whose binding in the active site is coordinated by glutamic acid. The catalytic thiol is provided by the cysteine residue enabling a nucleophilic attack on the carbonyl carbon of the aldehyde.

In this study, the conserved signature sequence of PAPNRGEIVRQIGD (residues 12-25) and NAD binding sites TAFN (residues 98-101) and FTGS (residues 179-182) confirmed the cDNA as a NAD^+ dependent ALDH. A variant, LELSGNNA, in the signature sequence of ALDH's glutamic acid active site was observed within residues 202 to 209. The catalytic thiol (Cys 228) and neighbor Arg are conserved in all ALDHs (Liu et al. 1997). The presence of a glycine-rich sequence Gly-X_4-Gly (residues 181-186) which is native to ALDH, is involved in an interaction with nicotinamide ring of NAD (Figure 4.17). Further study of the 239-a-long polypeptide also revealed 16 possible phosphorylation sites but no O-linked glycosylation sites. Multiple

sequence alignment of PLCc38 with homologous sequences of aldehydes dehydrogenase shows important conserved regions (Figure 4.18).

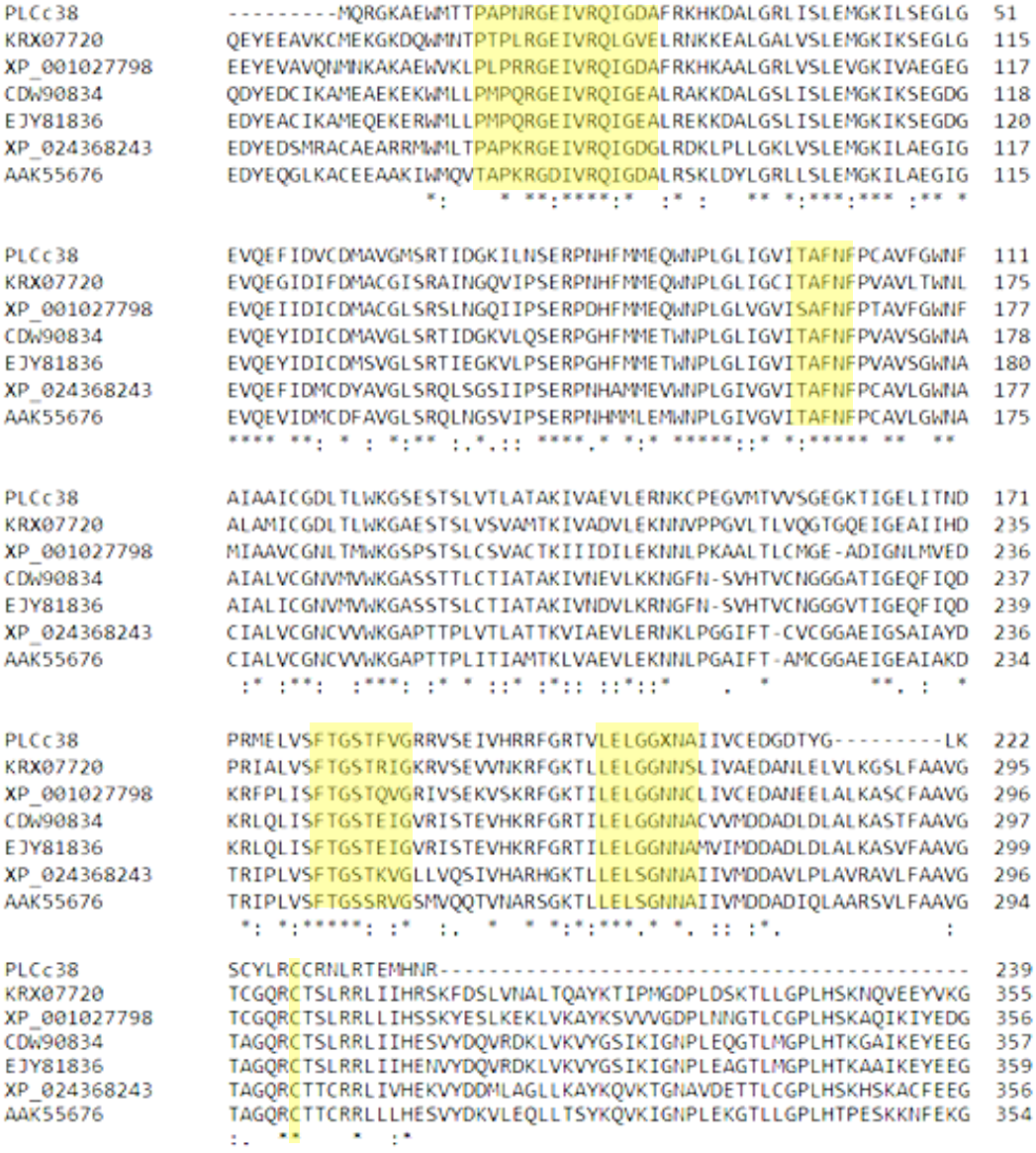


Figure 4.18: Multiple sequence alignment of PLCc38 gene with other homologous sequences retrieved by BLASTp analysis. Accession numbers are *Pseudocohnilembus persalinus* (KRX07720) *Tetrahymena thermophila* SB210 (XP001027798), *Stylonychia lemnae* (CDW90834), *Oxytricha trifallax* (EJY81836), *Physcomitrella patens* (XP024368243) and *Arabidopsis thaliana* (AAK55676). Important conserved regions are highlighted.

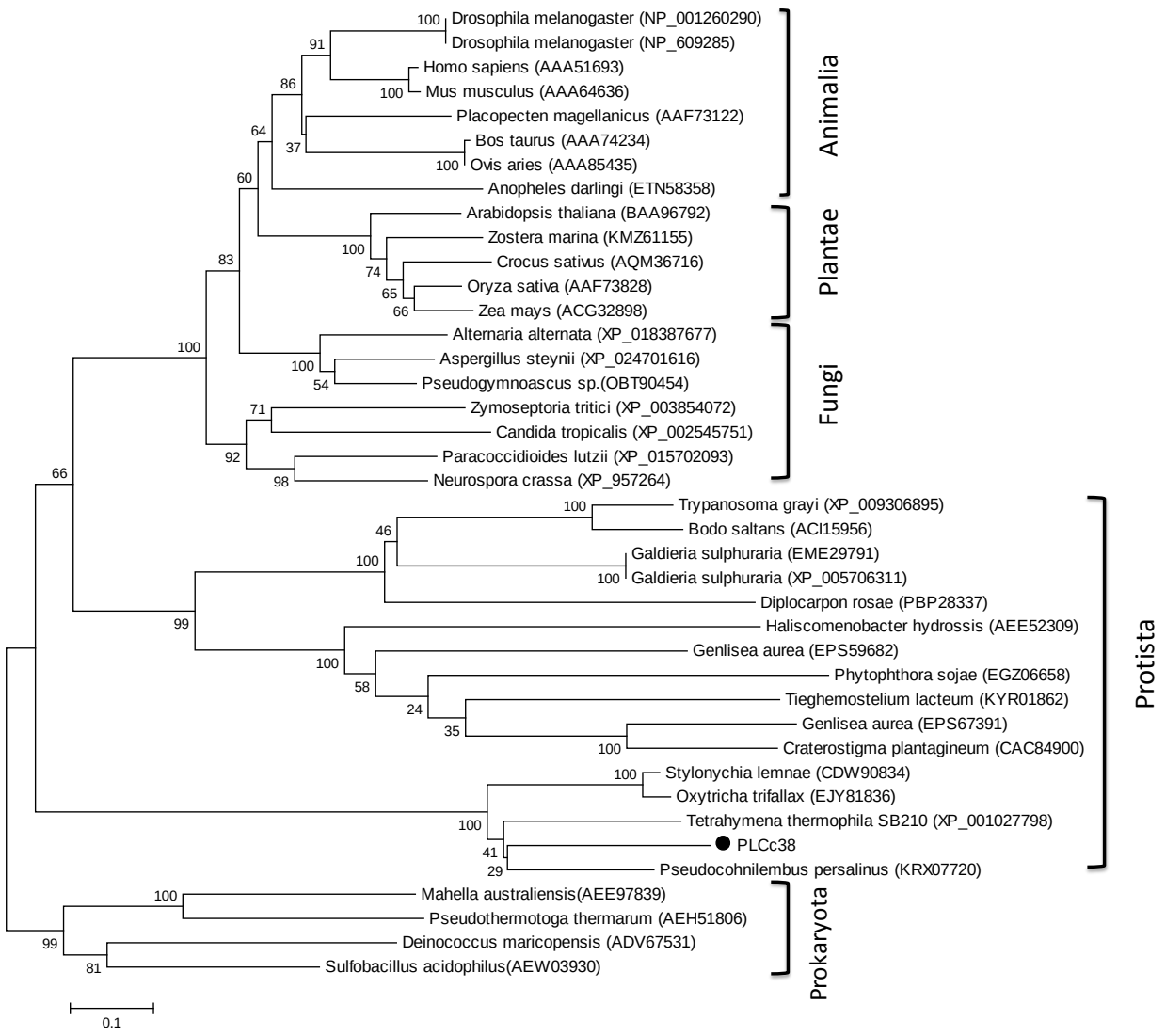


Figure 4.19: Phylogenetic relation of the transcript PLCc38 with ALDHs of other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses. Prokaryota ALDH sequences were included as outgroup. Branch lengths are proportional to evolutionary distances.

Phylogenetic analysis of ALDH from various taxons has distinctly divided them into different lineages such as Animalia, Plantae, Fungi, Protista and Prokaryota (Figure 4.30). The present study ALDH was clustered with Protista lineage confirming its origin might be from the protists.

Sequence of PLCc38 obtained in this study was submitted to National Center of Biotechnology Information (NCBI) database under the accession number MK077675 (Appendix II).

4.9. 2 Potential of PLCc38 to tolerate multiple metals

To test the cDNA PLCc38 for its multi metal tolerance ability transformation was carried out into a number of metal sensitive yeasts and were allowed to respond to different metal and at different concentrations.

Metal sensitive strains *cup1*$^{\Delta}$ *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PLCc38 were spotted on metal amended medium in drop assay and the mutant strains of *S. cerevisiae* transformed with PLCc38 were able to grow on SD-Ura containing 150 μM of Cu, 40 μM Cd, 10 mM Zn and 2 mM Co respectively. In contrast, no growth was observed when empty vector pFL61 transformed yeast mutants were tested for comparison. The metal-resistant parent strain BY4741 carrying an empty vector pFL61 was able to grow in all metal-supplemented media while *cup1*^{s} transformed with empty vector pFL61, failed to grow on 150 μM of Cu amended medium. In general, as yeast dilution increased, the decrease in the growth was observed for all yeast strains in metal amended medium (Figure. 4.20).

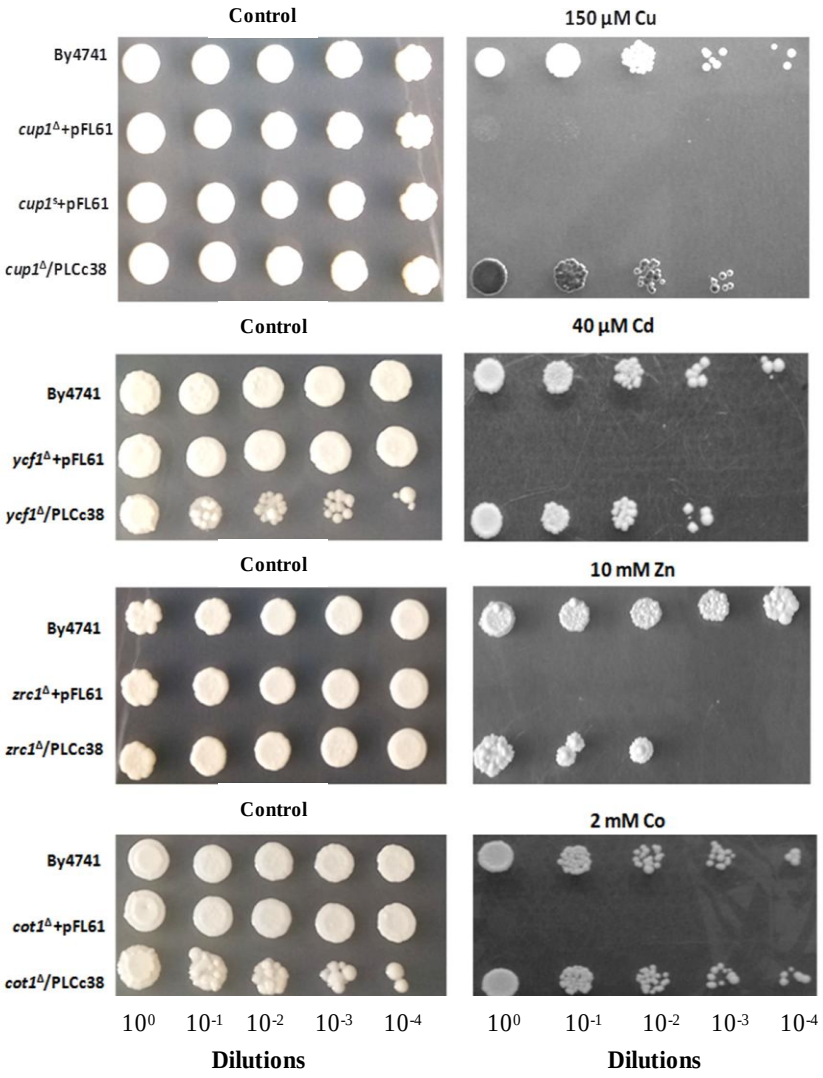


Figure 4.20: Drop assay of PlCc38 transformed *Saccharomyces cerevisae* mutants, *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ on SD-Ura agar. BY4741wild type and mutant strains transformed with empty vector pFL61 for positive and negative control. Serially diluted cultures were spotted on SD-Ura medium with or without metal supplement as indicated. Yeast strain *cup1*^{S} was also used as control.

To investigate the growth profile of yeast transformants bearing cDNA PLCc38 in presense of Cu, Cd, Zn and Co, the transformants were allowed to grow in SD-Ura broth supplemented with a wide range of metal concentrations (Table 4.6). The transformant *cup1*$^{\Delta}$/PLCc38 was tolerant towards high concentrations of Cu and was capable of growing in high Cu concentration. With

the increase in external Cu concentration, retardation of growth was observed although it was not majorly affected up to 500 μM Cu. Nevertheless, growth was significantly higher compared to wild-type strain BY4741. At 1000 μM Cu, the growth of both the *cup1*$^{\Delta}$/PLCc38 and wild type strain suffered severely. Mutant *cup1*$^{\Delta}$ transformed with empty pFL61 recorded no growth in presence of Cu (Figure 4.21A). The growth of *ycf1*$^{\Delta}$/PLCc38 substantially increased to 60 μM of Cd and subsequently decreased at higher concentrations, the growth being comparable with that of the wild type strain BY4741 for all the concentrations (Figure 4.21B). The transformant *zrc1*$^{\Delta}$/PLCc38 grew well up to 12 mM, with maximum growth observed at 10 mM of Zn. Thereafter, the growth significantly decreased at 13 mM of Zn concentration. For all the concentrations tested, the transformant *zrc1*$^{\Delta}$/PLCc38 registered higher growth than the wild type strain BY4741 (Figure 4.21C). In case of cobalt-ammended media, maximum growth was recorded at 2 mM of Co for PLCc38 transformed *cot1*$^{\Delta}$ cells and subsequently decreased with increase in concentration of Co. The observed growth of *cot1*$^{\Delta}$/PLCc38 was appreciably higher compared to BY4741 transformed with empty pFL61 vector for all concentrations of Co (Figure 4.21D). Heavy metals such as Cu and Cd can be involved in oxidative modifications and redox cycles resulting in extremely toxic reactive oxygen species (ROS) being hyperaccumulated (Romero-Puertas et al. 2002; Remans et al. 2010). There is growing evidence that links increased levels of ROS species to lipid peroxidation that leads to generation of large amounts of harmful aldehydes in the process (Chaoui et al. 1997), which are eliminated by ALDH. Presence of toxic levels of cadmium, cobalt and zinc can ultimately result in the inactivation of enzymes essential for life processes and also cause damage to key enzymes required for defense against superoxides (Geslin et al. 2001; Atici et al. 2005; Dietz et al. 1999).

Table 4.6: Growth response of *cup1*$^{\Delta}$,*ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PlCc38 compared to controls

	OD_{600} of Yeast samples		
Cu concentration(µM)	***cup1*$^{\Delta}$/PlCc38**	**BY4741**	***cup1*$^{\Delta}$/pFL61**
150	1.933±0.0345a	1.352±0.014b	0.080±0.010c
300	1.814±0.0345a	1.347±0.006b	0.045±0.005c
500	1.768±0.0345a	1.109±0.005b	0.015±0.005c
1000	0.146±0.0345a	0.095±0.004b	0.004±0.002c
Cd concentration(µM)	***ycf1*$^{\Delta}$/PlCc38**	**BY4741**	***ycf1*$^{\Delta}$/pFL61**
40	1.154±0.020a	1.122±0.004a	0.060±0.010b
60	1.008±0.006a	0.929±0.027a	0.035±0.005b
80	0.509±0.005a	0.496±0.006a	0.015±0.005b
100	0.253±0.025a	0.228±0.015a	0.002±0.000b
Zn concentration(mM)	***zrc1*$^{\Delta}$/PlCc38**	**BY4741**	***zrc1*$^{\Delta}$/pFL61**
10	1.866±0.101a	1.596±0.058b	0.175±0.005c
11	1.690±0.066a	1.438±0.086b	0.051±0.001c
12	1.395±0.161a	1.024±0.090b	0.031±0.002c
13	0.175±0.019a	0.095±0.004a	0.004±0.001b
Co concentration(mM)	***cot1*$^{\Delta}$/PlCc38**	**BY4741**	***cot1*$^{\Delta}$/pFL61**
2	0.260±0.025a	0.1625±0.019b	0.008±0.001c
10	0.141±0.006a	0.1145±0.005b	0.005±0.001c
25	0.141±0.005a	0.055±0.004b	0.003±0.000c
50	0.137±0.005a	0.042±0.001a	0.002±0.000b

O.D. values are Mean ± Standard deviation. Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (n=3)

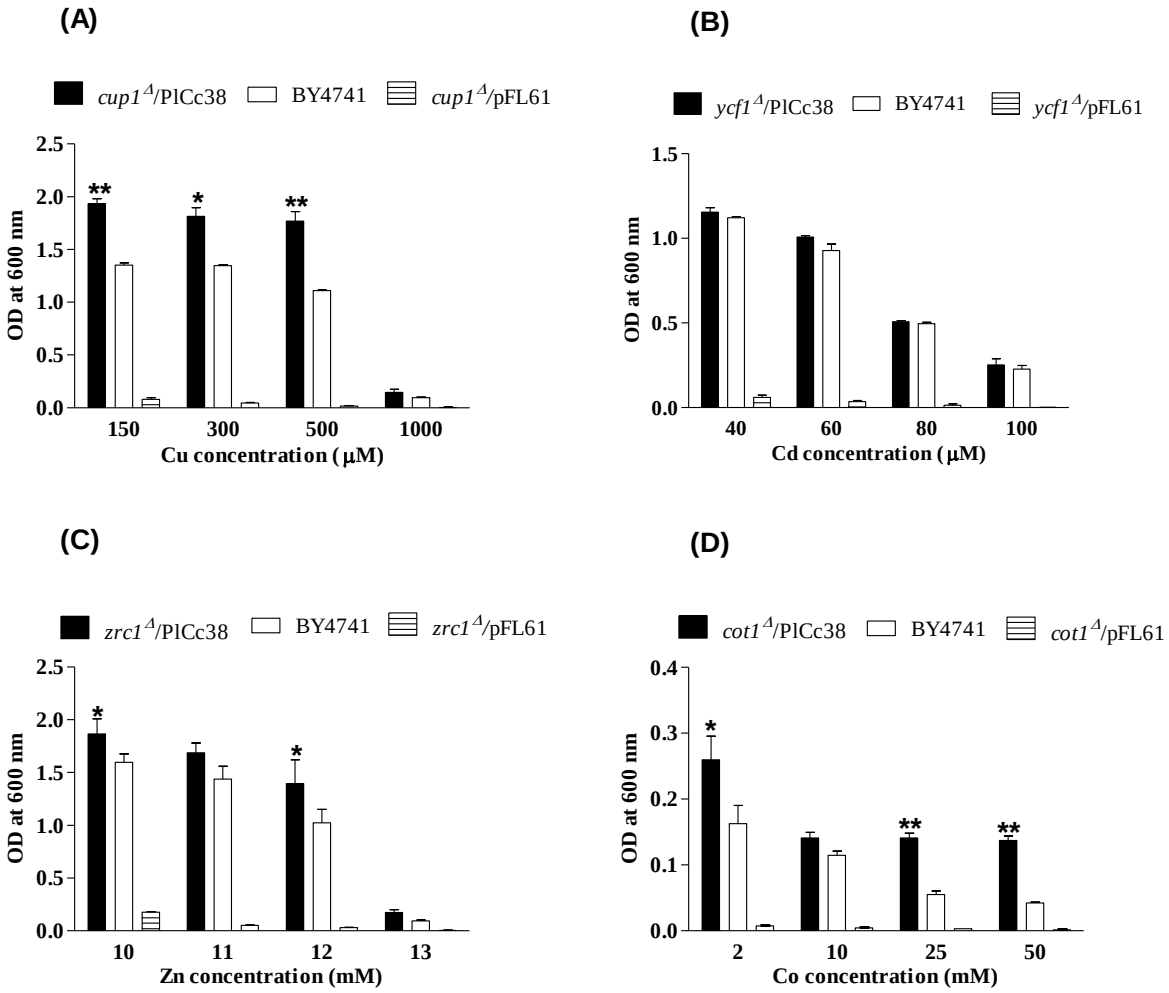


Figure 4.21: Growth response of (A) *cup1*$^{\Delta}$ (B) *ycf1*$^{\Delta}$ (C) *zrc1*$^{\Delta}$ and (D) *cot1*$^{\Delta}$ transformed with PlCc38, wild type BY4741 and mutants transformed with empty vector pFL61 in SD-Ura broth supplemented with different metals at different concentrations as shown. (* and ** are significant at $P<0.05$ and $P<0.01$ respectively).

4.9.3 Metal accumulation by PLCc38

The intracellular estimation of metals in yeast transformants growing in presence of different concentrations of Cu, Cd, Zn and Co was performed by ICP-MS (Table 4.7). Uptake of copper by *cup1*$^{\Delta}$/PLCc38 increased with increase in Cu concentration in the growth medium up to 500 μM and decreased significantly thereafter (Figure 4.22A). With the increase in cadmium

concentration in the medium, the intracellular Cd content of *ycf1*$^{\Delta}$/PLCc38 cells decreased with maximum being observed at 40 μM of Cd. The measured Cd content of *ycf1*$^{\Delta}$/PLCc38 was significantly higher for all the different concentrations when compared to the Cd content of the wild type strain BY4741 transformed with empty vector (Figure 4.22B). The *zrc1*$^{\Delta}$ /PLCc38 cells accumulated higher content of Zn up to 11 mM and decreased at higher concentrations. The Zn content was significantly higher in these cells compared to the wild type strain in all the concentrations of Zn (Figure 4.22C).

Table 4.7: Metal contents of yeast mutants *cup1*$^{\Delta}$,*ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PlCc38 compared to controls

	Metal uptake (mg^{-1}g dry weight)		
Cu concentration(μM)	***cup1*$^{\Delta}$/PlCc38**	**BY4741**	***cup1*$^{\Delta}$/pFL61**
150	0.084±0.007a	0.071±0.006a	0.005±0.000b
300	0.245±0.007a	0.146±0.008b	0.012±0.000c
500	0.307±0.020a	0.267±0.021b	0.009±0.007c
1000	0.151±0.005a	0.144±0.003a	0.005±0.001b
Cd concentration(μM)	***ycf1*$^{\Delta}$/PlCc38**	**BY4741**	***ycf1*$^{\Delta}$/pFL61**
40	0.054±0.003a	0.044±0.003b	0.0025±0.000c
60	0.042±0.001a	0.039±0.001b	0.001±0.000c
80	0.033±0.002a	0.028±0.003b	BDL
100	0.013±0.001a	0.010±0.001b	BDL
Zn concentration(mM)	***zrc1*$^{\Delta}$/PlCc38**	**BY4741**	***zrc1*$^{\Delta}$/pFL61**
10	1.617±0.023a	0.526±0.024b	0.463±0.053b
11	1.652±0.024a	0.243±0.005b	0.125±0.035c
12	1.041±0.015a	0.144±0.005b	0.060±0.014b
13	0.074±0.002a	0.052±0.001a	0.015±0.007a

Co concentration(mM)	*cot1*$^{\Delta}$/PlCc38	BY4741	*cot1*$^{\Delta}$/pFL61
2	0.084±0.002a	0.031±0.002b	0.008c
10	0.266±0.015a	0.241±0.002b	0.006c
25	0.098±0.003a	0.079±0.002b	BDL
50	0.015±0.000a	0.012±0.002a	BDL

Metal uptake values are Mean ± Standard deviation. Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (n=3). BDL-Below Detection Limit

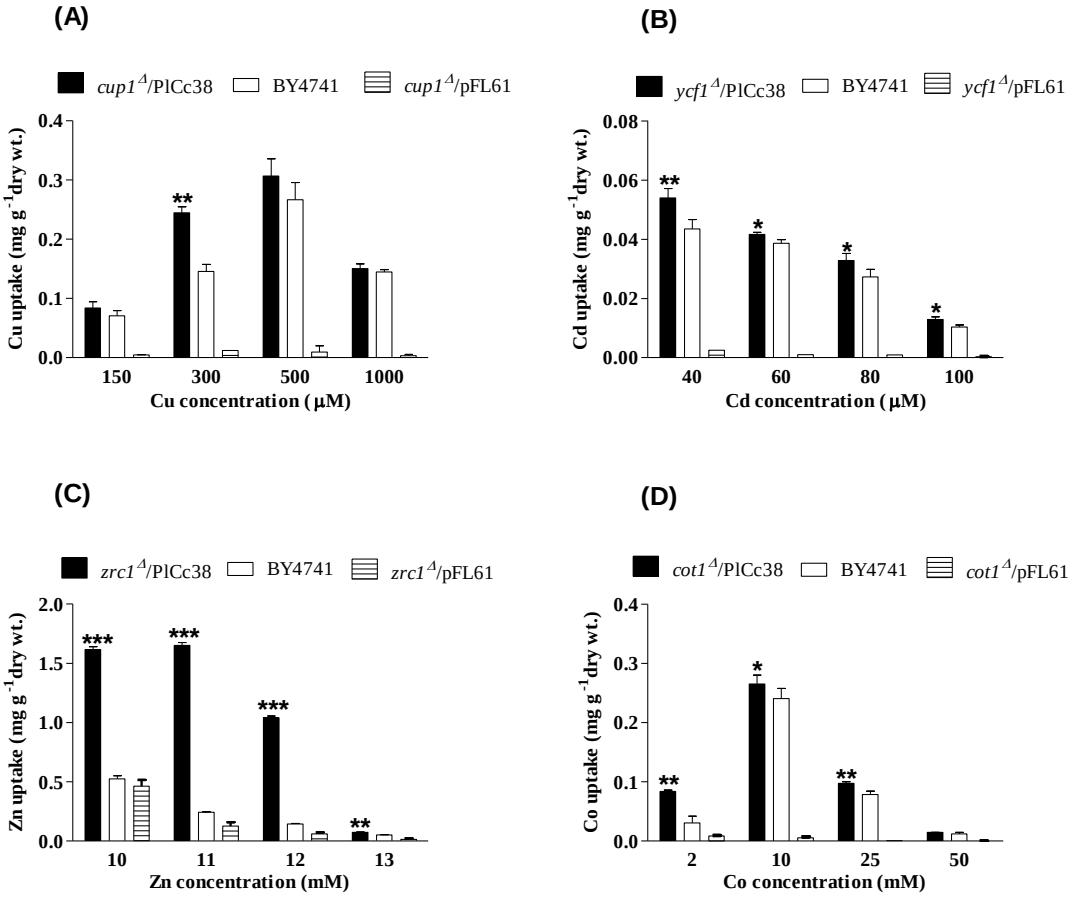


Figure 4.22: Effect of different concentration of (A) Copper (B) Cadmium (C) Zinc (D) Cobalt on metal uptake by PlCc38 when compared to wild type BY4741 after 40 hrs of growth. Error bars are ± Standard deviation (n=3). (*, ** and *** are significant at $P<0.05$, $P<0.01$ and $P<0.001$ respectively)

The *cot1*$^{\Delta}$/PLCc38 cells recorded highest metal uptake at 10 mM Co which was comparable to the Co content of the wild type strain BY4741 (Figure 4.22D). In a study on plant ALDH, considerable decrease in growth and survival was observed in wild type plants in comparison with transgenic plants containing ALDH gene when treated with 175 μM of copper and 300 μM of cadmium. Results have also shown that compared to wild type plants, the transgenic plants were less affected as far as germination, growth and lipid peroxidation are concerned (Sunkar et al. 2003). The results of this study are consistent with these findings and have depicted tolerance towards high concentrations of all four heavy metals, Cu, Cd, Zn and Co.

4.9.4 Oxidative stress response and tolerance

For oxidative stress tolerance test, the transcript PLCc38 was used to functionally complement superoxide dismutase yeast mutant *sod2*$^{\Delta}$ and rescue the oxidative stress sensitivity of the mutant strain. Different concentrations of H_2O_2 (0 to 3 mM) were used in the growth medium of the transformant *sod2*$^{\Delta}$/PLCc38 (Table 4.8). Under control condition (0 mM H_2O_2), growth pattern of BY4742, BY4742 transformed with PLCc38, *sod2*$^{\Delta}$ transformed with empty pFL61 and *sod2*$^{\Delta}$/PLCc38 was similar. When media was supplemented with 1 mM H_2O_2, growth of BY4742, BY4742/PLCc38 and *sod2*$^{\Delta}$/PLCc38 showed similar level of growth and was not much affected. However, significant retardation of the growth of *sod2*$^{\Delta}$ carrying empty pFL61 was observed after 20 hrs. The transcript could confer tolerance to *sod2*$^{\Delta}$ till 2 mM H_2O_2 but at this concentration, growth of *sod2*$^{\Delta}$ with empty pFL61 was significantly hindered. The growth of BY4742/PLCc38 was observed to be slightly higher than BY4742 alone at this concentration. At 3 mM H_2O_2, only the wild type and BY4742/PLCc38 survived till 15 hrs and thereafter showed sharp decline in their growth. As evident from the drop assay, BY4742/PLCc38 was able to grow till 3 dilutions; on the other hand, BY4742 alone could grow only till the first dilution. At this

concentration neither *sod2*$^{\Delta}$/PLCc38 nor *sod2*$^{\Delta}$ transformed with empty pFL61 could survive (Figure 4.23).

Heavy metals can damage cells in two ways; firstly, by binding to the active sites of the functional domain thereby inactivating biomolecules and secondly, by stimulating the formation of reactive oxygen species (ROS) and free radicals. ROS formation may be due to direct electron transfer involving metal cations, or as a result of metal-mediated metabolic reaction inhibition. If the unregulated oxidation process is not mediated, a higher rate of ROS formation and accumulation will result in severe oxidative stress. Studies have indicated that the most reactive aldehydes are produced from post-translation modifications or "protein carbonylation" induced by ROS. ALDH continues to oxidize free aldehydes into less toxic compounds, thus helping to detoxify those (Grimsrud et al. 2008). Cellular damages from abiotic stress have been studied mostly in plants, where the induced function of ALDH in response to drought stress, ROS-producing chemical treatment and heavy metal exposure was verified by transgenic technologies. Reports suggest, the oxidative stress in transgenic plants (Liu and Schnable, 2002) and animal tissues (Hlavacova et al. 2015) due to lipid peroxidation decreased considerably when the ALDH gene was expressed in the hosts. In a similar study, the metal-induced oxidative stress caused by heavy metal cadmium, significantly altered the proteome of a fungal pathogen *Trichosporon asahii* and as a result increased the abundance of aldehyde dehydrogenase-like protein (Ilyas et al. 2014). Regulation of aldehyde metabolism was also linked with oxidative stress in yeasts in response to a wide array of metals (Hosiner et al. 2014).

Table 4.8: Growth kinetics of *sod2*$^{\Delta}$ transformed with PLCc38 in response to oxidative stress

	Control (0 mM H_2O_2)				1 mM H_2O_2			
Time (Hrs)	**BY4742**	**BY4742/ PLCc38**	***sod2*$^{\Delta}$/pFL61**	***sod2*$^{\Delta}$/ PLCc38**	**BY4742**	**BY4742/ PLCc38**	***sod2*$^{\Delta}$/pFL61**	***sod2*$^{\Delta}$/ PLCc38**
0	0.020±0.002a	0.022±0.007a	0.024±0.001a	0.031±0.002a	0.027±0.003a	0.027±0.003a	0.031±0.002a	0.027±0.007a
5	0.123±0.008a	0.173±0.062a	0.117±0.004a	0.128±0.009a	0.127±0.004a	0.177±0.067a	0.093±0.002a	0.122±0.003a
10	0.458±0.046a	0.458±0.046a	0.452±0.028a	0.515±0.052a	0.400±0.054a	0.450±0.016a	0.248±0.000b	0.413±0.015a
15	0.831±0.067a	0.831±0.067a	0.813±0.028a	0.772±0.054a	0.585±0.017a	0.635±0.054a	0.407±0.001b	0.543±0.043a
20	1.855±0.092a	1.855±0.092a	1.806±0.022a	1.856±0.076a	1.425±0.006a	1.475±0.064a	0.718±0.000b	1.518±0.007a
25	2.081±0.069a	2.081±0.069a	2.031±0.140a	2.134±0.15a	1.724±0.015a	1.774±0.086a	0.555±0.002b	1.707±0.007a
30	1.929±0.008a	1.929±0.008a	1.908±0.120a	2.122±0.002a	1.712±0.013a	1.762±0.084a	0.500±0.003b	1.637±0.047a

	2 mM H_2O_2				3 mM H_2O_2			
Time (Hrs)	**BY4742**	**BY4742/ PLCc38**	***sod2*$^{\Delta}$/pFL61**	***sod2*$^{\Delta}$/ PLCc38**	**BY4742**	**BY4742/ PLCc38**	***sod2*$^{\Delta}$/pFL61**	***sod2*$^{\Delta}$/ PLCc38**
0	0.025±0.000a	0.025±0.001a	0.023±0.000a	0.022±0.042a	0.025±0.000a	0.025±0.000a	0.024±0.012a	0.029±0.012a
5	0.183±0.071a	0.183±0.075a	0.083±0.001a	0.120±0.014a	0.135±0.002a	0.143±0.006a	0.031±0.010b	0.042±0.010b
10	0.411±0.071a	0.411±0.056a	0.154±0.000b	0.302±0.002a	0.143±0.004b	0.164±0.007a	0.003±0.001c	0.005±0.01c
15	0.453±0.029a	0.536±0.053a	0.114±0.002b	0.424±0.003a	0.147±0.008a	0.167±0.008a	0.003±0.001b	0.004±0.02b
20	1.180±0.084a	1.330±0.013a	0.118±0.001c	0.661±0.101b	0.147±0.008a	0.143±0.004a	0.002±0.000b	0.003±0.001b
25	1.360±0.037a	1.454±0.028a	0.096±0.003c	0.767±0.064b	0.124±0.001a	0.126±0.004a	0.002±0.001b	0.002±0.000b
30	1.357±0.050a	1.356±0.049a	0.081±0.000c	0.625±0.078b	0.113±0.001a	0.115±0.001a	0.001±0.000b	0.002±0.002b

O.D. values are Mean ± Standard deviation. Values sharing common letters for a particular time interval between samples are not significantly different at $P<0.05$ (n=3)

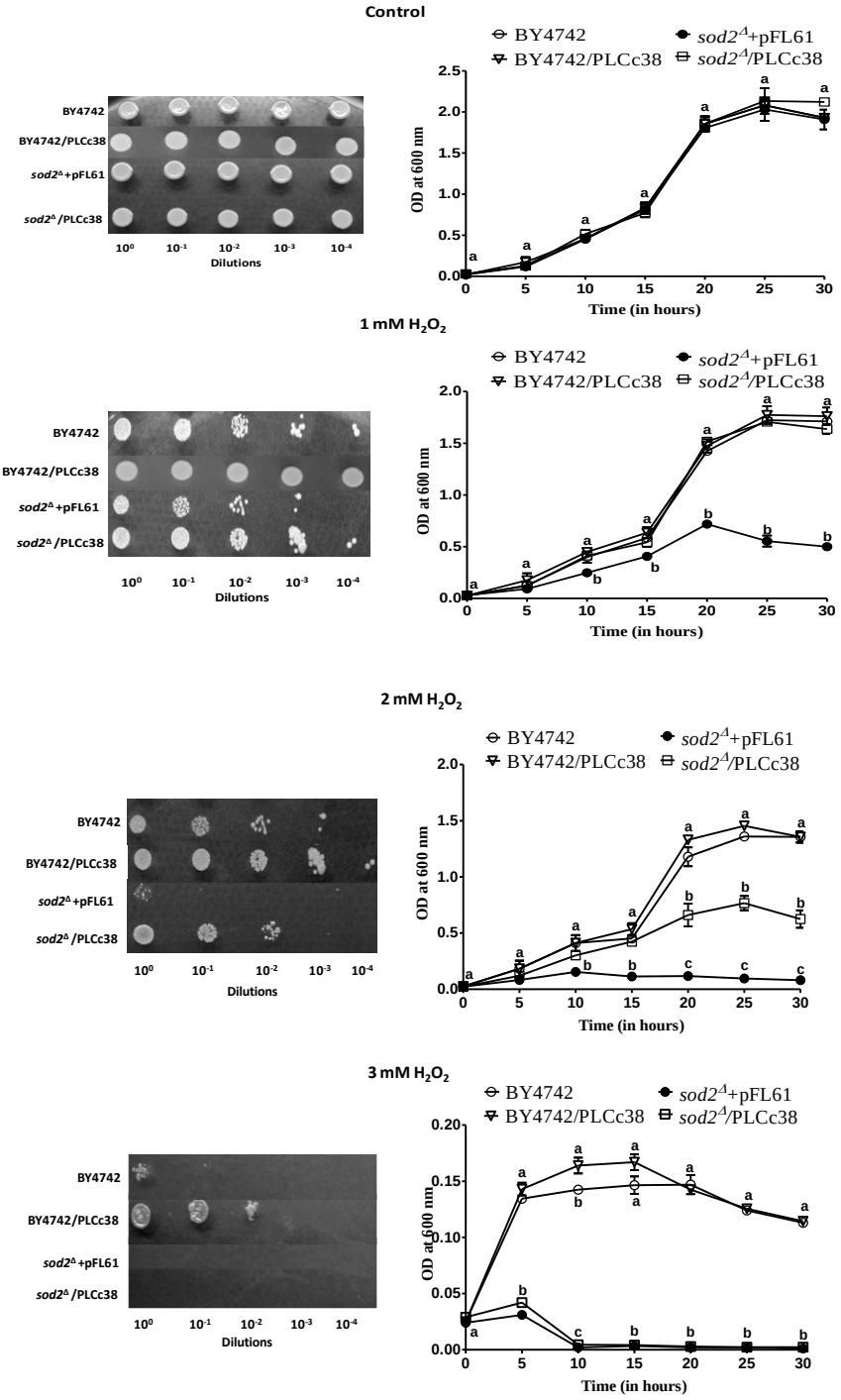


Figure 4.23: Drop assay and growth response of yeast mutant *sod2*$^{\Delta}$ transformed with PlCc38 in presence of different concentrations of H_2O_2 in the growth medium. Wild type BY4742 and *sod2*$^{\Delta}$ transformed with empty pFL61 were used as controls. Also BY4742 transformed with PlCc38 was used for comparison. Growth response was recorded till 30 hrs at an interval of 5

hours. Error bars are ±SD. Values sharing common letters on error bars for corresponding time interval between samples are not significantly different at $P<0.05$ ($n=3$)

From the results obtained in this analysis, it can be inferred that the cDNA PLCc38 can tolerate toxic levels of metals when expressed in yeasts. All the characteristic features of aldehyde dehydrogenase were evident in the translated protein which signified its role as a stress response protein having a vital role in abiotic stress management in metal polluted soil. The results suggest that several known and unknown genes along with their mechanisms can be discovered using metatranscriptomics approach, and can be utilized as a modern-day 'omics' technique for simplifying the high degree of diversity of natural phenomenon in any kind of ecosystem. Heterologous expression of this gene has shown that it can be used for a variety of biotechnological applications, such as revegetation of metal polluted sites, a biomarker for pollution studies and drafting successful bioremediation strategies.

4.10 Serine Protease Inhibitor (Serpin)

Another clone from library C of Cd contaminated soil from France, PLCg49, showed high tolerance to Cd was analyzed. Bioinformatic analysis and characterization of the gene was performed with respect to multi-metal tolerant capability.

4.10.1 Bioinformatic analysis of PLCg49

Sequence analysis of the cDNA PLCg49 revealed it be of size 1409 bp consisting of with an open reading frame of 1233 bp encoding a 410 amino acid polypeptide. Homology search of the deduced amino acid illustrated 37.94% sequence identity to Serine Protease Inhibitior (serpin) family of *Hypsibius dujardini* phylum *Tardigrada* (Figure 4.24 and Figure 4.25). The predicted molecular mass of PLCg49 was 45.57 kDa with pI of 6.47.

Sequences producing significant alignments Download Manage Columns Show 100

select all *0 sequences selected* GenPept Graphics

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
serine protease inhibitor-like protein [uncultured eukaryote]	852	852	87%	0.0	100.00%	QBQ70681.1
Serpin B9 [Hypsibius dujardini]	232	232	77%	5e-68	37.94%	OQV19906.1
serpin family protein [Aquisphaera sp. JC650]	237	237	77%	5e-70	37.84%	WP_152051297.1
uncharacterized protein Dsimw501_GD10292, isoform F [Drosophila simulans]	224	224	79%	6e-65	37.60%	XP_016026183.1
uncharacterized protein Dsimw501_GD10292, isoform A [Drosophila simulans]	224	224	79%	2e-64	37.60%	XP_002080346.1
Serpin (serine protease inhibitor) [Planctomycetes bacterium EIP]	228	228	77%	1e-66	37.57%	QDV35694.1
serpin family protein [Planctomycetes bacterium EIP]	229	229	77%	4e-66	37.57%	WP_145271431.1
serpin 42Da, isoform A [Drosophila melanogaster]	221	221	78%	1e-63	37.50%	NP_724512.1
Chain A, Serine Protease Inhibitor 4, Isoform B [Drosophila melanogaster]	221	221	78%	2e-63	37.50%	4P0F_A
serine protease inhibitor (serpin-4) [Drosophila melanogaster]	220	220	78%	3e-63	37.50%	CAB63099.1

Figure 4.24: Sequence homology search for cDNA PLCg49 using BLASTx through NCBI

```
gcggg ggt act ttt cgg cgg cca cgt tcg acg tct ctt ctt act ttg aga tta ttc gaa tca
   G   G   T   F   R   R   P   R   S   T   S   L   L   T   L   R   L   F   E   S
gct aag atg cat tcc tta ttc aag gga gtg atg atc atc ttt ggt ctg gct ttg gcc acc
 A   K   M   H   S   L   F   K   G   V   M   I   I   F   G   L   A   L   A   T
agc tcg cca gtt ccg ctg tgc cca gcg gca acg aga gag caa tcg cag caa gct gaa gca
 S   S   P   V   P   L   C   P   A   A   T   R   E   Q   S   Q   Q   A   E   A
gct ttg gct gac ttc gcc gtc aac ctt ttc aaa aat gtc tcg gct caa gac gaa gac aaa
 A   L   A   D   F   A   V   N   L   F   K   N   V   S   A   Q   D   E   D   K
aat caa gtc atg tcc ccg gtt tcc atc gct ctg gct ttg gcc atg ttg gaa aac gga gcc
 N   Q   V   M   S   P   V   S   I   A   L   A   L   A   M   L   E   N   G   A
gat gcc cac act cgc gaa caa ctc aaa cac gcc ttg gtc gga gtt gga tcc agc gcc gac
 D   A   H   T   R   E   Q   L   K   H   A   L   V   G   V   G   S   S   A   D
gtc ttg gcc gca tac cgc gaa atc caa aag cac ttg cgc gcc gat gac gag aaa acg aaa
 V   L   A   A   Y   R   E   I   Q   K   H   L   R   A   D   D   E   K   T   K
ttg acc gtc gcc aac ggc ctc ttc cag gac aag gat ctg aaa ttg aag gac cag tac gtc
 L   T   V   A   N   G   L   F   Q   D   K   D   L   K   L   K   D   Q   Y   V
agc act acc cgc gac tgc ttg gaa acc gaa gtg gac aac aag aac gac ttc tct gac cac
 S   T   T   R   D   C   L   E   T   E   V   D   N   K   N   D   F   S   D   H
ctc gag gac gcg cgt caa aaa atc aac aaa tgg gtc tcc caa aag acc tcc gag aag atc
 L   E   D   A   R   Q   K   I   N   K   W   V   S   Q   K   T   S   E   K   I
cca gag ctc ttc aag cga ggc gtc ttg aac gaa ggt gac cgt atg gtt ttg gcc aac gcc
 P   E   L   F   K   R   G   V   L   N   E   G   D   R   M   V   L   A   N   A
atc tac ttc aaa gct tcg tgg agg aac tcg ttc aac aaa gca atg acc aag caa gac acc
 I   Y   F   K   A   S   W   R   N   S   F   N   K   A   M   T   K   Q   D   T
ttc tac cgc aac ggt cac gac cag caa tct gtc cag ttc atg cac gaa act ggt gac cac
 F   Y   R   N   G   H   D   Q   Q   S   V   Q   F   M   H   E   T   G   D   H
cgt cac gtt tcg tcg aat gat ttg gat gcc gtc gaa ttg gcc tac tct cat ccc gat ttg
 R   H   V   S   S   N   D   L   D   A   V   E   L   A   Y   S   H   P   D   L
gcc atg gtc gtc gtc ttg ccg aaa gct cgc gac ggt ttg cgc aac ttg gag aaa cac ttg
 A   M   V   V   V   L   P   K   A   R   D   G   L   R   N   L   E   K   H   L
acc ggc aaa caa ttg agg gac atc gtc gct agc atg gaa caa cgc aag gtc aag gtc caa
 T   G   K   Q   L   R   D   I   V   A   S   M   E   Q   R   K   V   K   V   Q
ttg cca aaa ttc acc gtt cgc tcg ccc acc gat ttg aaa aat gtc ctg agc aaa atg ggt
 L   P   K   F   T   V   R   S   P   T   D   L   K   N   V   L   S   K   M   G
ttg gat cat atg ttc agc aac caa gcc aac ttc agc cgc atg agc gag acc cct ttg aag
 L   D   H   M   F   S   N   Q   A   N   F   S   R   M   S   E   T   P   L   K
gtc gac aac gct atc cat gaa gct tac atc gat gtc aac gaa aat gga acc gaa ggc gct
 V   D   N   A   I   H   E   A   Y   I   D   V   N   E   N   G   T   E   G   A
gct gca acc ggt gtt tcg atc aac atg cgt gcc atg ccc tcg atg cca gta gac cca gtg
 A   A   T   G   V   S   I   N   M   R   A   M   P   S   M   P   V   D   P   V
tcg ttc gtc gct gac cac cca tac ttg tac gcc gtt gtg cac aag cca acc ggc gcg atc
 S   F   V   A   D   H   P   Y   L   Y   A   V   V   H   K   P   T   G   A   I
gtt ttc ttg ggc aaa gtc aac tcg gtc gaa tcg cac tag act tcc agc gat tac atg aaa
 V   F   L   G   K   V   N   S   V   E   S   H   -   T   S   S   D   Y   M   K
```

Figure 4.25: Translation of cDNA PLCg49 encoding a protein which shows homology with serine protease inhibitor

Serine protease inhibitors are the most universal and the most widely distributed form of protease inhibitors with more than 1,500 members. A serpin's characteristic size typically ranges from 350 to 500 bp with 40-60 kDa molecular weight. Serpins typically have 3 β-sheets with 8 or 9 α-helices. Even with the extensive common characteristics for all serpins, the sequence identity of primary structures can be as little as 25%. Analysis of the reactive center loop (RCLs) of PLCg49 showed that the hinge region consisted of a slight variation in the consensus sequence "P_{17}[E] – P_{16}[E/K/R] – P_{15}[G] – P_{14}[T/S] – P_{13}[X] – $P_{12–9}$[AGSX] – $P_{8–1}$[X] – $P_{1'–4'}$" where X can be any amino acid and P represents position of amino acid residues (Gettins, 2002). At P_{16}, there is Asparagine (N) and at positions P_{12-10}, the amino acids are Glycine-Alanine-Alanine. The P_1-P_1'cleavage sites of the RCL is located near the C-terminal at position R368-A369 (Figure 4.26). The role of protease inhibition comes from the RCL located on the core structure forming an extended and exposed conformation above the serpin's body. The conserved regions observed based on sequence alignment are depicted in (Figure 4.27)

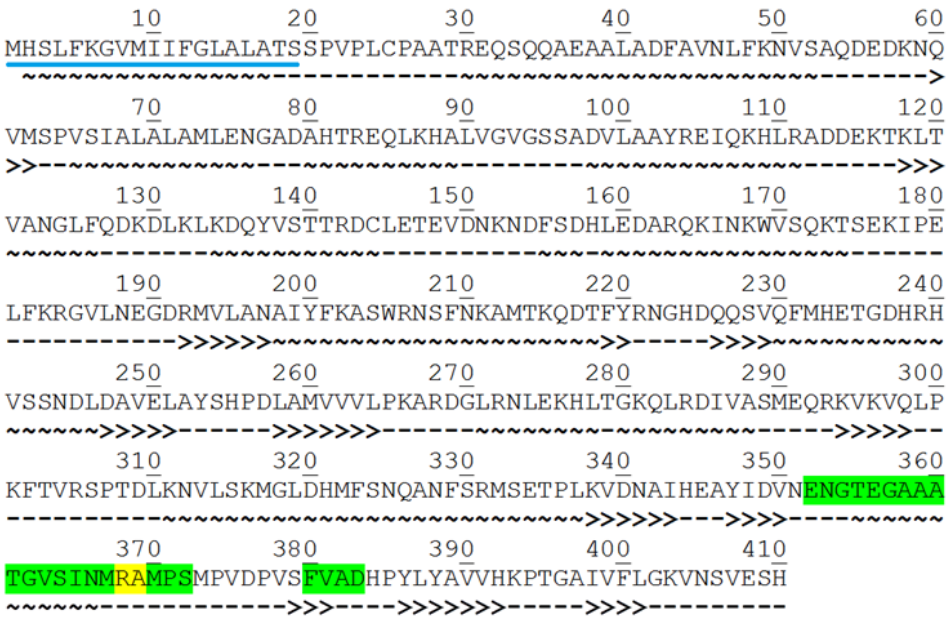


Figure 4.26: Amino acid sequence of PlCg49. Signature Reactive Center Loop (RCL) residues are highlighted in green. Protease cleavage site is marked in yellow. Blue underlined region

represents the signal peptide where cleavage site is between positions 19 and 20. Predicted secondary structures are represented as '~' for helices, '-' for coils and '>' for strands

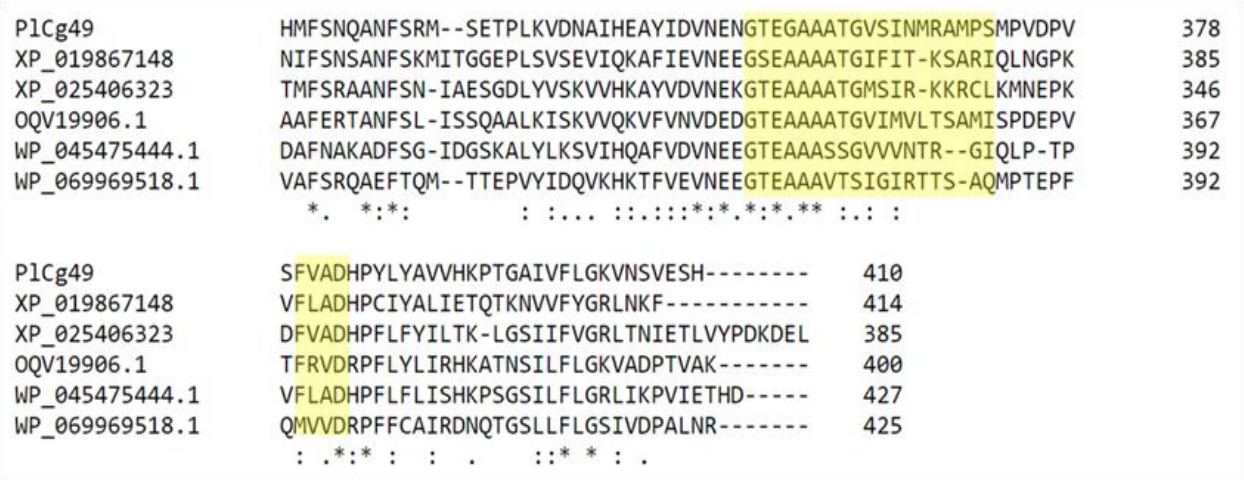


Figure 4.27: Multiple sequence alignment of PLCg49 gene with other homologous sequences retrieved by BLASTX analysis. Accession numbers are serpins of XP_019867148 (*Aethina tumida*), XP_025406323 (*Sipha flava*), OQV19906 (*Hypsibius dujardini*, WP_045475444 *Thioploca ingrica* and WP_069969518 (*Desertifilum sp.*). Highlighted regions are conserved Reactive Center Loop (RCL)

Phyllogenetic analysis with serine protease inhibitors from different taxons clearly clustered the cDNA PLCg49 into Animalia lineage thus confirming its origin from the kingdom Animalia (Figure 4.28). Sequence of PLCg49 obtained in this study was submitted to National Center of Biotechnology Information (NCBI) database under the accession number MK079356 (Appendix II).

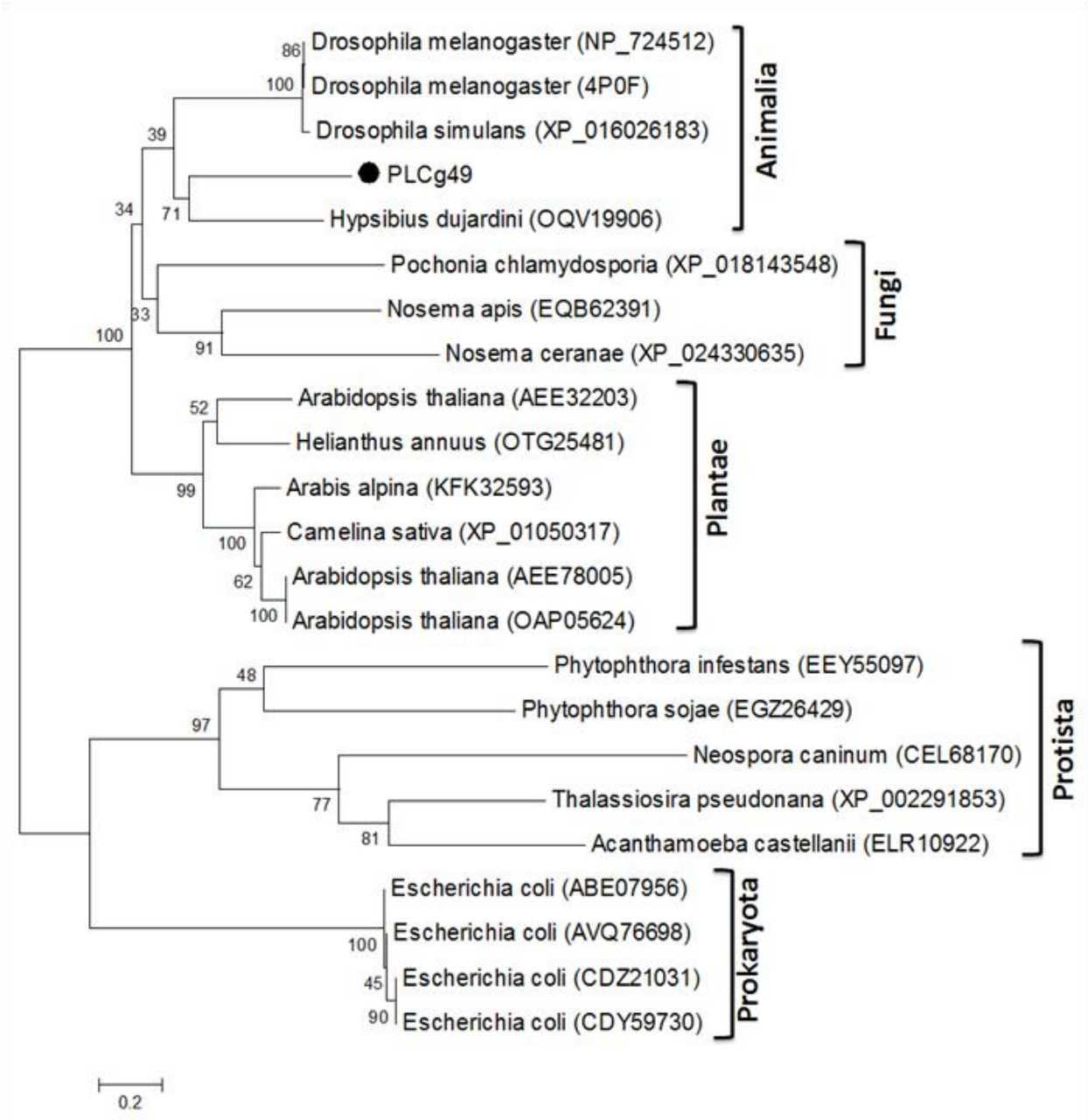


Figure 4.28: Phylogenetic relation of the transcript PLCg49 with SERPINs of other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses. Prokaryota Serpin sequences were included as outgroup. Branch lengths are proportional to evolutionary distances

4.10.2 Potential of PLCg49 to tolerate metals

To test the cDNA PLCg49 for its multi metal tolerance ability transformation was carried out into a number of metal sensitive yeasts and were allowed to respond to different metals at different concentrations. Metal sensitive mutant yeast cells *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PLCg49 were spotted on metal amended media and allowed to grow. It was observed that the cDNA was able to impart metal tolerance to the mutant yeasts. They were able to grow well on SD-Ura agar plates supplemented with 150 μM of Cu, 40 μM Cd, 10 mM Zn and 2 mM Co. On the other hand, yeast mutants transformed with empty vector pFL61 were unable to grow in any metal amended media. For drop assay, a general fall in growth was noticed for all the yeast strains in metal amended media (Figure 4.29).

In order to study the growth characteristics and metal tolerant properties of PLCg49 in the presence of Cu, Cd, Zn and Co, the transformants were grown in SD-Ura broth supplemented with a variety of these metals (Table 4.9). Tolerance towards copper for the transformant *cup1*$^{\Delta}$/PLCg49 was high, but decrease in the growth was observed with increase in Cu concentrations. The growth was not majorly affected up to 500 μM Cu (Figure 4.30A). The transformant *ycf1*$^{\Delta}$/PLCg49 was able to withstand and grow well at 40 μM Cd, but growth declined rapidly with increase in cadmium concentration in the medium. The wild type BY4741 displayed a comparable response to growth in relation to *ycf1*$^{\Delta}$/PLCg49 for all the concentrations (Figure 4.30B). The transformant *zrc1*$^{\Delta}$ /PLCg49 demonstrated higher growth than the control BY4741 in every concentration of Zn with maximum growth recorded at 10 mM. The growth of the transformant was significant up to 12 mM but decreased considerably thereafter. The wild type strain BY4741 showed similar trend in its growth (Figure 4.30C).

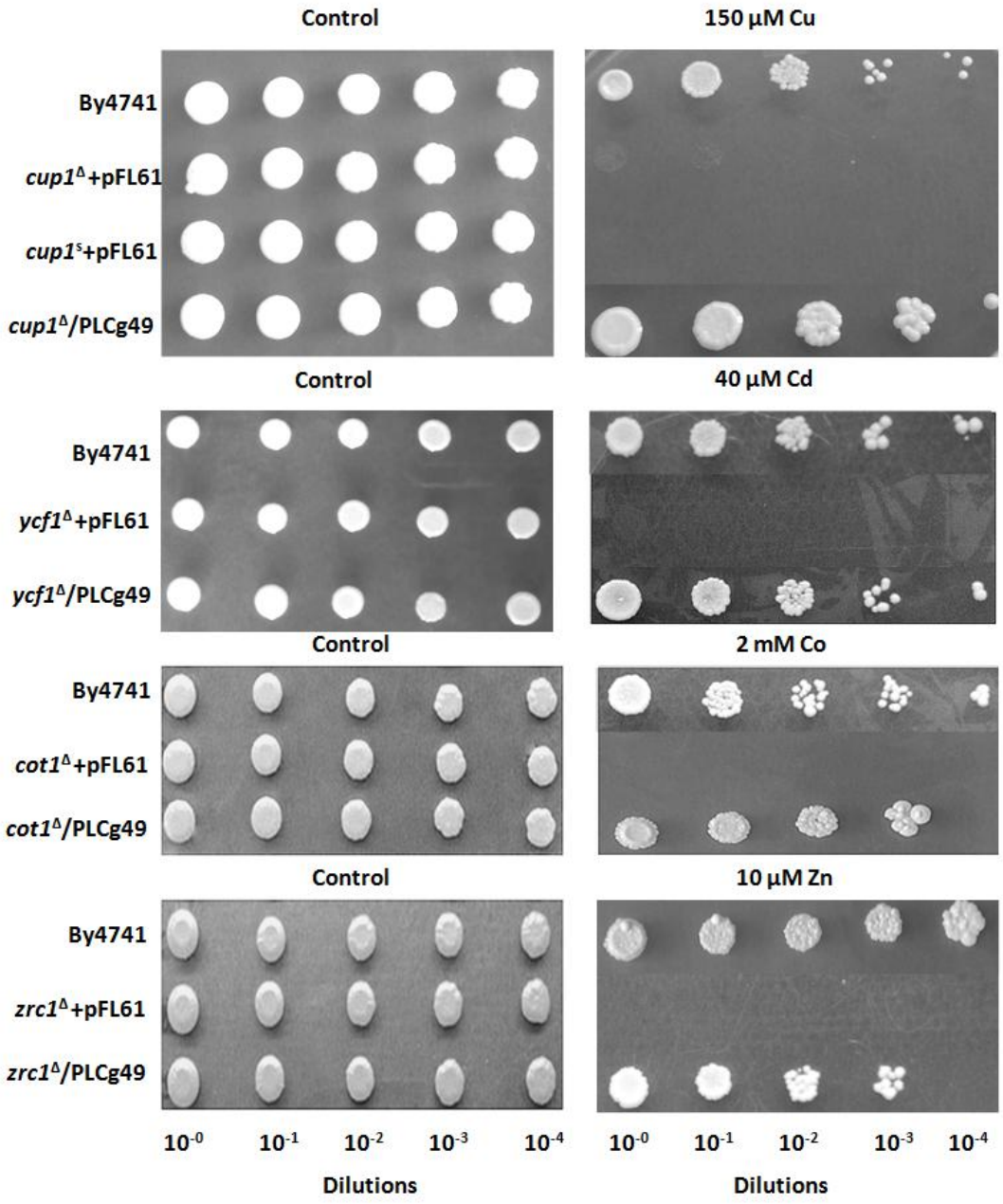


Figure 4.29: Drop assay of PlCg49 transformed *Saccharomyces cerevisae* mutants, *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ on SD-Ura agar. BY4741wild type and mutant strains transformed with empty vector pFL61 for positive and negative control. Serially diluted cultures were spotted on SD-Ura medium with or without metal supplement as indicated. Yeast strain *cup1*^{S} was also used as control.

Table 4.9: Growth response of *cup1*$^{\Delta}$,*ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PlCg49 compared to controls

	OD_{600} of Yeast samples		
Cu concentration(µM)	***cup1*$^{\Delta}$/PlCg49**	**BY4741**	***cup1*$^{\Delta}$/pFL61**
150	1.544±0.064a	1.352±0.020b	0.080±0.014c
300	1.389±0.023a	1.347±0.008a	0.045±0.007b
500	1.133±0.001a	1.109±0.006a	0.015±0.007b
1000	0.025±0.005a	0.095±0.006b	0.004±0.003a
Cd concentration(µM)	***ycf1*$^{\Delta}$/PlCg49**	**BY4741**	***ycf1*$^{\Delta}$/pFL61**
40	1.154±0.028a	1.122±0.006a	0.060±0.014b
60	0.985±0.019a	0.929±0.039a	0.035±0.007b
80	0.509±0.006a	0.496±0.009a	0.015±0.007b
100	0.253±0.036a	0.228±0.021a	0.002±0.000b
Zn concentration(mM)	***zrc1*$^{\Delta}$/PlCg49**	**BY4741**	***zrc1*$^{\Delta}$/pFL61**
10	1.875±0.073a	1.496±0.223b	0.175±0.007c
11	1.740±0.022a	1.438±0.122b	0.051±0.001c
12	1.295±0.086a	1.024±0.127b	0.031±0.001c
13	0.108±0.022a	0.095±0.006a	0.004±0.001b
Co concentration(mM)	***cot1*$^{\Delta}$/PlCg49**	**BY4741**	***cot1*$^{\Delta}$/pFL61**
2	0.188±0.007a	0.175±0.022a	0.008±0.002b
10	0.141±0.008a	0.115±0.022b	0.005±0.005c
25	0.068±0.007a	0.055±0.022b	0.003±0.007c
50	0.047±0.007a	0.042±0.022a	0.002±0.000b

O.D. values are Mean ± Standard deviation. Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (n=3)

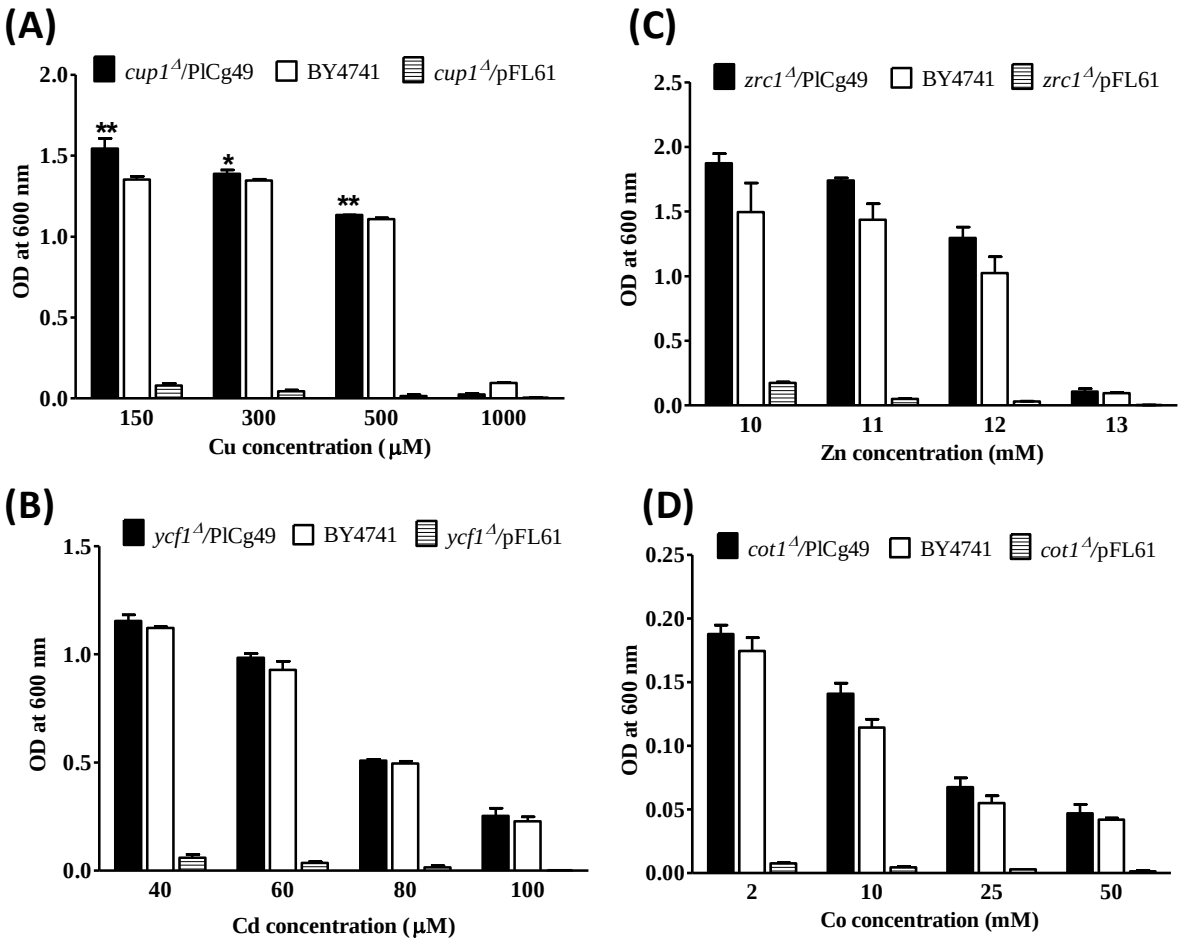


Figure 4.30: Growth response of (A) *cup1*$^{\Delta}$ (B) *ycf1*$^{\Delta}$ (C) *zrc1*$^{\Delta}$ and (D) *cot1*$^{\Delta}$ transformed with PlCg49 and wild type BY4741 and yeast mutants transformed with Empty Vector pFL61 in SD-Ura broth supplemented with different metals at different concentrations as shown (* and **are significant at $P<0.05$ and $P<0.01$ respectively).

For cobalt, the yeast transformant *cot1*$^{\Delta}$/PLCg49 demonstrated high tolerance with maximum growth recorded at 2 mM concentration. The growth of *cot1*$^{\Delta}$ /PLCg49 was not significantly higher than its wild type strain transformed with empty vector pFL61 in all concentrations of Co (Figure 4.30D). For all the metals, there was marginal growth in mutant yeast strains transformed with empty pFL61 vector.

Evidence from many recent studies shows that protease inhibitors are multifunctional and active in modulating plant's resistance to abiotic stress. Gene expression and metal accumulation studies in *Populus* roots when exposed to Cu stress were well documented by Guerra et al. (2015). The transcriptome of the plant was studied when subjected to a maximum of 60 μM of $CuSO_4$ and expression of defense proteins was observed out of which a considerable fold change was observed in Kunitz trypsin inhibitor genes. Furthermore, when protease inhibitor gene was cloned and expressed in mutant yeast strains as a part of functional complementation experiments, they were able to impart tolerance towards Cu, Cd, Zn and Co. In the same study, putative Cu binding sites were indicated which could chelate Cu and hence account for the uptake of metals (Guerra et al. 2009). In the present study, the results extensively correlate as the environmental transcript PLCg49 was also able to confer high concentrations against Cu, Cd, Zn and Co.

Protease inhibitors (PIs) have been comprehensively studied in plants with regard to different stress-inducing factors and have emerged as an important gene family which express in response to abiotic stresses. Among the different PIs, serine PIs are the most abundant and well known for their role as environmental response proteins. In a study by Shitan et al. (2007), a Bowman–Birk proteinase inhibitor was isolated by functional complementation which could rescue and allow the mutant yeast $ycf1^{\Delta}$ to grow in presence of heavy metal Cd. Further characterization of the PI was performed in relation to metal uptake and quantifying gene expression by qPCR analysis. It was found that the Cd tolerance potential depicted by the Bowman–Birk proteinase inhibitor was even higher than that of a metallothionein isolated in the same study. This proved that the PI could confer resistance to *Coptis japonica* plant in presence of high concentration of Cadmium. Correspondingly, mechanical injury which led to exposure to toxic concentrations of aluminium

resulted in the induction of a number of Bowman–Birk type protease inhibitors. Likewise, a number of studies have reported the induction of protease inhibitor genes in response to heavy metals exposure in different plants like wheat (Snowden et al. 1995), tobacco (Harada et al. 2010) maize (Cançado et al. 2008) and *Lupinus* and *Pisum* (Domash et al. 2008). Joshi et al. (2014) has well documented *Capsicum annuum* protease inhibitors that confer tolerance towards multiple stresses, including heavy metals cadmium and mercury in mutant yeasts and exhibited a delayed senescence in plants due to strong inhibition of cellular metacaspase activity. Also, prominent increase in growth of yeast cells in presence of heavy metals and protease inhibition activity was observed when expressed in yeast mutants. The expression of PIs in response to abiotic stresses such as drought, salinity, heavy metals, UV radiation and wounding suggests that there are numerous interlinked metabolic processes which ultimately lead to production of ROS species, leading to abnormalities and disruption of normal metabolic processes (Srinivasan and Kirti, 2012).

4.10.3 Yeast metal uptake studies of PLCg49

The cellular metal contents of mutant strains transformed with PLCg49 were estimated by ICP-MS by first allowing them to grow in media amended with different concentrations of metals (Table 4.10). With the increase in Cu in the growth medium, the cellular content of Cu decreased both in *cup1*$^{\Delta}$/PLCg49 and BY4741. Highest content of intracellular Cu was observed at 300 µM in *cup1*$^{\Delta}$/PLCg49 cells (Figure 4.31A). The levels of Cd significantly decreased both in *ycf1*$^{\Delta}$/PLCg49 as well as in the wild type strain as external Cd concentration in the growth medium increased (Figure 4.31B). The *zrc1*$^{\Delta}$/PLCg49 cells were able to accumulate Zn, highest being at 10 mM, but the cellular Zn contents decreased at higher concentrations (Figure 4.31C).

The cellular Co content also decreased with increased concentrations of Co in the growth medium both for *cot1*$^{\Delta}$ /PLCg49 as well as BY4741 cells (Figure 4.31D).

Table 4.10: Metal uptake by *cup1*$^{\Delta}$,*ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PlCg49 compared to controls

	Metal uptake (mg^{-1}g dry weight)		
Cu concentration(µM)	***cup1*$^{\Delta}$/PlCg49**	**BY4741**	***cup1*$^{\Delta}$/pFL61**
150	0.139±0.015a	0.084±0.009b	0.005±0.002c
300	0.303±0.019a	0.129±0.015b	0.011±0.002c
500	0.268±0.024a	0.246±0.027a	0.007±0.009b
1000	0.123±0.003a	0.128±0.015a	0.003±0.001b
Cd concentration(µM)	***ycf1*$^{\Delta}$/PlCg49**	**BY4741**	***ycf1*$^{\Delta}$/pFL61**
40	0.055±0.002a	0.043±0.003b	0.003±0.001c
60	0.042±0.001a	0.038±0.002b	0.001±0.000c
80	0.034±0.003a	0.026±0.003b	0.001±0.000c
100	0.013±0.001a	0.011±0.001b	BDL
Zn concentration(mM)	***zrc1*$^{\Delta}$/PlCg49**	**BY4741**	***zrc1*$^{\Delta}$/pFL61**
10	5.285±0.568a	0.784±0.348b	0.319±0.251c
11	1.658±0.020a	0.232±0.020b	0.123±0.025c
12	1.031±0.010a	0.145±0.004b	0.063±0.012c
13	0.075±0.004a	0.054±0.003b	0.013±0.006c
Co concentration(mM)	***cot1*$^{\Delta}$/PlCg49**	**BY4741**	***cot1*$^{\Delta}$/pFL61**
2	0.276±0.011a	0.035±0.011b	0.010±0.001c
10	0.088±0.006b	0.245±0.007a	0.005±0.000c
25	0.080±0.002a	0.080±0.004a	BDL
50	0.006±0.002a	0.014±0.003b	BDL

Metal uptake values are Mean ± Standard deviation. Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (n=3). BDL-Below Detection Limit

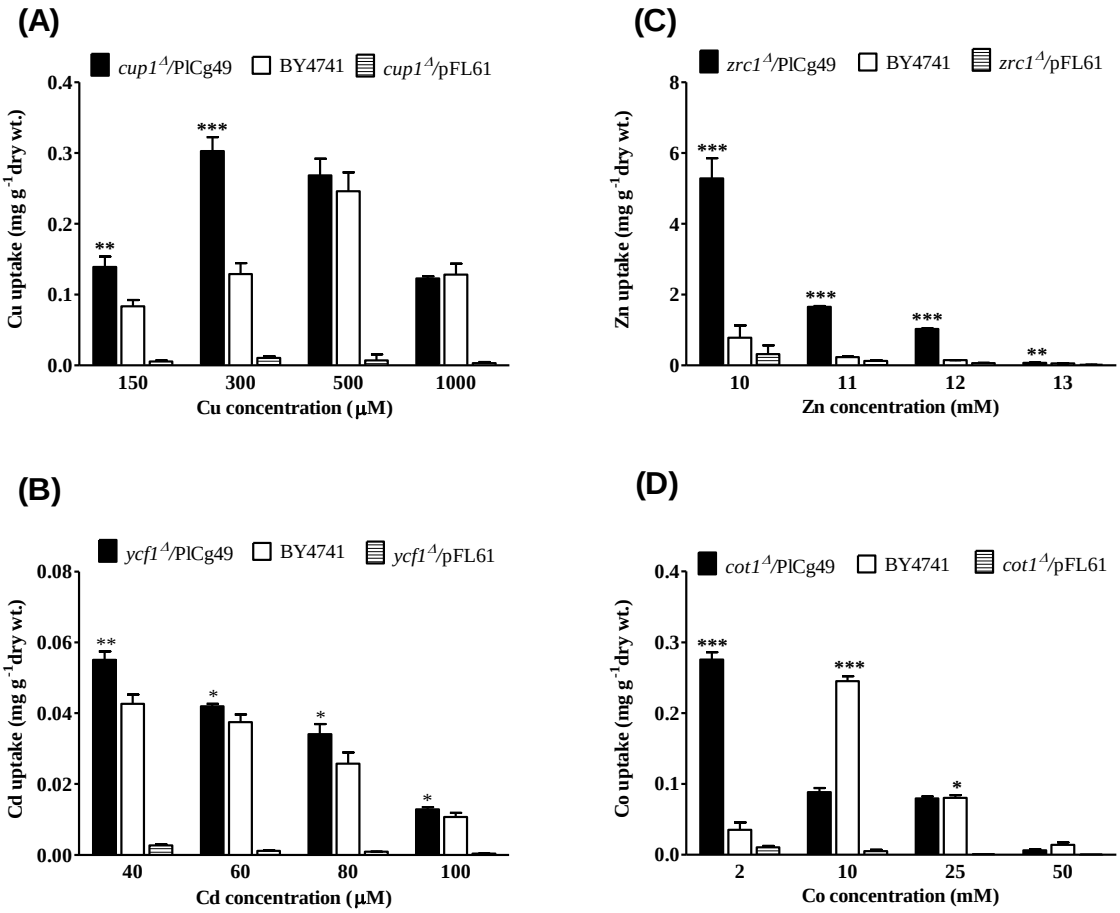


Figure 4.31: Effect of different concentration of (E) Copper (F) Cadmium (G) Zinc (H) Cobalt on metal uptake by PlCg49 when compared to wild type BY4741 after 40 hrs of growth. Error bars are ±SD (*n*=3). (*, ** and *** are significant at *P*<0.05, *P*<0.01 and *P*<0.001 respectively).

4.10.4 Effect of protease treatment

The role of the cDNA PLCg49 as a protease inhibitor was validated by measuring its ability to provide resistance against the lytic activity of serine proteases trypsin and proteinase K (Table 4.11). When compared to the mutant *pbi2*$^{\Delta}$ transformed with empty pFL61, the transformant *pbi2*$^{\Delta}$/PLCg49 showed reduced cellular lytic activity. This was significantly lower in all concentrations of trypsin. BY4741, on the other hand, had significantly higher cellular lytic activity at 500 μg. A considerable amount of resistance was demonstrated by the transformant

pbi2$^{\Delta}$/PLCg49 against cell lysis when treated with proteinase K enzyme, which was comparable to that of the wild type BY4741. The mutant *pbi2*$^{\Delta}$ bearing empty pFL61 documented significant cell lysis, the highest being recorded at 30 mg. BY4741/PLCg49 transformant reported a slight reduction in cell lysis compared to BY4741 alone for all concentrations of the proteases. However, this variation was not large (Figure 4.32). The slight reduction in cell lysis may be attributed to the synergistic effect of expression of the cDNA along with native protease inhibitors of wild type BY4741. Shimoi et al. (1991) documented the effects of different protease inhibitors on bacterial proteases by evaluating yeast cell lysis and enzyme activity. These results confirmed that serine protease inhibitors successfully inhibited protease activity and improved yeast cell viability. Similar results have been observed in the present study when the PLCg49 was expressed in yeast cells.

Table 4.11: Percentage cell lysis by protease of yeast mutant *pbi2*$^{\Delta}$ transformed with PlCg49 compared to controls

	Percentage cell lysis			
Trypsin conc.(µg)	**BY4741**	**BY4741/PLCg49**	***pbi2*$^{\Delta}$/pFL61**	***pbi2*$^{\Delta}$/PlCg49**
0	0.00±0.00a	0.00±0.00a	0.000±0.00a	0.000±0.00a
125	7.23±0.093b	6.91±0.092b	34.66±0.467a	7.72±1.625b
250	11.88±0.497b	10.66±0.561b	44.38±0.660a	10.42±0.394c
500	23.26±0.847b	21.59±0.368b	48.32±0.531a	17.28±0.080c
Proteinase K conc.(µM)	**BY4741**	**BY4741/PLCg49**	***pbi2*$^{\Delta}$/pFL61**	***pbi2*$^{\Delta}$/PlCg49**
0	0.000±0.00a	0.000±0.00a	0.000±0.00a	0.000±0.00a
10	13.32±0.118b	10.99±0.462b	24.59±0.105a	6.82±0.009c
20	16.30±0.265b	15.30±0.738b	33.95±1.114a	12.27±0.051c
300	22.55±0.731b	20.22±0.797b	37.77±0.611a	22.34±0.431b

Percentage cell lysis values are Mean±error. Values sharing common letters for a particular protease concentration between samples are not significantly different at $P<0.05$ (n=3).

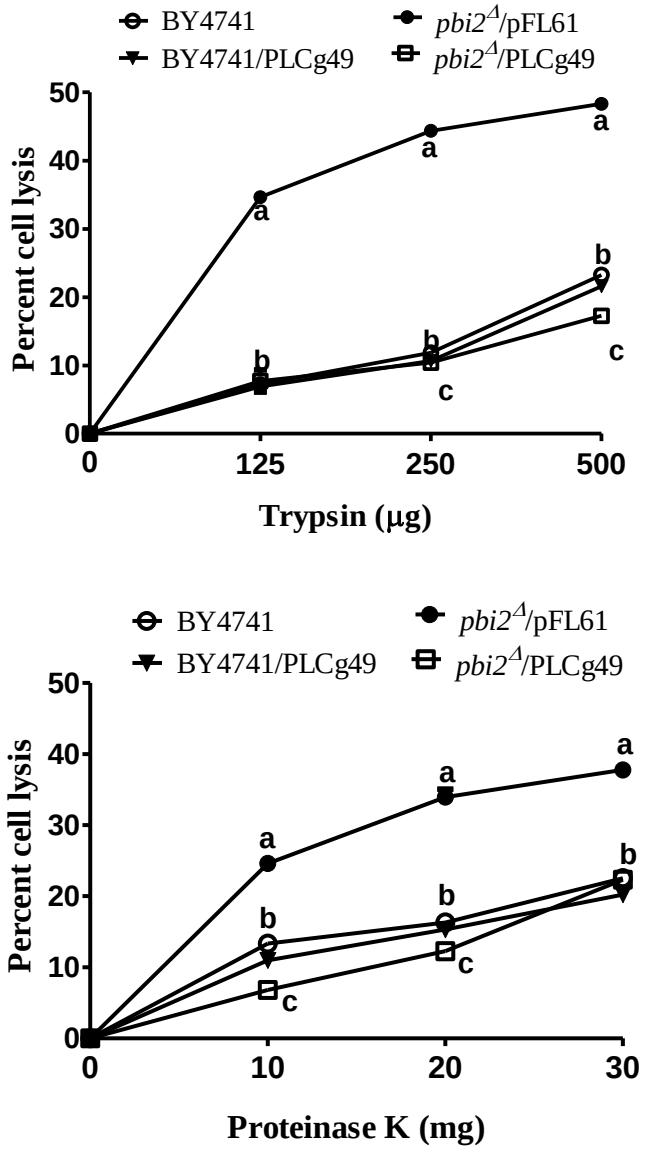


Figure 4.32: Percentage cell lysis of yeast mutant *pbi2*$^{\Delta}$ transformed with PlCg49 in presence of different concentrations of Trypsin and Proteinase K in the growth medium. Wild type BY4741 and *pbi2*$^{\Delta}$ transformed with empty pFL61 were used as controls. Also BY4741 transformed with PlCg49 was used for comparison. Error bars are ±SD. Values sharing common letters on error bars for corresponding protease concentrations between samples are not significantly different at $P<0.05$ (n=3)

It was inferred from the experimental observations that the cDNA PLCg49 can provide the capability to withstand toxic metal concentrations when expressed in yeasts. The presence of

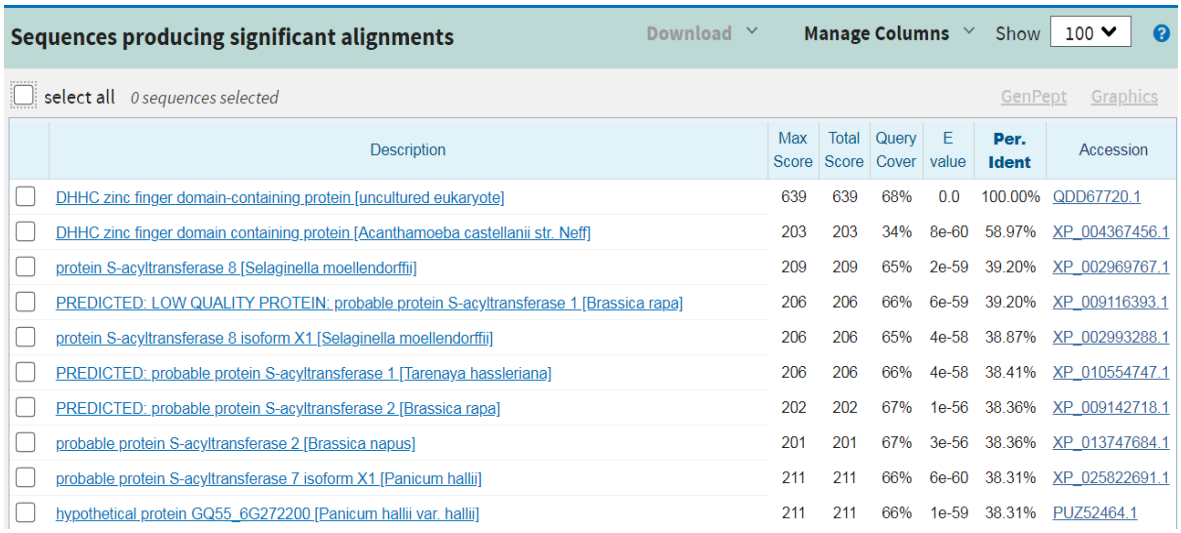
characteristic features of serine protease inhibitor (serpin) in the translated protein indicated that, this protein is induced in contaminated soil along with other stress response proteins as a result of abiotic stress. The approach of metatranscriptomics applied to complex ecosystems such as soil, can reveal a high degree of functional diversity. Previous studies focused mainly on host-specific serpins and the study of specific host transcriptomes that may or may not include complex interplay and the role of other serpins. The heterologous expression of this transcript can be deemed suitable for a variety of biotechnological applications like remediation of polluted land, biomarkers for ecotoxicological studies and developing strategies to improve heavy metal tolerant crop plants.

4.11 DHHC palmitoyl transferase

Another clone from library C of Cd contaminated soil from France, PLCg39 that demonstrated high tolerance to cadmium was analyzed. Characterization of the gene was performed and its response to multi-metal stress was analyzed. Sequence study indicated that PLCg39 demonstrated homology with DHHC palmitoyl transferase and shared common ancestry with phylum Amoebozoa.

4.11.1 Bioinformatic analysis of PLCg39

PLCg39 is a 1355 bp cDNA with open reading frame (ORF) of size 966 bp encoding for a protein of 321 amino acids. The predicted molecular mass and pI are 37.38 kDa and 9.39 respectively. Sequence analysis of the cDNA revealed homology with DHHC zinc finger domain containing protein of the Amoebozoa, *Acanthamoeba castellanii* with 58.97 % identity (Figure 4.33 and Figure 4.34).



Sequences producing significant alignments Download Manage Columns Show 100

select all *0 sequences selected* GenPept Graphics

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
DHHC zinc finger domain-containing protein [uncultured eukaryote]	639	639	68%	0.0	100.00%	QDD67720.1
DHHC zinc finger domain containing protein [Acanthamoeba castellanii str. Neff]	203	203	34%	8e-60	58.97%	XP_004367456.1
protein S-acyltransferase 8 [Selaginella moellendorffii]	209	209	65%	2e-59	39.20%	XP_002969767.1
PREDICTED: LOW QUALITY PROTEIN: probable protein S-acyltransferase 1 [Brassica rapa]	206	206	66%	6e-59	39.20%	XP_009116393.1
protein S-acyltransferase 8 isoform X1 [Selaginella moellendorffii]	206	206	65%	4e-58	38.87%	XP_002993288.1
PREDICTED: probable protein S-acyltransferase 1 [Tarenaya hassleriana]	206	206	66%	4e-58	38.41%	XP_010554747.1
PREDICTED: probable protein S-acyltransferase 2 [Brassica rapa]	202	202	67%	1e-56	38.36%	XP_009142718.1
probable protein S-acyltransferase 2 [Brassica napus]	201	201	67%	3e-56	38.36%	XP_013747684.1
probable protein S-acyltransferase 7 isoform X1 [Panicum hallii]	211	211	66%	6e-60	38.31%	XP_025822691.1
hypothetical protein GQ55_6G272200 [Panicum hallii var. hallii]	211	211	66%	1e-59	38.31%	PUZ52464.1

Figure 4.33: Sequence homology search for cDNA PLCg39 using BLASTx through NCBI

```
gcggg gtc cgc gaa aga act gtt gca aca aag aag gct gag ctt gta ggt gcg ggg tca gcg
   G   V   R   E   R   T   V   A   T   K   K   A   E   L   V   G   A   G   S   A
cga gtc gtt ttc tcc tcc ctc tct tgc ttg ggt ggg tgt gca cgg gcg aga caa atc gca
 R   V   V   F   S   S   L   S   C   L   G   G   C   A   R   A   R   Q   I   A
agc tag tac gat tcc ttc gat cga taa aac aag cca cgc aca tca aga gaa cag cta atg
 S   -   Y   D   S   F   D   R   -   N   K   P   R   T   S   R   E   Q   L   M
aag aga caa tat cag ttg tgg cca ggc aaa agc cac ttc ctc tgt tgg ggt cgt tgc gta
 K   R   Q   Y   Q   L   W   P   G   K   S   H   F   L   C   W   G   R   C   V
gtg gga cct gat aga aag tat ttc ctg ctt gcg gtg gcg ttt atc gtc gta cct gct gta
 V   G   P   D   R   K   Y   F   L   L   A   V   A   F   I   V   V   P   A   V
ttt tat tgt ggg ctt gta tgg cct tac ctg ttt tgg aaa ctt gag ccc tac ata tca gta
 F   Y   C   G   L   V   W   P   Y   L   F   W   K   L   E   P   Y   I   S   V
ccg tta gcg gtg ttc tac ttc tct att ctt ctt tgt atc ctt act ttc atg atg ctc aca
 P   L   A   V   F   Y   F   S   I   L   L   C   I   L   T   F   M   M   L   T
cga tat cga gac cca ggc atc att cct cgg ggt cgc gca atc gag tat gac cca gaa aat
 R   Y   R   D   P   G   I   I   P   R   G   R   A   I   E   Y   D   P   E   N
cct tgg gac tat gaa gag aaa aag cca gcg gag aca atg cga ctt aat ctg aaa gga gat
 P   W   D   Y   E   E   K   K   P   A   E   T   M   R   L   N   L   K   G   D
cca aaa ttt cgt gtc aaa tat tgc gaa aca tgt cat atc tac cgg ccg cct cga tgc acg
 P   K   F   R   V   K   Y   C   E   T   C   H   I   Y   R   P   P   R   C   T
cac tgc gcc gtg tgc aat aat tgc gta gag cgc ttt gat cac cac tgc cct tgg gtt ggc
 H   C   A   V   C   N   N   C   V   E   R   F   D   H   H   C   P   W   V   G
aat tgc ata ggt aga agg aat tac cag acg ttt ctc gcc ttt gtg tgg agc gtg att ttg
 N   C   I   G   R   R   N   Y   Q   T   F   L   A   F   V   W   S   V   I   L
ggg tgt gtc tac atc tgc ctg ttg agt atc gcc cac ctg gct atc gtt atc gta gag gcg
 G   C   V   Y   I   C   L   L   S   I   A   H   L   A   I   V   I   V   E   A
ata cat gaa gag ggt agc ggt gcc tat gtg ttc cta tac ata cta cga tat ggc ttt ttt
 I   H   E   E   G   S   G   A   Y   V   F   L   Y   I   L   R   Y   G   F   F
tcg gtt ata aca ttg ggg tat tgt gcg gtt gca ata ggg ctt gta gga ttc ttg ggc aca
 S   V   I   T   L   G   Y   C   A   V   A   I   G   L   V   G   F   L   G   T
ttt cat tgc ttg cta gtg tgc aaa aat cag act aca aac gag aag ctt aaa ggc ttg tac
 F   H   C   L   L   V   C   K   N   Q   T   T   N   E   K   L   K   G   L   Y
aag aag aaa atg aat ccg tac tca aaa ggg ccg tgt ctg aac ttt atc tct att ctg tgg
 K   K   K   M   N   P   Y   S   K   G   P   C   L   N   F   I   S   I   L   W
gca cct gtc ttc ccg aga cct gtc aat tgg cga gag cgt gtg gat gaa gat acg gcc aag
 A   P   V   F   P   R   P   V   N   W   R   E   R   V   D   E   D   T   A   K
ctg gtc aag cgg aaa gca tgg ccg aag gac aag agg aaa gca tca gca agg cgg aca aag
 L   V   K   R   K   A   W   P   K   D   K   R   K   A   S   A   R   R   T   K
taa aaa gcg gac caa gct tgt cct tcc tac cag cgg gta ccg cac tat gcg gcc cgt atc
 -   K   A   D   Q   A   C   P   S   Y   Q   R   V   P   H   Y   A   A   R   I
ttt ttt cga gag gac aga ggg tgg gat gtt aca tgt tgc ctg cca aac aac aac gac aac
 F   F   R   E   D   R   G   W   D   V   T   C   C   L   P   N   N   N   D   N
aag taa ggg aac cag tcg gtg tgt ata gag gat tac ctg gtg ttc tat tta ttg tat tta
 K   -   G   N   Q   S   V   C   I   E   D   Y   L   V   F   Y   L   L   Y   L
ttt ctt ttt aac ttt caa aaa aaa aaa aaa aaa
 F   L   F   N   F   Q   K   K   K   K   K
```

Figure 4.34: Translation of cDNA PLCg39 encoding a protein which shows homology with DHHC palmitoyltransferase

Highly conserved zinc finger domain was observed on the basis of amino acid sequence alignment of PLCg39 with other closely related DHHC zinc fingers from same phylum (Figure 4.35).

DHHC domain-containing family of proteins contains motif called DHHC-CRD (Asp-His-His-Cys cysteine-rich domain). This domain is responsible for the addition of a palmitoyl group to proteins as a part of post-translational modification in order to secure them to membranes (Putilina et al. 1999). The addition of modification happens via the side chain of cysteine residue of the protein forming a thioester linkage. According to *in silico* sequence analysis of PLCg39, it is polytopic protein having 4 transmembrane domains (TMD). The core DHHC motif of PLCg39 has a conserved region of 48 residues with a canonical sequence motif of $CX_2CX_9HCX_2CX_2CX_4DHHCX_5CX_4NX_3FX_4$. Other signature conserved regions are the DPG motif (Asp-Pro-Gly) adjacent to TMD2 and TTXE (Thr-Thr-Xaa-Glu) next to TMD4 (Montoro et al. 2009). Apart from the cysteine rich domain, the cDNA has 3 CXXXC motifs (Figure 4.36).

```
PLCg39          ------------E---EKKPAETMRLNLKGDPKFRVKYCETCHIYRPPRCTHCAVCNNC   149
XP_003292314.1  DEVEVERERKVNDNEKYQKDEKLYQEIIIKDYFKLKLKYCETCDIFRPPRSFHCSTCNNC   311
XP_004367456.1  ------------E---RKKPPETIKINVHGE-NLRIKYCDTCHIYRPPRAIHCSVCNNC   148
XP_004356236.1  HATHASPPGSIT-PSTSTQSIPFPDKTILIGDFPYTVKYCETCLIYRPPRSSHCSLCNAC   149
Q550R7.3        ------------D---HRQ-PLFKKI-TVKDTKQEIKWCETCCLYRPPRANHCGICNNC   137
XP_012756929.1  --ASVTDPGYITSE-----EGHSF-IQYQYDRLLYVKKQCETCGIAKPSRSKHCRVCDKC   159
KYQ90117.1      --SCNVNPGYINKL-----NHKYYMSLYDYDRYLYIKKKCETCSFNKPARSKHCRVCDKC   184
XP_020430454.1  QQQQQQNPNNIP-I---KKKKQSKKTININGE-SITIFYCKSCNIYRPPRCSHCSECNRC   177
                                        .                 *.:* : :* *. **  *: *

PLCg39          VERFDHHCPWVGNCIGRRNYQT-FLAFVWSVILGCVYICLLSIAHLAIVIVE---AIH--   203
XP_003292314.1  VQNFDHHCVWIGNCIGQRNYKY-FLFFIFSTLIYSTYICVMSIVFIVHHVNS---FISNN   367
XP_004367456.1  VERFDHHYILLHPPFASSSLLSSSLLFA--------------------------------   176
XP_004356236.1  ISRFDHHCPWVGNCVGQNNYKY-FFYFIASVALNILIVLITTIYHLDIIYKNTTIYPDNN   208
Q550R7.3        VERFDHHCPWVGNCIGRRNYQT-FLYFLYSLGFLCIWIMGFCVAHICIESAR---YRD--   191
XP_012756929.1  VARFDHHCPWINNCVGERNLRY-FLWFVGNTSALCFYGFYLCLCALLTIVETKNLFKLGY   218
KYQ90117.1      VSRFDHHCPWINNCVGEGNLKY-FLTFVFLTASLCLYGAYLCIYVIFSIIKVKNLLRLGY   243
XP_020430454.1  VMEFDHHCPWISNCVGKRNYRY-FVYFVWSAVGLSIMTMASSIVTIIKLT-------N--   227
                : .****   :   ..  .     . *
```

Figure 4.35: Multiple sequence alignment of PLCg39 with members of the phylum Amoebozoa. Accession numbers are DHHC proteins of *Dictyostelium purpureum* (XP_003292314),

Acanthamoeba castellanii str. *Neff* (XP_004367456), *Cavenderia fasciculate* (XP_004356236), *Dictyostelium discoideum* (Q550R7), *Acytostelium subglobosum* (XP_012756929), *Tieghemostelium lacteum* (KYQ90117), *Heterostelium album* (XP_020430454). The highly conserved zinc finger motif is highlighted in yellow and the DHHC motif in red

MKRQYQLWPGKSHFLCWGRCVVGPDRKYFLLAVAFIVVPAVFYCGLVWPYLFWKLEPYISVPLAVFYFSILLCILTFMM

LTRYRDPGIIPRGRAIEYDPENPWDYEEKKPAETMRLNLKGDPKFRVKYCETCHIYRPPRCTHCAVCNNCVERFDHHCP

WVGNCIGRRNYQTFLAFVWSVILGCVYICLLSIAHLAIVIVEAIHEEGSGAYVFLYILRYGFFSVITLGYCAVAIGLVGFLGTF

HCLLVCKNQTTNEKLKGLYKKKMNPYSKGPCLNFISILWAPVFPRPVNWRERVDEDTAKLVKRKAWPKDKRKASARRTI

Figure 4.36: Amino acid sequence of PlCg39 with transmenbrane helices underlined and highly conserved zinc finger motif highlighted in yellow

Phylogenetic analysis of DHHC palmitoyl transferase from various taxa has distinctively clustered into different lineages such as Yeast, Animalia, Fungi, Amoebozoa, Plantae and Prokaryota. In this study, PLCg39 was clustered with Amoebozoa lineage which possibly designates the origin of this eukaryotic gene from Amoeba (Figure 4.37). Sequence of PLCg39 obtained in this study was submitted to National Center of Biotechnology Information (NCBI) database under the accession number MK629535 (Appendix II).

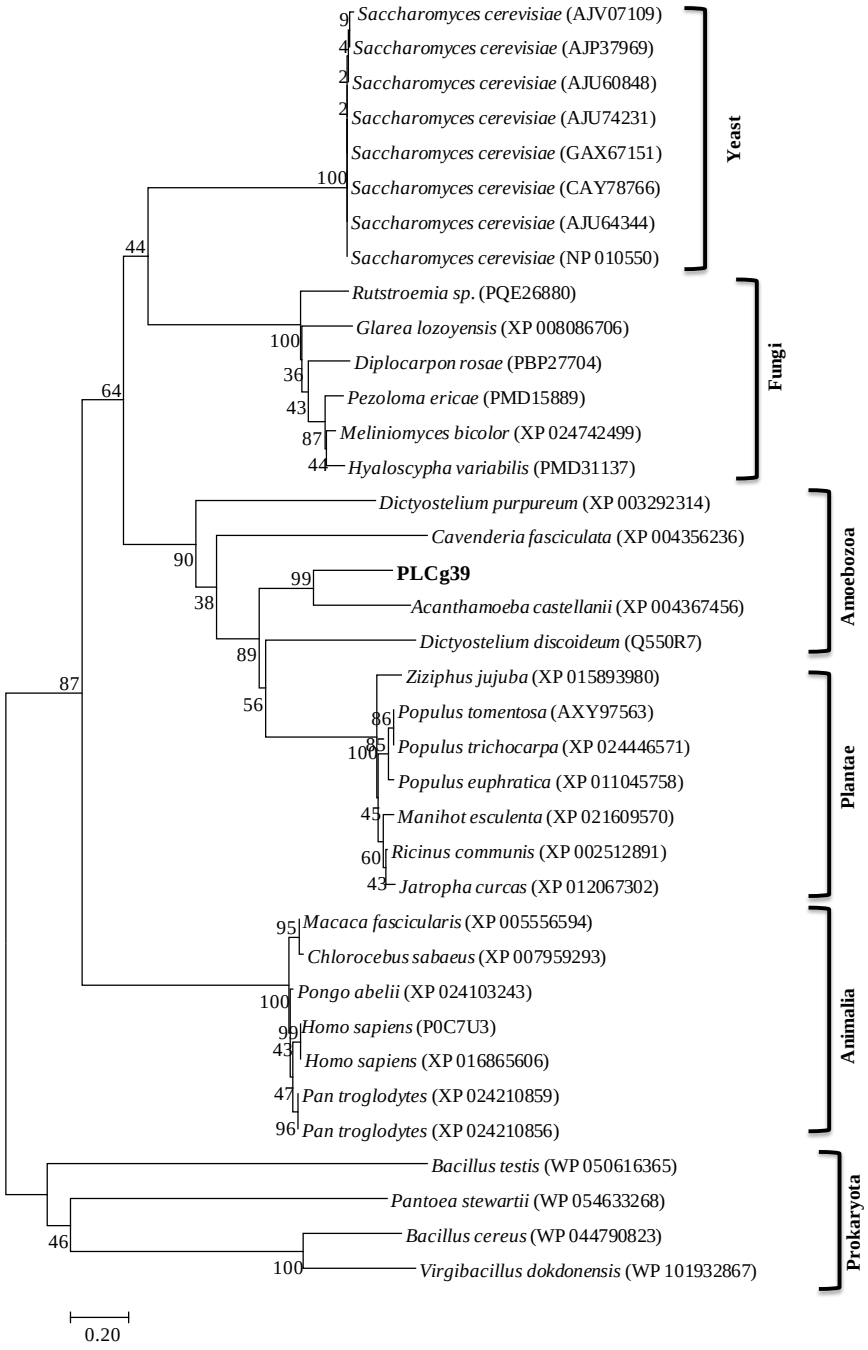


Figure 4.37: Phylogenetic relation of the transcript PLCg39 with DHHC palmitoyltrasferase of other taxa. The tree has been constructed using the neighbor-joining method. Bootstrap values were obtained with 1000 replicates are indicated at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses. Prokaryota DHHC palmitoyltrasferase sequences were included as outgroup. Branch lengths are proportional to evolutionary distances.

4.11.2 Metal stress tolerance by PLCg39

All yeast mutants, *cup1*$^{\Delta}$ *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PLCg39 demonstrated proficient growth in all the dilutions when spotted on SD-Ura media amended with 150 μM of Cu, 40 μM Cd,10 mM Zn and 2 mM Co. The mutant *cot1*$^{\Delta}$ transformed with PLCg39 could grow only till three dilutions. However, the growth of these mutants transformed with empty vector pFL61 were inhibited including the *cup1*^{s} transformed with pFL61 that was unable to grow in 150 μM of Cu. Overall, decrease in growth was observed with the increase in dilutions in metal amended medium (Figure 4.38).

Response of PLCg39 in presence of various concentrations of heavy metals (Cu, Cd, Zn and Co) was studied by observing their growth in SD-Ura medium supplemented with these heavy metals (Table 4.12). The transformant *cup1*$^{\Delta}$/PLCg39 showed high tolerance towards Cu in the medium. However, with increase in Cu concentration, cell viability was found to decrease. Beyond 500 μM Cu, the cell viability sharply declined. The growth was significantly higher when compared to the wild type BY4741 for all Cu concentrations although, at 1000 μM, both wild type and the transformant suffered severe cell growth inhibition (Figure 4.39A). In presence of Cd, both *ycf1*$^{\Delta}$/PLCg39 and BY4741 showed comparable cell viability which was highest at 40 μM but decreased with increasing concentration of Cadmium (Figure 4.39B). The cell viability of transformant *zrc1*$^{\Delta}$/PLCg39 was significantly higher than that of BY4741 in all concentrations with highest observed at 10 mM Zn. The transformant was able to maintain significant cell viability up to 12 mM of Zn. However, at 13 mM, a decline in the cell viability was observed for both *zrc1*$^{\Delta}$/PLCg39 and BY4741 (Figure 4.39C). The cell viability of all transformants in presence of Co declined rapidly at 2 mM (Figure 4.39D).

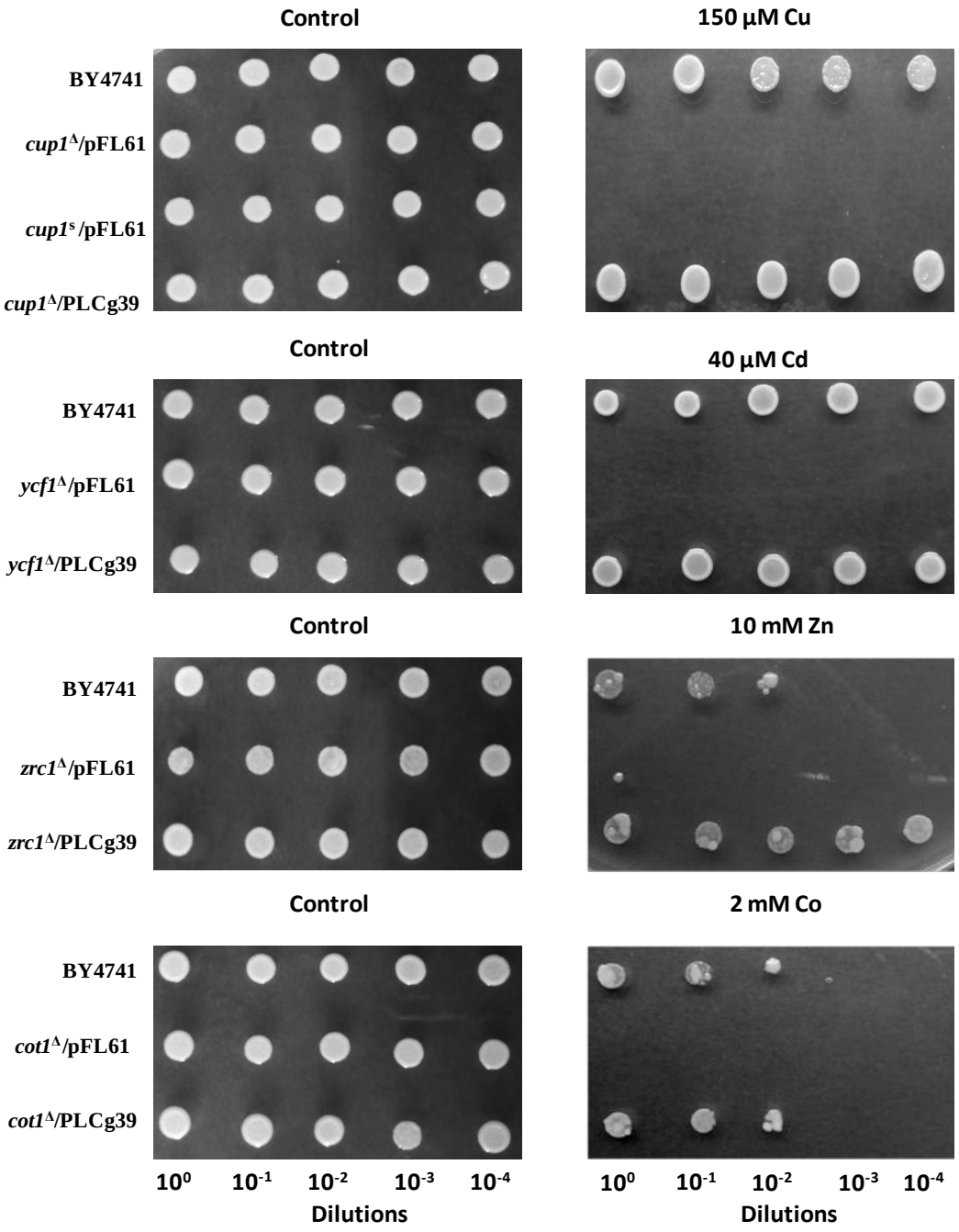


Figure 4.38: Drop assay of PlCg39 transformed *Saccharomyces cerevisae* mutants, *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$ and *zrc1*$^{\Delta}$ on SD-Ura agar. BY4741wild type and mutant strains transformed with empty vector pFL61 for positive and negative control. Serially diluted cultures were spotted on SD-Ura medium with or without metal supplement as indicated. Yeast strain *cup1*^{S} was also used as control.

Table 4.12: Percentage cell viability of *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PlCg49 compared to controls

	Percentage cell viability		
Cu concentration(µM)	***cup1*$^{\Delta}$/PlCg49**	**BY4741**	***cup1*$^{\Delta}$/pFL61**
0	100.00±0.00	100.00±0.00	100.00±0.00
150	76.45±4.023a	54.73±5.03b	3.16±0.183c
300	66.160±3.70a	52.64±3.26b	1.69±0.225c
500	58.475±1.76a	46.68±2.03b	0.56±0.068c
1000	19.390±1.48a	4.18±0.46b	0.13±0.037b
Cd concentration(µM)	***ycf1*$^{\Delta}$/PlCg49**	**BY4741**	***ycf1*$^{\Delta}$/pFL61**
0	100.00±0.00	100.00±0.00	100.00±0.00
40	41.07±0.771a	47.54±2.471a	2.75±0.399b
60	26.67±4.426a	40.62±3.848b	1.49±0.066c
80	13.79±2.044a	21.67±2.047b	0.57±0.068c
100	6.72±0.877a	7.41±2.684a	0.10±0.021a
Zn concentration(mM)	***zrc1*$^{\Delta}$/PlCg49**	**BY4741**	***zrc1*$^{\Delta}$/pFL61**
0	100.00±0.00	100.00±0.00	100.00±0.00
10	86.39±4.561a	62.66±2.284b	7.12±0.118c
11	75.56±2.80a	55.01±5.197b	2.07±0.049c
12	66.52±3.840a	43.45±2.362b	1.51±0.318c
13	26.37±1.372a	4.69±1.167b	0.16±0.023b
Co concentration(mM)	***cot1*$^{\Delta}$/PlCg49**	**BY4741**	***cot1*$^{\Delta}$/pFL61**
0	100.00±0.00	100.00±0.00	100.00±0.593
2	10.57±1.47a	6.71±0.593a	0.504±0.277a
10	8.09±0.636a	5.11±0.613a	0.188±0.003a
25	2.67±0.618a	3.18±1.326a	0.148±0.035a
50	1.96±0.113a	2.11±0.563a	0.070±0.012a

Percentage cell viability values are Mean ± Standard deviation. Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (n=3).

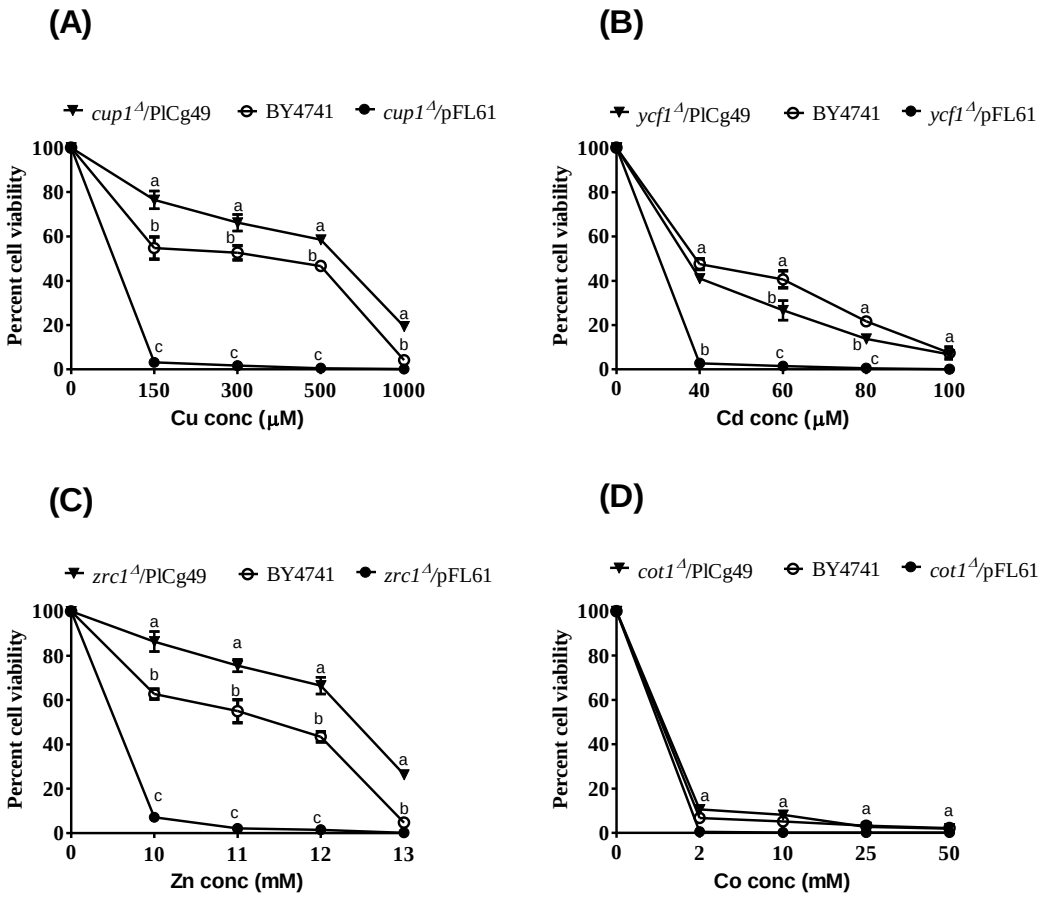


Figure 4.39: Growth response expressed as Percentage cell viability of (A) *cup1*$^{\Delta}$ (B) *ycf1*$^{\Delta}$ (C) *zrc1*$^{\Delta}$ and (D) *cot1*$^{\Delta}$ transformed with PlCg49 with wild type BY4741 and mutant yeasts transformed with Empty Vector pFL61 in SD-Ura broth supplemented with different metals at different concentrations as shown. Error bars are ±SD (*n*=3). Values sharing common letters on error bars for corresponding metal concentrations between samples are not significantly different at $P<0.05$ (n=3)

DHHC proteins play an important role as protein acyltransferases (PAT), as a vast majority of proteins are fatty acylated as a form of post-translational modifications. These proteins are primarily annotated containing a zinc finger domain on the basis of conserved pattern of cysteine and histidine residues. Studies have demonstrated that any mutation of the conserved region resulted in the loss of function of the protein as the zinc binding ability is lost. It has also been

reported that a selective pressure is essential for maintenance of the zinc finger pattern in DHHC protein (Gottlieb and Linder, 2017). Stoichiometric quantification has proved that there are two zinc binding sites per DHHC protein. Hence, binding of zinc has been associated with structural rather than functional role. There are little to no reports on the functional role of a DHHC protein in providing heavy metal tolerance to eukaryotic cells. However, the role of zinc finger has been studied extensively as a mediator in heavy metal homeostasis. The outcome of a range of abiotic stresses has been studied and found to be positively correlated with the increase in induction of zinc finger gene (Dai et al. 2007). Responses to heavy metal stresses have been extensively studied in zinc fingers and metals like cadmium and copper have shown to positively increase the zinc finger transcript abundance and confer resistance (Dixit and Dhankher, 2011; Sugano et al. 2003; Singh et al. 2015). A study by Chen et al. (2016) revealed the role of zinc finger transcriptional factor as a regulator for cadmium tolerance. Transgenic tobacco plants have been studied where overexpression of a zinc finger from rice conferred tolerance to a wide range of abiotic stresses including heavy metals (Mukhopadhyay et al. 2004).

4.11.3 Potential of PLCg39 to accumulate heavy metals

Metal estimation in the yeast transformants grown at various concentrations of Cu, Cd, Zn and Co was determined by ICP-MS analysis (Table 4.13). In the presence of external Cu, yeast transformant *cup1*$^{\Delta}$/PLCg39 was able to accumulate significant levels of Cu. Accumulation of Cu decreased with increase in Cu concentration and was found to be highest at 150 µM. For all the concentrations, the accumulation by *cup1*$^{\Delta}$/PLCg39 was always higher than the wild type BY471 (Figure 4.40A). The Cd content in *ycf1*$^{\Delta}$/PLCg39 was found to be comparable to that of the BY4741 for all the concentrations of Cd, highest being observed at 40 µM (Figure 4.40B). A significant amount of intracellular Zn accumulation was observed in case of *zrc1*$^{\Delta}$/ PLCg39

when compared to the wild type BY471. Highest accumulation was observed at 10 mM which decreased with the increase in external Zn concentration (Figure 4.40C). The cobalt content in the transformant *cot1*$^{\Delta}$/PLCg39 was significantly higher at 2 mM but as the concentration of Co in the medium increased, the accumulation of cobalt in the transformant *cot1*$^{\Delta}$/PLCg39 significantly decreased in comparison to the wild type BY4741(Figure 4.40D).

The accumulation of heavy metals by PLCg39 may be attributed to the DHHC-CRD containing a zinc finger. As mentioned earlier, the binding of zinc is necessary to maintain the functional role of the protein. However, the extent of metal accumulation in eukaryotic cells by a DHHC protein has not been studied earlier. Zinc fingers are associated with metal binding and are found to be flexible with metals other than zinc. In a recent study by Kluska et al. (2018), zinc fingers form stable complexes with cobalt, thus proving flexible coordination sites. The binding of cadmium and copper to zinc fingers was also studied as a replacement to zinc ions to regenerate the DNA binding activity of the zinc finger (Predki and Sarkar, 1994). In general, the zinc finger allows the binding of heavy metals which could explain the ability of PLCg39 for intracellular accumulation of toxic levels of heavy metals. In another recent study by Thakur et al. (2018), an environment transcript showing homology with a zinc finger fusion protein Ubiquitin containing the conserved cysteine and histidine residues demonstrated significant tolerance towards toxic concentrations of Cu, Cd and Zn. The present study has demonstrated the function of a DHHC protein containing a zinc motif which is able to tolerate toxic levels as well as accumulate heavy metals intracellularly.

Table 4.13: Metal uptake by *cup1*$^{\Delta}$,*ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ transformed with PlCg39 compared to controls

	Metal uptake (mg^{-1}g dry weight)		
Cu concentration(µM)	***cup1*$^{\Delta}$/PlCg39**	**BY4741**	***cup1*$^{\Delta}$/pFL61**
150	0.336±0.020a	0.084±0.009b	0.005±0.002c
300	0.309±0.016a	0.129±0.015b	0.011±0.002c
500	0.275±0.021a	0.246±0.027a	0.007±0.009b
1000	0.174±0.022a	0.128±0.015b	0.003±0.002c
Cd concentration(µM)	***ycf1*$^{\Delta}$/PlCg39**	**BY4741**	***ycf1*$^{\Delta}$/pFL61**
40	0.048±0.000a	0.043±0.003a	0.002±0.000b
60	0.040±0.001a	0.038±0.002a	0.001±0.000b
80	0.032±0.001a	0.026±0.003a	BDL
100	0.014±0.003a	0.011±0.001a	BDL
Zn concentration(mM)	***zrc1*$^{\Delta}$/PlCg39**	**BY4741**	***zrc1*$^{\Delta}$/pFL61**
10	5.388±0.424a	0.784±0.348b	0.319±0.251c
11	3.532±0.297a	0.232±0.020b	0.123±0.025c
12	1.961±0.081a	0.145±0.004b	0.063±0.012c
13	1.057±0.130a	0.054±0.003b	0.013±0.006c
Co concentration(mM)	***cot1*$^{\Delta}$/PlCg39**	**BY4741**	***cot1*$^{\Delta}$/pFL61**
2	0.076±0.011a	0.035±0.011b	0.010±0.001c
10	0.114±0.010b	0.245±0.007a	0.005±0.000c
25	0.070±0.002a	0.080±0.004a	BDL
50	0.006±0.002a	0.014±0.003a	BDL

Metal uptake values are Mean ± Standard deviation. Values sharing common letters on error bars for a particular metal concentration between samples are not significantly different at $P<0.05$ (n=3). BDL-Below Detection Limit

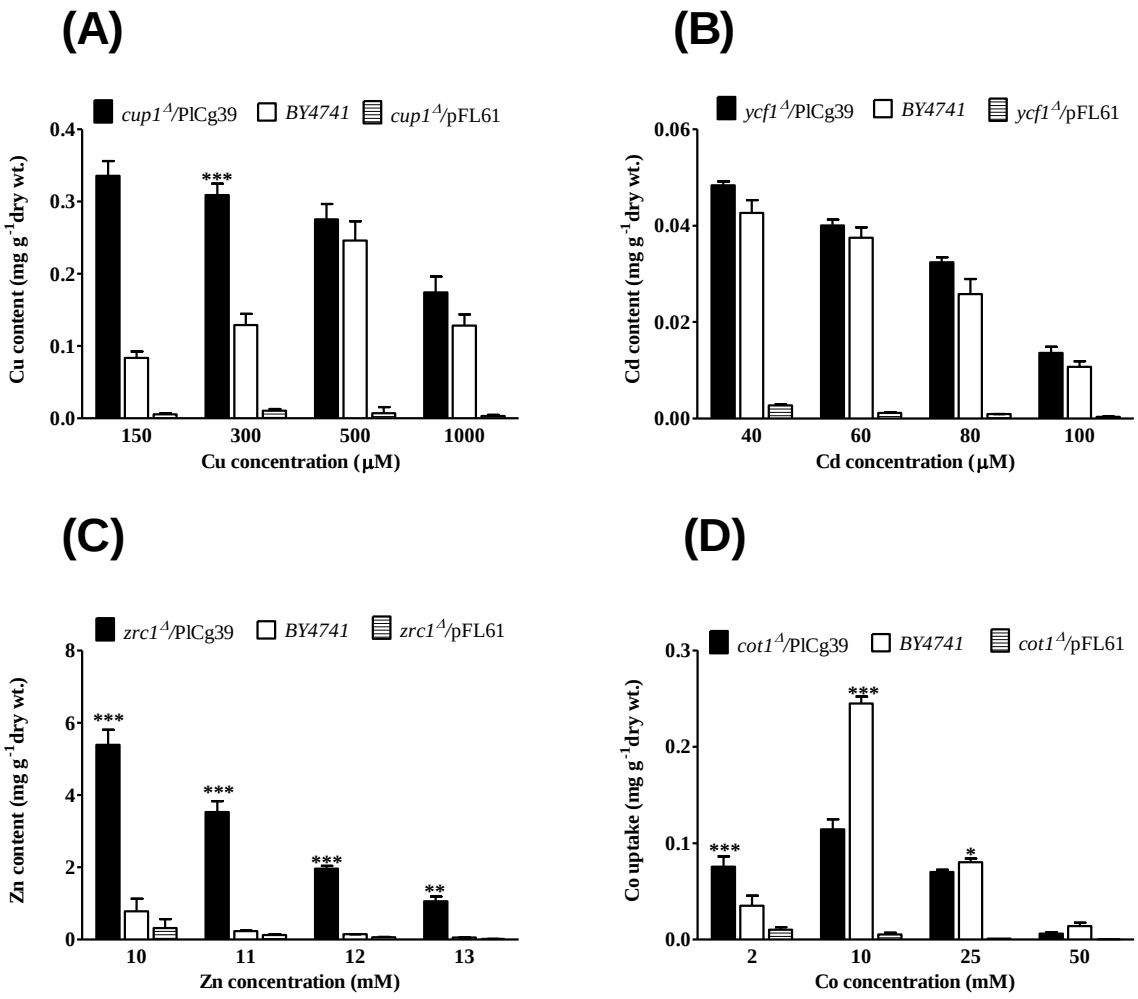


Figure 4.40: Effect of different concentration of (A) Copper (B) Cadmium (C) Zinc and (D) Cobalt on metal uptake by PlCg39 when compared to wild type BY4741 and yeast mutants transformed with empty pFL61 after 40 hrs of growth. Error bars are ±SD (*n*=3). (*, ** and *** are significant at $P<0.05$, $P<0.01$ and $P<0.001$ respectively).

Another probable explanation for the accumulation of heavy metals could be the presence of three CXXXC potential metal binding motifs in PLCg39. Various metallothioneins are found to contain CXXXC motifs and thus presumed to have the capability to ligate divalent heavy metals (Yang et al. 2014). Sequestration of copper by a cytochrome c oxidase has a highly conserved CXXXC domain which is a putative copper binding motif (Jaksch et al. 2001; Rosenzweig

2002). Metal ion binding motifs having a conserved Cys-X_{2-4}-Cys domain have been identified as effective chemical sensors for detection of toxic heavy metals like Cadmium and Arsenic in aqueous solutions. This renders the use of such proteins/peptides for developing effective biosensing strategies (Joshi et al. 2009).

4.11.4 DHHC Palmitoyl transferase activity for metal tolerance

The role of DHHC palmitoyl transferase as a stress response protein and rescuing the metal sensitivity in yeast was confirmed by transforming the cDNA PLCg39 in *akr1*$^{\Delta}$, a mutant of the DHHC palmitoyl transferase gene. The yeast transformant *akr1*$^{\Delta}$/PLCg39 was grown in SD-Ura medium amended with 0 μM, 150 μM, 300 μM and 500 μM of Cu. It was observed that the transformant *akr1*$^{\Delta}$/PLCg39 was able to grow well for all the dilutions till 300 μM of Cu. However, the growth was further decreased as the metal concentration increased. At 500 μM of Cu, the transformant *akr1*$^{\Delta}$/PLCg39 exhibited higher growth when compared to the mutant and comparable to the metal resistant wild type BY4741. This clearly indicates that the transformant *akr1*$^{\Delta}$/PLCg39 was able to rescue the DHHC palmitoyl transferase activity of the mutant yeast cells which resulted in enhanced high heavy metal tolerance (Figure 4.41).

The mutant lacks a functional DHHC palmitoyl transferase which makes it lose protein palmitoylation activity and hence form abnormality in cell morphology, reduced sporulation and respiratory functions. Overall, it displays reduced competitive fitness and becomes susceptible towards heat, heavy metals, hyperosmotic stress, chemical compound treatment etc. and causes overall decrease in vegetative growth (Dudley et al. 2005l; Enyenihi and Saunders, 2003; Pagani et al. 2007; van Bakel et al. 2005; Zhang et al. 2002). In a study on cadmium tolerance, results indicated that the gene *AKR1* proved important for genetic requirement for cadmium tolerance in yeast and contributed to the regulatory responses in presence of toxic concentrations of Cd^{2+}

(Serero et al. 2008). Similarly, when the effect of high concentration of Zn^{2+} on the cellular functioning of yeasts was evaluated, AKR1 was identified as major stress response gene as mutation caused hyper sensitivity towards the metal (Pagani et al. 2007). This encouraged us to test mutant *akr1*$^{\Delta}$ in presence of divalent cation Cu^{2+} and check the response by functionally complementing with cDNA PLCg39.

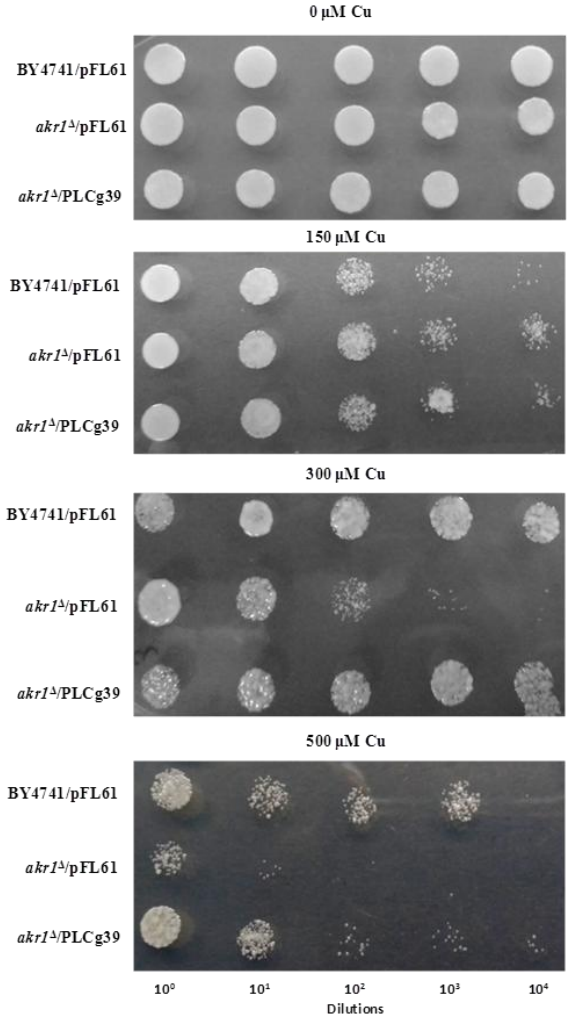


Figure 4.41: Drop assay of PlCg39 transformed yeast mutant *akr1*$^{\Delta}$ on SD-Ura agar. BY4741wild type and mutant strain *akr1*$^{\Delta}$ were transformed with empty vector pFL61 for positive and negative control. Serially diluted cultures were spotted on SD-Ura medium with or without metal supplement as indicated.

The potential of cDNA PLCg39 to tolerate toxic concentrations of heavy metals was quite evident from the experimental observations. The translated protein had the characteristics of a DHHC palmitoyl transferase. This protein has an important function as far as post-translational modification of other proteins is concerned.

From this study, a novel contribution in heavy metal stress tolerance has been identified which makes metatranscriptomics as an effective approach to reveal diversity in functions when applied to complex ecosystem like polluted soil. The heterologous expression of the transcript PLCg39 in suitable eukaryotic hosts can be useful for a variety of industrial applications like soil and remediation and developing improved variety of genetically modified heavy metal tolerant crop plants.

4.12 Maltose fermentation regulatory protein MAL33

A clone from library C of Cu contaminated soil from India, ARK3 that demonstrated high tolerance to copper was analyzed. Sequence study indicated that ARK3 demonstrated homology with maltose fermentation regulatory protein MAL33 and probably originated from the yeast *Saccharomyces cerevisiae*.

4.12.1 Bioinformatic analysis of ARK3

The plasmid carrying the cDNA from clone ARK3 was sequenced with primers NF and NR. The predicted molecular mass and pI of the protein are 14.17 kDa and 5.89 respectively. BLAST analysis revealed its homology with maltose fermentation regulatory protein MAL33 of *Saccharomyces cerevisiae* (Figure 4.42). The cDNA consists of 523 bp with an ORF of 375 bp encoding a protein of 124 amino acids (Figure 4.43).

Sequences producing significant alignments Download Manage Columns Show 100

select all *0 sequences selected* GenPept Graphics

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Maltose fermentation regulatory protein MAL33 [Saccharomyces cerevisiae]	261	261	72%	4e-85	99.21%	ONH79491.1
Mal33p [Saccharomyces cerevisiae YJM1250]	265	265	72%	5e-84	100.00%	AJP99130.1
Mal33p [Saccharomyces cerevisiae EC1118]	265	265	72%	5e-84	100.00%	CBK39375.1
Mal33p [Saccharomyces cerevisiae YJM326]	265	265	72%	5e-84	100.00%	AJQ02608.1
Mal33p [Saccharomyces cerevisiae YJM993]	265	265	72%	5e-84	100.00%	AHY74771.1
Mal33p [Saccharomyces cerevisiae P283]	265	265	72%	5e-84	100.00%	EWH19664.1
transcription factor MAL33 [Saccharomyces cerevisiae]	263	263	72%	6e-84	98.43%	PTN40364.1
Mal33p [Saccharomyces cerevisiae R103]	264	264	72%	6e-84	100.00%	EWG97481.1
Mal33p [Saccharomyces cerevisiae x Saccharomyces kudriavzevii VIN7]	264	264	72%	7e-84	100.00%	EHN08477.1
Mal33p [Saccharomyces cerevisiae YJM1252]	262	262	72%	4e-83	99.21%	AJQ08315.1
Mal33p [Saccharomyces cerevisiae YJM1463]	260	260	72%	2e-82	97.64%	AJP87251.1

Figure 4.42: Sequence homology search for cDNA ARK3 using BLASTx through NCBI

5'3' Frame 3

```
tagtg ttt cag atg aat ggt acc aag ttt ttc tca aac gcc aat aat gct cac ata ttg gtc
   V   F   Q   M   N   G   T   K   F   F   S   N   A   N   N   A   H   I   L   V
gaa att gca aag gat atg ctg gac gac ata ttc tta act cca aac aac ctg tat gta gtc
 E   I   A   K   D   M   L   D   D   I   F   L   T   P   N   N   L   Y   V   V
cat ggt cct gtg ata cca atg aag gca ctg gaa ata gcc aat gca ttg gta gat atc gta
 H   G   P   V   I   P   M   K   A   L   E   I   A   N   A   L   V   D   I   V
aat aag tat gat cac aat atg aag ttg gag gct tgg aat att ttg tgc gat gta tct aag
 N   K   Y   D   H   N   M   K   L   E   A   W   N   I   L   C   D   V   S   K
ttc gtt ttc tcc ctg aaa cat tgc aat cat aaa atg ttt caa agg ttt tca act aaa tgt
 F   V   F   S   L   K   H   C   N   H   K   M   F   Q   R   F   S   T   K   C
cag agt gct cta atc gat ttg cct atc tct aga cca ctg cgc cta aat gat gat tcc aaa
 Q   S   A   L   I   D   L   P   I   S   R   P   L   R   L   N   D   D   S   K
gat gaa gac gac ata att cct taa ttt att gtt cac gcc gtt cac tta tac gag ata gat
 D   E   D   D   I   I   P   -   F   I   V   H   A   V   H   L   Y   E   I   D
ata ctg ata gag tgt gag tga tat tct taa gtc ttg ctt ttc gag ggt gta aga agc tat
 I   L   I   E   C   E   -   Y   S   -   V   L   L   F   E   G   V   R   S   Y
gtt ctt cag gcg aga tta ttc tac tcc tgc ctt act tgt
 V   L   Q   A   R   L   F   Y   S   C   L   T   C
```

Figure 4.43: Translation of cDNA ARK3 encoding a protein showed homology with Maltose fermentation regulatory protein MAL33

Yeasts utilize a wide array of compounds as carbon and nitrogen sources and can make full use of the fermentable carbohydrates available. The process of selection of carbohydrates from diversely available sources is highly developed in the yeast system. The molecular mechanisms of sensing and regulation of the fermentation in yeast is controlled and orchestrated by several enzyme systems which follow induction and repression of genes associated with carbohydrate fermentation.

The fermentation of maltose requires one of the five MAL loci (*MAL1, MAL2, MAL3, MAL4* and *MAL6*) which are unlinked. Each of the loci encodes for three proteins which are essential for maltose fermentation. They are maltose permease, maltase and *MAL* transcriptional activator protein (Kim and Michels, 1988). The transcription of the structural genes is induced by maltose and repressed by glucose. The set of *MAL* genes include, a maltose transporter (MALT), maltase (MALS), and zinc finger-type transcription factors (MALR) that positively regulate *MALT* and *MALS* genes. The transcription of these genes are induced by maltose but are inhibited by the presence of glucose in the medium. The cDNA ARK3 revealed highly conserved regions towards the C-terminal end of the MALR genes of yeast *S. cerevisiae*. Multiple sequence analysis with homologous yeast sequences revealed highly conserved regions (Figure 4.44).

Figure 4.44: Multiple sequence alignment of ARK3 gene with other homologous sequences of *S. cerevisiae* retrieved by BLASTp analysis.

Phyllogenetic study of different Maltose fermentation regulatory Protein MAL33 belonging to different taxa has distinctly categorized them into different lineages such as Yeast and Fungi. The cDNA ARK3 was observed to cluster with Yeast indicating its possible taxonomic origin (Figure 4.45). Sequence of ARK3 obtained in this study was submitted to National Center of Biotechnology Information (NCBI) database under the accession number MN794026 (Appendix II).

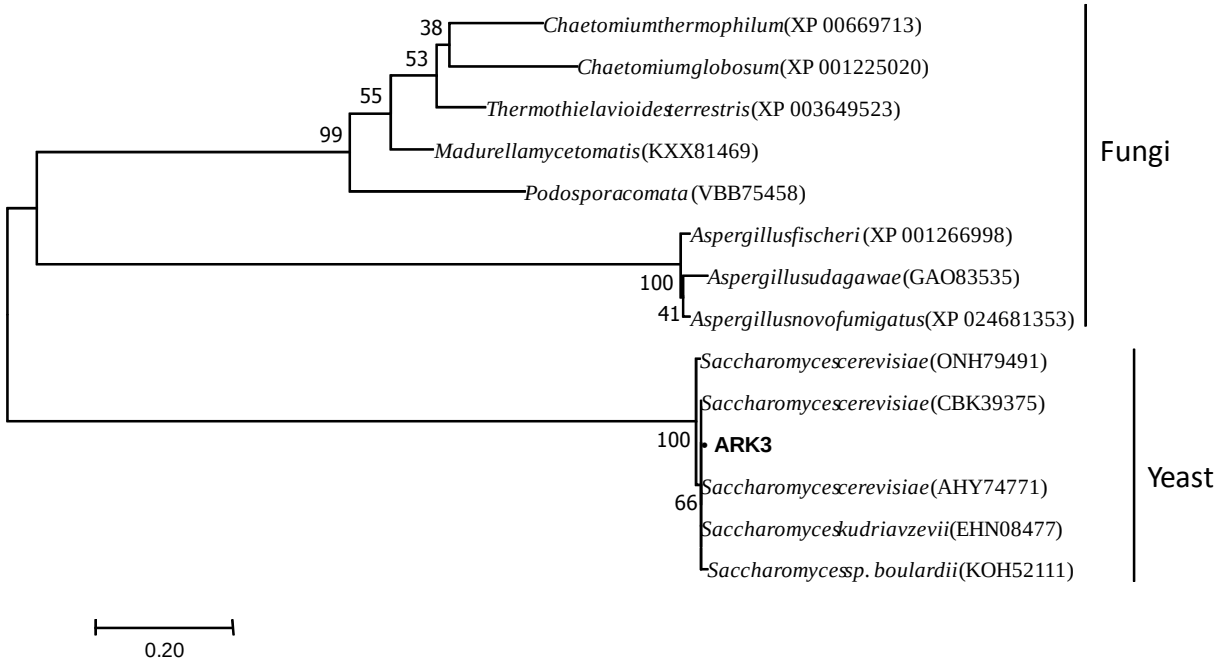


Figure 4.45: Phylogenetic relation of the transcript ARK3 with maltose fermentation regulatory protein of other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses.

4.12.2 Metal tolerance potential of ARK3

To study the metal tolerance potential imparted by ARK3, the copper sensitive mutant yeast *cup1*$^{\Delta}$ was transformed with ARK3 and was grown in media containing a range of concentrations of copper. The mutant yeast that was transformed with the cDNA ARK3, was able to maintain viability and was able to grow well on SD-Ura agar amended with 150, 300 and 500 µM of Cu. However, the growth of the yeast mutant *cup1*$^{\Delta}$ transformed with empty vector pFL61, was severely retarded in presence of metal amended media. In drop assay a decline in growth was observed for *cup1*$^{\Delta}$/ARK3 with the increase in metal concentration (Figure 4.46).

To study the growth pattern of the transformant *cup1*$^{\Delta}$/ARK3 in presence of copper, it was grown in SD-Ura broth amended with different concentrations of copper. The transformant

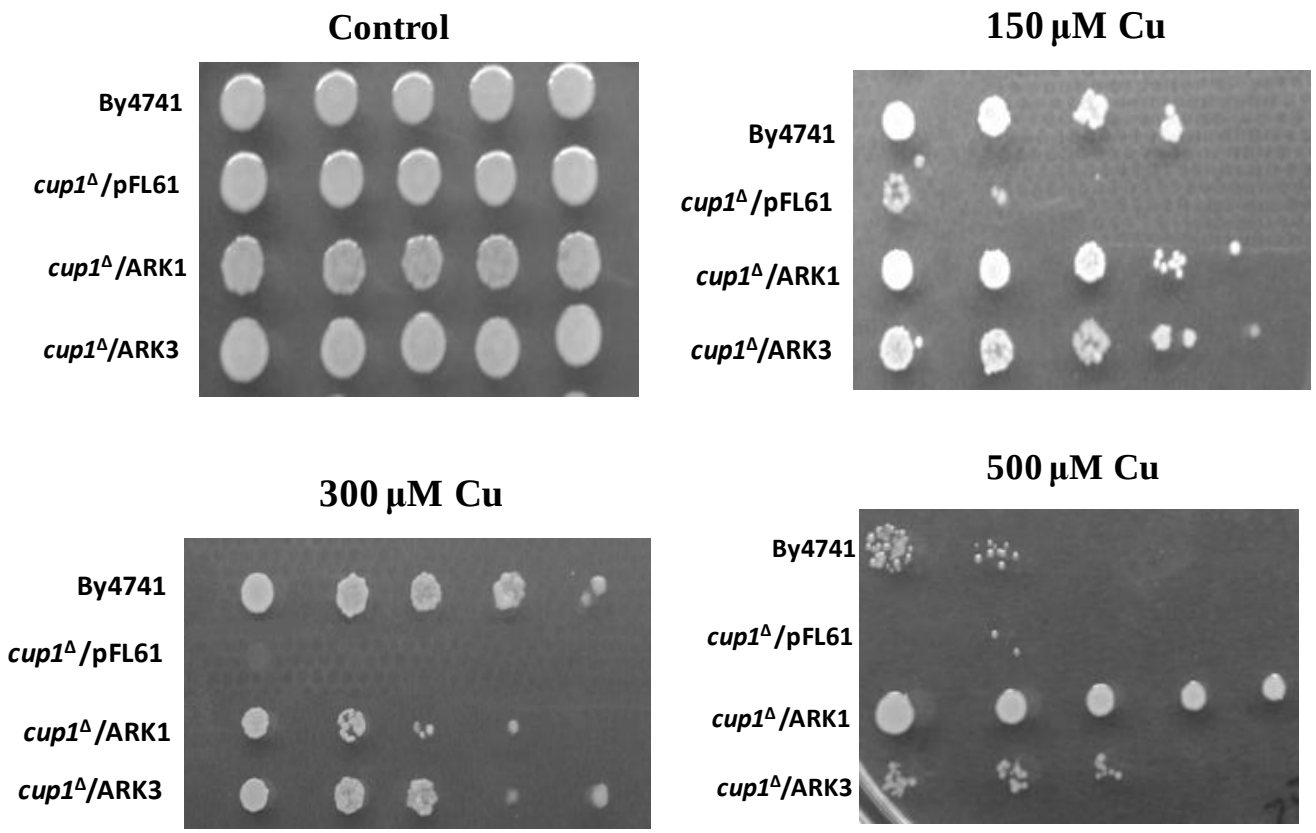


Figure 4.46: Drop assay of AKR1 (discussed in next section) and AKR3 transformed yeast mutant *cup1*$^{\Delta}$ on SD-Ura agar. BY4741 wild type and mutant *cup1*$^{\Delta}$ transformed with empty vector pFL61 were used for positive and negative control. Yeast strain *cup1*^{S} was also used as control

cup1$^{\Delta}$/ARK3 was able to grow well till 300 µM of Cu where the growth recorded was similar to the growth of the wild type BY4741. On the other hand, for all concentrations of Cu, *cup1*$^{\Delta}$ carrying the empty plasmid pFL61 recorded poor to no growth. At 500 µM of Cu, the growth of *cup1*$^{\Delta}$/ARK3 was registered to be significantly higher than the control BY4741 (Table 4.14 and Figure 4.47).

Table 4.14: Growth response of *cup1*$^{\Delta}$ transformed with ARK3 and controls in presence of Cu

Time (in hours)	Control (0 µM Cu)			150 µM Cu		
	BY4741	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK3**	**BY4741**	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK3**
0	0.019±0.002a	0.024±0.002a	0.031±0.002a	0.027±0.002a	0.031±0.00a	0.027±0.005a
5	0.123±0.006a	0.117±0.003a	0.128±0.007a	0.127±0.002a	0.093±0.00a	0.122±0.002a
10	0.458±0.032a	0.352±0.020a	0.565±0.013a	0.400±0.038a	0.248±0.005a	0.563±0.039a
15	0.831±0.048a	0.663±0.069a	0.722±0.088a	0.585±0.012a	0.257±0.055a	0.856±0.044a
20	1.855±0.065a	1.606±0.016a	1.856±0.054a	1.425±0.005a	0.268±0.056a	1.768±0.055a
25	2.081±0.049a	1.963±0.030a	2.134±0.111a	1.724±0.010a	0.255±0.375a	1.808±0.015a
30	1.929±0.006a	1.908±0.085a	2.122±0.002a	1.712±0.009a	0.238±0.024a	1.734±0.022a

Time (in hours)	300 µM Cu			500 µM Cu		
	BY4741	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK3**	**BY4741**	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK3**
0	0.025±0.000a	0.023±0.001a	0.022±0.001a	0.025±0.000a	0.029±0.000a	0.024±0.000a
5	0.120±0.000a	0.083±0.001a	0.183±0.050a	0.136±0.005a	0.131±0.002a	0.135±0.002a
10	0.302±0.000a	0.099±0.005a	0.411±0.050a	0.130±0.005a	0.068±0.003a	0.143±0.002a
15	0.511±0.013a	0.090±0.003a	0.623±0.049a	0.132±0.002a	0.014±0.000a	0.147±0.005a
20	1.011±0.010a	0.079±0.007a	1.130±0.010a	0.123±0.004a	0.013±0.001a	0.155±0.004a
25	1.311±0.007a	0.081±0.009a	1.360±0.026a	0.094±0.006a	0.002±0.000a	0.124±0.001a
30	1.226±0.025a	0.071±0.000a	1.357±0.035a	0.076±0.007a	0.002±0.000a	0.113±0.001a

O.D. values are Mean ± Standard deviation. Values sharing common letters for a particular time interval between samples are not significantly different at $P<0.05$ (n=3)

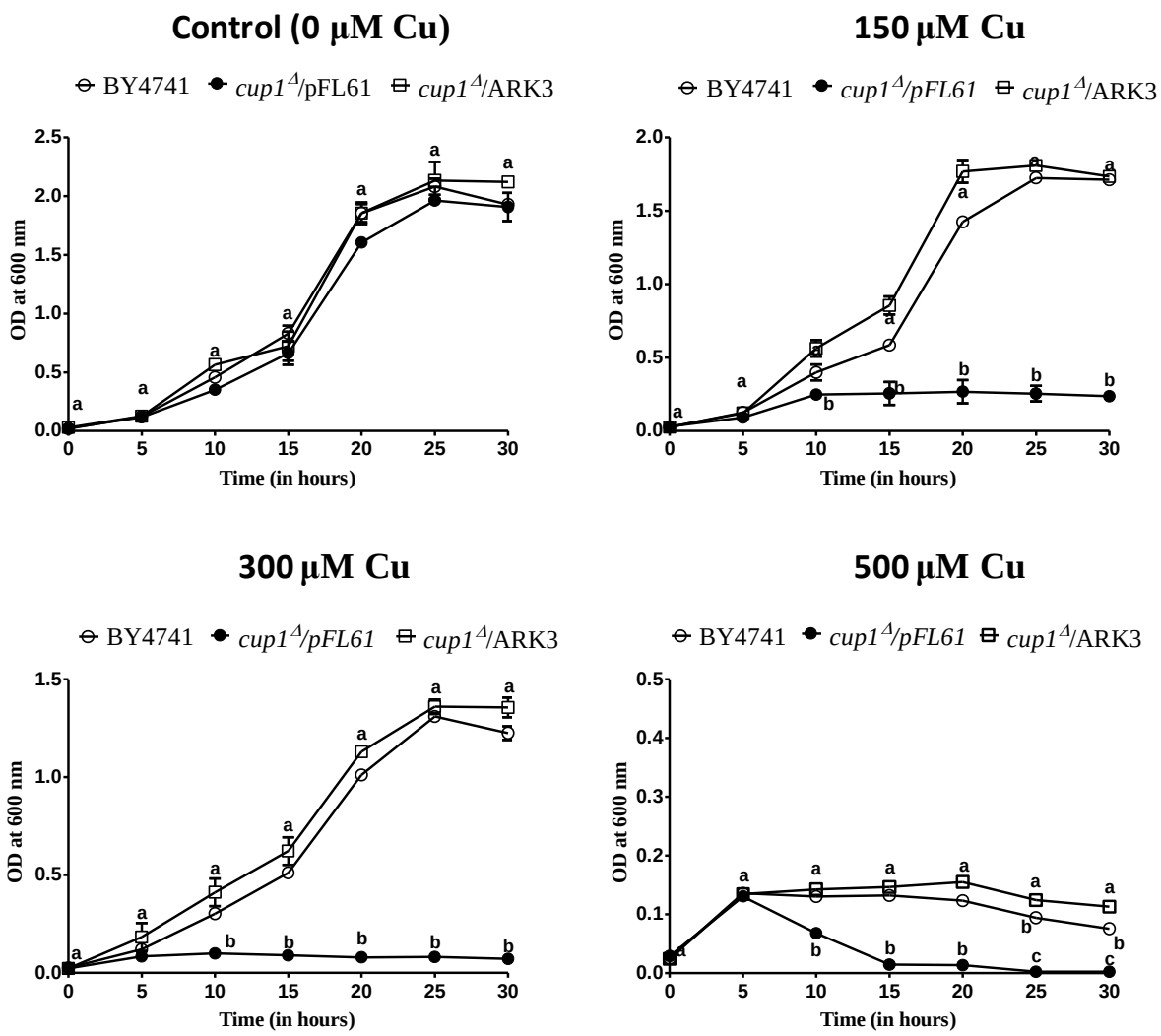


Figure 4.47: Growth response of *cup1*$^{\Delta}$ transformed with ARK3 compared with growth of parent type BY4741 and *cup1*$^{\Delta}$ transformed with empty pFL61 vector at control (0 µM Cu), 150, 300 and 500 µM of Cu. Error bars are ±SD. Values sharing common letters on error bars at a particular time between samples are not significantly different at $P<0.05$ (n=3)

Charron et al. (1989) reported on MAL63 that was from one of the family of almost identical genes each of which maps to one of the five complexes of MAL loci. They are referred to as MAL13, MAL23, MAL33, MAL43 or MAL63 according to the positions of the genes. In *Saccharomyces* species, the inducible *MAL63* has been reported to produce a protein which binds

to DNA and controls the expression of maltose fermentative enzymes. This feature of DNA binding is most widely noticed in transcriptional regulatory proteins. The N-terminal end of the protein MAL63 contains a cysteine-lysine-arginine rich region. Reports have suggested that these cysteine-zinc fingers of these proteins are responsible for DNA binding in general and also, the region is responsible for recognizing unique DNA sequences immediately towards the C-terminal (Kim and Michels, 1988). The conserved cysteine-rich region not only covers two pairs of cysteines but reaches beyond a sixth cysteine residue all of which contribute towards DNA binding. For the effective folding of the DNA binding proteins, they should contain two zinc binding sites. Three dimensional structure analysis by NMR of yeast GAL4 protein indicated the "zinc cluster" to be six cysteine residues bound to 2 zinc atoms which is similar to that of metallothioneins (Vallee et al. 1991). The DNA-binding domain of MAL-activator protein is basically the first 100 residues and a functional core area comprising of the transactivation domain, which is constituted from 60-283. Furthermore, the residues 251-299 help in inhibiting Mal63p activation.

The cDNA ARK3 was found to have highly conserved regions from the C-terminal region of the Mal-activator protein, which contains a maltose-responsive domain that functions as a unit to counter the inhibitory effect on the activation function. Reports have suggested a model system where the function of the C-terminal domain was explained. Approximately, a 90-residue region (residues 331–423) was held responsible for sensing of maltose, and approximately 50-residue region at the C-terminus of Mal63p (residues 420–470) was required for the formation and/or maintenance of the active state (Danzi et al. 2003). Hence, evidences have proved a clear role of the C-terminal domain in maintaining functionality of the DNA binding zinc-finger domain, and

binding of metals like zinc and cadmium are essential in maintaining the proper folding of the protein and thus maintaining its functionality (Freedman et al. 1988).

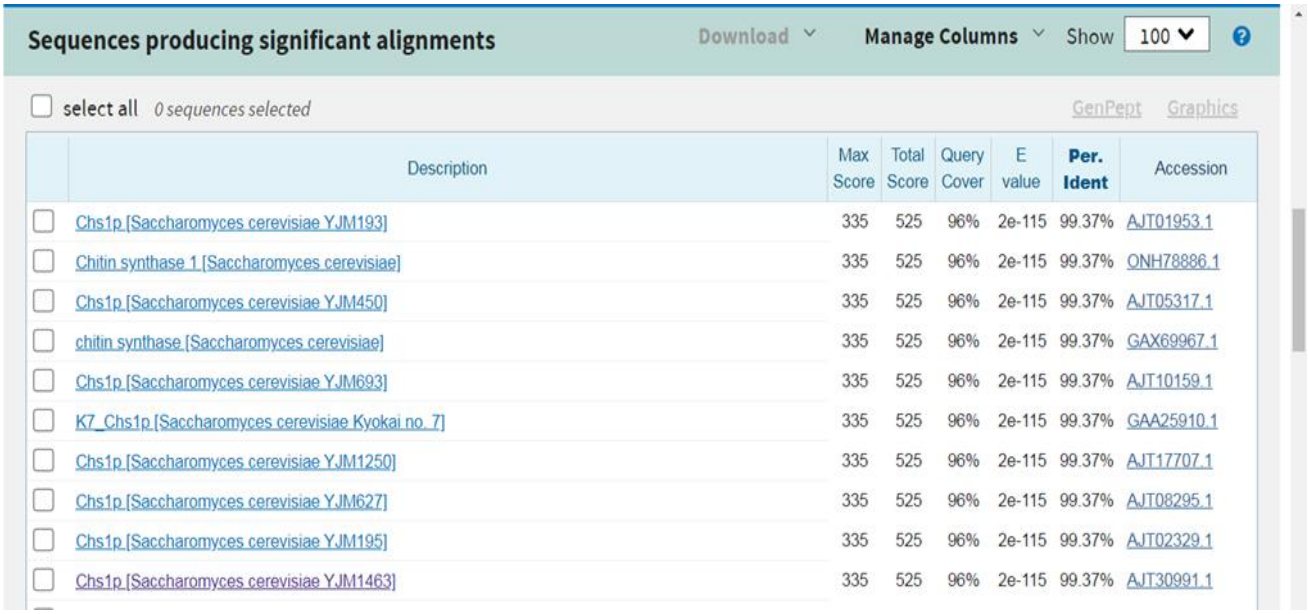
In conclusion, the results here indicate the role of the cDNA ARK3 in providing tolerance to metal sensitive yeasts when expressed in them. The cysteine-rich amino-terminal region is the most important domain of this protein which share homology with the zinc-cluster of DNA-binding proteins having cysteine-rich domains. The cDNA ARK3 has been found to be homologous to the C-terminal domain of the MAL-activator protein which is believed to have a positive regulation on the DNA binding domain. Heterologous expression of the gene ARK3 in yeast proved that the transcript could be used for providing tolerance towards toxic concentrations of copper and could possibly be used to develop bioremediation based strategies.

4.13 Chitin synthase

Another clone from library C of Cu contaminated soil from India, ARK1 that demonstrated high tolerance to copper was analyzed. Sequence study indicated that ARK1 demonstrated homology with chitin synthase protein and probably originated from the yeast *Saccharomyces cerevisiae*.

4.13.1 Bioinformatic analysis of ARK1

The cDNA ARK1 was sequenced with primers NF and NR and the size of the cDNA was found to be 849 bp with an ORF of 657 bp encoding a protein of size 218 amino acids. BLAST analysis showed the homology of ARK1 with chitin synthase of *S. cerevisiae* (99.37 %). The predicted molecular mass and pI of the protein are 24.97 kDa and 4.68 respectively. Figure 4.48 shows the sequence homology search result and Figure 4.49 depicts the deduced protein translation.

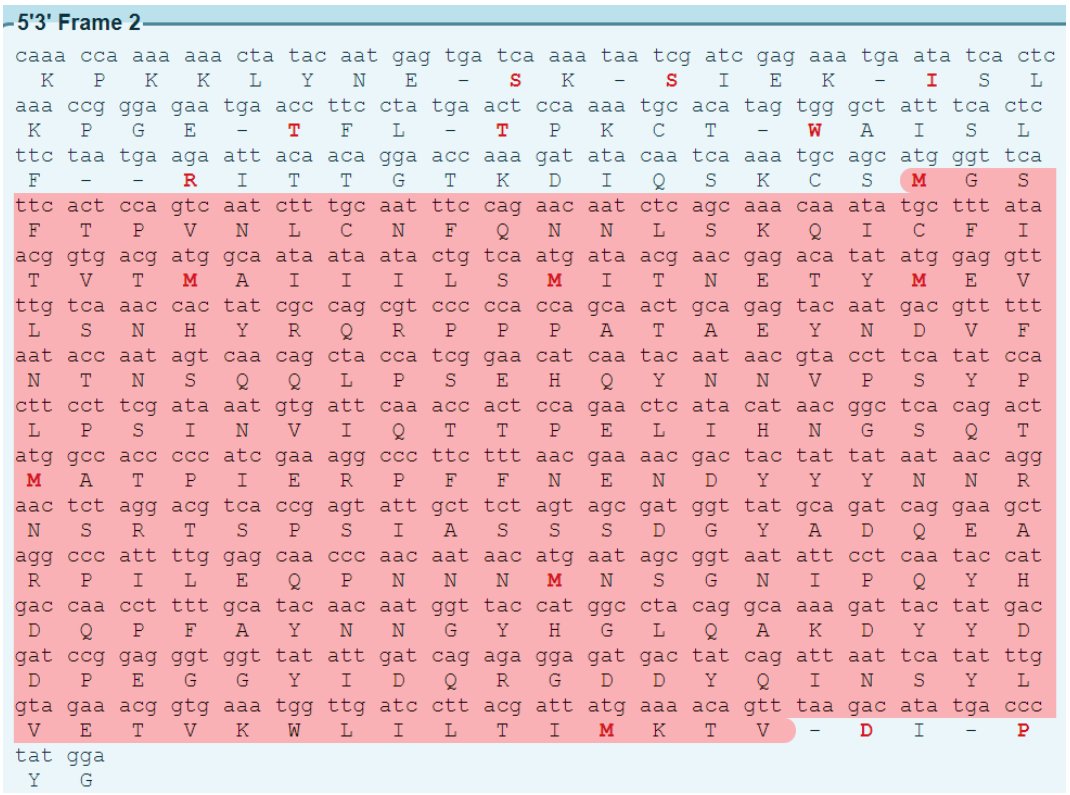


Sequences producing significant alignments Download Manage Columns Show 100

select all *0 sequences selected* GenPept Graphics

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☐	Chs1p [Saccharomyces cerevisiae YJM193]	335	525	96%	2e-115	99.37%	AJT01953.1
☐	Chitin synthase 1 [Saccharomyces cerevisiae]	335	525	96%	2e-115	99.37%	ONH78886.1
☐	Chs1p [Saccharomyces cerevisiae YJM450]	335	525	96%	2e-115	99.37%	AJT05317.1
☐	chitin synthase [Saccharomyces cerevisiae]	335	525	96%	2e-115	99.37%	GAX69967.1
☐	Chs1p [Saccharomyces cerevisiae YJM693]	335	525	96%	2e-115	99.37%	AJT10159.1
☐	K7_Chs1p [Saccharomyces cerevisiae Kyokai no. 7]	335	525	96%	2e-115	99.37%	GAA25910.1
☐	Chs1p [Saccharomyces cerevisiae YJM1250]	335	525	96%	2e-115	99.37%	AJT17707.1
☐	Chs1p [Saccharomyces cerevisiae YJM627]	335	525	96%	2e-115	99.37%	AJT08295.1
☐	Chs1p [Saccharomyces cerevisiae YJM195]	335	525	96%	2e-115	99.37%	AJT02329.1
☐	Chs1p [Saccharomyces cerevisiae YJM1463]	335	525	96%	2e-115	99.37%	AJT30991.1

Figure 4.48: Sequence homology search for ARK1 using BLASTx through NCBI



5'3' Frame 2

```
caaa cca aaa aaa cta tac aat gag tga tca aaa taa tcg atc gag aaa tga ata tca ctc
  K   P   K   K   L   Y   N   E   -   S   K   -   S   I   E   K   -   I   S   L
aaa ccg gga gaa tga acc ttc cta tga act cca aaa tgc aca tag tgg gct att tca ctc
 K   P   G   E   -   T   F   L   -   T   P   K   C   T   -   W   A   I   S   L
ttc taa tga aga att aca aca gga acc aaa gat ata caa tca aaa tgc agc atg ggt tca
 F   -   -   R   I   T   T   G   T   K   D   I   Q   S   K   C   S   M   G   S
ttc act cca gtc aat ctt tgc aat ttc cag aac aat ctc agc aaa caa ata tgc ttt ata
 F   T   P   V   N   L   C   N   F   Q   N   N   L   S   K   Q   I   C   F   I
acg gtg acg atg gca ata ata ata ctg tca atg ata acg aac gag aca tat atg gag gtt
 T   V   T   M   A   I   I   I   L   S   M   I   T   N   E   T   Y   M   E   V
ttg tca aac cac tat cgc cag cgt ccc cca cca gca act gca gag tac aat gac gtt ttt
 L   S   N   H   Y   R   Q   R   P   P   P   A   T   A   E   Y   N   D   V   F
aat acc aat agt caa cag cta cca tcg gaa cat caa tac aat aac gta cct tca tat cca
 N   T   N   S   Q   Q   L   P   S   E   H   Q   Y   N   N   V   P   S   Y   P
ctt cct tcg ata aat gtg att caa acc act cca gaa ctc ata cat aac ggc tca cag act
 L   P   S   I   N   V   I   Q   T   T   P   E   L   I   H   N   G   S   Q   T
atg gcc acc ccc atc gaa agg ccc ttc ttt aac gaa aac gac tac tat tat aat aac agg
 M   A   T   P   I   E   R   P   F   F   N   E   N   D   Y   Y   Y   N   N   R
aac tct agg acg tca ccg agt att gct tct agt agc gat ggt tat gca gat cag gaa gct
 N   S   R   T   S   P   S   I   A   S   S   S   D   G   Y   A   D   Q   E   A
agg ccc att ttg gag caa ccc aac aat aac atg aat agc ggt aat att cct caa tac cat
 R   P   I   L   E   Q   P   N   N   N   M   N   S   G   N   I   P   Q   Y   H
gac caa cct ttt gca tac aac aat ggt tac cat ggc cta cag gca aaa gat tac tat gac
 D   Q   P   F   A   Y   N   N   G   Y   H   G   L   Q   A   K   D   Y   Y   D
gat ccg gag ggt ggt tat att gat cag aga gga gat gac tat cag att aat tca tat ttg
 D   P   E   G   G   Y   I   D   Q   R   G   D   D   Y   Q   I   N   S   Y   L
gta gaa acg gtg aaa tgg ttg atc ctt acg att atg aaa aca gtt taa gac ata tga ccc
 V   E   T   V   K   W   L   I   L   T   I   M   K   T   V   -   D   I   -   P
tat gga
 Y   G
```

Figure 4.49: Translation of ARK1 cDNA encoding a protein which showed homology with chitin synthase

Chitin synthase activity is associated with the transmembrane of the yeast cell wall (Leal-Morales et al. 1988). Chitin synthase 1 (Chsp1) is the most abundant enzyme in yeast *S. cerevisiae* and was partially purified with digitonin. The molecular mass of a partially purified Chsp1 of is found to be 570 kDa (Kang et al. 1984). However the cDNA ARK1 homologous to Chsp1 of yeast *S. cerevisiae* was found to be much smaller. Transmembrane domain prediction revealed that ARK1 has one transmembrane domain from 20 to 41 amino acids. Multiple sequence analysis with homologous yeast sequences revealed highly conserved regions (Figure 4.50).

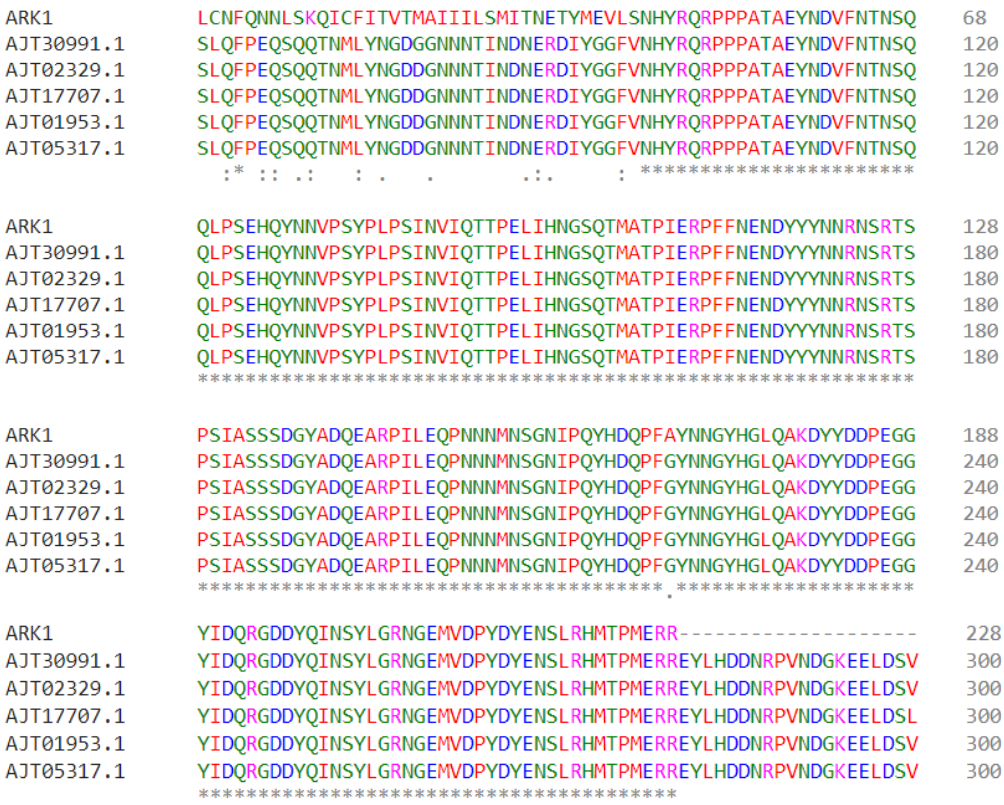


Figure 4.50: Multiple sequence alignment of ARK1 gene with other homologous sequences of *S. cerevisiae* retrieved by BLASTp analysis.

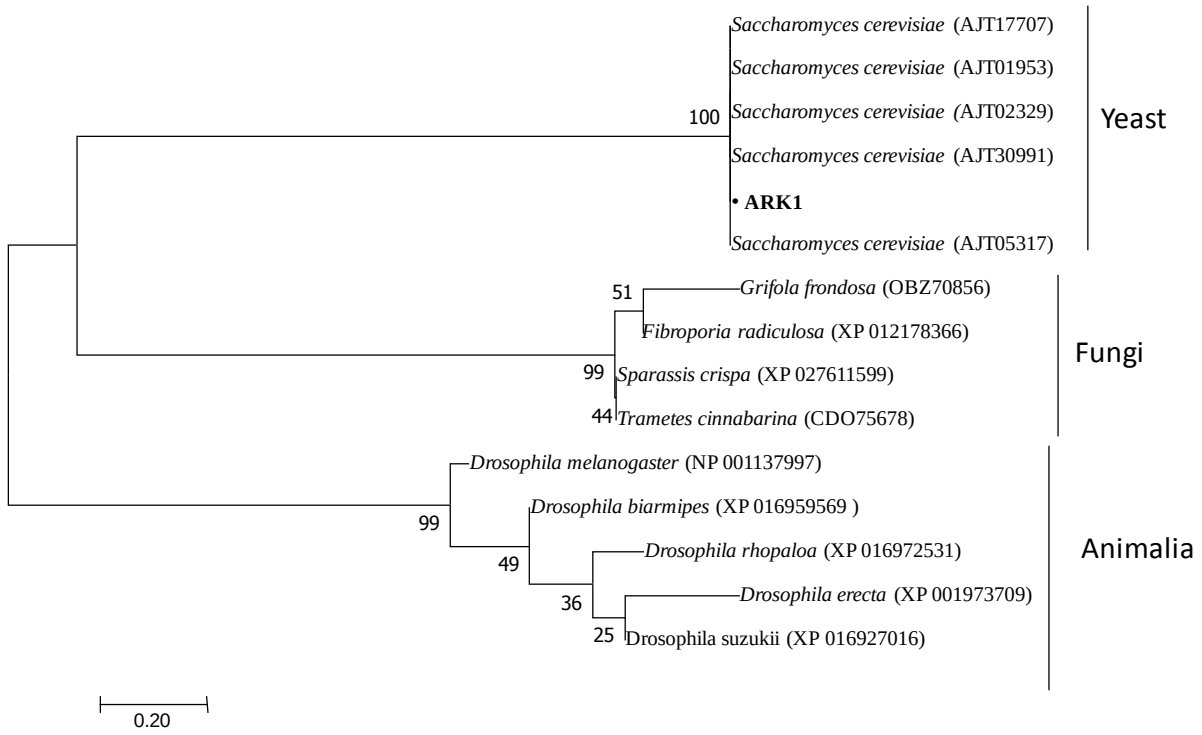


Figure 4.51: Phylogenetic relation of the transcript ARK1 with chitin synthase of other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses.

Phyllogenetic analysis of ARK1 with amino acids sequences of chitin synthase from various taxa has clearly clustered them into different lineages such as yeast, fungi and animalia. The cDNA ARK1 was observed to cluster with *Saccharomyces cerevisiae* indicating its possible taxonomic origin (Figure 4.51).

4.13.2 Metal tolerance potential of ARK1

The mutant yeast *cup1*$^{\Delta}$ was transformed with the the transcript ARK1 and was grown in media having different concentrations of copper. It was observed that the cDNA was able to confer tolerance to the mutant yeast and were able to grow well on SD-Ura agar plates supplemented with 150, 300, 500 μM of Cu. However, the yeast mutant *cup1*$^{\Delta}$ transformed with empty vector

pFL61 was unable to grow in metal amended media. In drop test, a general reduction in growth was observed as the concentration of copper in the media increased (Figure 4.46).

To measure the growth response of the transformant *cup1*$^{\Delta}$/ARK1 in presence of copper, it was grown in SD-Ura broth supplemented with varied concentrations of copper. The growth of the transformant *cup1*$^{\Delta}$/ARK1 was not affected till 300 μM of Cu and was comparable to the growth of the wild type BY4741. For all the concentrations of Cu, *cup1*$^{\Delta}$ carrying the empty plasmid pFL61 recorded poor to no growth. At 500 μM of Cu, the growth of all the transformant decreased, but *cup1*$^{\Delta}$/ARK1 recorded significantly higher growth than the control BY4741 (Table 4.15 and Figure 4.52).

Studies have reported that presence of heavy metals in the growth media generally results in a decrease in growth of yeasts. Metal sensitive *Ascomycete* confirmed reduction in growth compared to the other tolerant strains which proves that the presence of toxic concentrations of metals in the growth media negatively impacts the hyphal growth of the fungi (Martino et al. 2000). However, few of the studies pertaining to growth inhibition, do not include the known mechanisms involved in the decline in growth. Chitin, as the primary component of the cell wall structure, plays a primary role during the growth of a fungus. The modification in cell wall structure and chitin localization was studied in ericoid mycorrhizas by quantifying chitin content in the fungus when treated with varying concentrations of $ZnSO_4$. A higher content of chitin was found in the mycelium when it was treated with $ZnSO_4$ as compared to untreated samples, confirmed an unexpected increase in chitin at the molecular level (Lanfranco et al. 2002).

Table 4.15: Growth response of *cup1*$^{\Delta}$ transformed with ARK1 and controls in presence of Cu

Time (in hours)	Control (0 µM Cu)			150 µM Cu		
	BY4741	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK1**	**BY4741**	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK1**
0	0.0197±0.002a	0.024±0.001a	0.031±0.002a	0.027±0.002a	0.031±0.005a	0.027±0.005a
5	0.123±0.006a	0.117±0.003a	0.128±0.007a	0.127±0.002a	0.093±0.005a	0.122±0.002a
10	0.458±0.032a	0.2515±0.020a	0.515±0.037a	0.400±0.034a	0.248±0.006b	0.413±0.010a
15	0.830±0.048a	0.6625±0.069a	0.772±0.038a	0.585±0.012a	0.257±0.555b	0.543±0.030a
20	1.855±0.065a	1.6055±0.016a	1.856±0.054a	1.425±0.005a	0.268±0.056b	1.518±0.005a
25	2.081±0.049a	1.9625±0.030a	2.134±0.111a	1.724±0.010a	0.255±0.037b	1.707±0.005a
30	1.929±0.006a	1.908±0.085a	2.122±0.002a	1.712±0.009a	0.238±0.024b	1.637±0.033a

Time (in hours)	300 µM Cu			500 µM Cu		
	BY4741	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK1**	**BY4741**	***cup1*$^{\Delta}$/pFL61**	***cup1*$^{\Delta}$/ARK1**
0	0.025±0.000a	0.023±0.000a	0.022±0.000a	0.025±0.000a	0.029±0.000a	0.024±0.000a
5	0.183±0.050a	0.083±0.000a	0.120±0.000a	0.136±0.005a	0.1306±0.002a	0.135±0.002a
10	0.411±0.050a	0.099±0.005b	0.302±0.000a	0.131±0.005a	0.0679±0.003b	0.143±0.002a
15	0.6225±0.049a	0.091±0.003b	0.424±0.000a	0.132±0.002a	0.0142±0.000b	0.147±0.005a
20	1.1795±0.059b	0.079±0.007c	0.975±0.007a	0.123±0.004a	0.0134±0.001b	0.147±0.006a
25	1.360±0.026a	0.081±0.009b	1.271±0.041a	0.094±0.006b	0.0023±0.000c	0.124±0.001a
30	1.3565±0.035a	0.071±0.000b	1.226±0.025a	0.076±0.007b	0.0023±0.000c	0.113±0.000a

O.D. values are Mean ± error. Values sharing common letters for a particular time interval between samples are not significantly different at $P<0.05$ (n=3)

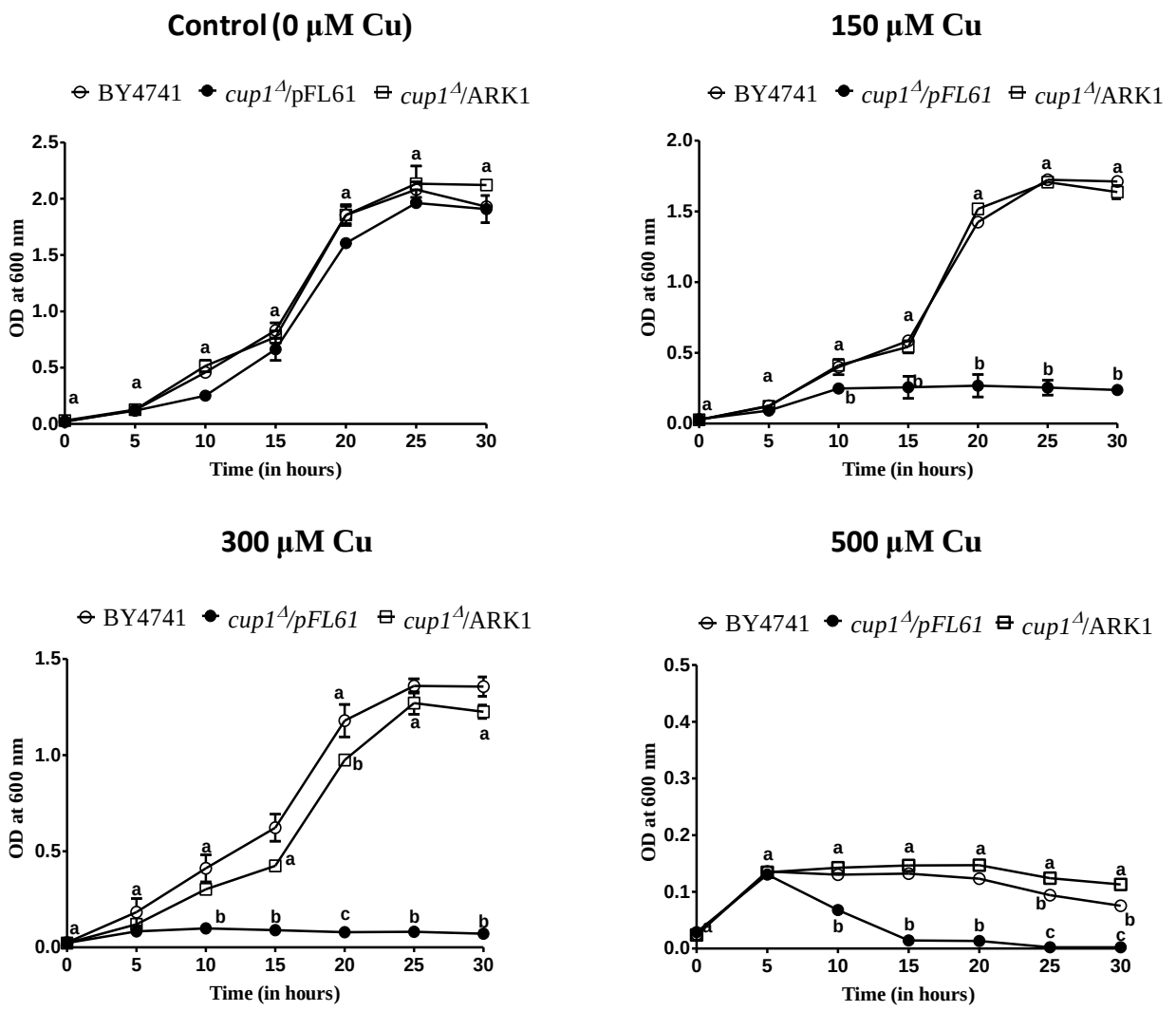


Figure 4.52: Growth response of *cup1*$^{\Delta}$ transformed with ARK1 compared with growth of parent type BY4741 and *cup1*$^{\Delta}$ transformed with empty pFL61 vector at control (0 µM Cu), 150, 300 and 500 µM of Cu. Error bars are ±SD. Values sharing common letters on error bars at a particular time between samples are not significantly different at $P<0.05$ (n=3)

The major fungal cell wall components, chitin and chitosan can sequester metal ions. Adsorbtion of Zn on *Rhizopus arrhizus* cell wall was particularly prominent on its chitin/chitosan components. A primary mechanism to fight heavy metal toxicity is to 'resist' the uptake of metals by extracellular chelation. Reports suggest that the metal ions are more actively

exchanged with calcium, hydrogen, magnesium and potassium ions present on the surface of fungal biomass (Viraraghavan and Srinivasan, 2011). The differential expression of chitin synthase (class IV) in presence of high concentrations of zinc was documented by Lafranco et al. (2004). The abundance of mRNA of a class IV chitin synthase was found to be upregulated in presense of high Zn concentration whereas class II chitin synthase was found to be unaltered. This differential expression indicated a discrete functional role for the two encoded enzymes.

It has been proved that chitin synthase occurs as a zymogen so it requires stimulation for it to become active (Bowen et al. 1992; Mayer et al. 1980). The role of endogenous trypsin has been responsible for chitin synthase to become active after proteolytic cleavage (Lorito et al. 1994). Some environmental conditions have been effective in the up-regulation of Chitin synthase mRNA. Abiotic stress conditions, like elevated incubation temperature and low pH influenced the transcript levels as gene expression was induced in the case of class III chitin synthase in *E. dermatitidis* (Wang and Szaniszlo, 2000). It was evident from a study by Gay et al. (1989) that metals like Mg^{2+}, are essential to stimulate the activity of chitin synthase, suggesting the role of metal ions in stimulating the enzyme activity. The reported size of *Saccharomyces cerevisiae* is about 1131 amino acids (Philippsen et al. 1997). In a report by Kang et al. (1984) when the zymogenic form was activated by trypsin treatment, few of the fractions showed positive chitan synthase activity. Hence the possibility of chitan synthase having multiple domains cannot be ruled out. Moreover, proteolytic processing of CHS from both fungal and insect sources has lead to discovery of polypeptides of lower molecular mass capable of chitin synthase activity. The molecular weight of chitin synthase derived from zygomycota *Absidia.glauca* was estimated to be 30 kDa which was visible on SDS-PAGE as a single band. Antibodies immunoprecipitated this 30-kDa polypeptide, which demonstrated that it exists as a single polypeptide species.

Furthermore, the N-terminal sequence was also similar to the N-terminal of the zymogenic form of the enzyme (Machida and Saito, 1993). In the present study, cDNA ARK1 also demonstrated sequence similarity with the amino-terminal of the homologous *Saccharomyces cerevisiae*. Studies on chitin synthase protein of zygomycota *Mucor rouxii* have revealed 4 polypeptide bands of molecular weights 21, 23, 33 and 39 kDa out of which 21 kDa showed the strongest activity (Lending et al. 1991). These molecular weights are quite similar to this study. Another study on *Phycomyces blakesleeanus* fungus suggested that a long mRNA of chitin synthase could encode for a more that one subunit of protein required for enzyme activity, or the protein could have multifunctional domains and exist in multiple forms (Miyazaki and Ootaki, 1997). The predicted 24.97 kDa, 218 amino acid polypeptide ARK1 isolated in this study could be a part of a subunit of chitin synthase or belong to a functional domain. More research and insights into the molecular structure and functioning of the cDNA is required to prove that the smaller coding sequence in comparison to the very large CHS1 gene is responsible for chitin synthesis function.

In conclusion, the experimental observations indicate that cDNA ARK1 has the ability to confer tolerance to metal sensitive yeast against toxic concentrations of copper. The homology depicted by ARK1 with chitin synthase gene shows that this gene may be upregulated in metal contaminated soil which is in accordance with studies where abiotic stresses were responsible for increase in chitin synthase gene expression and also stimulate its activity. It is apparent that the cell wall is the first line of defense and the primary barrier between the toxic environment and the microorganisms. Hence, the key metal detoxification mechanism comes from the ability of the eukaryote to form metal complexes within the cell wall. The approach of metatranscriptomics

to study a complex ecosystem like soil can be beneficial to understand the functional diversity as opposed to isolation based study which might not include all functionally active populations.

4.14 Functional diversity of eukaryotic microorganisms present in Copper polluted soil

To establish the link between the genes and the organisms which possess them, which are metabolically active eukaryotic population inhabiting the metal polluted soil, a functional diversity study was performed. To understand the eukaryotic functional diversity of copper polluted soil, 18S rRNA gene was amplified with universal degenerate primers and the amplicons were sequenced on Illumina HiSeq sequencing platform. The primers preferentially amplified the V4 region to encompass most of the eukaryotic diversity. Both the cDNA synthesized from total RNA and genomic DNA were amplified with the 18S rRNA gene primer yielding an amplification product of size 450 bp (Figure 4.53).

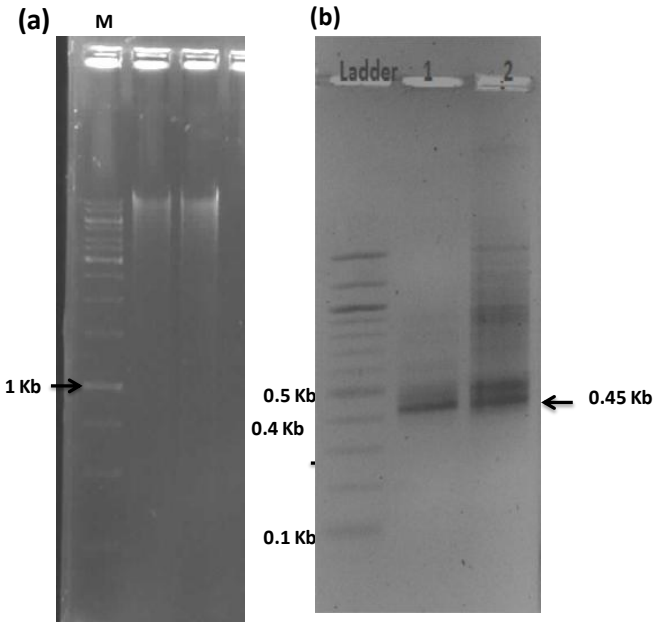


Figure 4.53: (a) cDNA synthesized from total RNA from soil visualized in 0,8 % agarose gel (b) Amplificaton by 18S degenerate primers 1- cDNA amplified by 18S rDNA degenerate primers 2- Total soil genomic DNA amplified by 18S rDNA degenerate primers

4.15 High-throughput amplicon sequencing for understanding eukaryotic diversity

Illumina HiSeq amplicon sequencing of both the cDNA and genomic DNA were produced paired end reads. When sequencing the cDNA, it generated 567223 paired reads of sequence lengths of 250 nucleotide bases on which preprocessing steps were performed to extract the V4 region from the paired end reads. All downstream sequence analysis steps were carried out using QIIME 1 software (Caporaso et al. 2010). Pre-processing of reads was carried out by first trimming the unwanted spacer sequences and filtering the sequences by removing conserved region from paired end reads using the FLASH program (Magoč and Salzberg, 2011). Following the mismatch filter, 565960 reads were obtained which passed all filters and were ready to be fed into the UCHIME implemented in the tool VSEARCH tool to remove Chimeric sequences. A total of 36,563 chimeric sequences were filtered out to result in 529397 pre-processed reads. Picking of OTUs was done based on the sequence similarity based on the UCLUST program integrated in the tool kit, where clustering was based on a similarity index of 97%. From the total 529397 reads, a total of 146433 OTUs were identified. Single-ton reads which appear in the OTUs not more than twice were filtered and a total of 20121 reads were obtained for further processing. In case of genomic DNA, amplicon sequencing resulted in a total of 182015 reads. Following pre-processing and chimera removal, a total of 107649 reads were obtained. After single-ton removal there were a total of 8592 reads.

4.16 Taxonomic annotations of eukaryotic communities in polluted soil

The eukaryotic community structure was analyzed by Next-Generation Sequencing (NGS). The results of taxonomic assignment indicated the wealth of eukaryotic diversity in heavy metal polluted soil.

The eukaryotic community was found to be distributed in approximately 29 known phyla with highest representation from the phyla *Amoebozoa* (3765 OTUs) followed by *Opisthokonta* (2108), SAR (943) (*Stramenopiles* (*heterokonts*), *Alveolates*, and *Rhizaria*) etc. At the class level of taxonomic rank three most abundant classes are *Gracilipodida* (1107), *Holozoa* (996) and *Nucletmycea* (939). At the order level of taxonomic rank, the abundance of OTUs was found to be highest for Metazoa (Animalia) (969) followed by Fungi (935), Filamoeba (642) etc. Family of *Filamoeba*, *Amphiesmenoptera*, *Dactylopodida* are prominent players in the polluted soil. From the existing data, it can be concluded that the most active and dominant population of eukaryotes in the polluted soil are primarily from the Animalia, Protista and Fungi kingdom.

The prokaryotic metagenomic profiling of Malanjkhand copper mine has been performed and analyzed extensively by Gupta et al. (2017). The highly acidic mine tailings hosted distinct microbial taxa, harboring Proteobacteria, Actinobacteria, Chloroflexiwhile. Abundance of species like Sulfobacillus, Leptospirillum, Acidithiobacillus, Acidiferrobacter etc indicates, the presence of extremophiles which have the ability to tolerate the harsh acidic environment. Studies pertaining to *in situ* bioremediation of acid mine drainage impacted soil focused on the enrichment and stimulation of anaerobic bacteria playing essential role in sulphate and metal removal thus addressing effective strategies towards bioremediation (Gupta et al. 2018).

The discovery of the vast unknown diversity of microbial species on earth is highly dependant on the curated reference databases for full-length small subunit (SSU) rRNA gene sequences. Over the past few years about 2 million full-length genes have been deposited in the SILVA SSU database (Quast et al. 2012). However such sequences account for just a fraction of the Earth's total microbial diversity and has been estimated to be in the range of billions to trillions (Karst et al. 2018). Apart from the sparse eukaryotic lineages represented in the reference database,

additional problems arise due to the presence of uneven distribution of 16S and 18S full-length sequences. The recent version of SILVA database (version 138) consists of 1,983,534 16S rRNA gene sequences as compared to only 172,450 18S rRNA gene sequences. An earlier version of SILVA (version 128) comprised of 10,789 Alveolates and 22,723 Animalia), with only 676 representing Amoebozoa (Popovic and Parkinson 2018). In a study on myxomycetes, a significant proportion of about 60-70 % of unknown myxomycetes sequences were found which did not cluster well enough to any reference sequences and were categorized as undescribed species within the reference database. This 'hidden diversity' of species creates methodological difficulties as it is difficult to demarcate whether this section is generated due o sequencing errors (e.g. chimera formation), lack of comparate sequences or truly belong to 'not-yet-discovered' (Schnittler et al. 2017).

There are fewer reports on the analysis of eukaryotic community structure of polluted soils. Few of the studies have focused in deciphering the abundance of eukaryotic diversity in environments affected by anthropogenic activity majorly acid mine drainage (AMD) affected zones. In a study on acid mine drainage in China, a large fraction of eukaryotic population belonged to the phylum *Alveolata* followed by *Nucleariidae* whereas about 6% of sequences were identified from *Ascomycota* phylum (Hao et al. 2010). A more extensive analysis on eukaryotic community structure in acid mine drainage was carried out by Mendez-Garcia et al. (2015). The authors have concluded that microbes dwelling in these environments belong primarily to *Archaeplastida*, SAR (*Stramenophiles, Alveolates, Rhizaria*), *Excavata* (protists) and *Opisthokonta* taxa. The prime inhabitants of a heavy metal site were Ascomycota and Basidiomycota fungi. Further, protists are also dominant in the eukaryotic community in heavy metal polluted areas along with fungi as they are more suited to the sequestration of heavy metals from soil (Gadanho et al.

2006). Thus, the population of metal sequesters play a major role in heavy metal polluted regions. Also, protozoa such as *Oxytricha sp*. predominantly found in AMD affected regions graze on acidophilic bacteria and archaea and thus alters the prokaryotic population by becoming the primary oxidizers of metals, thus altering the environment making it more acidic. In a recent report by Auld et al. (2013) the acid mine drainage sites were also inhabited by the number of algal species alongside fungi.

In another related study by Auld et al. (2017) eukaryotic communities in the mining region have been found to be significantly varying between seasons (high temperature to low winter temperatures). Three dominant classes were identified here as group of algae *Chrysophyceae*, *Microbotrymycetes*a group of fungi and LKM11, a novel fungal clade. In summer, *Chrysophyceae* algae were dominant in the mining area, however as the seasons changed there was a shift in community structure which lead to the proliferation of the fungal group LKM11. Foulon et al. (2016) investigated the eukaryotic microflora of two polluted forest soils in France by Illumina MiSeq sequencing. The metal polluted land was a phytomanagement site with Poplar plantation. Studies revealed, *Ascomycota* and *Basidiomycota* as the two most abundant fungal communities in the respective sites, which interestingly revealed significant differences between the inhabiting fungal communities. The success of phytomanagement by Poplar plantation was therefore clear as ectomycorrhizal communities as root symbionts were prevalent in the highly stressed soil.

Ciliates are single-celled eukaryotic species most suitable for ecotoxicological studies (Abraham et al. 2019) as they are highly responsive to various metals. As they have different copies of genes in the macronucleus making them ideal for characterizing the stress-induced gene and thereby to elucidate the mechanism of gene expression under heavy metal stress. Both

phenotypic and molecular characterization has been carried out on the ciliates and there are a number of studies on the effect of toxicity of heavy metals on ciliates which mainly focuses on accumulation, cell growth, viability and cellular structure. At the molecular level, induction of metallothionein (MT), superoxide dismutase, GPx and hsp70 genes in response to heavy metals is reported only in *Tetrahymena thermophila*, *Euplotes crassus* and *Oxytricha* (Sterkiella) nova (Kim et al. 2018; Anderson et al. 1996). While the induction of antioxidant enzymes and hsp70 gene has been used as molecular biomarkers for monitoring heavy metal stress in several living organisms, these studies are scarce in freshwater ciliates (Kim et al. 2014). Heavy metals like copper induce oxidative stress in the ciliated protozoa by increasing the levels of generated ROS. Metals with such redox potential like Cu, generate ROS directly by auto-oxidation and metals without redox activity like Cd generate ROS indirectly by affecting the antioxidant defense mechanism (Gutiérrez et al. 2008). Several antioxidant enzymes have been investigated in cliliates such as superoxide dismutase (SOD), ascorbate peroxidase, catalase, glutathione peroxidase and peroxiredoxin are known to be involved in ROS detoxification and in protecting the cells from oxidative stress along with several proteins responsible for the bioaccumulation of heavy metals (Kim et al. 2014;).

Tardigrades are microscopic marine creatures known for their resistance to harsh environmental conditions. Transcriptomic analysis revealed, high levels of expression of antioxidant enzymes, copper transporters, ATOX1, and a Cu-ATPase. Furthermore, the results from studies have indicated that tardigrades express well-known key osmoregulatory enzymes as revealed by transcriptome analysis following exposure to toxic levels of copper (Hygum et al. 2017).

The damage control mechanisms of tardigrade *Ramazzottius varieornatus* in extreme environmental conditions, was found to constitute of a unique nuclear protein termed Dsup, for

damage suppressor. Activity of the gene is specific towards damage control from ionizing radiation or hydrogen peroxide treatment, which generates hydroxyl radicals. A Dsup ortholog from the tardigrade *Hypsibius exemplaris* similarly binds to nucleosomes and protects DNA from hydroxyl radicals. These results indicate that Dsup promotes the survival of tardigrades through a direct mechanism that includes binding to nucleosomes and shielding chromosomal DNA from hydroxyl radicals under different conditions (Chavez et al. 2019). Remarkably, the transcriptome analysis upon exposure to copper, constituted well known copper-sensitive osmoregulatory enzymes in the *E. sigismundi* transcriptome. Specifically, α and β subunits of the Na/K ATPase as well as carbonic anhydrase, which seem highly expressed in this heterotardigrade. The sequences of these enzymes are also present in the eutardigrade *H. dujardini* where apart from osmoregulatory enzymes, antioxidant enzymes like Cu-Zn superoxide dismutase, Mn superoxide dismutase, glutathione reductase, glutathione peroxidase and several copper transporter proteins were also induced.

The amoebozoan genus *Acanthamoeba* shows an omnipresent worldwide distribution inhabiting aquatic and terrestrial environments, and it might even be one of the most dominant protists in soil with fundamental importance in nutrient cycling (Geisen et al. 2014). Several Acanthamoeba species were isolated and their phylogeny was investigated. Some of the sequence types were isolated from soils at high altitudes in Tibet. These environments are characterized by harsh, often extremely cold conditions (Geisen et al. 2014).

Generally, Acanthamoeba is known to have a broad ecological niche and exist as the most dominant protist communities in extreme environments such as deserts (Rodriguez-Zaragoza et al. 2005) or polluted soils (Lara et al. 2007). A large number of strains of Acanthamoeba were isolated and abundance and diversity analysis were performed on cultivable protozoa (flagellates

and amoebae) in a polycyclic aromatic hydrocarbon (PAH) as well as molecular diversity analysis by 18S rRNA gene. Noticeably, the number of cultivable protozoa was higher in the polluted soil; however, the polluted soil displayed an impoverished community, dominated by certain taxa, such as Acanthamoeba. It has also been evidenced that the co-existence of bacteria and protozoa may influence and speed up the pollutant degradation process (Kinner et al. 2002). A batch culture of a Pseudomonas strain could consume Toluene up to 7.5 times faster with the presence of soil flagellate *Heteromita globosa* (Mattison and Harayama, 2001), in another study, the grazing by the ciliate *Colpidium colpoda* enhanced the degradation of crude petroleum in batch culture (Rogerson and Berger, 1983). Still, the abundance and diversity of protozoan communities in hydrocarbon contaminated and polluted environments in general are poorly documented (Rogerson and Berger, 1981). However, the most crucial observation regarding soil Amoeba brings to prominence the fact that neither crude oil nor any kind of soil pollution could bring about a significant reduction in the diversity and abundance of communities of soil amoeba. The study performed here thus reinforces this claim that soil amoeba plays a major role in metal tolerance and takes part in mitigation of soil pollution.

The OTU abundance from the sequencing the DNA was found to be comparatively lesser than the OTUs obtained from the amplicon sequencing of cDNA prepared from RNA transcripts obtained from polluted soil. This method of amplicon sequencing enabled better understanding of the community functioning and directly compared the subset of the active population from the total microbial population. In a study reported by De Vrieze et al. (2018), a significantly higher proportion of functional population of archaea was found to survive and function when amplicon sequencing of RNA and DNA was carried out. A comparative study indicated the survival of methanogenic archaea rather than bacteria in the anaerobic digestion plant (Ju and Zhang 2015).

As a consequence, the 16S rRNA gene-based profiling approach could provide a broad indication of the community presence, activities and potential performance which are considered valuable when studying a particular dynamic ecosystem (De Vrieze et al. 2018). The dynamic nature of the soil ecosystem was studied in terms of biotic homogenization by Meyer et al. (2019). The effects of a rapid transition from rainforest to agricultural land on the microbial species were studied where results clearly indicate that the extent of community homogenization is larger in the RNA-inferred community than the DNA-inferred communities. It was observed that the RNA-inferred group monitors environmental variations and loses spatial structure as the environment changes whereas the DNA-inferred community did not respond to environmental variability to the same degree. The results indicate that complementing DNA-based surveys with RNA can provide insights into how microbial communities react to changes in the environment (Meyer et al. 2019). In the present study, the microbial communities which were active and dominant in the polluted soil were captured from the cDNA-based amplicon study which clearly showed a variation from the total diversity of microbial community which was captured by the DNA-based amplicon study. The functionally active and dominant organisms present in the soil ecosystem were Filamoeba, Fungi and Metazoa where significant representations of *Filamoeba-sp.-H9a_6E, Echinamoebida, Dactylopodida, Oligochaeta, Glomeromycetes,Vampyrellidae* and a considerable percentage of uncultured eukaryotes inhabiting the polluted soil. The major group of organisms which were derived from the genomic DNA-based amplicons was from the Metazoa order. It was deduced from the genomic data that a large section of the soil was inhabited by *Chromadorea* and *Tylenchida* from phylum Nematode and *Collembola* from phylum Arthropoda. The major taxonomic distribution analyzed for both the cDNA and genomic DNA is represented in Table 4.16 and Table 4.17 and Figure 4.54 to Figure 4.63.

Table 4.16: Major eukaryotic taxonomic distribution in cDNA

Phylum	Total OTUs	Class	Total OTUs	Order	Total OTUs	Family	Total OTUs	Genus	Total OTUs
Amoebozoa	3765	Gracilipodida	1107	Metazoa	969	Filamoeba-sp.-H9a_6E	563	Filamoeba-sp.-H9a_6E	563
Opisthokonta	2108	Holozoa	996	Fungi	935	Amphiesmenoptera	268	uncultured-eukaryote	306
SAR	943	Nucletmycea	939	Filamoeba	642	Dactylopodida	263	Lepidoptera	268
Proteobacteria	428	Myxogastria	861	Cercozoa	494	uncultured-eukaryote	263	Vermamoeba	215
Actinobacteria	281	Rhizaria	494	Flabellinia	297	Echinamoebida	248	Glomerales	163
Acidobacteria	169	Tubulinea	402	Arcellinida	259	Vampyrellidae	182	Uncultured	158
Archaeplastida	115	Discosea	334	Ciliophora	214	Glomeromycetes	165	Tylenchida	99
Excavata	48	Alveolata	258	Flamella	164	Chromadorea	141	uncultured-Eimeriidae	86
Chloroflexi	43	Alphaproteobacteria	256	LEMD267	147	Incertae-Sedis	122	Haplotaxida	80
Thaumarchaeota	36	Schizoplasmodiida	197	Rhizobiales	123	BOLA868	92	uncultured-Amoebozoa	77

Table 4.17: Major eukaryotic taxonomic distribution in genomic DNA

Phylum	Total OTUs	Class	Total OTUs	Order	Total OTUs	Family	Total OTUs	Genus	Total OTUs
Opisthokonta	2008	Holozoa	1007	Ciliophora	4862	Telaepolella-sp.-Tib190	6586	Filamoeba-sp.-H9a_6E	563
Amoebozoa	903	Nucletmycea	870	Arcellinida	993	Catenulida	288	uncultured-eukaryote	306
SAR	697	Gracilipodida	639	uncultured-Amoebozoa	867	Ancylistaceae	164	Lepidoptera	268
Archaeplastida	385	Chloroplastida	385	Myxococcales	322	Polysphondylium-violaceum	143	Vermamoeba	215
Centrohelida	19	Alveolata	372	Metazoa	288	Bdelloidea	140	Glomerales	163
Acidobacteria	6	Rhizaria	254	Cercozoa	254	Glissomonadida	133	uncultured	158
Excavata	4	Discosea	112	Unknown	250	Trichosporonaceae	126	Tylenchida	99
Thaumarchaeota	3	Incertae-Sedis	46	Nitrososphaerales	158	Echinamoebida	118	uncultured-Eimeriidae	86
Proteobacteria	3	Stramenopiles	38	Arboramoeba	157	Haptoria	102	Haplotaxida	80
				Fungi	103	Dorylaimia	99	uncultured-Amoebozoa	77

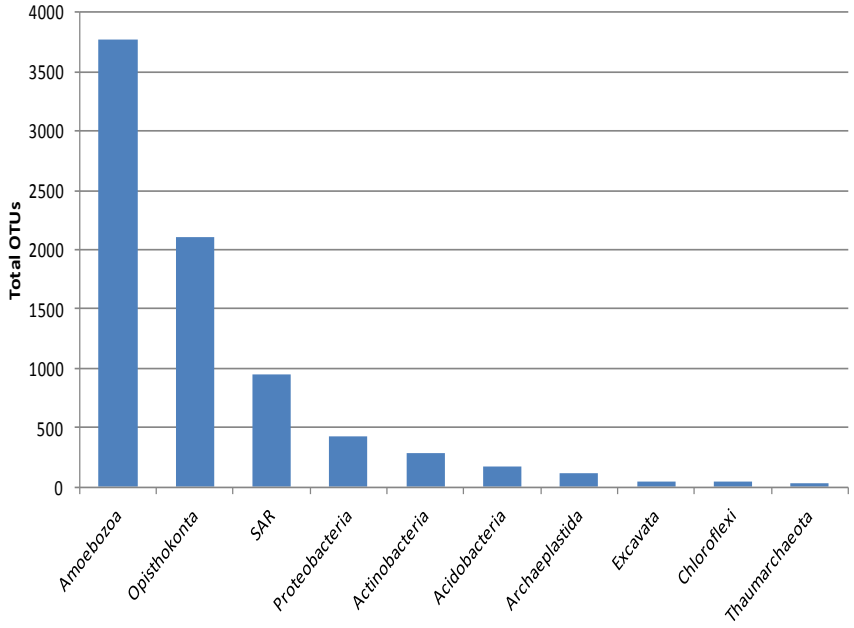


Figure 4.54: Major eukaryotic taxonomic distribution in cDNA at Phyllum level

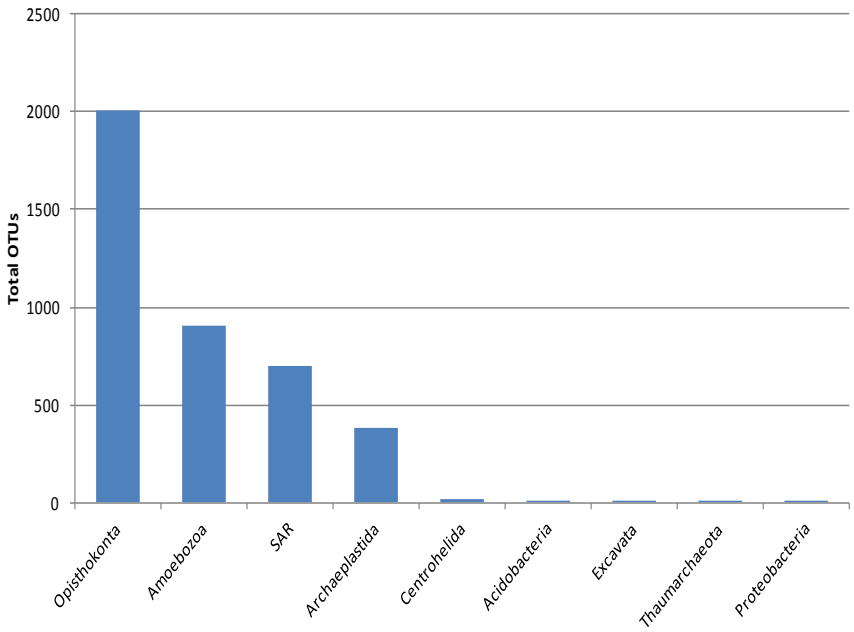


Figure 4.55: Major eukaryotic taxonomic distribution in genomic DNA at Phyllum level

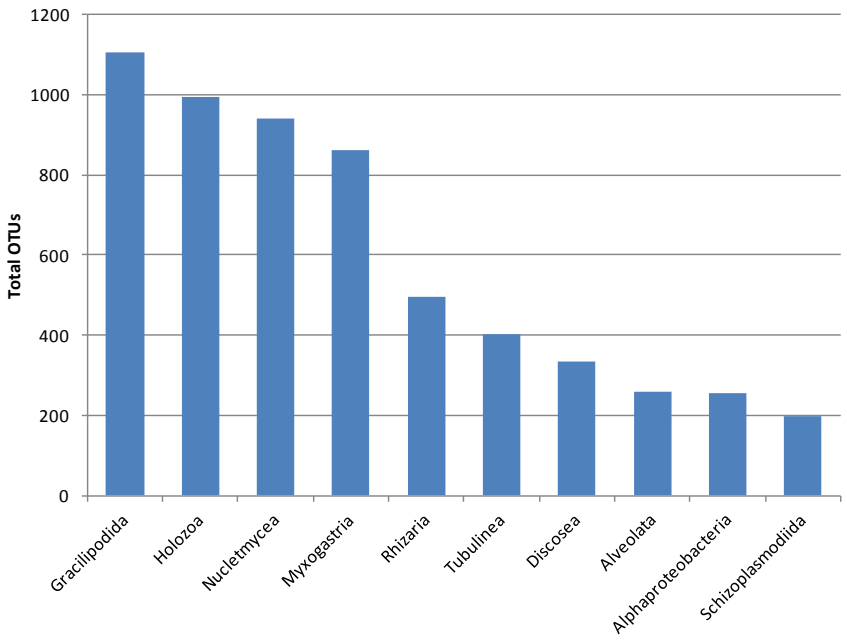


Figure 4.56: Major eukaryotic taxonomic distribution in cDNA at class level

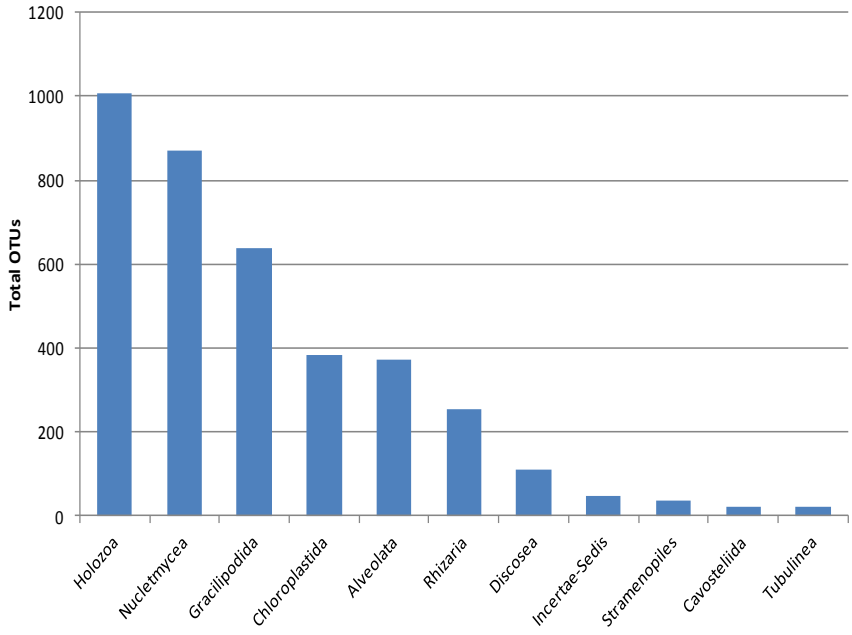


Figure 4.57: Major eukaryotic taxonomic distribution in genomic DNA at class level

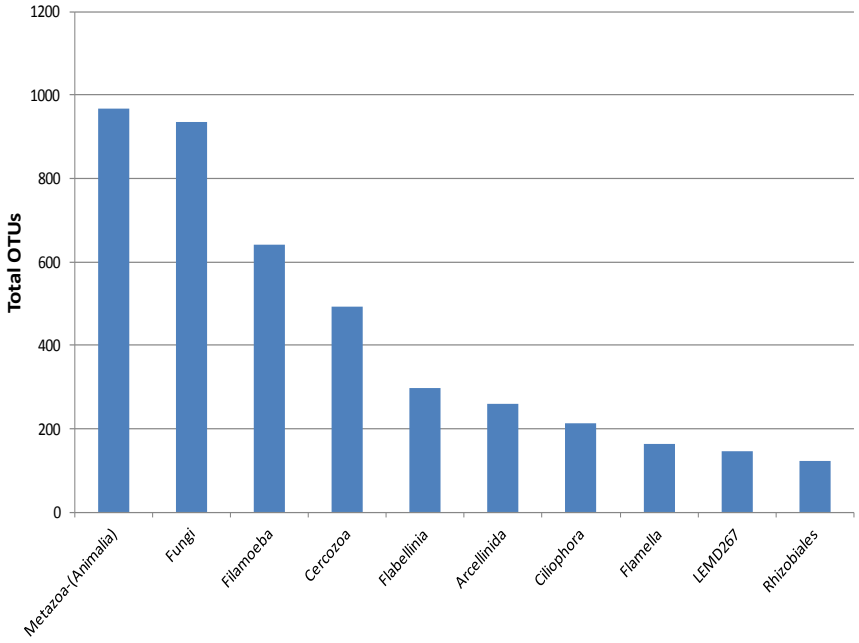


Figure 4.58: Major eukaryotic taxonomic distribution in cDNA at order level

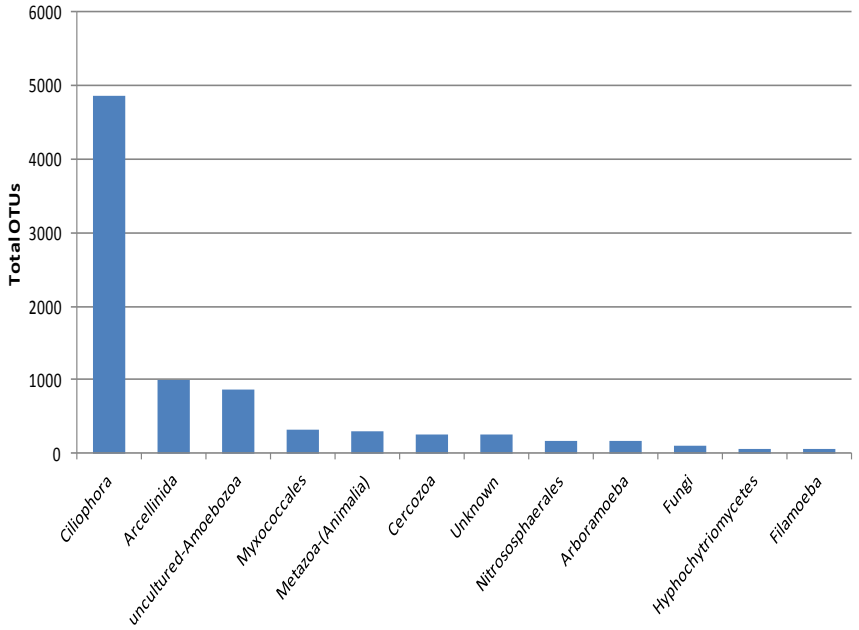


Figure 4.59: Major eukaryotic taxonomic distribution in genomic DNA at order level

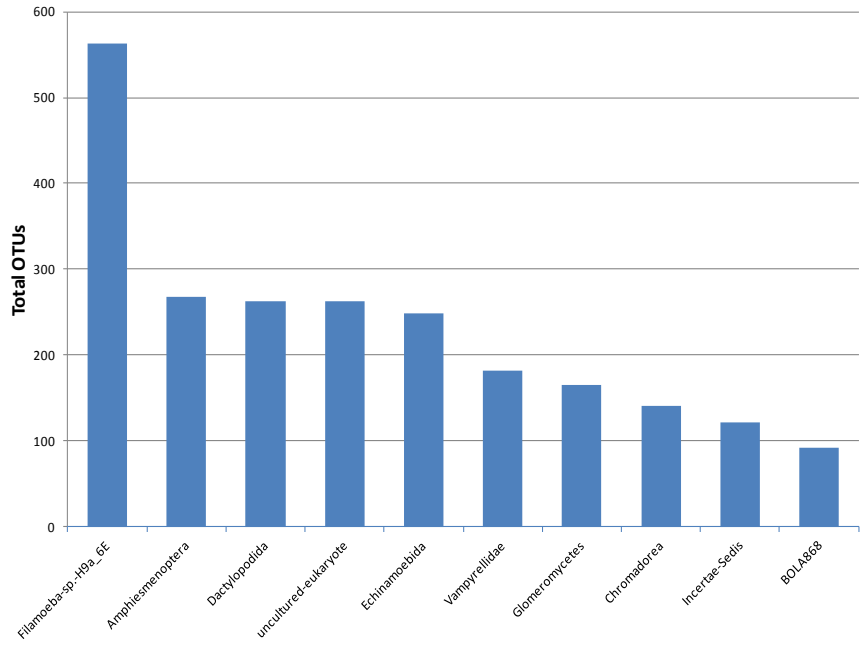


Figure 4.60: Major eukaryotic taxonomic distribution in cDNA at family level

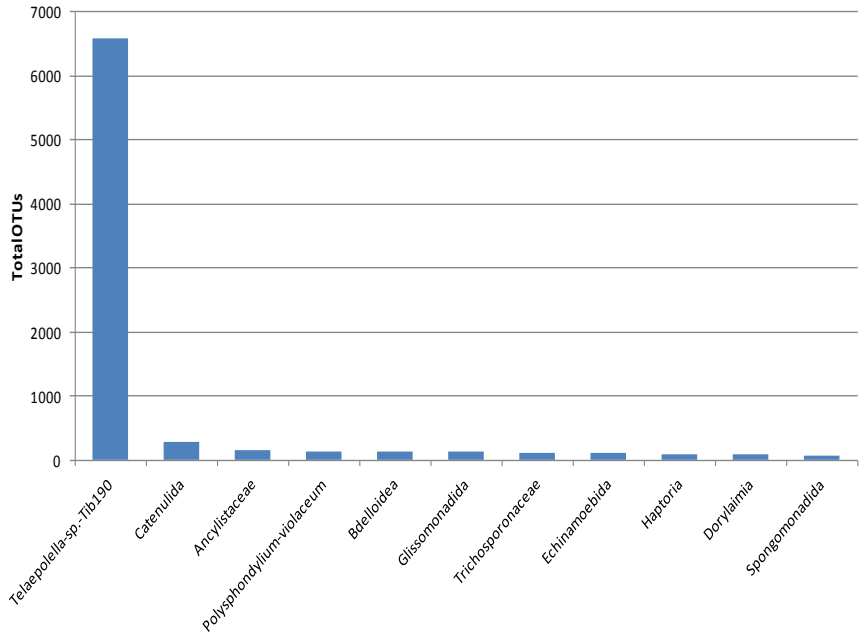


Figure 4.61: Major eukaryotic taxonomic distribution in genomic DNA at family level

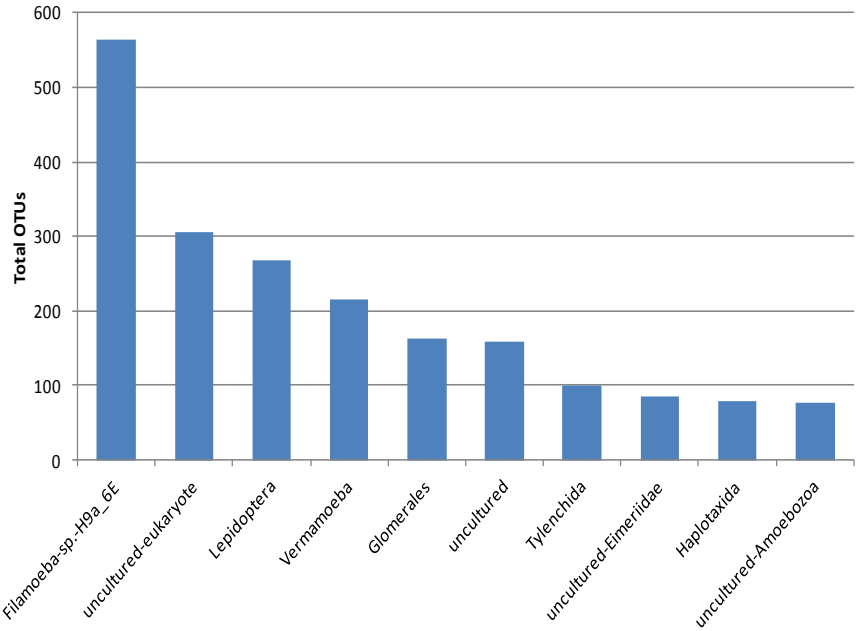


Figure 4.62: Major eukaryotic taxonomic distribution in cDNA at genus level

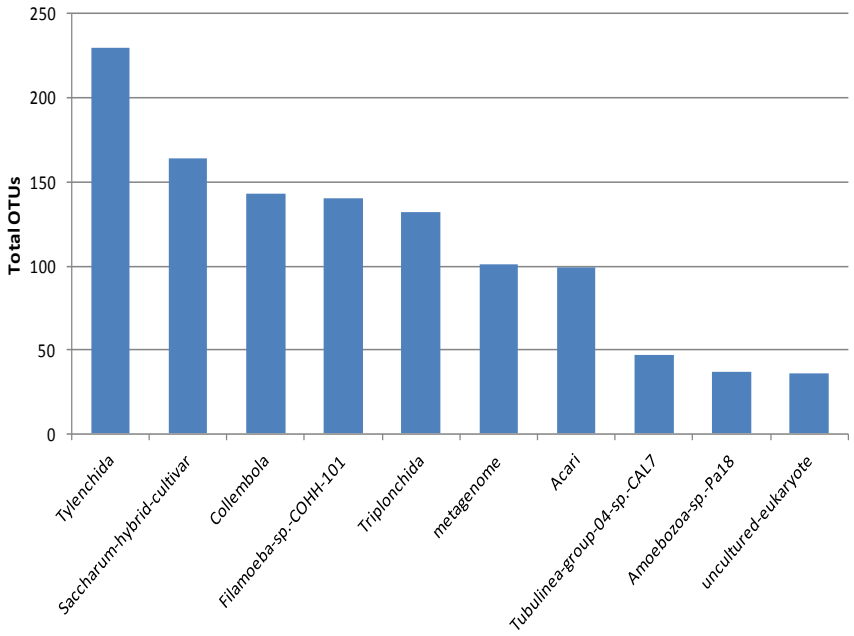


Figure 4.63: Major eukaryotic taxonomic distribution in genomic DNA at genus level

4.17 Alpha diversities with samples and rarefaction curves

In this section, eukaryotic community diversity within each sample was calculated by Shannon index, Chao1 and observed species metrics. The Shannon index is the measure the observed OTU abundances and the Chao1 metric estimates species richness and evenness. The number of unique OTUs measured in a sample is represented by observed species metric. The rarefaction curve for each of the metric is provided in Figure 64 to Figure 69. The metric calculation was performed using QIIME software.

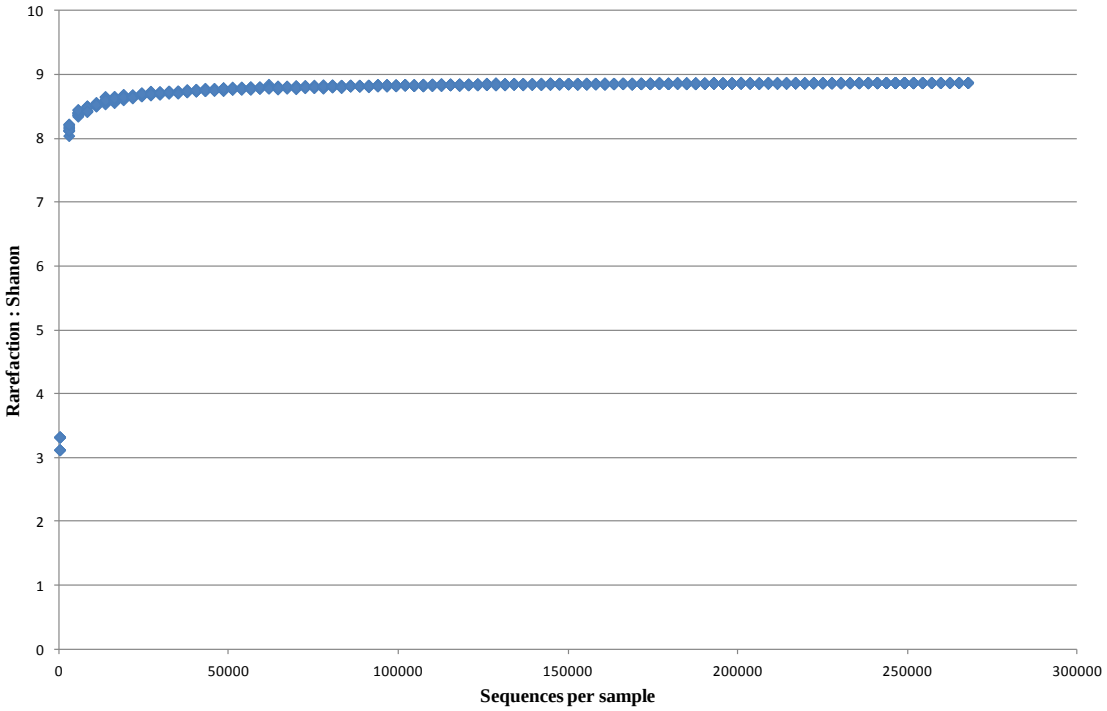


Figure 4.64: Shanon curve obtained from the sample cDNA

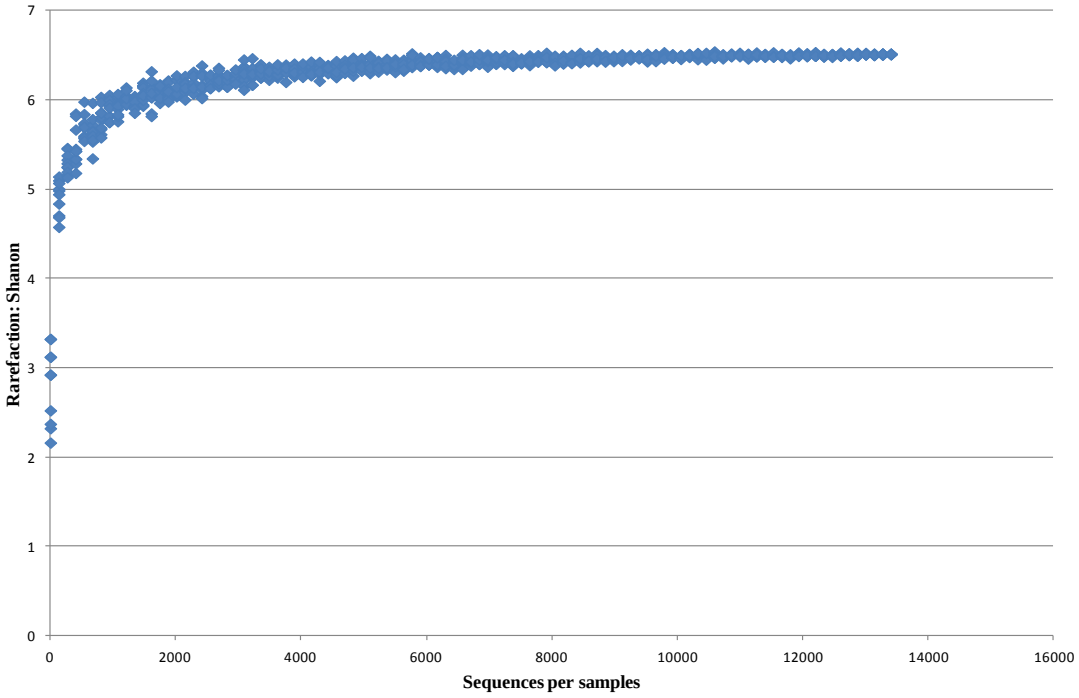


Figure 4.65: Shanon curve for sample genomic DNA

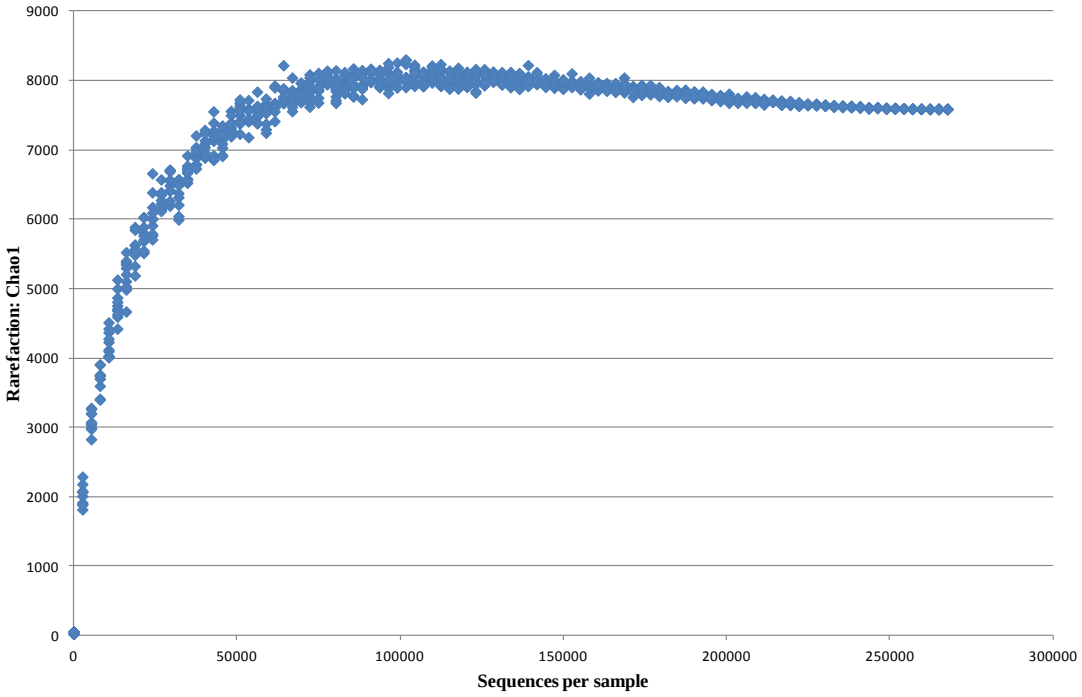


Figure 4.66: Chao1 curve obtained for the sample cDNA

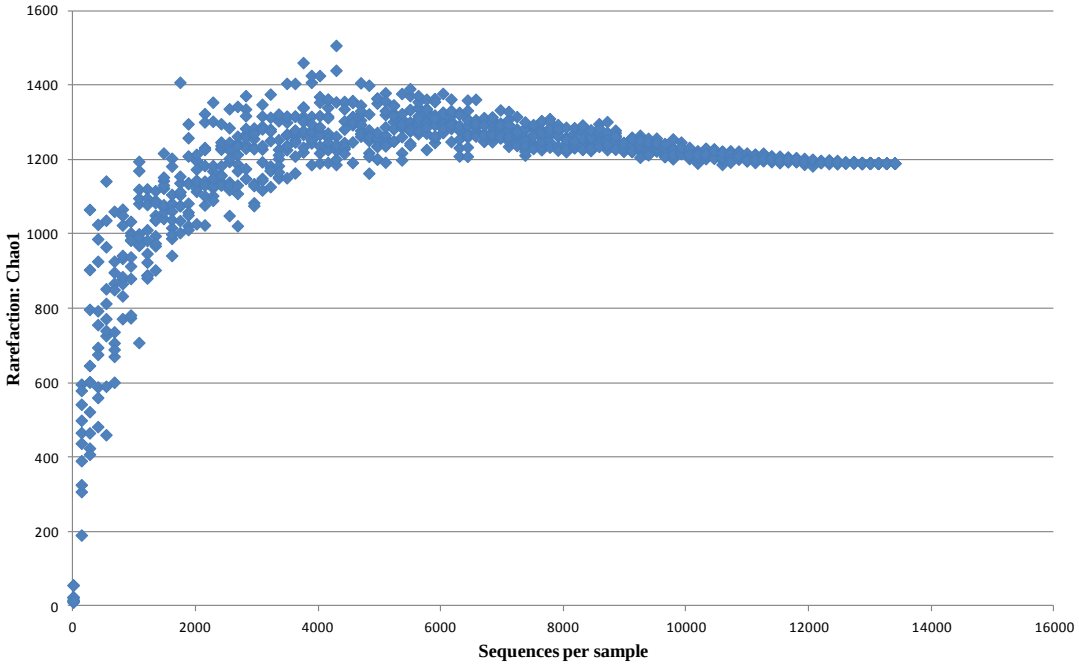


Figure 4.67: Chao1 curve obtained for the sample genomic DNA

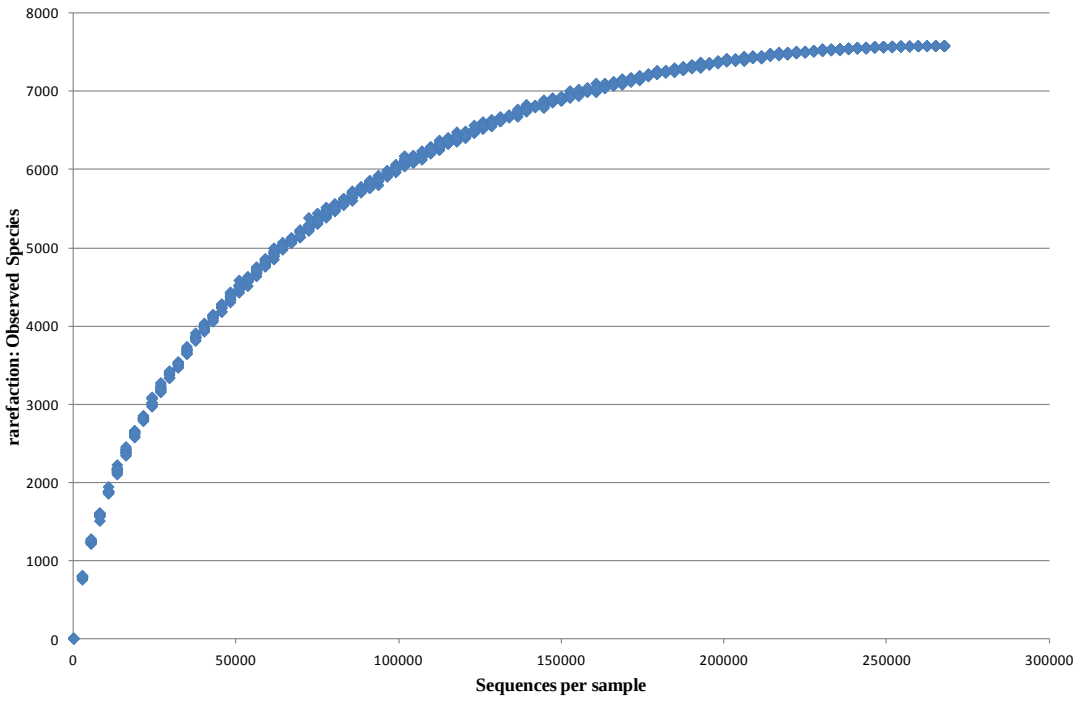


Figure 4.68: Observed species for the sample cDNA

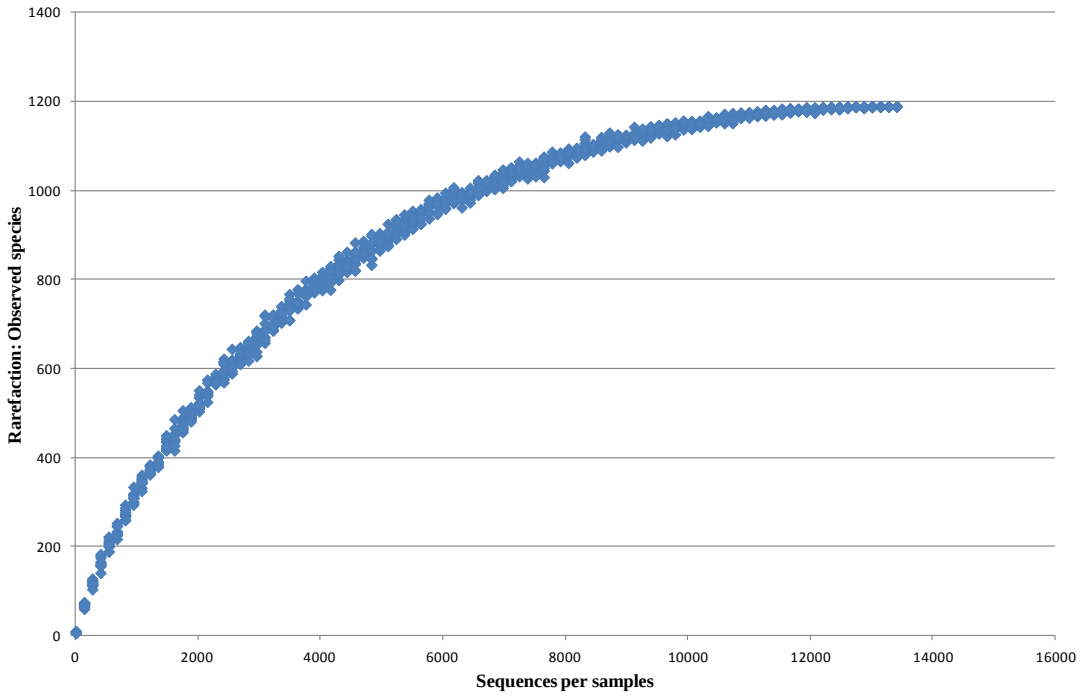


Figure 4.69: Observed species curve obtained for sample genomic DNA

Summary of various diversity indices is shown in Table 4.18

Table 4.18: Summary of various Diversity Indices

Diversity Index	**cDNA**	**Genomic DNA**
Chao1	7358.672	1182.604
Shanon	8.755	6.294
PD whole tree	293.986	72.37426
Observed species	5973.729	894.2505

Values are average of all the samples

The consequences of metal contamination of the soil can play a key role in determining the structuring of microbial communities and also aid in their functional adaptations. A metagenomic approach was employed to analyze the eukaryotic diversity of copper mining deposits to assess the actively transcribing population of microorganisms as well as the total

eukaryotic microbial population in general. To estimate the diversity within a sample, alpha diversity indices like Shannon indices, Chao1 and observed species were calculated using the rarefaction sampling. A high Shannon index of 8.5 and above was obtained for most of the samples in the cDNA which describes the richness of the sample in terms of biodiversity. However, in case of genomic DNA, the Shannon index was around 6.5. Chao1 rarefaction as well as observed species was lower in genomic DNA in contrast to cDNA samples. This might be due to the lower number of eukaryotic sequences that were represented in the genomic DNA sample as compared to the cDNA prepared from the total RNA of the soil sample, where a representation of the actively transcribing microbial population is taken into account.

Noticeable and signature eukaryotic communities and the hypothesized metal tolerance activities signify, that the microbial biodiversity in metal contaminated sites has a selective advantage in metal rich conditions by expression of certain genes which aid in the surviving in such toxic conditions. The eukaryotic diversity of metal contaminated sites has not yet been explored in depth. This study provides useful information about the mechanism of metal tolerance and which can be linked with the community functions, thus unearthing of rare genes and industrially important mechanism of metal tolerance.

Chapter 5: Summary

To understand the complex and dynamic nature of the functions of soil microbiota because of metal contamination, it is essential to focus on the expressed subset of total genes. Metatranscriptomics helps us to understand the responses of microbial population and the services performed by them with respect to the environmental modification or perturbation. The survival strategies in toxic metal polluted soil have encouraged taking up the present study and developing holistic approaches to discover new and underrepresented yet important eukaryotic microbial functions.

In view of this, an efficient way for constructing cDNA libraries of eukaryotic microbes residing in metal polluted soil samples was implemented in this study. In the first stage, total RNA was extracted from soil and cDNA library was synthesized utilizing the SMART (Switching Mechanism At 5' end of RNA) technique to achieve full-length enriched cDNA with large insert size of the library. cDNA libraries were constructed and screened using appropriate expression vectors by functional complementation of eukaryotic yeast cells to isolate and later characterize novel genes involved in providing tolerance to toxic metals.

Soil samples were collected from open cast Malanjkhand copper mine, which is located in Malanjkhand, Madhya Pradesh, India. Soil samples from a total of 15 different locations from the mining area and 2 from the tailing dump site were collected and were mixed to form a composite sample. Total RNA was isolated from the soil sample and cDNA was constructed by Mint-2 cDNA synthesis kit (Evrogen, Moscow). cDNAs were fractionated into different sizes (A < 0.5 kb, B 0.5–1.0 kb, and C 1.0–4.0 kb) by 2-dimensional agarose gel electrophoresis. Size fractionated cDNA were digested with SfiI restriction enzyme before ligation into yeast shuttle

plasmid vector pFL61 modified with the addition of SfiI restriction enzyme recognition sites. The sites are placed under the control of Phosphoglycerate kinase (PGK) promoter and its terminator from *S. cerevisiae*.

Plasmid DNA was isolated from the *E. coli* cells to screen the cDNA libraries (library A, B and C). Initially, the cDNA libraries were screened for genes involved in providing copper tolerance by functional complementation using the copper sensitive mutant *cup1*$^{\Delta}$ which is sensitive to copper at 150 μM on Synthetically Defined medium without uracil (SD-Ura). More than five times the original size of the libraries (10^6 clones) was screened. Approximately, 160 Cu tolerant yeast transformants were chosen from library C out of which 155 were tested on Cu amended media by drop assay. Seventy seven clones were identified and categorized into highly tolerant, 20 into moderately tolerant and 58 in low tolerant groups based on the appearance of growth in drop assay according to increasing dilutions. For library B, a total of 1 x 10^6 clones were screened and 17 clones were identified which exhibited high tolerance and 5 and 3 clones were categorized in moderately tolerant and low tolerant groups respectively. A total of 1.2 x 10^6 clones were screened for library A where a total of 52 Cu tolerant clones were identified. 23 clones were categorized into highly tolerant, 15 and 10 into moderately tolerant and low tolerant groups.

Some of the clones showing tolerance to Cu were further sequenced and analyzed for its homology and characterized further. Apart from these, the clones rescued from contaminated soil from France, were further characterized for its potential role in providing tolerance to Copper, Cadmium, Zinc and Cobalt.

One of the clones from library C of metal contaminated soil from France; PLCc38 showed high tolerance to Cd was analyzed. Sequence study indicated that PLCc38 share common ancestry

from protista kingdom, phylum Ciliophora. It encodes a 239 amino acid protein showing homology to aldehyde dehydrogenase. Aldeyde dehydrogenases are enzymes that can remove toxic aldehydes caused by abiotic environmental stress. like reactive oxygen species (ROS) and toxic levels of metals. The responses towards increasing concentrations of different heavy metals were recorded for ALDH and were compared to the metal resistant wild types. Most common forms of metal pollutants like Cu, Cd, Co and Zn were tested against them and tolerance to all the metals was shown by the mutant yeast cells. The cDNA PLCc38 was able to confer tolerance to 150 μM of Cu, 40 μM of Cd, 10 mM of Zn and 2 mM of Co. The expression of the gene was found to accumulate metals inside the cells higher when compared to the control. Accumulation was highest at 500 μM of Cu, 40 μM of Cd, 11 mM of Zn and 10 mM of Co. Furthermore, the role of PLCc38 in response to oxidative stress was investigated and it was able to provide tolerance to oxidative stress sensitive yeast mutant *sod2*$^{\Delta}$ against high concentrations of H_2O_2 in the external growth media.

The cDNA PLCg49 obtained from the library C of metal contaminated France soil consisted of 1409 bp with an ORF encoding a 410 amino acid protein and was found to originate from the serpin family of *Hypsibius dujardini* (PLCg49). Serpins are a large family of protease inhibitors that are mostly against reducing stresses that are usually caused by biotic and abiotic factors and help maintain cell viability. It was found to be efficiently tolerating high concentrations of metals like Cu, Cd, Zn and Co when transformed into metal sensitive yeast mutants. Compared to the wild type yeast strain, the transformed mutants depicted higher growth response in presence of different concentrations of metals. Highest growth was recorded at 150 μM of Cu, 40 μM of Cd, 10 mM of Zn and 2 mM of Co. Also accumulation studies proved that serpin was accountable for intracellular uptake of heavy metals significantly higher as compared to wild type. Cellular

uptake of metals was highest at 300 μM of Cu, 40 μM of Cd, 10 mM of Zn and 2 mM of Co. This revealed the importance of the serine protease inhibitor as a stress response protein. The function of serpin as a protease inhibitor was evident from the experiment where it was able to impart protection to the yeast cells against two different serine proteases, trypsin and proteinase K. Mutant yeast *pbi2*$^{\Delta}$ with a deleted serine protease inhibitor was transformed with the cDNA serpin and the gene was able to survive and grow in presence of the proteases.

The cDNA PLCg39 isolated from metatranscriptomic library C of metal contaminated soil from France is a zinc finger protein which also demonstrated enhanced tolerances towards different heavy metals particularly zinc. Sequence study and bioinformatic analysis revealed the cDNA had an open reading frame of 966 bp encoding a polypeptide of 321 amino acids showing close homology with Amoebozoa, *Acanthamoeba castellanii*. Conserved regions were found for a motif of size 48 residues which constituted the zinc finger and the DHHC palmitoyl transferase domain $CX_2CX_9HCX_2CX_2CX_4DHHCX_5CX_4NX_3FX_4$. DHHC pamitoyl transferases, which are responsible for addition of pamitoyl groups to proteins, usually possess metal coordination domains. This gene was used to transform metal sensitive mutant yeast cells and the growth response was studied in presence of different metals (Cu, Zn, and Cd). The cDNA PLCg39 was able to withstand 500 μM of Cu, 40 μM of Cd and 10 mM of Zn. Accumulation of metals was highest at 150 μM of Cu, 40 μM of Cd and 10 mM of Zn. The cDNA PLCg39 was also used to transform mutant strain *akr1*$^{\Delta}$ (deleted of a DHHC palmitoyl transferase gene) to understand the mechanism behind metal tolerance by DHHC palmitoyl transferase. It was observed that the cDNA was able to support the growth of the mutant fairly well up to 500 μM of Cu.

Sequence investigation of cDNA ARK3 revealed that the cDNA is 523 bp long with an open reading frame of 375 bp encoding a polypeptide of length 124 amino acids. This was homologous to the Maltose fermentation regulatory protein (MALR) MAL33 of *S. cerevisiae*. MALR genes are zinc–finger type transcription factor that positively regulate the fermentation of maltose. When tested for the capability of the cDNA to confer tolerance to Cu sensitive yeast *cup1*$^{\Delta}$, it was found that it was able to grow at 300 μM of Cu in the media. However the growth decreased with increase in copper concentration. For all the concentrations of Cu used in the study, the response in terms of growth for the mutant transformed with ARK3 was significantly more as compared to its wild type BY4741 which is naturally tolerant towards metals.

The cDNA ARK1 obtained from the library C of copper contaminated site was characterized in this study. The size of ARK1 cDNA was 849 bp with an ORF of size 657 encoding a protein of 218 amino acids. Sequence analysis revealed homology with chitin synthase of the yeast *S. cerevisiae*. Chitin synthase catalyses the formation of *N*-acetylglucosamine which is a derivative of glucose and a primary component of cell walls in fungi. Aside from its usual activity, a novel function of this enzyme was inferred from this study. The cDNA was able to provide tolerance towards Cu which was investigated by performing drop assay and studying its growth response in presence of increasing concentrations of Cu (0-1000 μM). It was found out that the cDNA could rescue the metal sensitive mutant yeast and support its growth even at 500 μM of Cu. The growth of the cells were significantly higher compared to the growth of the wild type BY4741, indicating its potential role in Cu toxicity.

Eukaryotic functional diversity study of copper polluted soil revealed that the genes isolated in this study responsible for conferring metal tolerance have clearly originated from organisms belonging to the fungi, metazoan and filamoeba order. Therefore, to establish the link between

these organisms having metal tolerant functions and the metabolically active eukaryotic population inhabiting the metal polluted soil, a functional diversity study was performed. This was achieved by Next Generation Sequencing followed by metagenomic analysis. The diversity was evaluated by two approaches. First, by constructing cDNA from total RNA and second, by isolation of metagenomic DNA from soil. In both the methods, amplicon sequencing was employed. The V4 region of the 18S rRNA gene from both the cDNA and genomic DNA was first amplified with eukaryotic universal degenerate primers and the amplicons were then sequenced on Illumina HiSeq sequencing platform. The amplicon size ranged between 0.4 and 0.5 kb. About 0.5 million raw reads of 250 bp were processed to extract information of the diversity of the soil. This produced a total of 20121 and 8592 Operational Taxonomic Units (OTUs) for the cDNA and genomic DNA respectively. They were further processed which lead to taxonomic assignments. The eukaryotic community was found to be distributed in 29 known phyla with the highest and most prominent being phyla *Amoebozoa* (3765 OTUs) followed by *Opisthokonta* (2108), SAR (943) (*Stramenopiles* (*heterokonts*), *Alveolates*, and *Rhizaria*).

The genes isolated in this study have a potential to be used in various industrial applications. The importance of the gene aldehyde dehydrogenase in bioremediation has been proved in a number of studies concerning removal from toxic chemicals from industrial effluent streams. It plays an important role as a detoxifying enzyme for mitigatng the adverse effects of aldehydes typically found in the environment, especially in industrial wastewaters (Jaureguibeitia et al. 2007). Aldehyde dehydrogenases can transform cytotoxic aldehydes into organic acids and the bioprocess is environmentally friendly, which replaces current chemical processes (Peng et al.2005). Additionally, in ethanol production, a vital step in microbial fermentation is the oxidation of aldehyde dehydrogenase (Jing et al. 2013). An aldehydes dehydrogenase gene

isolated from soil dwelling *Sphingomonas sp*. is used to biotransform a variety of aromatic aldehydes to their carboxylic acids thus proving its usefulness in industrial applications as an essential enzyme (Peng et al. 2005). As already mentioned, the primary role of the enzyme is inhibiting oxidative stress. In the field of medical science, the enzyme has been shown to protect against renal ischemia/reperfusion injury and further investigations revealed that it could reduce spinal cord reperfusion injury (Liu et al. 2017). This enzyme is a primary target for producing crop sturdy crop plants which are resistant to high salinity, dehydration, oxidative stress, or heavy metals, suggesting important roles in environmental adaptation (Zhao et al. 2017).

Future work

From each of the metatranscriptomic library A, B and C, several clones were identified and isolated. Each and every clone has a plasmid bearing a gene which has a role to play in metal tolerance. Systematic sequencing of the recombinant plasmid bearing clones for investigating specific sized proteins is an ongoing process and could produce a large set of data, which may link several proteins involved in the tolerance and management of toxic concentrations of metals in the environment. This will help us understand the network of genes and their associated functions as a collective endeavor to mitigate metal toxicity.

The genes need further characterization. Characterization involves cloning of the isolated genes in appropriate expression vectors and propagating into host bacterial cells for protein production followed by purification. This could lead to the understanding of the protein functions in a better manner. The interaction between the purified protein molecules and metal ions could help understand the mechanism of metal tolerance much clearly.

The isolated genes may have enormous potential in effective planning and preparation of transgenic crop plants. The introduction and over-expression of these genes could achieve abiotic

stress tolerance in the transgenic crops. Controlled microcosm based studies on metal accumulation and tolerance by plants transformed with the isolated genes from this study could help to formulize better bioremediation strategies.

References

Abdu, N., Abdullahi, A.A. and Abdulkadir, A., 2017. Heavy metals and soil microbes. Environ. Chem. Lett. 15, 65-84.

Abraham, J.S., Sripoorna, S., Maurya, S., Makhija, S., Gupta, R. and Toteja, R., 2019. Techniques and tools for species identification in ciliates: a review. Int. J. Syst. Evol. Microbiol. 69, 877-894.

Adriaensen, K., Vrålstad, T., Noben, J.P., Vangronsveld, J. and Colpaert, J.V., 2005. Copper-adapted Suillus luteus, a symbiotic solution for pines colonizing Cu mine spoils. Appl. Environ. Microbiol. 71, 7279-7284.

Agarwal, S., Macfarlan, T.S., Sartor, M.A. and Iwase, S., 2015. Sequencing of first-strand cDNA library reveals full-length transcriptomes. Nat. Commun. 6, 1-12.

Alessio, J.M. and Gerard, G.F., 1988. Second-strand cDNA synthesis with E. coli DNA polymerase I and RNase H: the fate of information at the mRNA 5'terminus and the effect of E. coli DNA ligase. Nucleic acids Res. 16, 1999-2014.

Allison, S.D. and Martiny, J.B., 2008. Resistance, resilience, and redundancy in microbial communities. Proc. Natl. Acad. Sci. 105(Supplement 1), 11512-11519.

Anderson, R.C., Lindauer, K.R. and Prescott, D.M., 1996. A gene-sized DNA molecule encoding heat-shock protein 70 in Oxytricha nova. Gene, 168, 103-107.

Atafar, Z., Mesdaghinia, A., Nouri, J., Homaee, M., Yunesian, M., Ahmadimoghaddam, M. and Mahvi, A.H., 2010. Effect of fertilizer application on soil heavy metal concentration. Environ Monit Assess. 160, p.83.

Atici, Ö., Ağar, G. and Battal, P., 2005. Changes in phytohormone contents in chickpea seeds germinating under lead or zinc stress. Biol. Plant. 49, 215-222.

Auld, R.R., Mykytczuk, N.C., Leduc, L.G. and Merritt, T.J., 2017. Seasonal variation in an acid mine drainage microbial community. Can. J. Microbiol. 63, 137-152.

Auld, R.R., Myre, M., Mykytczuk, N.C., Leduc, L.G. and Merritt, T.J., 2013. Characterization of the microbial acid mine drainage microbial community using culturing and direct sequencing techniques. J. Microbiol. Methods 93, 108-115.

Azizan, M.S., Zamani, A.I., Stahmann, K.P. and Ng, C.L., 2016. Fungal metabolites and their industrial importance: A brief review. Malaysian J. Biochem. Mol. Biol. 19, 15-23.

Azizi, G., Akodad, M., Baghour, M., Layachi, M. and Moumen, A., 2018. The use of Mytilus spp. mussels as bioindicators of heavy metal pollution in the coastal environment. A review. J. Mater. Environ. Sci. 9, 1170-1181.

Bailly, J., Fraissinet-Tachet, L., Verner, M.C., Debaud, J.C., Lemaire, M., Wésolowski-Louvel, M. and Marmeisse, R., 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME J. 1, 632.

Ballatori, N., 2002. Transport of toxic metals by molecular mimicry. Environ. Health Perspect., 110(suppl 5), 689-694.

Bashiardes, S., Zilberman-Schapira, G. and Elinav, E., 2016. Use of metatranscriptomics in microbiome research. Bioinform. Biol. Insights, 10, BBI-S34610.

Bellion, M., Courbot, M., Jacob, C., Guinet, F., Blaudez, D. and Chalot, M., 2007. Metal induction of a Paxillus involutus metallothionein and its heterologous expression in Hebeloma cylindrosporum. New Phytol., 174, 151-158.

Bhanoori, M. and Venkateswerlu, G., 2000. In vivo chitin–cadmium complexation in cell wall of Neurospora crassa. Biochim. Biophys. Acta, 1523, 21-28.

Biddle, J.F., Fitz-Gibbon, S., Schuster, S.C., Brenchley, J.E. and House, C.H., 2008. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. Proc. Natl. Acad. Sci. U.S.A., 105, 10583-10588.

Binz, P.A. and Kägi, J.H., 1999. Metallothionein: molecular evolution and classification. In: Klaassen, C.D. (Ed.), Metallothionein IV. Advances in Life Sciences. Birkhäuser, Basel, 7-13.

Blazewicz, S.J., Barnard, R.L., Daly, R.A. and Firestone, M.K., 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. 7, 2061.

Blindauer, C.A., 2013. Lessons on the critical interplay between zinc binding and protein structure and dynamics. J. Inorg. Biochem. 121, 145-155.

Boeke, J.D., Trueheart, J., Natsoulis, G. and Fink, G.R., 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. In: Ray, W. and Grossman, L. (Eds.), Methods in enzymology. Academic Press, Cambridge, 164-175.

Borer, B., Tecon, R. and Or, D., 2018. Spatial organization of bacterial populations in response to oxygen and carbon counter-gradients in pore networks. Nat. Commun. 9, 1-11.

Botero, L.M., D'imperio, S., Burr, M., McDermott, T.R., Young, M. and Hassett, D.J., 2005. Poly (A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. Appl. Environ. Microbiol. 71, 1267-1275.

Bowen, A.R., Chen-Wu, J.L., Momany, M., Young, R., Szaniszlo, P.J. and Robbins, P.W., 1992. Classification of fungal chitin synthases. Proc. Natl. Acad. Sci. USA. 89, 519-523.

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P. and Boeke, J.D., 1998. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast, 14, 115-132.

Brady, D., Stoll, A.D., Starke, L. and Duncan, J.R., 1994. Chemical and enzymatic extraction of heavy metal binding polymers from isolated cell walls of Saccharomyces cerevisiae. Biotechnol. Bioeng. 44, 297-302.

Bragalini, C., Ribiere, C., Parisot, N., Vallon, L., Prudent, E., Peyretaillade, E., Girlanda, M., Peyret, P., Marmeisse, R. and Luis, P., 2014. Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. DNA Res. 21, 685-694.

Brim, H., Heyndrickx, M., De Vos, P., Wilmotte, A., Springael, D., Schlegel, H.G. and Mergeay, M., 1999. Amplified rDNA restriction analysis and further genotypic characterisation of metal-resistant soil bacteria and related facultative hydrogenotrophs. Syst. Appl. Microbiol. 22, 258-268.

Brodie, E.L., DeSantis, T.Z., Parker, J.P.M., Zubietta, I.X., Piceno, Y.M. and Andersen, G.L., 2007. Urban aerosols harbor diverse and dynamic bacterial populations. Proc. Natl. Acad. Sci. U.S.A. 104, 299-304.

Brown, M.T. and Wilkins, D.A., 1985. Zinc tolerance of Amanita and Paxillus. T. Brit. Mycol. Soc. 84, 367-369.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E.V.L. and Schulze-Lefert, P., 2013. Structure and functions of the bacterial microbiota of plants. Annu. Rev. Plant Biol. 64, 807-838.

Buss, J.E. and Sefton, B.M., 1986. Direct identification of palmitic acid as the lipid attached to p21ras. Mol. Cell. Biol. 6, 116-122.

Buttigieg, P.L., Hankeln, W., Kostadinov, I., Kottmann, R., Yilmaz, P., Duhaime, M.B. and Glöckner, F.O., 2013. Ecogenomic perspectives on domains of unknown function: correlation-based exploration of marine metagenomes. PloS one, 8, p.e50869.

Cabib, E., Roberts, R. and Bowers, B., 1982. Synthesis of the yeast cell wall and its regulation. Annu. Rev. Biochem. 51, 763-793.

Cançado, G.M., Nogueira, F.T., Camargo, S.R., Drummond, R.D., Jorge, R.A. and Menossi, M., 2008. Gene expression profiling in maize roots under aluminum stress. Biol. Plant. 52, 475–485.

Capdevila, M and Atrian, S., 2011. Metallothionein protein evolution: a miniassay. J. Biol. Inorg. Chem. 16, 977-989.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I. and Huttley, G.A., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 7, p.335.

Carvalhais, L.C., Dennis, P.G., Tyson, G.W. and Schenk, P.M., 2012. Application of metatranscriptomics to soil environments. J. Microbiol. Meth. 91, 246-251.

Cervantes, C. and Gutierrez-Corona, F., 1994. Copper resistance mechanisms in bacteria and fungi. FEMS Microbiol. Rev. 14, 121-137.

Chao-Rong, G.E. and Zhang, Q.C., 2011. Microbial community structure and enzyme activities in a sequence of copper-polluted soils. Pedosphere, 21, 164-169.

Chaoui, A., Mazhoudi, S., Ghorbal, M.H. and El Ferjani, E., 1997. Cadmium and zinc induction of lipid peroxidation and effects on antioxidant enzyme activities in bean (Phaseolus vulgaris L.). Plant Sci. 127, 139-147.

Charron, M.J., Read, E., Haut, S.R. and Michels, C.A., 1989. Molecular evolution of the telomere-associated MAL loci of Saccharomyces. Genetics, 122, 307-316.

Chavez, C., Cruz-Becerra, G., Fei, J., Kassavetis, G.A. and Kadonaga, J.T., 2019. The tardigrade damage suppressor protein binds to nucleosomes and protects DNA from hydroxyl radicals. Elife, 8, p.e47682.

Cheeke, T.E., Phillips, R.P., Brzostek, E.R., Rosling, A., Bever, J.D. and Fransson, P., 2017. Dominant mycorrhizal association of trees alters carbon and nutrient cycling by selecting for microbial groups with distinct enzyme function. New Phytol. 214, 432-442.

Chen, J., Yang, L., Yan, X., Liu, Y., Wang, R., Fan, T., Ren, Y., Tang, X., Xiao, F., Liu, Y. and Cao, S., 2016. Zinc-finger transcription factor ZAT6 positively regulates cadmium tolerance through the glutathione-dependent pathway in Arabidopsis. Plant. Physiol. 171, 707-719.

Chenchik, A., Zhu, Y.Y., Diatchenko, L., Li, R., Hill, J., Siebert, P.D., 1998. Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. Gene cloning and analysis by RT-PCR. 305-319.

Choi, W.J. and Cabib, E., 1994. The use of divalent cations and pH for the determination of specific yeast chitin synthetases. Anal. Biochem. 219, 368-372.

Clement, B.G., Kehl, L.E., DeBord, K.L. and Kitts, C.L., 1998. Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. J. Microbiol. Methods, 31, 135-142.

Cohen, M., Davydov, O. and Fluhr, R., 2019. Plant serpin protease inhibitors: specificity and duality of function. J. Exp. Bot. 70, 2077-2085.

Coldsbrough, P., 1999. Metal tolerance in plants: the role of phytochelatins and metallothioneins. In: Terry, N. and Banuelos, G.S. (Eds.), Phytoremediation of contaminated soil and water. Lewis Publishers, USA, 221.

Colpaert, J.V., Wevers, J.H., Krznaric, E. and Adriaensen, K., 2011. How metal-tolerant ecotypes of ectomycorrhizal fungi protect plants from heavy metal pollution. Ann. For. Sci. 68, 17-24.

Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretaillade, E., Bayat, A.R., Shingfield, K.J., Peyret, P. and Forano, E., 2017. Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the rumen of dairy cow fed a mixed diet. Front. Microbiol. 8, p.67.

Conklin, D.J., Bhatnagar, A., Cowley, H.R., Johnson, G.H., Wiechmann, R.J., Sayre, L.M., Trent, M.B. and Boor, P.J., 2006. Acrolein generation stimulates hypercontraction in isolated human blood vessels. Toxicol. Appl.Pharmacol. 217, 277–288.

Conklin, D.S., McMaster, J.A., Culbertson, M.R. and Kung, C.H., 1992. COT1, a gene involved in cobalt accumulation in Saccharomyces cerevisiae. Mol. Cell. Biol. 12, 3678-88.

Cook, T., Gebelein, B. and Urrutia, R., 1999. Sp1 and its likes: biochemical and functional predictions for a growing family of zinc finger transcription factors. Ann. N. Y. Acad. Sci. 880, 94-102.

Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., Liu, L., Luo, Y., Liu, D., Zheng, H., Wang, J., Dong, Z., Hu, S. and Huang, L., 2015. Metatranscriptomic Analyses of Plant Cell Wall

Polysaccharide Degradation by Microorganisms in the Cow Rumen. Appl. Environ. Microbiol. 81, 1375-1386.

Dai, X., Xu, Y., Ma, Q., Xu, W., Wang, T., Xue, Y., and Chong, K., 2007. Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis. Plant. Physiol. 143, 1739-51.

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L. and Marmeisse, R., 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoS ONE, 7, p.e28967.

Damon, C., Vallon, L., Zimmermann, S., Haider, M.Z., Galeote, V., Dequin, S., Luis, P., Fraissinet-Tachet, L. and Marmeisse, R., 2011. A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. ISME J. 5, p.1871.

Daniel, R., 2002. Construction of Environmental Libraries for Functional Screening of Enzyme Activity. In: Brakmann, S. and Johnsson, K. (Eds.), Directed Molecular Evolution of Proteins: or How to Improve Enzymes for Biocatalysis. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 63-78.

Daniel, R., 2005. The metagenomics of soil. Nat. Rev. Microbiol. 3, 470-478.

Danzi, S.E., Bali, M. and Michels, C.A., 2003. Clustered-charge to alanine scanning mutagenesis of the Mal63 MAL-activator C-terminal regulatory domain. Curr. Genet. 44, 173-183.

De Vrieze, J., Pinto, A.J., Sloan, W.T. and Ijaz, U.Z., 2018. The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool. Microbiome, 6, p.63.

Degens, B.P., Schipper, L.A., Sparling, G.P. and Duncan, L.C., 2001. Is the microbial community in a soil with reduced catabolic diversity less resistant to stress or disturbance?. Soil Biol. Biochem. 33, 1143-1153.

Dell'Amico, E., Mazzocchi, M., Cavalca, L., Allievi, L. and Andreoni, V., 2008. Assessment of bacterial community structure in a long-term copper-polluted ex-vineyard soil. Microbiol. Res. 163, 671-683.

Deng, W., Xi, D., Mao, H. and Wanapat, M., 2008. The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. Mol. Biol. Rep. 35, 265-274.

Dietz, K.J., Bair, M., Kramer, U., 1999. Free radical and reactive oxygen species as mediators of heavy metal toxicity in plants. In: Prasad, M.N.V. and Hagemeyer, J. (Eds.), Heavy Metal stress in Plants from Molecules to Ecosystems. Spinger-Verlag, Berlin, 73–79.

Dixit, A.R., and Dhankher, O.P., 2011. A novel stress-associated protein 'AtSAP10'from Arabidopsis thaliana confers tolerance to nickel, manganese, zinc, and high temperature stress. PLoS. One. 6, e20921.

Domash, V.I., Sharpio, T.P., Zabreiko, S.A. and Sosnovskaya, T.F., 2008. Proteolytic enzymes and trypsin inhibitors of higher plants under stress conditions. Russ. J. Bioorg. Chem. 34, 318–322.

Dombrowski, J.E., 2003. Salt stress activation of wound-related genes in tomato plants. Plant Physiol. 132, 2098-2107.

Dominati, E., Patterson, M. and Mackay, A., 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol. Eng. 69, 1858-1868.

Doran, J.W. and Zeiss, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. Appl. Soil. Ecol. 15, 3-11.

Downing, W.L., Mauxion, F., Fauvarque, M.O., Reviron, M.P., De Vienne, D., Vartanian, N. and Giraudat, J., 1992. A Brassica napus transcript encoding a protein related to the Künitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. Plant J. 2, 685-693.

Dudley, A.M., Janse, D.M., Tanay, A., Shamir, R. and Church, G.M., 2005. A global view of pleiotropy and phenotypically derived gene function in yeast. Mol. Syst. Biol. 1, 1-11.

Edery, I., Chu, L.L., Sonenberg, N. and Pelletier, J., 1995. An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). Mol. Cell. Biol. 15, 3363-3371.

Eichner, C.A., Erb, R.W., Timmis, K.N. and Wagner-Döbler, I., 1999. Thermal gradient gel electrophoresis analysis of bioprotection from pollutant shocks in the activated sludge microbial community. Appl. Environ. Microbiol. 65, 102-109.

Emamverdian, A., Ding, Y., Mokhberdoran, F. and Xie, Y., 2015. Heavy metal stress and some mechanisms of plant defense response. ScientificWorldJournal, 1-18.

Enyenihi, A.H. and Saunders, W.S., 2003. Large-scale functional genomic analysis of sporulation and meiosis in Saccharomyces cerevisiae. Genetics. 163, 47-54.

Feng, G., Xie, T., Wang, X., Bai, J., Tang, L., Zhao, H., Wei, W., Wang, M. and Zhao, Y., 2018. Metagenomic analysis of microbial community and function involved in cd-contaminated soil. BMC Microbiol. 18, p.11.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O.V. and Golyshin, P.N., 2016. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. Microb. Biotechnol. 9, 22-34.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S. and Knight, R., 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microbiol. 73, 7059-7066.

Foulon, J., Zappelini, C., Durand, A., Valot, B., Girardclos, O., Blaudez, D. and Chalot, M., 2016. Environmental metabarcoding reveals contrasting microbial communities at two poplar phytomanagement sites. Sci. Total Environ. 571, 1230-1240.

Frankel, A.D., Berg, J.M. and Pabo, C.O., 1987. Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc. Natl. Acad. Sci. U.S.A. 84, 4841-4845.

Freedman, L.P., Luisi, B.F., Korszun, Z.R., Basavappa, R., Sigler, P.B. and Yamamoto, K.R., 1988. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. Nature, 334, p.543.

Gadanho, M., Libkind, D. and Sampaio, J.P., 2006. Yeast diversity in the extreme acidic environments of the Iberian Pyrite Belt. Microb. Ecol. 52, 552-563.

Gadd, G.M. and de Rome, L., 1988. Biosorption of copper by fungal melanin. Appl. Microbiol. Biotechnol. 29, 610-617.

Gallego, S.M., Benavides, M.P. and Tomaro, M.L., 1996. Effect of heavy metal ion excess on sunflower leaves: evidence for involvement of oxidative stress. Plant Sci. 121, 151-159.

Garcia-Toledo, A., Babich, H. and Stotzky, G., 1985. Training of Rhizopus stolonifer and Cunninghamella blakesleeana to copper: cotolerance to cadmium, cobalt, nickel, and lead. Can. J. Microbiol. 31, 485-492.

Gardner, K.H., Pan, T., Narula, S., Rivera, E. and Coleman, J.E., 1991. Structure of the binuclear metal-binding site in the GAL4 transcription factor. Biochem. 30, 11292-11302.

Gay, L., Hébraud, M., Girard, V. and Fevre, M., 1989. Chitin synthase activity from Neocallimastix frontalis, an anaerobic rumen fungus. Microbiol. 135, 279-283.

Gechev, T.S., Van Breusegem, F., Stone, J.M., Denev, I. and Laloi, C., 2006. Reactive oxygen species as signals that modulate plant stress responses and programmed cell death. Bioessays, 28, 1091-1101.

Geisen, S., Fiore-Donno, A.M., Walochnik, J. and Bonkowski, M., 2014. Acanthamoeba everywhere: high diversity of *Acanthamoeba* in soils. Parasitology Res. 113, 3151-3158.

Gelsomino, A., Keijzer-Wolters, A.C., Cacco, G. and van Elsas, J.D., 1999. Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. J. Microbiol. Meth. 38, 1-15.

Geoghegan, I., Steinberg, G. and Gurr, S., 2017. The role of the fungal cell wall in the infection of plants. Trends Microbiol. 25, 957-967.

Geslin, C., Llanos, J., Prieur, D. and Jeanthon, C., 2001. The manganese and iron superoxide dismutases protect Escherichia coli from heavy metal toxicity. Res. Microbiol. 152, 901-905.

Gettins, P.G., 2002. Serpin structure, mechanism, and function. Chem. Rev. 102, 4751–4804.

Giaever, G. and Nislow, C., 2014. The yeast deletion collection: a decade of functional genomics. Genetics, 197, 451-465.

Gietz, R.D., Schiestl, R.H., Willems, A.R. and Woods, R.A., 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast, 11, 355-360.

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M. and Nelson, K.E., 2006. Metagenomic analysis of the human distal gut microbiome. Science, 312, 1355-1359.

Giller, K. E., Witter, E. and Mcgrath, S. P., 1998. Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. Soil. Biol. Biochem. 30, 1389-1414.

Goldmann, K., Schöning, I., Buscot, F. and Wubet, T., 2015. Forest management type influences diversity and community composition of soil fungi across temperate forest ecosystems. Front. Microbiol. 6, 1300.

Gosti, F., Bertauche, N., Vartanian, N. and Giraudat, J., 1995. Abscisic acid-dependent and-independent regulation of gene expression by progressive drought in Arabidopsis thaliana. Mol. Gen. Genet. 246, 10-18.

Gottlieb, C.D. and Linder, M.E., 2017. Structure and function of DHHC protein S-acyltransferases. Biochem. Soc. Trans. 45, 923-8.

Grant, S., Grant, W.D., Cowan, D.A., Jones, B.E., Ma, Y., Ventosa, A. and Heaphy, S., 2006. Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. Appl. Environ. Microbiol. 72, 135-143.

Greaves, J., Prescott, G.R., Gorleku, O.A. and Chamberlain, L.H., 2009. The fat controller: roles of palmitoylation in intracellular protein trafficking and targeting to membrane microdomains. Mol. Membr. Biol. 26, 67-79.

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G. and Bailey, M.J., 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl. Environ. Microbiol. 66, 5488-5491.

Grimsrud, P.A., Xie, H., Griffin, T.J. and Bernlohr, D.A., 2008. Oxidative stress and covalent modification of protein with bioactive aldehydes. J. Biol. Chem. 283, 21837-21841.

Gruninger, R.J., Nguyen, T.T.M., Reid, I.D., Yanke, J., Wang, P., Abbott, D.W., Tsang, A. and McAllister, T., 2018. Application of transcriptomics to compare the carbohydrate active enzymes that are expressed by diverse genera of anaerobic fungi to degrade plant cell wall carbohydrates. Front. Microbiol. 9, 1581.

Gubler, U. and Hoffman, B.J., 1983. A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269.

Gudynaite-Savitch, L. and White, T.C., 2016. Fungal biotechnology for industrial enzyme production: focus on (Hemi) cellulase production strategies, advances and challenges. In: Schmoll, M. and Dattenböck, C. (Eds.), Gene Expression Systems in Fungi: Advancements and Applications. Springer, Cham, 395-439.

Guerra, F., Duplessis, S., Kohler, A., Martin, F., Tapia, J., Lebed, P., Zamudio, F. and González, E., 2009. Gene expression analysis of Populus deltoides roots subjected to copper stress. Environ. Exp. Bot. 67, 335-344.

Guerra, F.P., Reyes, L., Vergara-Jaque, A., Campos-Hernández, C., Gutiérrez, A., Pérez-Díaz, J., Pérez-Díaz, R., Blaudez, D. and Ruíz-Lara, S., 2015. Populus deltoides Kunitz trypsin inhibitor 3 confers metal tolerance and binds copper, revealing a new defensive role against heavy metal stress. Environ. Exp. Bot. 115, 28-37.

Guerriero, G., Hausman, J.F., Strauss, J., Ertan, H. and Siddiqui, K.S., 2016. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. Eng. Life Sci. 16, 1-16.

Gupta, A., Dutta, A., Sarkar, J., Panigrahi, M.K. and Sar, P., 2018. Low-Abundance Members of the Firmicutes Facilitate Bioremediation of Soil Impacted by Highly Acidic Mine Drainage From the Malanjkhand Copper Project, India. Front Microbiol. 9, p.2882.

Gupta, A., Dutta, A., Sarkar, J., Paul, D., Panigrahi, M.K. and Sar, P., 2017. Metagenomic exploration of microbial community in mine tailings of Malanjkhand copper project, India. Genom. Data, 12, 11-13.

Gutiérrez, J.C., Martin-Gonzalez, A., Diaz, S., Amaro, F., Ortega, R., Gallego, A. and de Lucas, M.P., 2008. Ciliates as cellular tools to study the eukaryotic cell-heavy metal interactions. Heavy metal pollution, pp.1-44.

Hall, J.L., 2002. Cellular mechanisms for heavy metal detoxification and tolerance. J. Exp. Bot. 53, 1-11.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. and Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5, R245-R249.

Hao, C., Wang, L., Gao, Y., Zhang, L. and Dong, H., 2010. Microbial diversity in acid mine drainage of Xiang Mountain sulfide mine, Anhui Province, China. Extremophiles, 14, 465-474.

Harada, E., Kim, J.A., Meyer, A.J., Hell, R., Clemens, S. and Choi, Y.E., 2010. Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. Plant Cell Physiol. 51, 1627–1637.

Harris, J.L. and Szaniszlo, P.J., 1986. Localization of chitin in walls of Wangiella dermatitidis using colloidal gold-labeled chitinase. Mycologia, 78, 853-857.

Hartwig, A., 2001. Zinc finger proteins as potential targets for toxic metal ions: differential effects on structure and function. Antioxid. Redox Signal. 3, 625-634.

He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., Wu, W., Gu, B., Jardine, P. and Criddle, C., 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. 1, p.67.

Hesse, C.N., Mueller, R.C., Vuyisich, M., Gallegos-Graves, L.V., Gleasner, C.D., Zak, D.R. and Kuske, C.R., 2015. Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. Front. Microbiol. 6, 337.

Higashimoto, M., Isoyama, N., Ishibashi, S., Inoue, M., Takiguchi, M., Suzuki, S., Ohnishi, Y. and Sato, M., 2009. Tissue-dependent preventive effect of metallothionein against DNA damage in dyslipidemic mice under repeated stresses of fasting or restraint. Life Sciences, 84(17-18), pp.569-575.

Hirsch, P.R., Mauchline, T.H. and Clark, I.M., 2010. Culture-independent molecular techniques for soil microbial ecology. Soil Biol. Biochem. 42, 878-887.

Hlavacova, M., Gumulec, J., Stracina, T., Fojtu, M., Raudenská, M., Masarik, M., Nováková, M. and Paulová, H., 2015. Different doxorubicin formulations affect plasma 4-hydroxy-2-nonenal and gene expression of aldehyde dehydrogenase 3A1 and thioredoxin reductase 2 in rat. Physiol. Res. 64, p.S653.

Hosiner, D., Gerber, S., Lichtenberg-Frate, H., Glaser, W., Schüller, C. and Klipp, E., 2014. Impact of acute metal stress in Saccharomyces cerevisiae. PloS ONE, 9, p.e83330.

Hu, Z., Gibson, A.W., Kim, J.H., Wojciechowicz, L.A., Zhang, B. and Michels, C.A., 1999. Functional domain analysis of the Saccharomyces MAL-activator. Curr. Genet. 36, 1-12.

Huang, Y., Xiao, B. and Xiong, L., 2007. Characterization of a stress responsive proteinase inhibitor gene with positive effect in improving drought resistance in rice. Planta, 226, 73-85.

Huntington, J.A., Read, R.J. and Carrell, R.W., 2000. Structure of a serpin–protease complex shows inhibition by deformation. Nature, 407, 923.

Hygum, T.L., Fobian, D., Kamilari, M., Jørgensen, A., Schiøtt, M., Grosell, M. and Møbjerg, N., 2017. Comparative investigation of copper tolerance and identification of putative tolerance related genes in tardigrades. Front. Physiol. 8, p.95.

Idera, F., Omotola, O., Paul, U.J. and Adedayo, A., 2014. Evaluation of the effectiveness of different acid digestion on sediments. IOSR. J. Appl. Chem. 12, 39-47.

Ilyas, S., Rehman, A., Varela, A.C. and Sheehan, D., 2014. Redox proteomics changes in the fungal pathogen Trichosporon asahii on arsenic exposure: identification of protein responses to metal-induced oxidative stress in an environmentally-sampled isolate. PloS one, 9, p.e102340.

Inouhe, M., 2005. Phytochelatins. Braz. J. Plant Physiol. 17, 65-78.

Iost, I. and Dreyfus, M., 1995. The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. EMBO J. 14, 3252.

Irving, J.A., Steenbakkers, P.J., Lesk, A.M., Op den Camp, H.J., Pike, R.N. and Whisstock, J.C., 2002. Serpins in prokaryotes. Mol. Biol. Evol. 19, 1881-1890.

Ishitani, R., Miyakawa, A. and Iwamoto, T., 1977. Elemental analyses of autoradiographic grains by X-ray microanalysis. Experientia, 33, 440-441.

Jaishankar, M., Tseten, T., Anbalagan, N., Mathew, B.B. and Beeregowda, K.N., 2014. Toxicity, mechanism and health effects of some heavy metals. Interdiscip. Toxicol. 7, 60-72.

Jaksch, M., Paret, C., Stucka, R., Horn, N., Müller-Höcker, J., Horvath, R., Trepesch, N., Stecker, G., Freisinger, P., Thirion, C. and Müller, J., 2001. Cytochrome c oxidase deficiency due to mutations in SCO2, encoding a mitochondrial copper-binding protein, is rescued by copper in human myoblasts. Hum. Mol. Genet. 10, 3025-35.

Jarque, S., Bittner, M., Blaha, L. and Hilscherova, K., 2016. Yeast biosensors for detection of environmental pollutants: current state and limitations. Trends Biotechnol. 34, 408-419.

Jaureguibeitia, A., Saá, L., Llama, M.J. and Serra, J.L., 2007. Purification, characterization and cloning of aldehyde dehydrogenase from Rhodococcus erythropolis UPV-1. Appl. Microbiol. Biotechnol. 73, 1073-1086.

Jing, Q., Wang, J. and Wu, G., 2013. Teth137, a conserved factor of unknown function from thermoanaerobacter ethanolicus JW200, represses the transcription of the adhE gene in vitro. Indian J. Microbiol. 53, 149-154.

Jones, M.D. and Hutchinson, T.C., 1986. The effect of mycorrhizal infection on the response of Betula papyrifera to nickel and copper. New Phytol. 102, 429-442.

Joshi, B.P., Park, J., Lee, W.I. and Lee, K.H., 2009. Ratiometric and turn-on monitoring for heavy and transition metal ions in aqueous solution with a fluorescent peptide sensor. Talanta, 78, 903-9.

Joshi, R.S., Tanpure, R.S., Singh, R.K., Gupta, V.S. and Gir, A.P., 2014. Resistance through inhibition: ectopic expression of serine protease inhibitor offers stress tolerance via delayed senescence in yeast cell. Biochem. Biophys. Res. Commun. 452, 361–368.

Jozefczak, M., Remans, T., Vangronsveld, J. and Cuypers, A., 2012. Glutathione is a key player in metal-induced oxidative stress defenses. Int. J. Mol. Sci. 13, 3145-3175.

Ju, F. and Zhang, T., 2015. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. The ISME J. 9, 683-695.

Kalsotra, T., Khullar, S., Agnihotri, R. and Reddy, M.S., 2018. Metal induction of two metallothionein genes in the ectomycorrhizal fungus Suillus himalayensis and their role in metal tolerance. Microbiology, 164, 868-876.

Kang, M.S., Elango, N., Mattia, E., Au-Young, J., Robbins, P.W. and Cabib, E., 1984. Isolation of chitin synthetase from Saccharomyces cerevisiae. Purification of an enzyme by entrapment in the reaction product. J. Biol. Chem 259, 14966-14972.

Karst, S.M., Dueholm, M.S., McIlroy, S.J., Kirkegaard, R.H., Nielsen, P.H. and Albertsen, M., 2018. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat. Biotechnol. 36, p.190.

Kartal, S.N. and Imamura, Y., 2005. Removal of copper, chromium, and arsenic from CCA-treated wood onto chitin and chitosan. Bioresour. Technol. 96, 389-392.

Kazuo, M. and Sumio, S., 1994. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene, 138, 171-174.

Kellner, H., Luis, P., Portetelle, D. and Vandenbol, M., 2011. Screening of a soil metatranscriptomic library by functional complementation of Saccharomyces cerevisiae mutants. Microbiol. Res. 166, 360-368.

Khan, A.G., 2005. Role of soil microbes in the rhizospheres of plants growing on trace metal contaminated soils in phytoremediation. J. Trace. Elem. Med. Biol. 18, 55-364.

Kim, B.M., Rhee, J.S., Jeong, C.B., Seo, J.S., Park, G.S., Lee, Y.M. and Lee, J.S., 2014. Heavy metals induce oxidative stress and trigger oxidative stress-mediated heat shock protein (hsp) modulation in the intertidal copepod Tigriopus japonicus. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 166, pp.65-74.

Kim, J. and Michels, C.A., 1988. The MAL63 gene of Saccharomyces encodes a cysteine-zinc finger protein. Curr. Genet. 14, 319-323.

Kim, J.S., Kim, H., Yim, B., Rhee, J.S., Won, E.J. and Lee, Y.M., 2018. Identification and molecular characterization of two Cu/Zn-SODs and Mn-SOD in the marine ciliate Euplotes crassus: modulation of enzyme activity and transcripts in response to copper and cadmium. Aquat. Toxicol. 199, 296-304.

Kinner, N.E., Harvey, R.W., Shay, D.M., Metge, D.W. and Warren, A., 2002. Field evidence for a protistan role in an organically-contaminated aquifer. Environ. Sci. Technol. 36, 4312-4318.

Kirch, H.H., Bartels, D., Wei, Y., Schnable, P.S. and Wood, A.J., 2004. The ALDH gene superfamily of Arabidopsis. Trends Plant Sci. 9, 371-377.

Kitson, R.E. and Mellon, M.G., 1944. Colorimetric determination of phosphorus as molybdivanadophosphoric acid. Ind. Eng. Chem. 16, 379-383.

Kluska, K., Adamczyk, J. and Krężel, A., 2018. Metal binding properties of zinc fingers with a naturally altered metal binding site. Metallomics. 10, 248-263.

Knietsch, A., Bowien, S., Whited, G., Gottschalk, G. and Daniel, R., 2003. Identification and characterization of coenzyme B12-dependent glycerol dehydratase-and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl. Environ. Microbiol. 69, 3048-3060.

Koringa, P.G., Thakkar, J.R., Pandit, R.J., Hinsu, A.T., Parekh, M.J., Shah, R.K., Jakhesara, S.J. and Joshi, C.G., 2019. Metagenomic characterisation of ruminal bacterial diversity in buffaloes from birth to adulthood using 16S rRNA gene amplicon sequencing. Funct. Integr. Genomics, 19, 237-247.

Korycka, J., Łach, A., Heger, E., Bogusławska, D.M., Wolny, M., Toporkiewicz, M., Augoff, K., Korzeniewski, J. and Sikorski, A.F., 2012. Human DHHC proteins: a spotlight on the hidden player of palmitoylation. Eur. J. Cell Biol. 91, 107-117.

Kozdrój, J. and van Elsas, J.D., 2000. Response of the bacterial community to root exudates in soil polluted with heavy metals assessed by molecular and cultural approaches. Soil Biol. Biochem. 32, 1405-1417.

Krizek, B.A. and Berg, J.M., 1992. Complexes of zinc finger peptides with nickel (2^+) and iron (2^+). Inorg. Chem. 31, 2984-2986.

Krizek, B.A., Merkle, D.L. and Berg, J.M., 1993. Ligand variation and metal ion binding specificity in zinc finger peptides. Inorg. Chem. 32, 937-940.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870-1874.

Kunitz, M., 1945. Crystallization of a trypsin inhibitor from soybean. Science, 101, 668-669.

Labidi, A., Salaberria, A.M., Fernandes, S.C., Labidi, J. and Abderrabba, M., 2016. Adsorption of copper on chitin-based materials: Kinetic and thermodynamic studies. J. Taiwan Inst. Chem. Eng. 65,140-148.

Lanfranco, L., Balsamo, R., Martino, E., Bonfante, P. and Perotto, S., 2004. Zinc ions differentially affect chitin synthase gene expression in an ericoid mycorrhizal fungus. Plant Biosyst. 138, 271-277.

Lanfranco, L., Balsamo, R., Martino, E., Perotto, S. and Bonfante, P., 2002. Zinc ions alter morphology and chitin deposition in an ericoid fungus. Eur. J. Histochem. 46, 341-50.

Lara, E., Berney, C., Harms, H. and Chatzinotas, A., 2007. Cultivation-independent analysis reveals a shift in ciliate 18S rRNA gene diversity in a polycyclic aromatic hydrocarbon-polluted soil. FEMS Microbiol. Ecol., 62, 365-373.

Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P. and Rossi, J.P., 2006. Soil invertebrates and ecosystem services. Eur. J. Soil Biol. 42, S3-S15.

Leal-Morales, C.A., Bracker, C.E. and Bartnicki-Garcia, S., 1988. Localization of chitin synthetase in cell-free homogenates of Saccharomyces cerevisiae: chitosomes and plasma membrane. Proc. Natl. Acad. Sci. USA. 85, 8516-8520.

Lehembre, F., Doillon, D., David, E., Perrotto, S., Baude, J., Foulon, J., Harfouche, L., Vallon, L., Poulain, J., Da Silva, C. and Wincker, P., 2013. Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. Environ. Microbiol. 15, 2829-2840.

Lehmann, M., Riedel, K., Adler, K. and Kunze, G., 2000. Amperometric measurement of copper ions with a deputy substrate using a novel Saccharomyces cerevisiae sensor. Biosens. Bioelectron. 15, 211-219.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C. and Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature, 442, 806.

Lending, C.R., Leal-Morales, C.A., Flores-Martinez, A., Bracker, C.E. and Bartnicki_Garcia, S., 1991. Purification and characterization of 16 S chitin synthetase particles from cell walls of Mucor rouxii. Exp Mycol. 15, pp.11-25.

Lesage, G., Shapiro, J., Specht, C.A., Sdicu, A.M., Ménard, P., Hussein, S., Tong, A.H.Y., Boone, C. and Bussey, H., 2005. An interactional network of genes involved in chitin synthesis in Saccharomyces cerevisiae. BMC Genet. 6, 8.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25, 1754-1760.

Li, L. and Kaplan, J., 2001. The yeast gene MSC2, a member of the cation diffusion facilitator family, affects the cellular distribution of zinc. J. Biol. Chem. 276, 5036-43.

Li, Y., Nakazono, M., Tsutsumi, N. and Hirai, A., 2000. Molecular and cellular characterizations of a cDNA clone encoding a novel isozyme of aldehyde dehydrogenase from rice. Gene, 249, 67-74.

Li, Z.S., Lu, Y.P., Zhen, R.G., Szczypka, M., Thiele, D.J. and Rea, P.A., 1997. A new pathway for vacuolar cadmium sequestration in Saccharomyces cerevisiae: YCF1-catalyzed transport of bis (glutathionato) cadmium. Proc. Natl. Acad. Sci. 94, 42-47.

Liang, X., Nazarian, A., Erdjument-Bromage, H., Bornmann, W., Tempst, P. and Resh, M.D., 2001. Heterogeneous fatty acylation of Src family kinases with polyunsaturated fatty acids regulates raft localization and signal transduction. J. Biol. Chem. 276, 30987-30994.

Liles, M.R., Manske, B.F., Bintrim, S.B., Handelsman, J. and Goodman, R.M., 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl. Environ. Microbiol. 69, 2684-2691.

Liu, F. and Schnable, P.S., 2002. Functional specialization of maize mitochondrial aldehyde dehydrogenases. Plant Physiol. 130, 1657-1674.

Liu, W.T. and Stahl, D.A., 2007. Molecular approaches for the measurement of density, diversity, and phylogeny. In: Hurst, C., Crawford, R., Garland, J., Lipson, D., Mills, A. and

Stetzenbach, L. (Eds.), Manual of Environmental Microbiology. American Society of Microbiology Press, Washington DC, 139-156.

Liu, W.T., Marsh, T.L., Cheng, H. and Forney, L.J., 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63, 4516-4522.

Liu, X.Z., Sun, X., Shen, K.P., Jin, W.J., Fu, Z.Y., Tao, H.R. and Xu, Z.X., 2017. Aldehyde dehydrogenase 2 overexpression inhibits neuronal apoptosis after spinal cord ischemia/reperfusion injury. Neural Regen. Res. 12, p.1166.

Liu, Z.J., Sun, Y.J., Rose, J., Chung, Y.J., Hsiao, C.D., Chang, W.R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. and Wang, B.C., 1997. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat. Struct. Biol. 4, 317.

Lobo, S., Greentree, W.K., Linder, M.E. and Deschenes, R.J., 2002. Identification of a Ras palmitoyltransferase in Saccharomyces cerevisiae. J. Biol. Chem. 277, 41268-41273.

Longo, V.D., Gralla, E.B. and Valentine, J.S., 1996. Superoxide dismutase activity is essential for stationary phase survival in Saccharomyces cerevisiae Mitochondrial production of toxic oxygen species in vivo. J. Biol. Chem. 271, 12275-12280.

Lorito, M., Broadway, R.M., Hayes, C.K., Woo, S.L., Noviello, C., Williams, D.L., Harman, G.E., Grings, E.E., Adams, D.C. and Short, R.E., 1994. Proteinase inhibitors from plants as a novel class of fungicides. Mol. Plant Microbe Interact. 7, 525-527.

Luis, P., Kellner, H., Martin, F. and Buscot, F., 2005. A molecular method to evaluate basidiomycete laccase gene expression in forest soils. Geoderma, 128, 18-27.

Ma, J., Wang, Z., Li, H., Park, H.D. and Wu, Z., 2016. Metagenomes reveal microbial structures, functional potentials, and biofouling-related genes in a membrane bioreactor. Appl. Microbiol. Biotechnol. 100, 5109-5121.

Machida, S. and Saito, M., 1993. Purification and characterization of membrane-bound chitin synthase. J Biol Chem. 268, pp.1702-1707.

MacPherson, S., Larochelle, M. and Turcotte, B., 2006. A fungal family of transcriptional regulators: the zinc cluster proteins. Microbiol. Mol. Biol. Rev. 70, 583-604.

Magoč, T. and Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 27, 2957-2963.

Makowski, G.S. and Sunderman Jr, F.W., 1992. The interactions of zinc, nickel, and cadmium with Xenopus transcription factor IIIA, assessed by equilibrium dialysis. J. Inorg. Biochem. 48, 107-119.

Marchitti, S.A., Brocker, C., Stagos, D. and Vasiliou, V., 2008. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert Opin. Drug Metab. Toxicol. 4, 697-720.

Marmeisse, R., Kellner, H., Fraissinet-Tachet, L. and Luis, P., 2017. Discovering protein-coding genes from the environment: time for the eukaryotes?. Trends Biotechnol. 35, 824-835.

Marschner, P., Jentschke, G. and Godbold, D.L., 1998. Cation exchange capacity and lead sorption in ectomycorrhizal fungi. Plant Soil, 205, 93-98.

Marsh, T.L., 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. Curr. Opin. Microbiol. 2, 323-327.

Martin, B.R., Wang, C., Adibekian, A., Tully, S.E. and Cravatt, B.F., 2012. Global profiling of dynamic protein palmitoylation. Nat. Methods, 9, 84.

Martino, E., Turnau, K., Girlanda, M., Bonfante, P. and Perotto, S., 2000. Ericoid mycorrhizal fungi from heavy metal polluted soils: their identification and growth in the presence of zinc ions. Mycol. Res. 104, 338-344.

Matsuura, H., Yamamoto, Y., Muraoka, M., Akaishi, K., Hori, Y., Uemura, K., Tsuji, N., Harada, K., Hirata, K., Bamba, T. and Miyasaka, H., 2013. Development of surface-engineered yeast cells displaying phytochelatin synthase and their application to cadmium biosensors by the combined use of pyrene-excimer fluorescence. Biotechnol. Prog. 29, 1197-1202.

Mattison, R.G. and Harayama, S., 2001. The predatory soil flagellate Heteromita globosa stimulates toluene biodegradation by a Pseudomonas sp. FEMS Microbiol. Lett. 194, pp.39-45.

Mayer, R.T., Chen, A.C. and DeLoach, J.R., 1980. Characterization of a chitin synthase from the stable fly, Stomoxys calcitrans (L.). Insect Biochem. Mol. Biol. 10, 549-556.

McLeary, F.A., Rcom-H'cheo-Gauthier, A.N., Goulding, M., Radford, R.A., Okita, Y., Faller, P., Chung, R.S. and Pountney, D.L., 2019. Switching on Endogenous Metal Binding Proteins in Parkinson's Disease. Cells, 8, 179.

MD'Alessio, J. and Gerard, G.F., 1988. Second-strand cDNA synthesis with *E. coli* DNA polymerase I and RNase H: the fate of information at the mRNA 5'terminus and the effect of E. coli DNA ligase. Nucleic Acids Res. 16, 1999-2014.

Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R., Falkow, S. and Rappuoli, R., 2008. Microbiology in the post-genomic era. Nat. Rev. Microbiol. 6, 419.

Mehra, R.K., Winge, D.R., 1991. Metal ion resistance in fungi: molecular mechanisms and their regulated expression. J. Cell Biochem. 45, 30-40.

Mendes, A.M.S., Duda, G.P., Nascimento, C.W.A.D. and Silva, M.O., 2006. Bioavailability of cadmium and lead in a soil amended with phosphorus fertilizers. Sci Agr. 63, pp.328-332.

Méndez-García, C., Peláez, A.I., Mesa, V., Sánchez, J., Golyshina, O.V. and Ferrer, M., 2015. Microbial diversity and metabolic networks in acid mine drainage habitats. Front Microbiol. 6, 475.

Meyer, K.M., Petersen, I.A., Tobi, E., Korte, L. and Bohannan, B.J., 2019. Use of RNA and DNA to identify mechanisms of bacterial community homogenization. Front. Microbiol. 10, p.2066.

Miller, D.N., Bryant, J.E., Madsen, E.L. and Ghiorse, W.C., 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl. Environ. Microbiol. 65, 4715-4724.

Minet, M., Dufour, M.E. and Lacroute, F., 1992. Complementation of Saccharomyces cerevisiae auxotrophic mutants by Arabidopsis thaliana cDNAs. Plant J. 2, 417-422.

Mishra, S. and Dubey, R.S., 2006. Heavy metal uptake and detoxification mechanisms in plants. Int. J. Agric. Res. 1, 122-141.

Missihoun, T.D., Schmitz, J., Klug, R., Kirch, H.H. and Bartels, D., 2011. Betaine aldehyde dehydrogenase genes from Arabidopsis with different sub-cellular localization affect stress responses. Planta, 233, 369-382.

Mitchell, D.A., Mitchell, G., Ling, Y., Budde, C. and Deschenes, R.J., 2010. Mutational analysis of Saccharomyces cerevisiae Erf2 reveals a two-step reaction mechanism for protein palmitoylation by DHHC enzymes. J. Biol. Chem. 285, 38104-38114.

Mitchell, E., Frisbie, S. and Sarkar, B., 2011. Exposure to multiple metals from groundwater—a global crisis: Geology, climate change, health effects, testing, and mitigation. Metallomics, 3, 874-908.

Miyayama, T., Arai, Y., Suzuki, N. and Hirano, S., 2013. Mitochondrial electron transport is inhibited by disappearance of metallothionein in human bronchial epithelial cells following exposure to silver nitrate. Toxicology, 305, 20-29.

Miyazaki, A. and Ootaki, T., 1997. Multiple genes for chitin synthase in the zygomycete fungus Phycomyces blakesleeanus. J Gen Appl Microbiol. 43, pp.333-340.

Moltó, E., Bonzón-Kulichenko, E., Gallardo, N. and Andrés, A., 2007. MTPA: A crustacean metallothionein that affects hepatopancreatic mitochondrial functions. Arch. Biochem. Biophys. 467, 31-40.

Montoro, A.G., Quiroga, R., Maccioni, H.J., and Taubas, J.V., 2009. A novel motif at the C-terminus of palmitoyltransferases is essential for Swf1 and Pfa3 function in vivo. Biochem. J. 419, 301-8

Moran, M.A., 2009. Metatranscriptomics: Eavesdropping on Complex Microbial Communities-Large-scale sequencing of mRNAs retrieved from natural communities provides insights into microbial activities and how they are regulated. Microbe, 4, 329.

Mukhopadhyay, A., Vij, S. and Tyagi, A.K., 2004. Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. Proc. Natl. Acad. Sci. USA, 101, 6309-14.

Müller, A.K., Westergaard, K., Christensen, S. and Sørensen, S.J., 2001. The effect of long-term mercury pollution on the soil microbial community. FEMS Microbiol. Ecol. 36, 11-19.

Murasugi, A., Nakagawa, C.W. and Hayashi, Y., 1981. Formation of cadmium-binding peptide allomorphs in fission yeast. J. Biochem. 96, 1375-1379.

Muyzer, G. and Smalla, K., 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie van Leeuwenhoek, 73, 127-141.

Nacke, H., Engelhaupt, M., Brady, S., Fischer, C., Tautzt, J. and Daniel, R., 2012. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol. Lett. 34, 663-675.

Nannipieri, P., Ascher, J., Ceccherini, M., Landi, L., Pietramellara, G. and Renella, G., 2003. Microbial diversity and soil functions. Eur. J. Soil Sci. 54, 655-670.

Nannipieri, P., Pietramellara, G. and Renella, G., 2014. Omics in soil science. Caster Academic Press, Norfolk

Needleman, R., 1991. Control of maltase synthesis in yeast. Mol. Microbiol. 5, 2079-2084.

Niehaus, F., Gabor, E., Wieland, S., Siegert, P., Maurer, K.H. and Eck, J., 2011. Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. Microb. Biotechnol. 4, 767–776.

Nishi, N., Maekita, Y., Nishimura, S.I., Hasegawa, O. and Tokura, S., 1987. Highly phosphorylated derivatives of chitin, partially deacetylated chitin and chitosan as new functional polymers: metal binding property of the insolubilized materials. Int. J. Biol. Macromol. 9, 109-114.

Nüsslein, K. and Tiedje, J.M., 1998. Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from

DNA fractionated on the basis of its guanine and cytosine composition. Appl. Environ. Microbiol. 64, 1283-1289.

O'Brien, P.J. and Zatz, M., 1984. Acylation of bovine rhodopsin by [3H] palmitic acid. J. Biol. Chem. 259, 5054-5057.

Ochiai, E.I., 2012. General principles of biochemistry of the elements. Springer Science & Business Media.

Olsen, S.R., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate (No. 939). US Dept. of Agriculture.

Opel, K.L., Chung, D. and McCord, B.R., 2010. A study of PCR inhibition mechanisms using real time PCR. J Forensic Sci. 55, 25-33.

Orlean, P., 1987. Two chitin synthases in Saccharomyces cerevisiae. J. Biol. Chem. 262, 5732-5739.

Osobová, M., Urban, V., Jedelský, P.L., Borovička, J., Gryndler, M., Ruml, T. and Kotrba, P., 2011. Three metallothionein isoforms and sequestration of intracellular silver in the hyperaccumulator Amanita strobiliformis. New Phytol. 190, 916-926.

Pagani, M.A., Casamayor, A., Serrano, R., Atrian, S. and Ariño, J., 2007. Disruption of iron homeostasis in Saccharomyces cerevisiae by high zinc levels: a genome-wide study. Mol. Microbiol. 65, 521-537.

Panda, S.K., Chaudhury, I. and Khan, M.H., 2003. Heavy metals induce lipid peroxidation and affect antioxidants in wheat leaves. Biol. Plant. 46, 289-294.

Pandey, P.K., Sharma, R., Roy, M. and Pandey, M., 2007. Toxic mine drainage from Asia's biggest copper mine at Malanjkhand, India. Environ. Geochem. Health, 29, 237-248.

Patel, D.D., Patel, A.K., Parmar, N.R., Shah, T.M., Patel, J.B., Pandya, P.R. and Joshi, C.G., 2014. Microbial and Carbohydrate Active Enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet. Gene, 545, 88-94.

Pemberton, P.A., Stein, P.E., Pepys, M.B., Potter, J.M. and Carrell, R.W., 1988. Hormone binding globulins undergo serpin conformational change in inflammation. Nature, 336, 257.

Peng, X., Shindo, K., Kanoh, K., Inomata, Y., Choi, S.K. and Misawa, N., 2005. Characterization of Sphingomonas aldehyde dehydrogenase catalyzing the conversion of various aromatic aldehydes to their carboxylic acids. Appl. Microbiol. Biotechnol. 69, 141-150.

Peng, X., Shindo, K., Kanoh, K., Inomata, Y., Choi, S.K. and Misawa, N., 2005. Characterization of Sphingomonas aldehyde dehydrogenase catalyzing the conversion of various aromatic aldehydes to their carboxylic acids. Appl. Microbiol. Biotechnol. 69, 141-150.

Pham, A.N., Xing, G., Miller, C.J. and Waite, T.D., 2013. Fenton-like copper redox chemistry revisited: hydrogen peroxide and superoxide mediation of copper-catalyzed oxidant production. J. Catal. 301, 54-64.

Philippsen, P., Kleine, K., Pöhlmann, R., Düsterhöft, A., Hamberg, K., Hegemann, J.H., Obermaier, B., Urrestarazu, L.A., Aert, R., Albermann, K. and Altmann, R., 1997. The nucleotide sequence of Saccharomyces cerevisiae chromosome XIV and its evolutionary implications. Nature, 387, pp.93-98.

Piper, C.S., 1966. Soil and plant analysis: a laboratory manual of methods for the examination of soils and the determination of the inorganic constituents of plants. Hans Publications, Bombay.

Pointing, S.B. and Belnap, J., 2012. Microbial colonization and controls in dryland systems. Nat. Rev. Microbiol. 10, 551-562.

Politis, E.G., Roth, A.F. and Davis, N.G., 2005. Transmembrane topology of the protein palmitoyl transferase Akr1. J. Biol. Chem. 280, 10156-10163.

Popovic, A. and Parkinson, J., 2018. Characterization of eukaryotic microbiome using 18S amplicon sequencing. In Microbiome Analysis. Humana Press, New York, NY. 29-48

Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A. and Hollibaugh, J.T., 2005. Analysis of microbial gene transcripts in environmental samples. Appl. Environ. Microbiol. 71, 4121-4126.

Predki, P.F. and Sarkar, B., 1994. Metal replacement in" zinc finger" and its effect on DNA binding. Environ. Health. Perspect. 102(suppl 3), 195-8.

Putilina, T., Wong, P. and Gentleman, S., 1999. The DHHC domain: a new highly conserved cysteine-rich motif. Mol. Cell. Biochem. 195, 219-26.

Qing, X., Yutong, Z. and Shenggao, L., 2015. Assessment of heavy metal pollution and human health risk in urban soils of steel industrial city (Anshan), Liaoning, Northeast China. Ecotoxicol. Environ. Saf. 120, 377-385.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590-D596.

Radhika, V., Milkevitch, M., Audigé, V., Proikas-Cezanne, T. and Dhanasekaran, N., 2005. Engineered Saccharomyces cerevisiae strain BioS-1, for the detection of water-borne toxic metal contaminants. Biotechnol. Bioeng. 90, 29-35.

Rajkumar, M., Sandhya, S., Prasad, M.N.V. and Freitas, H., 2012. Perspectives of plant-associated microbes in heavy metal phytoremediation. Biotechnol Adv. 30, pp.1562-1574.

Ramesh, G., Podila, G.K., Gay, G., Marmeisse, R. and Reddy, M.S., 2009. Copper and cadmium metallothioneins of the ectomycorrhizal fungus Hebeloma cylindrosporum have different patterns of regulation. Appl. Environ. Microbiol. 2266-2274.

Ranjard, L., Nazaret, S., Gourbière, F., Thioulouse, J., Linet, P. and Richaume, A., 2000. A soil microscale study to reveal the heterogeneity of Hg (II) impact on indigenous bacteria by quantification of adapted phenotypes and analysis of community DNA fingerprints. FEMS Microbiol. Ecol. 31, 107-115.

Redon, E., Loubière, P., Cocaign-Bousquet, M., 2005. Role of mRNA stability during genome-wide adaptation of Lactococcus lactis to carbon starvation. J. Biol. Chem. 280, 36380-36385.

Rehman, S., Aziz, E., Akhtar, W., Ilyas, M. and Mahmood, T., 2017. Structural and functional characteristics of plant proteinase inhibitor-II (PI-II) family. Biotechnol Lett. 39, 647-666.

Remans, T., Opdenakker, K., Smeets, K., Mathijsen, D., Vangronsveld, J. and Cuypers, A., 2010. Metal-specific and NADPH oxidase dependent changes in lipoxygenase and NADPH oxidase gene expression in Arabidopsis thaliana exposed to cadmium or excess copper. Funct. Plant Biol. 37, 532-544.

Roane, T.M., Pepper, I.L., 1999. Microbial responses to environmentally toxic cadmium. Microb. Ecol. 38, 358-364.

Robinson, L.J., Busconi, L. and Michel, T., 1995. Agonist-modulated palmitoylation of endothelial nitric oxide synthase. J. Biol. Chem. 270, 995-998.

Rocks, O., Gerauer, M., Vartak, N., Koch, S., Huang, Z.P., Pechlivanis, M., Kuhlmann, J., Brunsveld, L., Chandra, A., Ellinger, B. and Waldmann, H., 2010. The palmitoylation

machinery is a spatially organizing system for peripheral membrane proteins. Cell, 141, 458-471.

Roda, A., Roda, B., Cevenini, L., Michelini, E., Mezzanotte, L., Reschiglian, P., Hakkila, K. and Virta, M., 2011. Analytical strategies for improving the robustness and reproducibility of bioluminescent microbial bioreporters. Anal. Bioanal. Chem. 401, 201-211.

Rodrigues, S.M., Andrade, M.O., Gomes, A.P.S., DaMatta, F.M., Baracat-Pereira, M.C. and Fontes, E.P., 2006. Arabidopsis and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. J. Exp. Bot. 57, 1909-1918.

Rodriguez-Zaragoza, S., Mayzlish, E. and Steinberger, Y., 2005. Vertical distribution of the free-living amoeba population in soil under desert shrubs in the Negev Desert, Israel. Appl. Environ. Microbiol. 71, 2053-2060.

Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K., Kent, A.D., Daroub, S.H., Camargo, F.A., Farmerie, W.G. and Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1, 283.

Rogerson, A. and Berger, J., 1981. The effects of cold temperatures and crude oil on the abundance and activity of protozoa in a garden soil. Can. J. Zool. 59, 1554-1560.

Romero-Puertas, M.C., Palma, J.M., Gómez, M., Del Rio, L.A. and Sandalio, L.M., 2002. Cadmium causes the oxidative modification of proteins in pea plants. Plant Cell Environ. 25, 677-686.

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., Minor, C. and Tiong, C.L., 2000. Cloning the

soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66, 2541-2547.

Rosenzweig, A.C. 2002, Metallochaperones: bind and deliver. Chem. Biol. 9, 673-7

Rossman, A.Y., 1998. Protocols for an all taxa biodiversity inventory of fungi in a Costa Rican conservation area. Parkway Publishers, Inc.

Roth, A.F., Feng, Y., Chen, L. and Davis, N.G., 2002. The yeast DHHC cysteine-rich domain protein Akr1p is a palmitoyl transferase. J. Cell Biol. 159, 23-28.

Russo, D.A., Couto, N., Beckerman, A.P. and Pandhal, J., 2016. A metaproteomic analysis of the response of a freshwater microbial community under nutrient enrichment. Front. Microbiol. 7, p.1172.

Sandaa, R.A., Torsvik, V., Enger, Ø., Daae, F.L., Castberg, T. and Hahn, D., 1999. Analysis of bacterial communities in heavy metal-contaminated soils at different levels of resolution. FEMS Microbiol. Ecol. 30, 237-251.

Sanger, F., Nicklen, S. and Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.

Sarwar, N., Imran, M., Shaheen, M.R., Ishaque, W., Kamran, M.A., Matloob, A., Rehim, A. and Hussain, S., 2017. Phytoremediation strategies for soils contaminated with heavy metals: modifications and future perspectives. Chemosphere, 171, pp.710-721.

Satoh, H., Uchida, A., Nakayama, K. and Okada, M., 2001. Water-soluble chlorophyll protein in Brassicaceae plants is a stress-induced chlorophyll-binding protein. Plant Cell Physiol. 42, 906-911.

Sburlati, A. and Cabib, E., 1986. Chitin synthetase 2, a presumptive participant in septum formation in Saccharomyces cerevisiae. J. Biol. Chem. 261, 15147-15152.

Schjerling, P. and Holmberg, S., 1996. Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. Nucleic Acids Res. 24, 4599-4607.

Schloss, P.D. and Handelsman, J., 2003. Biotechnological prospects from metagenomics. Curr. Opin. Biotechnol. 14, 303-310.

Schnittler, M., Shchepin, O.N., Dagamac, N.H.A., Borg Dahl, M. and Novozhilov, Y.K., 2017. Barcoding myxomycetes with molecular markers: challenges and opportunities. Nova Hedwigia, 104, 323-341.

Schu, P., Suarez Rendueles, P. and Wolf, D.H., 1991. The proteinase yscB inhibitor (PBI2) gene of yeast and studies on the function of its protein product. Eur. J. Biochem. 197, 1-7.

Seishi, K., Shingo, S., Su-Wan, O., Nam-Soon, K., Yuri, U., Naoto, A., Midori, Y.K. and Takashi, A., 1994. Construction of a human full-length cDNA bank. Gene, 150, 243-250.

Serero, A., Lopes, J., Nicolas, A. Boiteux, S., 2008. Yeast genes involved in cadmium tolerance: identification of DNA replication as a target of cadmium toxicity. DNA Repair, 7, 1262-1275.

Seshadri, R., Leahy, S.C., Attwood, G.T., Teh, K.H., Lambie, S.C., Cookson, A.L., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Hadjithomas, M., Varghese, N.J. and Paez-Espino, D., 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat. Biotechnol. 36, 359.

Shahid, M., Dumat, C., Khalid, S., Schreck, E., Xiong, T. and Niazi, N.K., 2017. Foliar heavy metal uptake, toxicity and detoxification in plants: A comparison of foliar and root metal uptake. J Hazard Mater. 325, pp.36-58.

Sharma, R. and Sharma, P.K., 2018. Metatranscriptome sequencing and analysis of agriculture soil provided significant insights about the microbial community structure and function. Ecol. Genet. Genomics, 6, 9-15.

Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5, p.209.

Shetty, R.S., Deo, S.K., Liu, Y. and Daunert, S., 2004. Fluorescence-based sensing system for copper using genetically engineered living yeast cells. Biotechnol. Bioeng. 88, 664-670.

Shimoi, H., Muranaka, Y., Sato, S.I., Saito, K., Tadenuma, M., 1991. Purification of the enzymes responsible for the lysis of yeast cells by Rarobacter faecitabidus. Agric. Biol. Chem. 55, 371–378.

Shitan, N., Horiuchi, K.I., Sato, F., Yazaki, K., 2007. Bowman–Birk proteinase inhibitor confers heavy metal and multiple drug tolerance in yeast. Plant Cell Physiol. 48, 193–197.

Simpkins, C., Lloyd, T., Li, S. and Balderman, S., 1998. Metallothionein-induced increase in mitochondrial inner membrane permeability. J. Surg. Res. 75, 30-34.

Singh, D.P., Prabha, R., Gupta, V.K. and Verma, M.K., 2018. Metatranscriptome analysis deciphers multifunctional genes and enzymes linked with the degradation of aromatic compounds and pesticides in the wheat rhizosphere. Front. Microbiol. 9, 1331.

Singh, S., Parihar, P., Singh, R., Singh, V.P. and Prasad, S.M., 2015. Heavy metal tolerance in plants: role of transcriptomics, proteomics, metabolomics, and ionomics. Front. Plant. Sci. 6, 1143.

Smit, E., Leeflang, P. and Wernars, K., 1997. Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. FEMS Microbiol Ecol 23, 249-261.

Smith, C.J. and Osborn, A.M., 2009. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol. Ecol. 67, 6-20.

Smotrys, J.E., Schoenfish, M.J., Stutz, M.A. and Linder, M.E., 2005. The vacuolar DHHC-CRD protein Pfa3p is a protein acyltransferase for Vac8p. J. Cell Biol. 170, 1091-1099.

Snowden, K.C., Richards, K.D. and Gardner, R.C., 1995. Aluminum-induced genes (induction by toxic metals, low calcium, and wounding and pattern of expression in root tips. Plant Physiol. 107, 341–348.

Söllinger, A., Tveit, A.T., Poulsen, M., Noel, S.J., Bengtsso, M., Bernhardt, J., Hellwing, A.L.F., Lund, P., Riedel, K., Schleper, C. and Højberg, O., 2018. Holistic assessment of rumen microbiome dynamics through quantitative metatranscriptomics reveals multifunctional redundancy during key steps of anaerobic feed degradation. MSystems, 3, e00038-18.

Srinivasan, T. and Kirti, P.B., 2012. Protease inhibitors and stress tolerance. Funct. Plant Sci. Biotechnol. 6, 59–66.

Srivastav, R., Sharma, R., Tandon, S. and Tandon, C., 2018. Role of DHH superfamily proteins in nucleic acids metabolism and stress tolerance in prokaryotes and eukaryotes. Int. J.Biol. Macromol. 127, 66-75.

Stein, P.E. and Carrell, R.W., 1995. What do dysfunctional serpins tell us about molecular mobility and disease?. Nat. Struct. Biol. 2, 96.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W. and Richards, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. 19, pp.21-31.

Strandberg, G.W., Shumate, S.E. and Parrott, J.R., 1981. Microbial cells as biosorbents for heavy metals: accumulation of uranium by Saccharomyces cerevisiae and Pseudomonas aeruginosa. Appl. Environ. Microbiol. 41, 237-245.

Štursová, M., Žifčáková, L., Leigh, M.B., Burgess, R. and Baldrian, P., 2012. Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. FEMS Microbiol. Ecol. 80, 735-746.

Sugano, S., Kaminaka, H., Rybka, Z., Catala, R., Salinas, J., Matsui, K., Ohme-Takagi, M. and Takatsuji, H., 2003. Stress-responsive zinc finger gene ZPT2-3 plays a role in drought tolerance in petunia. Plant J. 36, 830-41.

Sunkar, R., Bartels, D. and Kirch, H.H., 2003. Overexpression of a stress-inducible aldehyde dehydrogenase gene from Arabidopsis thaliana in transgenic plants improves stress tolerance. Plant J. 35, 452-464.

Tabaczar, S., Czogalla, A., Podkalicka, J., Biernatowska, A. and Sikorski, A.F., 2017. Protein palmitoylation: Palmitoyltransferases and their specificity. Exp. Biol. Med. 242, 1150-1157.

Takasaki, K., Miura, T., Kanno, M., Tamaki, H., Hanada, S., Kamagata, Y. and Kimura, N., 2013. Discovery of glycoside hydrolase enzymes in an avicel-adapted forest soil fungal community by a metatranscriptomic approach. PLoS One, 8, p.e55485.

Tecon, R. and Or, D., 2017. Biophysical processes supporting the diversity of microbial life in soil. FEMS Microbio. Rev. 41, 599-623.

Terry, S.A., Badhan, A., Wang, Y., Chaves, A.V. and McAllister, T.A., 2019. Fibre digestion by rumen microbiota - a review of recent metagenomic and metatranscriptomic studies. Can. J. Anim. Sci. 99, 678-692.

Thakur, B., Yadav, R., Fraissinet-Tachet, L., Marmeisse, R. and Reddy, M.S., 2018. Isolation of multi-metal tolerant ubiquitin fusion protein from metal polluted soil by metatranscriptomic approach. J. Microbiol. Methods 152, 119-25.

Thakur, B., Yadav, R., Vallon, L., Marmeisse, R., Fraissinet-Tachet, L., Reddy, M.S., 2019. Multi-metal tolerance of von Willebrand factor type D domain isolated from metal contaminated site by metatranscriptomics approach. Sci. Total Environ 661, 432-440.

Todd, R.B. and Andrianopoulos, A., 1997. Evolution of a fungal regulatory gene family: the Zn (II) 2Cys6 binuclear cluster DNA binding motif. Fungal Genet. Biol. 21, 388-405.

Torsvik, V. and Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5, 240-245.

Torsvik, V., Sørheim, R. and Goksøyr, J., 1996. Total bacterial diversity in soil and sediment communities—a review. J. Ind. Microbiol. Biotechnol. 17, 170-178.

Tringe, S.G. and Rubin, E.M., 2005. Metagenomics: DNA sequencing of environmental samples. Nat. Rev. Genet. 6, 805-814.

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C. and Schuster, S.C., 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PloS One, 3, p.e2527.

Valdez-Taubas, J. and Pelham, H., 2005. Swf1-dependent palmitoylation of the SNARE Tlg1 prevents its ubiquitination and degradation. EMBO J. 24, 2524-2532.

Vallee, B.L., Coleman, J.E. and Auld, D.S., 1991. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. Proc. Natl. Acad. Sci. U.S.A. 88, 99-1003.

van Bakel, H., Strengman, E., Wijmenga, C. and Holstege, F.C., 2005. Gene expression profiling and phenotype analyses of S. cerevisiae in response to changing copper reveals six genes with new roles in copper and iron metabolism. Physiol. Genomics, 22, 356-367.

Van Raamsdonk, J.M. and Hekimi, S., 2009. Deletion of the mitochondrial superoxide dismutase sod-2 extends lifespan in Caenorhabditis elegans. PLoS Genet. 5, e1000361.

Vasiliou, V., Pappa, A. and Estey, T., 2004. Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. Drug Metab. Rev. 36, 279-299.

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

Viraraghavan, T. and Srinivasan, A., 2011. Fungal biosorption and biosorbents. In: Kotrba, P., Mackova, M. and Macek, T. (Eds.), Microbial Biosorption of Metals. Springer, Dordrecht, 143-158.

Vopálenská, I., Váchová, L. and Palková, Z., 2015. New biosensor for detection of copper ions in water based on immobilized genetically modified yeast cells. Biosens. Bioelectron. 72, 160-167.

Walkley, A., 1947. A critical examination of a rapid method for determining organic carbon in soils - effect of variations in digestion conditions and of inorganic soil constituents. Soil Sci. 63, 251-264.

Wang, G.W., Klein, J.B. and Kang, Y.J., 2001. Metallothionein inhibits doxorubicin-induced mitochondrial cytochrome c release and caspase-3 activation in cardiomyocytes. J. Pharmacol. Exp. Ther. 298, 461-468.

Wang, Z. and Szaniszlo, P.J., 2000. WdCHS3, a gene that encodes a class III chitin synthase in Wangiella (Exophiala) dermatitidis, is expressed differentially under stress conditions. J. Bacteriol. 182, 874-881.

Warnecke, F. and Hess, M., 2009. A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. J. Biotech. 142, 91-95.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N. and Sorek, R., 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature, 450, 560.

Wasserman, G.A., Liu, X., Factor-Litvak, P., Gardner, J.M. and Graziano, J.H., 2008. Developmental impacts of heavy metals and undernutrition. Basic Clin. Pharmacol. Toxicol. 102, 212-217.

Weidner, S., Arnold, W. and Puhler, A., 1996. Diversity of uncultured microorganisms associated with the seagrass Halophila stipulacea estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. Appl. Environ. Microbiol. 62, 766-771.

Welch, J., Fogel, S., Buchman, C. and Karin, M., 1989. The CUP2 gene product regulates the expression of the CUP1 gene, coding for yeast metallothionein. EMBO J. 8, 255-260.

Wellenreuther, R., Schupp, I., Poustka, A. and Wiemann, S., 2004. SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. BMC Genom. 5, 36.

Widjojoatmodjo, M.N., Fluit, A.C. and Verhoef, J., 1995. Molecular identification of bacteria by fluorescence-based PCR-single-strand conformation polymorphism analysis of the 16S rRNA gene. J. Clin. Microbiol. 33, 2601-2606.

Williams, L.E., Pittman, J.K. and Hall, J.L., 2000. Emerging mechanisms for heavy metal transport in plants. Biochim. Biophys. Acta, Biomembranes, 1465, 104-126.

Wong, H.L., Sakamoto, T., Kawasaki, T., Umemura, K. and Shimamoto, K., 2004. Down-regulation of metallothionein, a reactive oxygen scavenger, by the small GTPase OsRac1 in rice. Plant Physiol. 135, 447-1456.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S. and Basham, D., 2002. The genome sequence of Schizosaccharomyces pombe. Nature, 415, 871.

Wuana, R.A. and Okieimen, F.E., 2011. Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. ISRN Ecol. 2011, 1-20.

Xue, T., Li, X., Zhu, W., Wu, C., Yang, G. and Zheng, C., 2008. Cotton metallothionein GhMT3a, a reactive oxygen species scavenger, increased tolerance against abiotic stress in transgenic tobacco and yeast. J. Exp. Bot. 60, 339-349.

Yadav, R.K., Barbi, F., Ziller, A., Luis, P., Marmeisse, R., Reddy, M.S. and Fraissinet-Tachet, L., 2014. Construction of sized eukaryotic cDNA libraries using low input of total environmental metatranscriptomic RNA. BMC Biotechnol. 14, 80.

Yang, M., Pan, C., Bo, J. and Wang, K., 2014. Regulation of metallothionein gene expression in response to benzo [a] pyrene exposure and bacterial challenge in marine cultured black porgy (Acanthopagrus schlegelii). Chin. J. Geochem. 33, 404-10.

Zhang, J., Schneider, C., Ottmers, L., Rodriguez, R., Day, A., Markwardt, J. and Schneider, B.L., 2002. Genomic scale mutant hunt identifies cell size homeostasis genes in S. cerevisiae. Curr. Biol. 12, 1992-2001.

Zhang, M., Liu, S., Takano, T. and Zhang, X., 2019. The interaction between AtMT2b and AtVDAC3 affects the mitochondrial membrane potential and reactive oxygen species generation under NaCl stress in Arabidopsis. Planta, 249, 417-429.

Zhang, Y., Zhang, X., Zhang, H., He, Q., Zhou, Q., Su, Z. and Zhang, C., 2009. Responses of soil bacteria to long-term and short-term cadmium stress as revealed by microbial community analysis. Bull. Environ. Contam. Toxicol. 82, 367-372.

Zhao, J., Missihoun, T.D. and Bartels, D., 2017. The role of Arabidopsis aldehyde dehydrogenase genes in response to high temperature and stress combinations. J. Exp. Bot. 68, 4295-4308.

Zhou, J., Xia, B., Treves, D.S., Wu, L.Y., Marsh, T.L., O'Neill, R.V., Palumbo, A.V. and Tiedje, J.M., 2002. Spatial and resource factors influencing high microbial diversity in soil. Appl. Environ. Microbiol. 68, 326-334.

Zhou, Q., Snipas, S., Orth, K., Muzio, M., Dixit, V.M. and Salvesen, G.S., 1997. Target protease specificity of the viral serpin CrmA Analysis of five caspases. J. Biol. Chem. 272, 7797-7800.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. and Siebert, P.D., 2001. Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. Biotechniques, 30, 892-897.

Ziller, A., Yadav, R.K., Capdevila, M., Reddy, M.S., Vallon, L., Marmeisse, R., Atrian, S., Palacios, Ò. and Fraissinet-Tachet, L., 2017. Metagenomics analysis reveals a new

metallothionein family: Sequence and metal-binding features of new environmental cysteine-rich proteins. .J Inorg. Biochem. 167, 1-11.

Ziman, M., Chuang, J.S. and Schekman, R.W., 1996. Chs1p and Chs3p, two proteins involved in chitin synthesis, populate a compartment of the Saccharomyces cerevisiae endocytic pathway. Mol. Biol. Cell, 7, 1909-1919.

Appendix I

Media

Synthetic defined media w/o uracil

Composition	***Quantity (g/L)***
Yeast Nitrogen Base w/o amino acids	6.7
Glucose	20
1%Adenine Stock	1 ml
1% Tyrosine stock	5 ml
Dropout supplement –Ura	0.77
Agar	15
pH	5.9 ± 0.2

LB broth/agar

Composition	***Quantity (g/L)***
NaCl	10.0
Beef extract	5.0
Tryptone	10.0
Agar (for agar plates)	10.0

SOC broth

Composition	***Quantity (g/L)***
Casein enzymic hydrolysate	20.0
Yeast extract	5.0
Sodium chloride	0.5
Magnesium sulphate	2.4
Potassium chloride	0.186
Final pH (at 25°C)	7.0± 0.2

YPD broth/agar

Composition	*Quantity (g/L)*
Peptic digest of animal tissue	20.0
Yeast extract	10.0
Dextrose	20.0
Final pH (at 25°C)	6.5±0.2
Agar (for agar plates)	10

Reagents and Buffers

50X TAE Electrophoresis buffer

Composition	*Quantity (g/L)*
Tris free base	242 g
Disodium EDTA	18.61 g
Glacial Acetic Acid	57.1 ml
Distilled water	to 1000 ml

Add the Tris free base and EDTA to approximately 700 ml distilled water and stir until the Tris and EDTA are dissolved. Add the acetic acid and adjust the volume to 1 liter.

The 1x TAE solution is 40 mM Tris, 20 mM Acetate and 1 mM EDTA.

Agarose gel loading dye (6X)

Bromophenol blue	0.25%
Glycerol in water	30.0%
TE buffer 10X	0.1 M (pH 8)
Tris-HCl	10 M (pH 8)
EDTA disodium salt	100 mM

1M Potassium Phosphate Buffer

To make 100 mL of 1M potassium phosphate:

1M Monopotassium phosphate	13.4 ml
1M Dipotassium phosphate	86.6 ml
pH	7.6±0.2

Plasmid DNA isolation buffers

Soluton I (Resuspension buffer)

Glucose	50 mM
EDTA(Ethylenediaminetetraacetic acid)	10 mM
Tris (pH 8.0)	25 mM

Soluton II (Lysis solution)

NaOH	0.2 N
SDS (Sodium dodecyl sulphte)	100 mM

Soluton III (Neutralizing solution)

KOAc (pH 6.0)	3 M
Glacial acetic acid	11.5 ml
Distilled H_20	28.5 ml

Yeast breaking buffer

TritonX-100	10 ml
SDS (10%)	50 ml
NaCl (5 M)	10 ml
Tris-Cl (1 M, pH 8.0)EDTA (0.5 M, pH 8.0)	5 ml
EDTA (0.5 M, pH 8.0)	1 ml
Distilled H_20	Upto 500 ml

Appendix II

Uncultured eukaryote clone PLCc38 aldehyde dehydrogenase mRNA, complete cds

GenBank: MK077675.1

```
LOCUS       MK077675                1571 bp    mRNA    linear   ENV 02-
APR-2019
DEFINITION  Uncultured eukaryote clone PLCc38 aldehyde dehydrogenase
            mRNA,complete cds.
ACCESSION   MK077675
VERSION     MK077675.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 1571)
  AUTHORS   Mukherjee,A., Yadav,R., Marmeisse,R., Fraissinet-Tachet,L.and
            Reddy,M.S.
  TITLE     Heavy metal hypertolerant eukaryotic aldehyde dehydrogenase
            isolated from metal contaminated soil by metatranscriptomics
            approach
  JOURNAL   Biochimie (2019) In press
   PUBMED   30905733
  REMARK    Publication Status: Available-Online prior to print
REFERENCE   2  (bases 1 to 1571)
  AUTHORS   Mukherjee,A., Yadav,R., Marmeisse,R., Fraissinet-Tachet,L.and
            Reddy,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2018) Department of Biotechnology, Thapar
            Institute of Engineering & Technology, Bhadson Road, Patiala,
            Punjab 147004, India
COMMENT     ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..1571
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="soil"
                     /db_xref="taxon:100272"
                     /clone="PLCc38"
                     /environmental_sample
                     /country="France: northwest of Paris, Pierrelaye,
                     agro-forestry site"
     CDS             113..832
                     /EC_number="1.2.1.3"
                     /function="involved in metal tolerance"
```

```
                     /codon_start=1
                     /product="aldehyde dehydrogenase"
                     /protein_id="QBQ70680.1"

/translation="MQRGKAEWMTTPAPNRGEIVRQIGDAFRKHKDALGRLISLEMGKILSEGLGEVQEFIDVC
DMAVGMSRTIDGKILNSERPNHFMMEQWNPLGLIGVITAFNFPCAVFGWNFAIAAICGDLTLWKGSESTSLVTL
ATAKIVAEVLERNKCPEGVMTVVSGEGKTIGELITNDPRMELVSFTGSTFVGRRVSEIVHRRFGRTVLELGGXN
AIIVCEDGDTYGLKSCYLRCCRNLRTEMHNR"

ORIGIN
   1 gcgggggatc ggtgaaggtg atgtttacaa tagcatcaac cctcacaatg gagaggtgat
  61 cgccaccatc cgctacggaa ccaaaaccaa ctacgaagaa gctgtacagt ccatgcaacg
 121 aggaaaggct gagtggatga ccacccctgc tccaaacaga ggagaaatcg tcagacagat
 181 cggtgacgcc ttcagaaaac acaaggacgc attgggaaga ctcatctcgc ttgaaatggg
 241 aaagatcttg agtgaaggcc ttggagaggt ccaagaattc attgacgtct gcgacatggc
 301 agtcggcatg agcagaacta tcgatggaaa gatcttgaac tccgaaagac caaaccactt
 361 catgatggaa caatggaatc ctcttggtct catcggtgtc atcactgcct tcaacttccc
 421 ttgcgccgtc ttcggctgga actttgctat cgctgctatc tgcggtgact tgactctttg
 481 gaaaggctct gaaagcactt cattggtgac actcgccact gcaaaaatcg ttgctgaagt
 541 ccttgagaga aacaagtgcc cagaaggtgt catgaccgtc gtttctggtg aaggaaagac
 601 aattggagag cttatcacca atgacccaag aatggaactc gtctcattca ctggatcaac
 661 ctttgtggga agaagagtca gcgagatcgt ccacagaaga ttcggaagaa ctgtcctcga
 721 actcggagga nacaatgcta ttattgtttg cgaggacgga gatacatatg gccttaagag
 781 ctgttacctt cgctgctgtc ggaacctgcg gacagagatg cacaaccgtt agaagactga
 841 tcgtccacga gaagcactat gacggtttcg tcaagcagtt gcttgaagtc tacccaacta
 901 tcaagatcgg caacccaatg gaaagaaggt actctctgcg gtccacttca caccaagaac
 961 gctgttaaga tgtatgccca aggaatcgaa aagatcaaat ctcaaggagg aaaggtcctg
1021 actggaggca aagtccttga taatgaattc cctggaggca attacgtggc tccactattg
1081 tcgaaatcga ttctagtgct gaaatcgtca aggaagaatt gttcgctcca gtcctctacg
1141 ttctcaagtt caaaactttc gaagaggcca ttaagatcaa caatggcgtc ccacaaggat
1201 tgagcagtgc tatcttcaca aaagatctcc aaaagttgtt caagtggatt ggtcctcttg
1261 gctctgattg cggtattgtg aactgcaaca ttggtacatc aggtgctgag atcggaggtg
1321 ccttcggagg tgagaaagag actggaggtg gccgagaatc tggttctgac tcttggaagc
1381 agtacatgaa aagagcgaca tgcacgatta acttcggaaa tactcttcca cttgcacagg
1441 gagttaaatt cgatttaaaa ccaaagattt agggctatga tcagatagaa ttacatgaat
1501 aatgatgagt gagagtcttc cagaattgaa taagtgaggg atatgaatct atttaagcca
1561 gaataacaca c
//
```

Uncultured eukaryote clone PLCg49 serine protease inhibitor-like protein (serpin) mRNA, complete cds

GenBank: MK079356.1

```
LOCUS       MK079356                1409 bp    mRNA    linear   ENV 02-
APR-2019
DEFINITION  Uncultured eukaryote clone PLCg49 serine protease inhibitor-
like
            protein (serpin) mRNA, complete cds.
ACCESSION   MK079356
VERSION     MK079356.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 1409)
  AUTHORS   Mukherjee,A., Yadav,R., Marmeisse,R., Fraissinet-Tachet,L.
      and Reddy,S.M.
  TITLE     Metatranscriptomics approach to isolate a eukaryotic serine
            protease inhibitor from metal contaminated soil having high
           heavy metal tolerant properties
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1409)
  AUTHORS   Mukherjee,A., Yadav,R., Marmeisse,R., Fraissinet-Tachet,L.and
            Reddy,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2018) Department of Biotechnology, Thapar
            Institute of Engineering & Technology, Bhadson Road, Patiala,
            Punjab 147004, India
COMMENT     ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..1409
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="soil"
                     /db_xref="taxon:100272"
                     /clone="PLCg49"
                     /environmental_sample
                     /country="France: Pierrelaye, North-West of Paris"
     gene            1..1409
                     /gene="serpin"
     CDS             69..1301
                     /gene="serpin"
                     /function="involved in metal tolerance"
                     /codon_start=1
                     /product="serine protease inhibitor-like protein"
                     /protein_id="QBQ70681.1"

/translation="MHSLFKGVMIIFGLALATSSPVPLCPAATREQSQQAEAALADFAVNLFKNVSAQDEDKNQ
VMSPVSIALALAMLENGADAHTREQLKHALVGVGSSADVLAAYREIQKHLRADDEKTKLTVANGLFQDKDLKLK
```

DQYVSTTRDCLETEVDNKNDFSDHLEDARQKINKWVSQKTSEKIPELFKRGVLNEGDRMVLANAIYFKASWRNS
FNKAMTKQDTFYRNGHDQQSVQFMHETGDHRHVSSNDLDAVELAYSHPDLAMVVVLPKARDGLRNLEKHLTGKQ
LRDIVASMEQRKVKVQLPKFTVRSPTDLKNVLSKMGLDHMFSNQANFSRMSETPLKVDNAIHEAYIDVNENGTE
GAAATGVSINMRAMPSMPVDPVSFVADHPYLYAVVHKPTGAIVFLGKVNSVESH"
ORIGIN

```
   1 gcgggggtac ttttcggcgg ccacgttcga cgtctcttct tactttgaga ttattcgaat
  61 cagctaagat gcattcctta ttcaagggag tgatgatcat ctttggtctg gctttggcca
 121 ccagctcgcc agttccgctg tgcccagcgg caacgagaga gcaatcgcag caagctgaag
 181 cagctttggc tgacttcgcc gtcaaccttt tcaaaaatgt ctcggctcaa gacgaagaca
 241 aaaatcaagt catgtccccg gtttccatcg ctctggcttt ggccatgttg gaaaacggag
 301 ccgatgccca cactcgcgaa caactcaaac acgccttggt cggagttgga tccagcgccg
 361 acgtcttggc cgcataccgc gaaatccaaa agcacttgcg cgccgatgac gagaaaacga
 421 aattgaccgt cgccaacggc ctcttccagg acaaggatct gaaattgaag gaccagtacg
 481 tcagcactac ccgcgactgc ttggaaaccg aagtggacaa caagaacgac ttctctgacc
 541 acctcgagga cgcgcgtcaa aaaatcaaca aatgggtctc ccaaaagacc tccgagaaga
 601 tcccagagct cttcaagcga ggcgtcttga acgaaggtga ccgtatggtt ttggccaacg
 661 ccatctactt caaagcttcg tggaggaact cgttcaacaa agcaatgacc aagcaagaca
 721 ccttctaccg caacggtcac gaccagcaat ctgtccagtt catgcacgaa actggtgacc
 781 accgtcacgt ttcgtcgaat gatttggatg ccgtcgaatt ggcctactct catcccgatt
 841 tggccatggt cgtcgtcttg ccgaaagctc gcgacggttt gcgcaacttg gagaaacact
 901 tgaccggcaa acaattgagg gacatcgtcg ctagcatgga acaacgcaag gtcaaggtcc
 961 aattgccaaa attcaccgtt cgctcgccca ccgatttgaa aaatgtcctg agcaaaatgg
1021 gtttggatca tatgttcagc aaccaagcca acttcagccg catgagcgag acccctttga
1081 aggtcgacaa cgctatccat gaagcttaca tcgatgtcaa cgaaaatgga accgaaggcg
1141 ctgctgcaac cggtgtttcg atcaacatgc gtgccatgcc ctcgatgcca gtagacccag
1201 tgtcgttcgt cgctgaccac ccatacttgt acgccgttgt gcacaagcca accggcgcga
1261 tcgttttctt gggcaaagtc aactcggtcg aatcgcacta gacttccagc gattacatga
1321 aaacgctaat gcttttatga cgtcttgttt ggtaaattct aacagccaat aaaagcagtt
1381 gacgaaaaaa aaaaaaaaaa aaaacaaaa
//
```

Uncultured eukaryote clone PLCg39 DHHC zinc finger domain-containing protein mRNA, complete cds

GenBank: MK629535.1

```
LOCUS       MK629535                1355 bp    mRNA    linear   ENV 24-
JUN-2019
DEFINITION  Uncultured eukaryote clone PLCg39 DHHC zinc finger
            domain-containing protein mRNA, complete cds.
ACCESSION   MK629535
VERSION     MK629535.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 1355)
  AUTHORS   Mukherjee,A., Yadav,R., Marmeisse,R., Fraissinet-Tachet,L.and
       Reddy,M.S.
  TITLE     DHHC domain containing protein isolated from metal polluted
       soil having high heavy metal tolerant properties
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1355)
  AUTHORS   Mukherjee,A., Yadav,R., Marmeisse,R., Fraissinet-Tachet,L.and
            Reddy,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2019) Department of Biotechnology, Thapar
            Institute of Engineering & Technology, Bhadson Road, Patiala,
            Punjab 147004, India
COMMENT     ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..1355
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="Agro-Forestry Soil"
                     /db_xref="taxon:100272"
                     /clone="PLCg39"
                     /environmental_sample
                     /country="France: Pierrelaye, Northwest of Paris"
     CDS             180..1145
                     /note="involved in metal tolerance; DHHC
                     palmitoyltransferase"
                     /codon_start=1
                     /product="DHHC zinc finger domain-containing protein"
                     /protein_id="QDD67720.1"

/translation="MKRQYQLWPGKSHFLCWGRCVVGPDRKYFLLAVAFIVVPAVFYCGLVWPYLFWKLEPYIS
VPLAVFYFSILLCILTFMMLTRYRDPGIIPRGRAIEYDPENPWDYEEKKPAETMRLNLKGDPKFRVKYCETCHI
YRPPRCTHCAVCNNCVERFDHHCPWVGNCIGRRNYQTFLAFVWSVILGCVYICLLSIAHLAIVIVEAIHEEGSG
AYVFLYILRYGFFSVITLGYCAVAIGLVGFLGTFHCLLVCKNQTTNEKLKGLYKKKMNPYSKGPCLNFISILWA
PVFPRPVNWRERVDEDTAKLVKRKAWPKDKRKASARRTK"
ORIGIN
```

```
   1 gcggggtccg cgaaagaact gttgcaacaa agaaggctga gcttgtaggt gcggggtcag
  61 cgcgagtcgt tttctcctcc ctctcttgct tgggtgggtg tgcacgggcg agacaaatcg
 121 caagctagta cgattccttc gatcgataaa acaagccacg cacatcaaga gaacagctaa
 181 tgaagagaca atatcagttg tggccaggca aaagccactt cctctgttgg ggtcgttgcg
 241 tagtgggacc tgatagaaag tatttcctgc ttgcggtggc gtttatcgtc gtacctgctg
 301 tattttattg tgggcttgta tggccttacc tgttttggaa acttgagccc tacatatcag
 361 taccgttagc ggtgttctac ttctctattc ttctttgtat ccttactttc atgatgctca
 421 cacgatatcg agacccaggc atcattcctc ggggtcgcgc aatcgagtat gacccagaaa
 481 atccttggga ctatgaagag aaaaagccag cggagacaat gcgacttaat ctgaaaggag
 541 atccaaaatt tcgtgtcaaa tattgcgaaa catgtcatat ctaccggccg cctcgatgca
 601 cgcactgcgc cgtgtgcaat aattgcgtag agcgctttga tcaccactgc ccttgggttg
 661 gcaattgcat aggtagaagg aattaccaga cgtttctcgc ctttgtgtgg agcgtgattt
 721 tggggtgtgt ctacatctgc ctgttgagta tcgcccacct ggctatcgtt atcgtagagg
 781 cgatacatga agagggtagc ggtgcctatg tgttcctata catactacga tatggctttt
 841 tttcggttat aacattgggg tattgtgcgg ttgcaatagg gcttgtagga ttcttgggca
 901 catttcattg cttgctagtg tgcaaaaatc agactacaaa cgagaagctt aaaggcttgt
 961 acaagaagaa aatgaatccg tactcaaaag ggccgtgtct gaactttatc tctattctgt
1021 gggcacctgt cttcccgaga cctgtcaatt ggcgagagcg tgtggatgaa gatacggcca
1081 agctggtcaa gcggaaagca tggccgaagg acaagaggaa agcatcagca aggcggacaa
1141 agtaaaaagc ggaccaagct tgtccttcct accagcgggt accgcactat gcggcccgta
1201 tctttttttcg agaggacaga gggtgggatg ttacatgttg cctgccaaac aacaacgaca
1261 acaagtaagg gaaccagtcg gtgtgtatag aggattacct ggtgttctat ttattgtatt
1321 tatttctttt taactttcaa aaaaaaaaaa aaaaa
//
```

Uncultured eukaryote clone ARK3 Maltose fermentation regulatory protein MAL33-like protein, complete cds

Genbank: MN794026

```
LOCUS       ARK3                    523 bp    mRNA    linear   ENV 05-
DEC-2019
DEFINITION  Uncultured eukaryote clone ARK3.
ACCESSION   ARK3
VERSION
KEYWORDS    .
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Mukherjee,A. and Reddy,M.S.
  TITLE     Maltose fermentation regulatory protein MAL33-like protein
            isolated
            from copper mine soil having high copper tolerant properties
  JOURNAL   unpublished
REFERENCE   2  (bases 1 to 523)
  AUTHORS   Mukherjee,A. and Reddy,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2019) Department of Biotechnology, Thapar
            Institute of Engineering & Technology, Bhadson Road, Patiala,
            Punjab 147004, India
COMMENT     Bankit Comment: ALT EMAIL:vasu70@yahoo.com
            Bankit Comment: TOTAL # OF SEQS:1

            ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..523
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="Copper mine soil"
                     /db_xref="taxon:100272"
                     /clone="ARK3"
                     /country="India"
                     /note="Copper mine soil, Malanjkhand copper mine,
                     Balaghat, Madhya Pradesh, India"
     gene            <12..386
                     /gene="MAL33"
                     /note="Involved in metal tolerance"
     CDS             <12..386
                     /gene="MAL33"
                     /note="Mal33p"
                     /codon_start=1
                     /product="Maltose fermentation regulatory protein
MAL33"

/translation="MNGTKFFSNANNAHILVEIAKDMLDDIFLTPNNLYVVHGPVIPMKALEIANALVDIVNKY
DHNMKLEAWNILCDVSKFVFSLKHCNHKMFQRFSTKCQSALIDLPISRPLRLNDDSKDEDDIIP"
```

```
BASE COUNT      155 a     94 c    102 g    172 t
ORIGIN
  1 tagtgtttca gatgaatggt accaagtttt tctcaaacgc caataatgct cacatattgg
 61 tcgaaattgc aaaggatatg ctggacgaca tattcttaac tccaaacaac ctgtatgtag
121 tccatggtcc tgtgatacca atgaaggcac tggaaatagc caatgcattg gtagatatcg
181 taaataagta tgatcacaat atgaagttgg aggcttggaa tattttgtgc gatgtatcta
241 agttcgtttt ctcccctgaaa cattgcaatc ataaaatgtt tcaaaggttt tcaactaaat
301 gtcagagtgc tctaatcgat ttgcctatct ctagaccact gcgcctaaat gatgattcca
361 aagatgaaga cgacataatt ccttaattta ttgttcacgc cgttcactta tacgagatag
421 atatactgat agagtgtgag tgatattctt aagtcttgct tttcgagggt gtaagaagct
481 atgttcttca ggcgagatta ttctactcct gccttacttg ttt
//
```

International Biodeterioration & Biodegradation 143 (2019) 104718

Contents lists available at ScienceDirect

International Biodeterioration & Biodegradation

journal homepage: www.elsevier.com/locate/ibiod

Detoxification of toxic heavy metals by serine protease inhibitor isolated from polluted soil

Arkadeep Mukherjee[a], Rajiv Yadav[b,1], Roland Marmeisse[b], Laurence Fraissinet-Tachet[b], M. Sudhakara Reddy[a,*]

[a] Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab, 147004, India
[b] Ecologie Microbienne, UMR CNRS, UMR INRA, Université Claude Bernard Lyon 1 Université de Lyon, F-69622 Villeurbanne, France

ARTICLE INFO

Keywords:
Metatranscriptomics
Eukaryotes
Heavy metal tolerance
Yeast complementation
Serine protease inhibitor

ABSTRACT

Various anthropogenic activities attribute to the addition and accumulation of heavy metals in soil which creates an environment unsuitable for all organisms. Despite such conditions, some microbial populations are able to survive and grow due to functioning of certain genes. Investigating the role of eukaryotic microbes in metal contaminated soil was the focus of this study. Culture independent approach was employed and size fractionated eukaryotic cDNA libraries (library A < 0.5 kb, library B 0.5–1.0 kb, and library C 1.0–4.0 kb) of soil were constructed from soil RNA and were screened for cadmium (Cd) tolerance by functional complementation assay by using Cd sensitive yeast mutant *ycf1*$^{\Delta}$. Screening of the library C yielded clones capable of growing in presence of Cd amended media. In the present study, one of the clones, PLCg49 was selected because of its ability to grow at high concentrations of Cd (40–100 μM). PLCg49 was also found to be highly inducible in presence of Cd where the highest fold expression was observed at 60 μM of Cd. Sequence analysis of PLCg49 showed homology with Serine Protease Inhibitor (Serpin) of the eukaryote *Hypsibius dujardini*. Serpins are a broad family of protease inhibitors which are majorly stress-inducible genes usually triggered by biotic and abiotic stresses and help in maintaining cell viability. Copper-sensitive (*cup1*$^{\Delta}$), Zn-sensitive (*zrc1*$^{\Delta}$) and Cobalt-sensitive (*cot1*$^{\Delta}$) yeast mutants rescued their sensitivity when transformed with PLCg49, indicating its tolerance to different heavy metals. The protease sensitive yeast mutant (*pbi2*$^{\Delta}$) transformed with PLCg49 protected from the cell lysis when grown in presence of trypsin and proteinase K confirming the role of PLCg49 as an effective serpin. Present study results suggested that PLCg49 serves as a potential metal tolerant gene, helping us to understand new pathways participating in metal tolerance.

1. Introduction

Soil provides a solid foundation to numerous forms of life to survive and grow. Soil, which is rich in essential nutrients, serves as a reservoir of diverse microbial population and their associated biological processes. Any condition which alters the property of soil would have a direct impact on the inhabiting microorganisms. One such condition is contamination of soil with heavy metals due to various anthropogenic activities which causes toxic effects on the soil microorganisms thus disturbing their normal functioning (Damodaran et al., 2013; Pennanen et al., 1996; Utgikar et al., 2003). Heavy metals cause blocking of functional groups of enzymes causing conformational modifications, denaturation and thus inactivation of enzymes (Ochiai, 2012; Li et al., 2015). In addition, toxic levels of heavy metals in soil may stimulate the generation of free radicals and reactive oxygen species, leading to oxidative stress (Dietz et al., 1999). This causes an overall decline in microbial biomass of the soil ecosystem. However, despite the high levels of heavy metals, some microorganisms have adopted various survival mechanisms to attain resistance or tolerance mostly through metal homeostasis and metabolism (Oyetibo et al., 2017). Generally, they 'avoid' external metals from damaging the cells by extracellular precipitation, reduced uptake, biosorption to cell walls etc., or 'tolerate' metals by intracellular chelation through metallothioenins, phytochelatins and or compartmentation within vacuoles (Clemens, 2001; Blindauer and Leszczyszyn, 2010; Wysocki and Tamás, 2010). Understanding these survival mechanisms in metal polluted soil at the cellular level faces a major challenge since majority of the microbial population is uncultivable under standard laboratory conditions (Torsvik and

* Corresponding author. Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, 147 004, Punjab, India.
E-mail address: msreddy@thapar.edu (M.S. Reddy).
[1] Presently at Department of Botany, University of Allahabad, Allahabad, Uttar Pradesh, India.

https://doi.org/10.1016/j.ibiod.2019.104718
Received 15 March 2019; Received in revised form 17 May 2019; Accepted 13 June 2019

Øvreås, 2002). A bulk of the uncultivable microorganisms are eukaryotes which provides a rich genetic pool of novel genes involved in metal tolerance. Discovering these genes and their mechanisms in metal tolerance would be of immense importance and could be directly implemented in biotechnological applications.

In order to study the functions of microbial communities from the environmental samples, metagenomics approach has been practiced initially which involves the isolation of genomic DNA followed by sequencing and annotations. Though this approach provide information on the potential of a gene to express but cannot be related to the actual microbial activity in soil at a particular time or during a change in environmental conditions. To overcome this situation, metatranscriptomics approach can be used to study the diversity of the active genes in microbial communities, to quantify their expression levels and to monitor how the expression levels change in different conditions. This method starts with the extraction of RNA from environmental sample followed by reverse transcribing to cDNA which allows specific characterization of genes expressed by microorganisms. This technology has been applied to study complex ecosystems like gut microbiome of humans and animals (Bikel et al., 2015; Bashiardes et al., 2016), extreme environments (Dopson and Holmes, 2014; Tveit et al., 2014), in soil to tap the genetic resources for industrially important enzymes (Damon et al., 2011; Kellner et al., 2011) and for metal tolerance (Thakur et al., 2018).

In the present study, size fractionated eukaryotic cDNA libraries (library A: 0.1–0.5 kb, library B: 0.5–1 kb and library C: 1–4 kb) were prepared from the metal polluted soil and screened through yeast functional complementation to isolate genes involved in metal tolerance. Screening of one of the size-fractionated libraries (library C) led to the isolation of several genes involved in metal tolerance including the cDNA PLCg49 which showed homology to serine protease inhibitors (Serpin) family with high tolerance towards toxic concentrations of Cd, Cu, Zn and Co. Protease inhibitors are present extensively in plants, animals and insects. They form a complex with protease thus inhibiting their proteolytic activity. These are grouped into four major classes, i.e. cysteine, serine, aspartic and metallo protease inhibitors. Out of these, serine protease inhibitors are most abundant and responsive towards environmental stresses (Rehman et al., 2017). A number of abiotic stress conditions like hypersalinity, radiation, osmotic stress and heavy metal stress lead to the induction of protease inhibitor genes (Lopez et al., 1994; Conconi et al., 1996; Huang et al., 2007; Domash et al., 2008). This study is intended to characterize the Serpin with respect to its potential role in heavy metal tolerance.

2. Materials and methods

2.1. Soil sampling and total RNA extraction

Soil samples were collected from the agro-forest soil populated by Poplar trees in Pierrelaye (PL) (49°1′45″N, 2°10′32″E) near Paris, France contaminated with heavy metals (Thakur et al., 2018). To encompass a large diversity of soil microbial activity, composite soil was prepared by mixing equal volumes from each core sample. They were then sieved (2 mm mesh), transferred in sterile plastic bags and preserved at −70 °C until RNA extraction. Soil sample was analyzed for its physicochemical properties i.e. pH, organic carbon (Walkley, 1947), total P (Kitson and Mellon, 1944) available P (Olsen, 1954) and total nitrogen (Piper, 1966). To determine the metal content in the soil, the soil samples were digested with nitric acid and perchloric acid in a ratio 2:1 (v/v) (Idera et al., 2014). The content of metals were analyzed by ICP-MS (Element XR, ThermoFisher Scientific, Germany). Soil sample was crushed into a fine powder by using liquid nitrogen and total RNA was extracted using the RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories, Carlsbad, CA) (Damon et al., 2011). The integrity and purity of RNA samples were validated by electrophoresis and by nanodrop spectrophotometry (Thermo Fisher, USA) respectively.

Table 1
Primers used for qPCR analysis in this study.

Gene/Sequence name	Primer name and sequence (5′- 3′)
α-Actin (YFL039C)	qACT1F CGAATTGAGAGTTGCCCCAG
	qACT1R CAAGGACAAAACGGCTTGGA
TATA-Box Binding Protein Associated Factor 10 (YDR167W)	qTAF10F GCTAACAACAGTCAGGCGAG
	qTAF10R GAGCCCGTATTCAGCAACAG
Transcription initiation factor 1 (YBR123C)	qTFCF ACGACGCTGCTTTGGAAAAT
	qTFCR GCTTTTCATTGTTTCCGCGG
Serine Protease Inhibitor	qSERF ATGAGCGAGACCCCTTTGAA
	qSERR CAACGGCGTACAAGTATGGG

Table 2
Physicochemical properties of the soil samples collected from Pierrelaye, northwest of Paris, France used in this study for metatranscriptomic analysis.

Parameters	Content
pH	7.15
Organic carbon (%)	1.6
Available P (mg kg^{-1})	14.2
Total P (mg kg^{-1})	291
Total Nitrogen (%)	0.12
Metal Content (mg kg^{-1})	
Cd	2.5
Cu	64
Zn	385

2.2. Biological materials

All yeast mutant and wild type strains were procured from Euroscarf (http://www.euroscarf.de). Cadmium-sensitive *ycf1*$^{\Delta}$, Copper-sensitive *cup1*$^{\Delta}$ (DTY4), Cobalt-sensitive cot*1*$^{\Delta}$ and Zinc-sensitive *zrc1*$^{\Delta}$ mutant strains of *Saccharomyces cerevisiae* were used in this study. *ycf1*$^{\Delta}$, cot*1*$^{\Delta}$ and *zrc1*$^{\Delta}$ strains were derived from the wild-type strain BY4741 (*MATα, his3*$^{\Delta}$*1, leu2Δ0, met15*$^{\Delta}$*0, ura3Δ0*) and the *cup1*$^{\Delta}$ strain from DTY3 (*cup1*^{s}) (*MATα, leu2-3, 112his3*$^{\Delta}$*1, trp1-1, ura3-50, gal1, CUP1*^{S}). *ycf1*$^{\Delta}$ is devoid of an ABC transporter gene YCF1 (Li et al., 1997), deletion of the ZRC1 gene encoding for transporter protein in *zrc1*$^{\Delta}$ (Li and Kaplan, 2001) and deletion of COT1 gene in cot1$^{\Delta}$ (Conklin et al., 1992) imparted hypersensitivity to these strains towards their respective metals. The *cup1*$^{\Delta}$ (DTY4) is completely devoid of the copperthionein gene (Longo et al., 1996). Also, serine protease sensitive mutant strain *pbi2*$^{\Delta}$ (BY4741;*YNL015w:KanMX*) was used for protease treatment experiments. Disruption of the YNL015w gene led to sensitivity towards serine proteases (Schu et al., 1991). Growth medium for these yeasts was synthetic defined medium without uracil (SD-Ura), which is a minimal medium supplemented with 2% glucose (Hi-media laboratories, India) and drop out supplement without uracil (Takara-Clontech, Japan). For bacterial transformations, One Shot® *TOP10* Electrocomp™ *E. coli* (Thermofisher, USA) and Luria broth (LB) medium (Hi-media laboratories, India) was used. For protease treatment experiments, trypsin from bovine pancreas (EC 3.4.21.4: 2000 U/mg) and proteinase K (EC 3.4.21.64: 30 U/mg) were obtained from Hi-media laboratories, India.

2.3. Screening of cDNA library C for metal tolerant genes

Synthesis of cDNAs from the eukaryotic mRNA and construction of size fractionated libraries (library A: 0.1–0.5 kb, library B: 0.5–1 kb and library C: 1–4 kb) were performed as described in Yadav et al. (2014). The library C was further transformed into Cd-sensitive yeast mutant *ycf1*$^{\Delta}$ using lithium acetate method (Gietz et al., 1992) and screened for genes involved in Cd tolerance by growing the transformants on SD-Ura

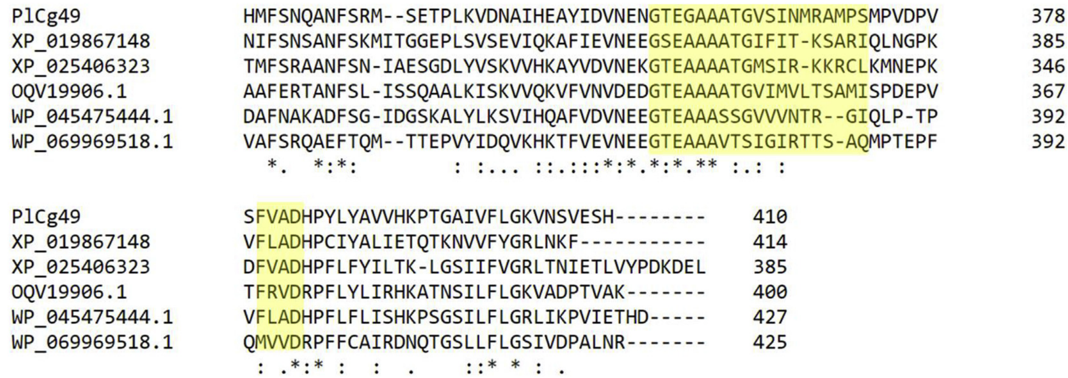


Fig. 1. Multiple sequence alignment of PLCg49 gene with other homologous sequences retrieved by BLASTX analysis. Accession numbers are serpins of XP_019867148 (*Aethina tumida*), XP_025406323 (*Sipha flava*), OQV19906 (*Hypsibius dujardini*, WP_045475444 *Thioploca ingrica* and WP_069969518 (*Desertifilum* sp.). Highlighted regions are conserved Reactive Center Loop (RCL).

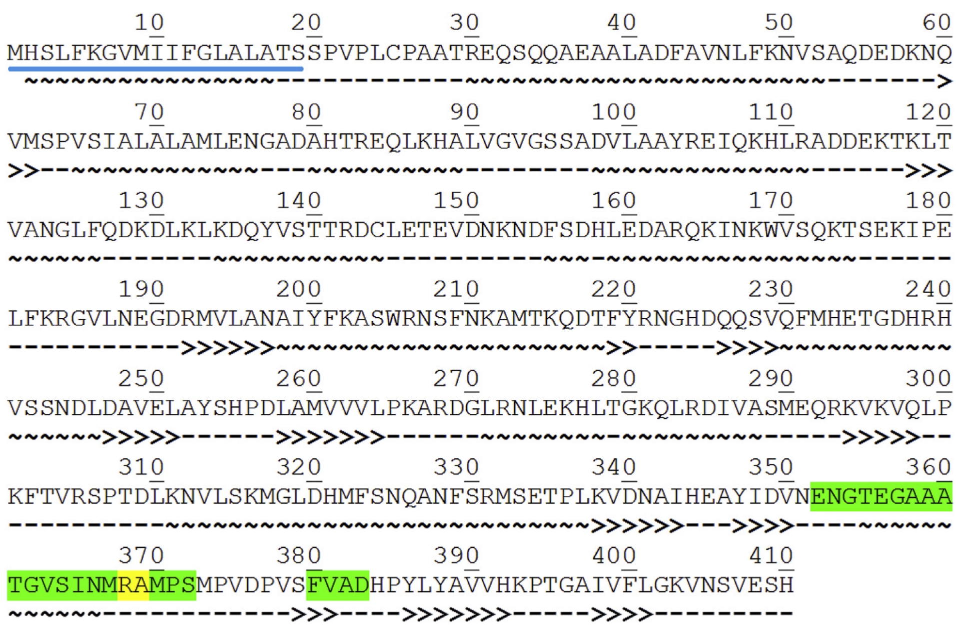


Fig. 2. Deduced amino acid sequence of PlCg49. Signature Reactive Center Loop (RCL) residues are highlighted in green. Protease cleavage site is marked in yellow. Blue underlined region represents the signal peptide where cleavage site is between positions 19 and 20. Predicted secondary structures are represented as '~' for helices, '-' for coils and ' > ' for strands. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

medium supplemented with 40 μM of $CdCl_2.H_2O$ (Cd). Many transformants were obtained on SD-Ura medium supplemented with 40 μM of Cd showing tolerance to Cd. One of the clones, PLCg49 was selected for further studies based on its tolerance to higher concentrations of cadmium and other metals such as copper, zinc and cobalt.

2.4. Sequence analysis

Plasmid DNA was extracted from PFL61/PLCg49 using Zymoprep yeast plasmid Miniprep II (Zymo Research) according to manufacturer's protocol. Insert cDNA PLCg49 was sequenced with vector specific primers NF (5′-CAGATCATCAAGGAAGTAATTATCTAC-3′) and NR (5′-CAGAAAAGCAGGCTGGGAAGC-3′) using Sanger method. Homology search for PLCg49 sequence was performed with BLASTX through NCBI web server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the GenBank nr protein database. Homologous sequences were aligned with PLCg49 using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Transmembrane domain prediction in PLCg49 amino acid sequence was performed using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and prediction of transmembrane helices was performed using HMMTOP server (http://www.enzim.hu/hmmtop/). The tools and databases of CBS Prediction Servers (http://www.cbs.dtu.dk/services/) were used for glycosylation (NetOGlyc 4.0 Server) and phosphorylation sites predictions (NetPhos 3.1 Server). Secondary structure prediction was done by PSIPRED v3.3 (http://bioinf.cs.ucl.ac.uk/psipred/). For phylogenetic analysis, serine protease inhibitor sequences from various taxons were aligned using Clustal W program in MEGA version 7.0 (Kumar et al., 2016) and the phylogenetic tree was reconstructed using neighbor-joining algorithm employing 1000 bootstrap replicates.

2.5. Metal tolerance potential of PLCg49

Yeast mutants *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and cot1$^{\Delta}$ susceptible to copper, cadmium, zinc and cobalt respectively were transformed by the recombinant plasmid pFL61 carrying PLCg49 to study the potential metal tolerance. The transformants were tested for their metal tolerance potential on SD-Ura agar by drop assay. Yeast cultures *cup1*$^{\Delta}$/PLCg49, *ycf1*$^{\Delta}$/PLCg49, *zrc1*$^{\Delta}$/PLCg49 and cot1$^{\Delta}$/PLCg49 were grown in SD-Ura medium overnight and were centrifuged. The culture pellet was washed with sterile water and the OD_{600} was adjusted to 1.0. Serial dilutions of 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} were prepared from these cultures and five μl from all these dilutions were spotted on SD-Ura medium agar plates containing 150 μM $CuSO_4.5H_2O$ (Cu), 40 μM $CdCl_2.H_2O$ (Cd), 10 mM $ZnCl_2$ (Zn) and 2 mM $CoCl_2.6H_2O$ (Co), respectively. Growth assays of *cup1*$^{\Delta}$/PLCg49, *ycf1*$^{\Delta}$/PLCg49, *zrc1*$^{\Delta}$/PLCg49 and cot1$^{\Delta}$/PLCg49 were performed to study the growth pattern in presence of varying concentrations of metals. For growth assay, 50 ml of SD-Ura broth was

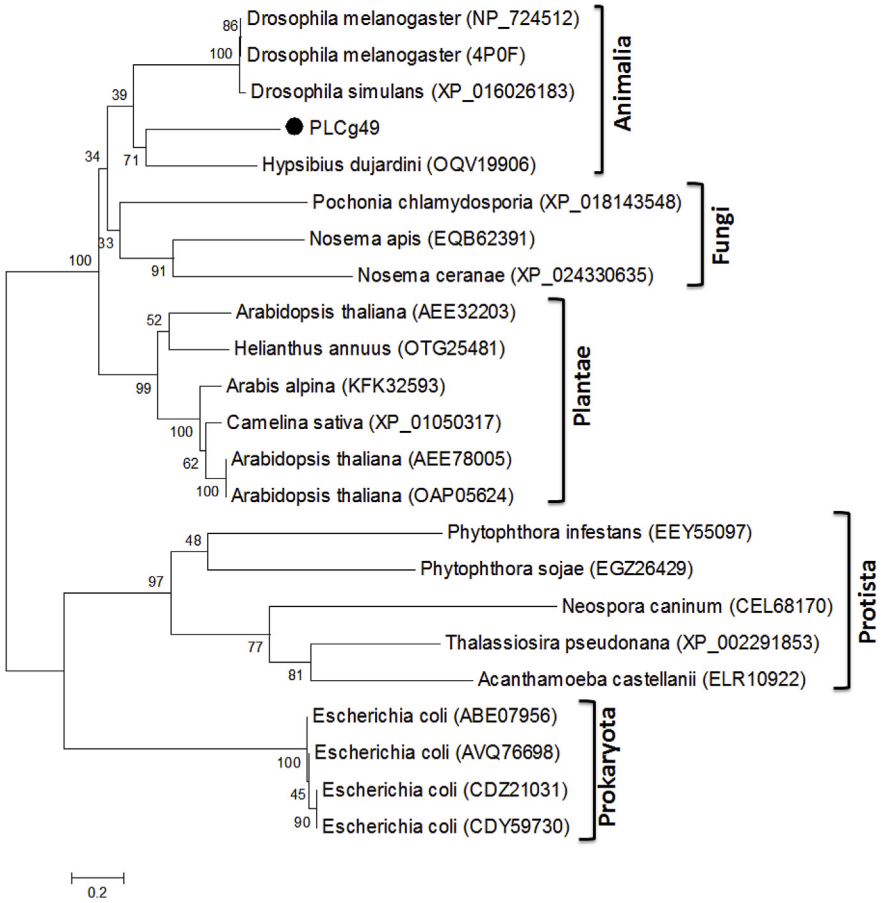


Fig. 3. Phylogenetic relation of the transcript PLCg49 with SERPINs of other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses. Prokaryota Serpin sequences were included as outgroup. Branch lengths are proportional to evolutionary distances.

inoculated with mid-log pre-cultures of yeast cells to achieve an initial OD_{600} = 0.02 and were allowed to grow at 30 °C at 220 rpm for 3 h. Then, varying concentration of $CuSO_4.5H_2O$ (0–1000 μM), $CdCl_2.H_2O$ (0–100 μM), $ZnCl_2$ (0–13 mM) and $CoCl_2.6H_2O$ (0–50 mM) were supplemented separately to the respective mutant yeast cells carrying PLCg49 and their growth was monitored till 40 h at an interval of 3 h. For comparative analysis, all yeast cultures were grown in SD-Ura in the absence of metals. Wild type BY4741 and all mutant yeast strains transformed with empty vector pFL61 were used as controls.

2.6. qPCR analysis of PLCg49 in response to Cd stress

The *ycf1*$^{\Delta}$/PLCg49 cells were grown in separate Erlenmeyer flasks containing 50 ml SD-Ura medium supplemented with 0, 40, 60, 80 and 100 μM of Cd. After 40 h of growth, yeast cells were harvested, washed with sterile distilled water and ground in liquid nitrogen. About 100 mg of the homogenized cells were used for total RNA extraction using TRIzol reagent (Invitrogen, USA) according to manufacturer's instruction. One μg of total RNA was reverse transcribed into cDNA using Primescript first strand cDNA synthesis kit (Takara, Japan) as per the manufacturer's instructions. The relative expression of PLCg49 was studied by Real time PCR analysis using SYBR® Premix Ex Taq (Takara) in Mastercycler®ep realplex system (Eppendorf; Hamburg, Germany) using following program: initial denaturation at 95 °C for 20 s followed by 35 cycles of 95 °C for 15 s, 55 °C for 15 s and 72 °C for 20 s. The E value was obtained in the range of 0.93–1.10 using equation $E = \{10^{(-1/slope)}\}$-1. The E value was used to calculate C_{t1} values using the equation $C_{t1} = C_{te} \times [\log (1 + E)/\log 2]$. Three reference genes α-Actin, TATA-Box Binding Protein Associated Factor 10 and Transcription initiation factor 1 were used as reference genes which were amplified by designing the gene specific primers (Table 1) and their stability values were calculated using NormFinder algorithm (Andersen et al., 2004). Since the stability value of α-actin was minimum, it was used as the reference gene for further calculating the relative gene expression of PLCg49 using $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). This study was performed on RNA extracted from three independent biological samples (experimental replicates) and for each isolated RNA sample, three technical replicates were performed.

2.7. Accumulation of heavy metals in yeast transformants

Yeast cells were grown overnight in 2 ml of SD-Ura medium till the optical density (OD_{600}) reached to 1.0. Cells were washed and re-suspended with sterile Milli-Q water and 500 μl of the re-suspended cells were inoculated into 250 ml Erlenmeyer flasks containing 50 ml each of SD-Ura broth and incubated at 30 °C at 200 rpm. The growth media were spiked with different concentrations of heavy metals (Cu, Cd, Co and Zn) after 5 h similar to the growth kinetics study. Cells were allowed to grow till their log phase and were harvested and washed thrice with 10 mM EDTA solution. Finally, cells were washed with Milli-Q water to remove EDTA and dried for 24 h at 80 °C. Dry yeast biomass was digested with a combination of nitric acid and perchloric acid (2:1 v/v) and the content of metals were analyzed by ICP-MS.

2.8. Response of PLCg49 towards protease treatment

The yeast mutant *pbi2*$^{\Delta}$ was transformed with PLCg49 as well as the empty vector pFL61. Wild type BY4741 transformed with empty pFL61 was used as positive control. Also, PLCg49 was incorporated in wild type BY4741 to understand the combined effect of PLCg49 and the wild type BY4741 in response to protease treatment. These transformants were subjected to the treatment of trypsin and proteinase K to investigate the role of PLCg49 as a serine protease inhibitor. Yeast cell lysis assay followed the protocol as described by Grenier (1994) with

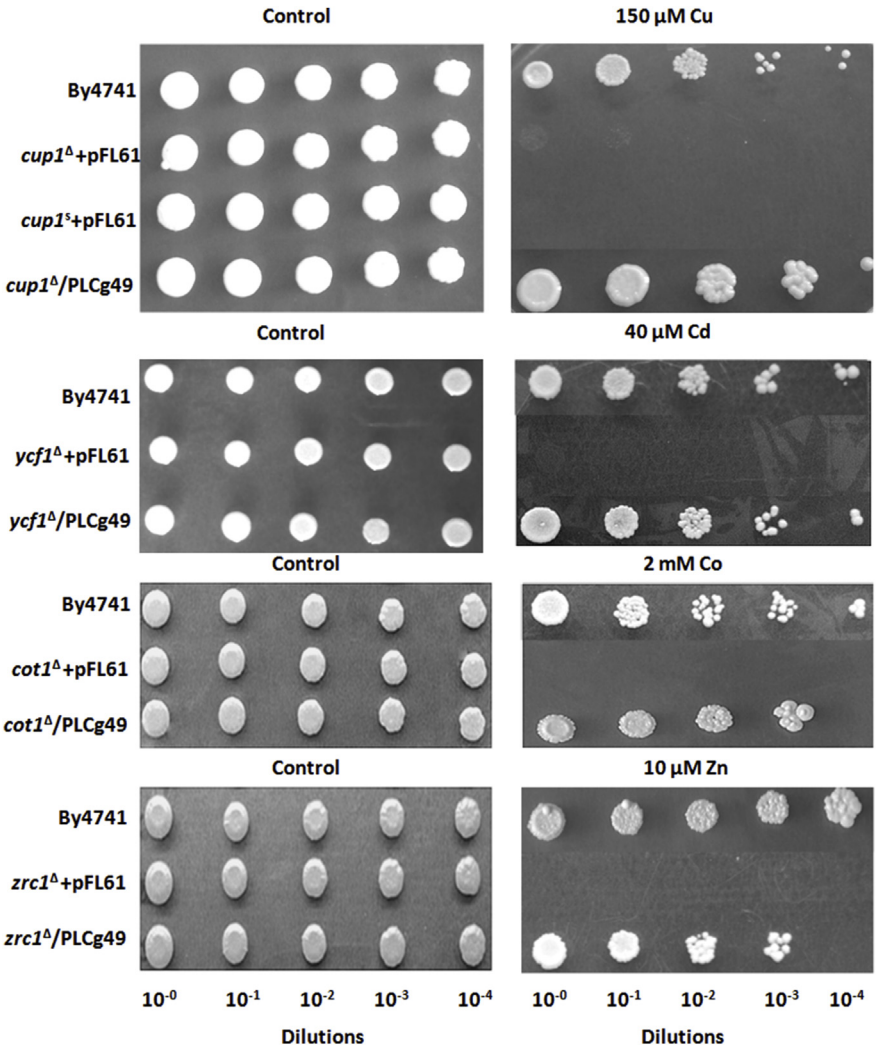


Fig. 4. Drop assay of PlCg49 transformed *Saccharomyces cerevisae* mutants, *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ *and* cot1$^{\Delta}$ on SD-Ura agar. BY4741wild type and mutant strains transformed with empty vector pFL61 for positive and negative control. Serially diluted cultures were spotted on SD-Ura medium with or without metal supplement as indicated. Yeast strain *cup1*^{S} was also used as control.

some modifications. Transformed yeast cells were grown in SD-Ura medium till optical density (OD_{600}) reached 1.0. Cells were harvested by centrifugation at 700 × *g* for 5 min at room temperature, washed thrice with 0.1 M Potassium Phosphate buffer (pH 7) and re-suspended in the same buffer. Varying concentrations of trypsin (0–500 µg/ml) and proteinase K (0–30 mg/ml) were prepared in 0.1 M Potassium Phosphate buffer (pH 7). Equal volumes of yeast cell cultures and proteolytic enzymes were mixed and OD_{660} was measured immediately. The mixture was incubated at 30 °C for 2.5 h and OD_{660} was measured again. Yeast cultures mixed with equal volumes of buffer instead of proteolytic enzymes served as control. The percentage of yeast cell lysis was calculated as described by Grenier (1994).

3. Results and discussion

The soil collected from Pierrelaye region is sandy luvisol. The region was primarily a maize cultivation site which was irrigated with untreated wastewater for an extended period and as a result, led to heavy metal contamination. This region was later converted into Agroforestry land by plantation of different genotypes of Poplar trees (Foulon et al., 2016). Soil analysis for various physicochemical properties and the metal content determined in this study are given in Table 2.

3.1. Screening of library for cadmium tolerance

Yeast mutant *ycf1*$^{\Delta}$ was used for screening the cDNA library C, synthesized from total soil RNA. Functional complementation restored the Cd tolerant phenotype of the mutant strain due to the transformed environmental cDNAs having metal tolerance property. A large number of clones were rescued on this basis indicating the role of different genes in providing metal tolerance. One of the clones PLCg49 was selected for further analysis based on its high tolerance towards Cd.

3.2. Sequence analysis

Sequence analysis revealed that PLCg49 consists of 1409 bp cDNA with an open reading frame of 1233 bp encoding a 410 amino acid polypeptide. The predicted molecular mass of PLCg49 was 45.57 kDa with pI of 6.47. The deduced amino acid sequence showed 38% identity to Serine Protease Inhibitior (serpin) family of *Hypsibius dujardini*. Serine protease inhibitors are the largest and most broadly distributed superfamily of protease inhibitors with more than 1500 members identified. The characteristic size range of a serpin is usually between 350 and 500 bp with molecular weight of 40–60 kDa. The serpins usually have 3 β-sheets with 8 or 9 α-helices. The protease inhibition function comes from the reactive-center loop (RCL), which is found on the core structure forming an extended and exposed conformation

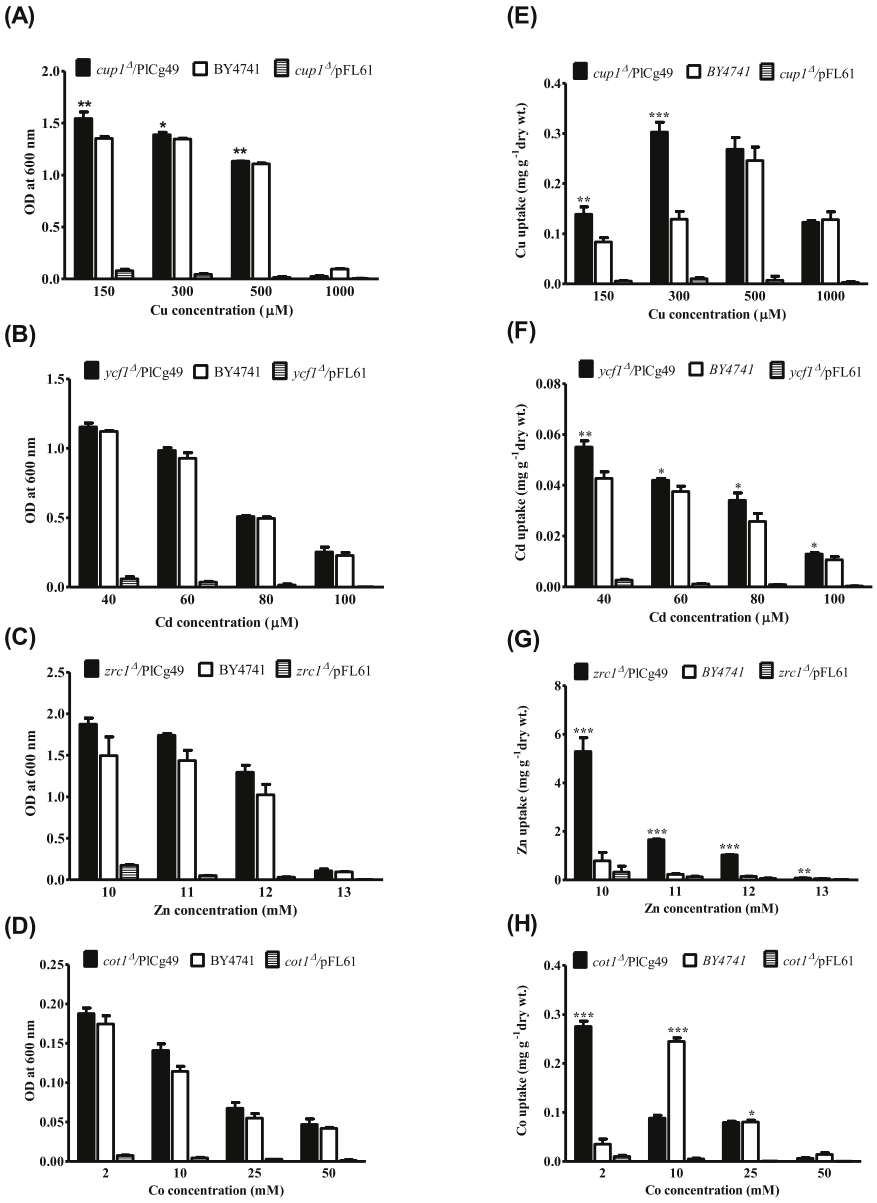


Fig. 5. Growth response of (A) *cup1*$^{\Delta}$ (B) *ycf1*$^{\Delta}$ (C) *zrc1*$^{\Delta}$ and (D) cot1$^{\Delta}$ transformed with PlCg49 and wild type BY4741 transformed with Empty Vector pFL61 in SD-Ura broth supplemented with different metals at different concentrations as shown. Effect of different concentration of (E) Copper (F) Cadmium (G) Zinc (H) Cobalt on metal uptake by PlCg49 when compared to wild type BY4741 after 40 h of growth. Error bars are ± SD (n = 3). (*, ** and *** are significant at $P < 0.05$, $P < 0.01$ and $P < 0.001$ respectively).

above the body of the serpin. Conserved regions observed based on sequence alignment are depicted in Fig. 1. Even with the extensive common features in all serpins, the sequence identity of primary structures can be as low as 25%. Analysis of PLCg49 sequence of RCLs showed that the hinge region consisted of a slight variation of the consensus sequences, "P_{17}[E] – P_{16}[E/K/R] – P_{15}[G] – P_{14}[T/S] – P_{13}[X] – P_{12-9}[AGSX] – P_{8-1}[X] – $P_{1'-4'}$" where X can be any amino acids and P represents position of amino acid residues (Gettins, 2002). At P_{16}, there is Asparagine (N) and at positions P_{12-10}, the amino acids are Glycine-Alanine-Alanine. The P_1–P_1'cleavage sites of the RCL is located near the C-terminal at position R368-A369 (Fig. 2).

Almost 16 clades of serpins have been identified and it has been found that the inhibitory serpins are 'single use' or 'suicidal' which undergoes extensive conformational change of its active site during protease inhibition and become inactive (Law et al., 2006). Further, bioinformatic characterization of the 410 aa protein predicted the presence of 36 potential phosphorylation sites and 2 N-Glycosylation sites but no transmembrane helices. Presence of a signal peptide suggested its transport via the Endoplasmic Reticulum (Guerra et al., 2015). Phylogenetic analysis of serine protease inhibitors from various taxons have clearly grouped into different lineages such as Animalia, Plantae, Fungi, Protista and Prokaryota (Fig. 3). In the present study, serpin was clustered with Animalia lineage confirming its origin from kingdom Animalia. The sequence of PLCg49 obtained in this study was submitted to National Center of Biotechnology Information (NCBI) database under the accession number MK079356.

3.3. Metal tolerance potential of PLCg49

All mutant cells, *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ *and* cot1$^{\Delta}$ transformed with PLCg49 were able to grow well in all four dilutions on SD-Ura containing 150 μM of Cu, 40 μM Cd, 10 mM Zn and 2 mM Co respectively. However, the growth of yeast mutants transformed with empty vector pFL61 was inhibited in these concentrations including *cup1*^{s} which was

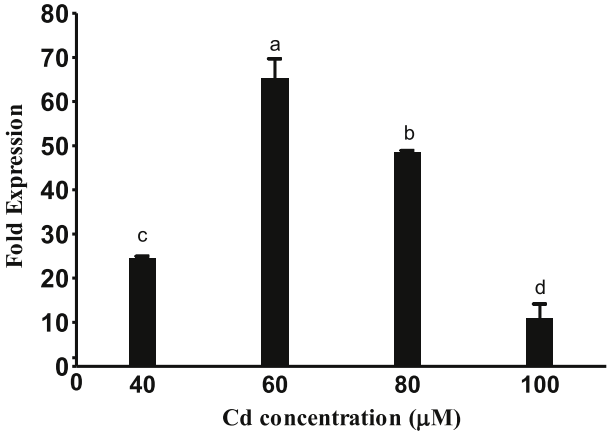


Fig. 6. Fold increase in level of expression of PlCg49 after 40 h of incubation with different concentrations of Cd calculated with reference to the control condition (no metal exposure). All values are mean of three biological replicates ± SD For each sample, values sharing common letters on error bars are not significantly different at $P < 0.05$ ($n = 3$).

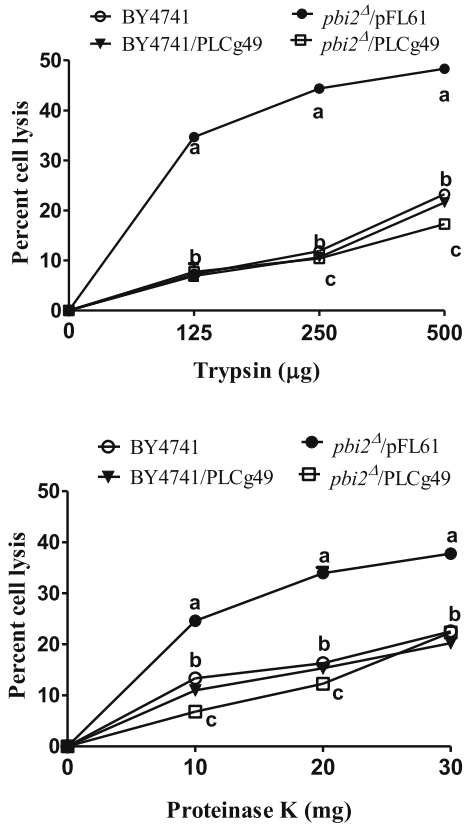


Fig. 7. Percentage cell lysis of yeast mutant *pbi2*$^{\Delta}$ transformed with PlCg49 in presence of different concentrations of Trypsin and Proteinase K in the growth medium. Wild type BY4741 and *pbi2*$^{\Delta}$ transformed with empty pFL61 were used as controls. Also BY4741 transformed with PlCg49 was used for comparison. Error bars are ± SD. Values sharing common letters on error bars for corresponding protease concentrations between samples are not significantly different at $P < 0.05$ (n = 3).

unable to grow in 150 μM of Cu. The wild type BY4741 transformed with empty pFL61 was able to grow in all metal amended media. In general, reduction in growth was observed for all yeast strains with increase in dilution in medium supplemented with metal. (Fig. 4). To study heavy metal tolerant attribute of PLCg49 in presence of Cu, Cd, Zn and Co, the transformants were allowed to grow in SD-Ura broth supplemented with a wide concentrations of these metals. The selected transformant *cup1*$^{\Delta}$/PLCg49 could tolerate high Cu concentrations but with increase in concentration of Cu, inhibition in the growth was observed. The growth was not affected much up to 500 μM Cu (Fig. 5A). In presence of Cd, the transformant *ycf1*$^{\Delta}$/PLCg49 was able to grow well at 40 μM, but the growth decreased rapidly with increase in Cd concentration. The wild type BY4741 displayed comparable growth response with respect to *ycf1*$^{\Delta}$/PLCg49 for all the concentrations (Fig. 5B). The growth of *zrc1*$^{\Delta}$/PLCg49 cells was higher than BY4741 in all the concentrations of Zn with maximum growth recorded at 10 mM. The transformant was able to grow well up to 12 mM but decreased considerably when Zn concentration increased to 13 mM. The wild type strain BY4741 showed similar trend in its growth (Fig. 5C). The Co-sensitive yeast transformant bearing cDNA PLCg49 had also displayed tolerance towards Co with maximum growth recorded at 2 mM concentration. The growth of *cot1*$^{\Delta}$/PLCg49 was not significantly higher than its wild type strain transformed with empty vector pFL61 in all concentrations of Co (Fig. 5D). For all the metals, mutant yeast strains transformed with empty pFL61 vector depicted negligible growth.

Evidences from many recent studies suggest that protease inhibitors are multifunctional and are positively responsible in modulating tolerance of plants against abiotic stresses. Gene expression studies and accumulation of metal in *Populus* roots when exposed to Cu stress was well documented by Guerra et al. (2015). With maximum of 60 μM of $CuSO_4$, the transcriptome expressed defense proteins out of which a considerable fold change was observed in Kunitz trypsin inhibitor genes. Additionally, when this gene was cloned and expressed in complementation experiments using mutant yeast strains, it was able to impart tolerance towards Cu, Cd, Zn and Co. In the same study, putative Cu binding sites were indicated which could chelate Cu and hence account for the uptake of metals (Guerra et al., 2009). In the present study, the environmental transcript was also able to withstand high concentrations of Cu, Cd, Zn and Co.

3.4. Yeast metal uptake and qPCR analysis

Metal sensitive yeast mutants transformed with PLCg49 were grown in presence of different concentrations of metals along with the wild type strain BY4741 and the metal contents were determined in the yeast cells by using ICP-MS analysis. The cellular content of Cu decreased with increase in Cu concentration in the growth medium both in *cup1*$^{\Delta}$/PLCg49 and wild type strain. Maximum Cu content was observed at 300 μM of Cu in *cup1*$^{\Delta}$/PLCg49 cells (Fig. 5E). With increase in Cd concentration in the growth medium, the Cd levels were significantly decreased both in *ycf1*$^{\Delta}$/PLCg49 as well as the wild type strains (Fig. 5F). Accumulation of Zn also decreased with increase in Zn concentration in the growth medium with maximum Zn uptake observed at 10 mM in *zrc1*$^{\Delta}$/PLCg49 (Fig. 5G). The Co content in the cellular levels also decreased with increase in Co concentrations in the growth medium both in *cot1*$^{\Delta}$/PLCg49 cells as well as BY4741 cells (Fig. 5H).

To determine the accumulation of transcripts of the PLCg49 in response to Cd, qPCR analysis was performed by exposing *ycf1*$^{\Delta}$/PLCg49 cells to varied concertation of Cd. With increase in Cd concentration, aaccumulation of mRNA transcripts of PLCg49 significantly increased up to 60 μM of Cd and decreased thereafter. The induction levels of PLCg49 was 65 fold higher at 60 μM of Cd followed by 48 fold higher at 80 μM of Cd (Fig. 6). In the wild type strain BY4741, no amplification of PLCg49 was observed indicating the absence of this gene in yeast.

Protease inhbitiors (PIs) have been studied in plants with respect to various stress-inducible factors and it has emerged as an important family of genes which are expressed in response to abiotic stresses. Out of the various PIs, the most abundant are serine PIs and are well known for the role in environmental response. In a study by Shitan et al. (2007), a Bowman–Birk proteinase inhibitor was isolated by functional complementation which could rescue and allow the mutant yeast *ycf1*$^{\Delta}$

to grow in presence of heavy metal Cd. Further, the PI was characterized by metal uptake studies and qPCR analysis and it was found that the Cd tolerance potential depicted by the Bowman–Birk proteinase inhibitor was even higher that a metallothionein isolated in the same study and proved that this PI could confer resistance to *Coptis japonica* plant in presence of high concentration of Cadmium. Similarly with the wounding of plant tissue and exposing with toxic concentrations of aluminium led to the induction of a number of Bowman–Birk type protease inhibitors. Likewise, a number of studies have reported the induction of protease inhibitor genes in response to heavy metals exposure in different plants like wheat (Snowden et al., 1995), tobacco (Harada et al., 2010) maize (Cançado et al., 2008), *Lupinus* and *Pisum* (Domash et al., 2008). Potential protease inhibitors from *Capsicum annuum* conferring tolerance to multiple stresses including heavy metals Cd and Hg in mutant yeast has been well documented by Joshi et al. (2014) where strong inhibition of cellular metacaspase activity leading to delayed senescence was observed. When expressed in yeast mutants, significant increase in growth of yeast cells in presence of heavy metals as well as protease inhibition activity was observed. The expression of PIs in response to abiotic stresses such as drought, salinity, heavy metals, UV radiation and wounding suggests that there are numerous interlinked metabolic processes which ultimately lead to their induction. A common factor being the production of damaging ROS species which arises due to both biotic and abiotic stresses leading to abnormalities and disruption of normal metabolic processes (Mukherjee et al., 2019; Srinivasan and Kirti, 2012).

3.5. Effect of protease on yeast cells

To validate the role of the transcript PLCg49 in providing resistance to protease treatment, the lytic activity was measured in presence of trypsin and proteinase K. The transformant *pbi2*$^{\Delta}$/PLCg49 showed reduced lytic activity when compared to the *pbi2*$^{\Delta}$ transformed with empty pFL61 which was significantly higher in all concentrations of trypsin. The cell lytic activity at 500 µg was significantly higher in BY4741. The transformant *pbi2*$^{\Delta}$/PLCg49 showed resistance in cell lysis when treated with proteinase K enzyme and the cell lysis was comparable to the wild type BY4741. However, for all concentrations of proteinase K, *pbi2*$^{\Delta}$ transformed with empty pFL61 recorded significant cell lysis highest being at 30 mg. The transformant BY4741/PLCg49 showed slightly reduced cell lysis when compared to BY4741 alone for all concentrations of the proteases. However, this difference was not significant (Fig. 7). Shimoi et al. (1991) reported the effects of various protease inhibitors on bacterial proteases by determining the yeast cell lysis and enzyme activity. Their results showed that serine protease inhibitors effectively inhibited protease activity and improved yeast cell viability. In the present study, similar results were observed when the PLCg49 was expressed in yeast cells.

The results in the present study also support the existing research data and have depicted tolerance towards high concentrations of heavy metals like Cu, Cd, Zn and Co. We are reporting for the first time the characteristics of an environmental Serine Protease Inhibitor isolated using metatranscriptomics approach from polluted soil capable of expressing, tolerating and accumulating heavy metals inside the cells.

4. Conclusion

From the experimental observations, it was concluded that the cDNA PLCg49 has the potential to tolerate toxic levels of metals when expressed in yeast. The presence of characteristic features of Serine protease inhibitor (Serpin) in the translated protein signified that this protein is induced as a result of abiotic stresses along with other stress response proteins in contaminated soil. The method of metatranscriptomics applied to complex ecosystem like soil can reveal a high degree of diversity of functions. Previous studies have mostly focused of host specific serpins and studying the transcriptome of specific host which may or may not include the complex interplay and the role of other serpins. Heterologous expression of this transcript proved that it could be used for a variety of biotechnological applications like bioremediation of contaminated land, revegetation, biomarkers for contamination studies and developing strategies to improve heavy metal tolerance potential of crop plants.

Conflicts of interest

Authors declare no conflict of interest.

Acknowledgements

Authors are thankful to Council of Scientific and Industrial Research (CSIR), Govt. of India, New Delhi for providing the financial support under CSIR grant No. 38(1425)/16/EMR-II. The part of this work also supported by Indo-French Centre for the Promotion of Advanced Research, New Delhi, India under the project No. 4709-1.

References

Andersen, C.L., Jensen, J., Ørntoft, T., 2004. Normalization of real-time quantitative RT-PCR data: a model based variance estimation approach to identify genes suited for normalization-applied to bladder-and colon-cancer data-sets. Cancer Res. 64, 5245–5250. https://doi.org/10.1158/0008-5472.CAN-04-0496.

Bashiardes, S., Zilberman-Schapira, G., Elinav, E., 2016. Use of metatranscriptomics in microbiome research. Bioinf. Biol. Insights 10, 19–25. https://doi.org/10.4137/BBI.S34610.

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., Del Pozo-Yauner, L., Ochoa-Leyva, A., 2015. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. Comput. Struct. Biotechnol. J. 13, 390–401. https://doi.org/10.1016/j.csbj.2015.06.001.

Blindauer, C.A., Leszczyszyn, O.I., 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. Nat. Prod. Rep. 27, 720–741. https://doi.org/10.1039/B906685N.

Cançado, G.M., Nogueira, F.T., Camargo, S.R., Drummond, R.D., Jorge, R.A., Menossi, M., 2008. Gene expression profiling in maize roots under aluminum stress. Biol. Plant. 52, 475–485. https://doi.org/10.1007/s10535-008-0093-5.

Clemens, S., 2001. Molecular mechanisms of plant metal tolerance and homeostasis. Planta 212, 475–486. https://doi.org/10.1007/s004250000458.

Conconi, A., Smerdon, M.J., Howe, G.A., Ryan, C.A., 1996. The octadecanoid signalling pathway in plants mediates a response to ultraviolet radiation. Nature 383, 826–829. https://doi.org/10.1038/383826a0.

Conklin, D.S., McMaster, J.A., Culbertson, M.R., Kung, C.H., 1992. COT1, a gene involved in cobalt accumulation in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12, 3678–3688. https://doi.org/10.1128/MCB.12.9.3678.

Damodaran, D., Shetty, K.V., Mohan, B.R., 2013. Effect of chelaters on bioaccumulation of Cd (II), Cu (II), Cr (VI), Pb (II) and Zn (II) in Galerina vittiformis from soil. Int. Biodeterior. Biodegrad. 85, 182–188. https://doi.org/10.1016/j.ibiod.2013.05.031.

Damon, C., Vallon, L., Zimmermann, S., Haider, M.Z., Galeote, V., Dequin, S., Luis, P., Fraissinet-Tachet, L., Marmeisse, R., 2011. A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. ISME J. 5, 1871–1880. https://doi.org/10.1038/ismej.2011.67.

Dietz, K.J., Bair, M., Kramer, U., 1999. Free radical and reactive oxygen species as mediators of heavy metal toxicity in plants. In: Prasad, M.N.V., Hagemeyer, J. (Eds.), Heavy Metal Stress in Plants from Molecules to Ecosystems. Spinger-Verlag, Berlin, pp. 73–79. https://doi.org/10.1007/978-3-662-07745-0_4.

Domash, V.I., Sharpio, T.P., Zabreiko, S.A., Sosnovskaya, T.F., 2008. Proteolytic enzymes and trypsin inhibitors of higher plants under stress conditions. Russ. J. Bioorg. Chem. 34, 318–322. https://doi.org/10.1134/S1068162008030114.

Dopson, M., Holmes, D.S., 2014. Metal resistance in acidophilic microorganisms and its significance for biotechnologies. Appl. Microbiol. Biotechnol. 98, 8133–8144. https://doi.org/10.1007/s00253-014-5982-2.

Foulon, J., Zappelini, C., Durand, A., Valot, B., Girardclos, O., Blaudez, D., Chalot, M., 2016. Environmental metabarcoding reveals contrasting microbial communities at two poplar phytomanagement sites. Sci. Total Environ. 571, 1230–1240. https://doi.org/10.1016/j.scitotenv.2016.07.151.

Gettins, P.G., 2002. Serpin structure, mechanism, and function. Chem. Rev. 102, 4751–4804. https://doi.org/10.1021/cr010170+.

Gietz, D., St Jean, A., Woods, R.A., Schiestl, R.H., 1992. Improved method for high-efficiency transformation of intact yeast-cells. Nucleic Acids Res. 20 1425–1425.

Grenier, D., 1994. Effect of proteolytic enzymes on the lysis and growth of oral bacteria. Oral Microbiol. Immunol. 9, 224–228. https://doi.org/10.1111/j.1399-302X.1994.tb00062.x.

Guerra, F., Duplessis, S., Kohler, A., Martin, F., Tapia, J., Lebed, P., Zamudio, F., González, E., 2009. Gene expression analysis of Populus deltoides roots subjected to copper stress. Environ. Exp. Bot. 67, 335–344. https://doi.org/10.1016/j.envexpbot.2009.08.004.

Guerra, F.P., Reyes, L., Vergara-Jaque, A., Campos-Hernández, C., Gutiérrez, A., Pérez-Díaz, J., Pérez-Díaz, R., Blaudez, D., Ruíz-Lara, S., 2015. Populus deltoides Kunitz trypsin inhibitor 3 confers metal tolerance and binds copper, revealing a new defensive role against heavy metal stress. Environ. Exp. Bot. 115, 28–37. https://doi.org/10.1016/j.envexpbot.2015.02.005.

Harada, E., Kim, J.A., Meyer, A.J., Hell, R., Clemens, S., Choi, Y.E., 2010. Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. Plant Cell Physiol. 51, 1627–1637. https://doi.org/10.1093/pcp/pcq118.

Huang, Y., Xiao, B., Xiong, L., 2007. Characterization of a stress responsive proteinase inhibitor gene with positive effect in improving drought resistance in rice. Planta 226, 73–85. https://doi.org/10.1007/s00425-006-0469-8.

Idera, F., Omotola, O., Paul, U.J., Adedayo, A., 2014. Evaluation of the effectiveness of different acid digestion on sediments. IOSR J. Appl. Chem. 12, 39–47. https://doi.org/10.9790/5736-071213947.

Joshi, R.S., Tanpure, R.S., Singh, R.K., Gupta, V.S., Gir, A.P., 2014. Resistance through inhibition: ectopic expression of serine protease inhibitor offers stress tolerance via delayed senescence in yeast cell. Biochem. Biophys. Res. Commun. 452, 361–368. https://doi.org/10.1016/j.bbrc.2014.08.075.

Kellner, H., Luis, P., Portetelle, D., Vandenbol, M., 2011. Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. Microbiol. Res. 166, 360–368. https://doi.org/10.1016/j.micres.2010.07.006.

Kitson, R.E., Mellon, M.G., 1944. Colorimetric determination of phosphorus as molybdivanadophosphoric acid. Ind. Eng. Chem. 16, 379–383.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

Law, R.H., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., Whisstock, J.C., 2006. An overview of the serpin superfamily. Genome Biol. 7, 216. https://doi.org/10.1186/gb-2006-7-5-216.

Li, J., Ma, Y.B., Hu, H.W., Wang, J.T., Liu, Y.R., He, J.Z., 2015. Field-based evidence for consistent responses of bacterial communities to copper contamination in two contrasting agricultural soils. Front. Microbiol. 6, 31. https://doi.org/10.3389/fmicb.2015.00031.

Li, L., Kaplan, J., 2001. The yeast gene MSC2, a member of the cation diffusion facilitator family, affects the cellular distribution of zinc. J. Biol. Chem. 276, 5036–5043. https://doi.org/10.1074/jbc.M008969200.

Li, Z.S., Lu, Y.P., Zhen, R.G., Szczypka, M., Thiele, D.J., Rea, P.A., 1997. A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis (glutathionato) cadmium. Proc. Natl. Acad. Sci. 94, 42–47. https://doi.org/10.1073/pnas.94.1.42.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262.

Longo, V.D., Gralla, E.B., Valentine, J.S., 1996. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae* mitochondrial production of toxic oxygen species in vivo. J. Biol. Chem. 271, 12275–12280. https://doi.org/10.1074/jbc.271.21.12275.

Lopez, F., Vansuyt, G., Derancourt, J., Fourcroy, P., Casse-Delbart, F., 1994. Identification by 2D-page analysis of salt-stress induced proteins in radish (Raphanus sativus). Cell. Mol. Biol. 40, 85–90.

Mukherjee, A., Yadav, R., Marmeisse, R., Fraissinet-Tachet, L., Reddy, M.S., 2019. Heavy metal hypertolerant eukaryotic aldehyde dehydrogenase isolated from metal contaminated soil by metatranscriptomics approach. Biochimie 160, 183–192. https://doi.org/10.1016/j.biochi.2019.03.010.

Ochiai, E.I., 2012. General Principles of Biochemistry of the Elements. Springer Science & Business Media. https://doi.org/10.1007/978-1-4684-5371-3.

Olsen, S.R., 1954. Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate. United States Department of Agriculture, Washington.

Oyetibo, G.O., Miyauchi, K., Huang, Y., Chien, M.F., Ilori, M.O., Amund, O.O., Endo, G., 2017. Biotechnological remedies for the estuarine environment polluted with heavy metals and persistent organic pollutants. Int. Biodeterior. Biodegrad. 119, 614–625. https://doi.org/10.1016/j.ibiod.2016.10.005.

Pennanen, T., Frostegard, A.S., Fritze, H., Baath, E., 1996. Phospholipid fatty acid composition and heavy metal tolerance of soil microbial communities along two heavy metal-polluted gradients in coniferous forests. Appl. Environ. Microbiol. 62, 420–428.

Piper, C.S., 1966. Soil and Plant Analysis: a Laboratory Manual of Methods for the Examination of Soils and the Determination of the Inorganic Constituents of Plants. Hans Publications, Bombay.

Rehman, S., Aziz, E., Akhtar, W., Ilyas, M., Mahmood, T., 2017. Structural and functional characteristics of plant proteinase inhibitor-II (PI-II) family. Biotechnol. Lett. 39, 647–666. https://doi.org/10.1007/s10529-017-2298-1.

Schu, P., Suarez Rendueles, P., Wolf, D.H., 1991. The proteinase yscB inhibitor (PBI2) gene of yeast and studies on the function of its protein product. Eur. J. Biochem. 197, 1–7. https://doi.org/10.1111/j.1432-1033.1991.tb15874.x.

Shimoi, H., Muranaka, Y., Sato, S.I., Saito, K., Tadenuma, M., 1991. Purification of the enzymes responsible for the lysis of yeast cells by Rarobacter faecitabidus. Agric. Biol. Chem. 55, 371–378. https://doi.org/10.1080/00021369.1991.10870610.

Shitan, N., Horiuchi, K.I., Sato, F., Yazaki, K., 2007. Bowman–Birk proteinase inhibitor confers heavy metal and multiple drug tolerance in yeast. Plant Cell Physiol. 48, 193–197. https://doi.org/10.1093/pcp/pcl051.

Snowden, K.C., Richards, K.D., Gardner, R.C., 1995. Aluminum-induced genes (induction by toxic metals, low calcium, and wounding and pattern of expression in root tips. Plant Physiol. 107, 341–348. https://doi.org/10.1104/pp.107.2.341.

Srinivasan, T., Kirti, P.B., 2012. Protease inhibitors and stress tolerance. Funct. Plant Sci. Biotechnol. 6, 59–66.

Thakur, B., Yadav, R., Fraissinet-Tachet, L., Marmeisse, R., Reddy, M.S., 2018. Isolation of multi-metal tolerant ubiquitin fusion protein from metal polluted soil by metatranscriptomic approach. J. Microbiol. Methods 152, 119–125. https://doi.org/10.1016/j.mimet.2018.08.001.

Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5, 240–245. https://doi.org/10.1016/S1369-5274(02)00324-7.

Tveit, A.T., Urich, T., Svenning, M.M., 2014. Metatranscriptomic analysis of arctic peat soil microbiota. Appl. Environ. Microbiol. 80, 5761–5772. https://doi.org/10.1128/AEM.01030-14.

Utgikar, V.P., Tabak, H.H., Haines, J.R., Govind, R., 2003. Quantification of toxic and inhibitory impact of copper and zinc on mixed cultures of sulfate-reducing bacteria. Biotechnol. Bioeng. 82, 306–312. https://doi.org/10.1002/bit.10575.

Walkley, A., 1947. A critical examination of a rapid method for determining organic carbon in soils - effect of variations in digestion conditions and of inorganic soil constituents. Soil Sci. 63, 251–264.

Wysocki, R., Tamás, M.J., 2010. How Saccharomyces cerevisiae copes with toxic metals and metalloids. FEMS Microbiol. Rev. 34, 925–951. https://doi.org/10.1111/j.1574-6976.2010.00217.x.

Yadav, R.K., Barbi, F., Ziller, A., Luis, P., Marmeisse, R., Reddy, M.S., Fraissinet-Tachet, L., 2014. Construction of sized eukaryotic cDNA libraries using low input of total environmental metatranscriptomic RNA. BMC Biotechnol. 14, 80–86. https://doi.org/10.1186/1472-6750-14-80.

Biochimie 160 (2019) 183–192

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Heavy metal hypertolerant eukaryotic aldehyde dehydrogenase isolated from metal contaminated soil by metatranscriptomics approach

Arkadeep Mukherjee [a], Rajiv Yadav [b], Roland Marmeisse [b], Laurence Fraissinet-Tachet [b], M. Sudhakara Reddy [a, *]

[a] Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab, 147004, India
[b] Ecologie Microbienne, UMR CNRS, UMR INRA, Université Claude Bernard Lyon 1 Université de Lyon, F-69622 Villeurbanne, France

ARTICLE INFO

Article history:
Received 3 November 2018
Accepted 18 March 2019
Available online 21 March 2019

Keywords:
Metatranscriptomics
Eukaryotes
Heavy metal tolerance
Yeast complementation
Aldehyde dehydrogenase
Copper tolerance

ABSTRACT

Constant addition of heavy metal pollutants in soil resulting from anthropogenic activities can prove detrimental to both macro and micro life forms inhabiting the ecosystem. The potential functional roles of eukaryotic microbes in such environment were explored in this study by metatranscriptomics approach. Sized eukaryotic cDNA libraries, library A (<0.5 kb), library B (0.5–1.0 kb), and library C (>1 kb) were constructed from the soil RNA and screened for copper (Cu) tolerance genes by using copper sensitive yeast mutant strain *cup1*$^{\Delta}$. Screening of the cDNA libraries yielded different clones capable of growing in Cu amended medium. In the present investigation, one of the transcripts PLCc38 from the library C was characterized and tested for its ability to tolerate different heavy metals by using metal sensitive yeast mutants. Sequence analysis PLCc38 showed homology with aldehyde dehydrogenase (ALDH) and capable of tolerating high concentrations of Cu (150–1000 μM). Aldeyde dehydrogenases are stress response enzymes capable of eliminating toxic levels of aldehydes generated due to abiotic environmental stresses. The cDNA PLCc38 also provided tolerance to wide range of Cd (40–100 μM), Zn (10–13 mM) and Co (2–50 mM) concentrations. Oxidative stress tolerance potential of PLCc38 was also confirmed in presence of different concentrations of H_2O_2. This study proves that PLCc38 is a potent gene associated with metal tolerance which could be used to revegetate heavy metal polluted soil or as a biomarker for detection of metal contamination.

1. Introduction

Industrial metal mining zones have considerable area of land contaminated with heavy metals. Although metals like Zn, Cu, Ni, Fe etc., are naturally present in soil and also found as trace elements in plants and animals, their presence at higher concentration could cause detrimental effects to the environment leading to reduction in soil quality making it unsuitable for agricultural practices [1]. Microorganisms in metal polluted soil are constantly exposed to toxic levels of heavy metals which cause a decline in soil diversity due to loss of more sensitive population. This leads to lower enzymatic activity and poor utilization of nutrients which ultimately creates an ecological imbalance [2]. However, some soil microorganisms, even at elevated levels of heavy metal exposure over a long-term, are able to maintain the overall biomass as in unpolluted soils. To lower the toxic effect of heavy metals, these microorganisms might have developed efficient resistance mechanisms [3]. Broadly these mechanisms can be described as efflux of metals from cells, enzyme coupled transformations of metals, cellular sequestration of heavy metals into complexes by metal-chelating proteins and vacuolar compartmentalization. These mechanisms have been studied in plants, fungi and bacteria inhabiting metal contaminated soil and their potential molecular targets have been identified [4,5].

The soil holds a vast pool of genetic information contributed by microorganisms. This can represent a significant fraction of the microbial biomass; most of which are uncultivable under normal

* Corresponding author. Department of Biotechnology Thapar Institute of Engineering & Technology Patiala, 147 004, Punjab, India.
E-mail address: msreddy@thapar.edu (M.S. Reddy).

https://doi.org/10.1016/j.biochi.2019.03.010
0300-9084/

laboratory conditions. Studying this genetic resource would allow us to unravel novel biological molecules and their key functional roles in metal tolerance or resistance. Novel biocatalysts isolated from metal polluted environment could be implemented directly as biomarkers or for bioremediation and development of metal tolerant crop plants [6]. Metatranscriptomics approach play an important role in isolation of genes of ecological importance. This approach involves isolation of total RNA from soil, synthesis of cDNA particularly from eukaryotic polyadenylated mRNA followed by cDNA library preparation [7]. This library would represent a large fraction of protein coding genes contributed by the inhabiting eukaryotes and could be exploited for accessing genes of interest [8]. This approach has been applied to study complex ecosystems like gut microbiome of humans and animals [9,10], soil [11–13] and extreme environments [14,15].

In this study, sized eukaryotic cDNA libraries were constructed from the metal polluted soil and screened by yeast functional complementation to isolate genes associated with metal tolerance. Screening of one of the size-fractionated libraries (library C) led to the detection of several genes involved in metal tolerance including the cDNA PLCc38 which showed homology to aldehyde dehydrogenase (ALDH). Aldehyde dehydrogenases are family of NAD (P^+)-dependent enzymes that irreversibly catalyze the oxidation of endogenously and exogenously produced aldehydes to their respective carboxylic acids [16]. ALDH family genes are also known for their roles in drought resistance and other oxidative stress related challenges [17,18]. Very few reports are available on the role of ALDH gene in metal tolerance [19]. In this study we have characterized the ALDH clone PLCc38 isolated from the metal contaminated soil through metatranscriptomic approach and tested its ability to rescue different metal sensitive yeast mutants under metal stress.

2. Material and methods

2.1. Soil sample collection

Soil sampling site was Pierrelaye (PL) (49°1′45″N, 2°10′32″E) situated northwest of Paris, France. The sampling region was primarily a maize cultivation site which was later turned into Agroforestry land by trees of Poplar. Extended use of untreated wastewater for irrigation led to heavy metal contamination of this site. Twenty soil cores were collected and equal volumes of each core were mixed together to form composite samples, sieved with 2 mm mesh, packed in sterile plastic bags and transported in dry ice to laboratory and preserved at −70 °C until RNA extraction. Soil sample was analyzed for its physicochemical properties i.e. pH, organic carbon [20], total P [21], available P [22] and total nitrogen [23]. The metal contents of the soil samples were determined by ICP-AES (ARCOS, Simultaneous ICP Spectrometer, SPECTRO Analytical Instruments GmbH, Germany).

2.2. Biological materials

Copper-sensitive *cup1*$^{\Delta}$ (DTY4), cadmium -sensitive *ycf1*$^{\Delta}$, zinc-sensitive *zrc1*$^{\Delta}$ and cobalt-sensitive *cot1*$^{\Delta}$ mutant strains of *Saccharomyces cerevisiae* along with their respective wild type strains were used in this study. *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ strains were derived from the wild-type strain BY4741 (*MATα, his3*$^{\Delta}$*1, leu2*$^{\Delta}$*0, met15*$^{\Delta}$*0, ura3*$^{\Delta}$*0*). DTY3 (*cup1*^{s}) (*MATα, leu2-3, 112his3*$^{\Delta}$*1, trp1-1, ura3-50, gal1, CUP1*^{S}) having single copy of *cup1* was used as control. Yeast strains BY4742 (*MATα his3*$^{\Delta}$*1 leu2*$^{\Delta}$*0 lys2*$^{\Delta}$*0 ura3*$^{\Delta}$*0*) and superoxide dismutase mutant *sod2*$^{\Delta}$ (BY4742 *sod2*::*KanMX*) were used for oxidative stress experiments. All yeast mutant and wild type strains were acquired from Euroscarf (http://www.euroscarf.de). The copper-sensitive *cup1*$^{\Delta}$ (DTY4) is lacking the copperthionein gene [24]. Cadmium sensitive *ycf1*$^{\Delta}$ lacks an ABC transporter gene YCF1 [25] while deletion of the ZRC1, a gene encoding for transporter protein in *zrc1*$^{\Delta}$ [26] and deletion of COT1 gene in *cot1*$^{\Delta}$ [27] has led to the hypersensitivity of these strains towards their respective metals. Deletion of the mitochondrial superoxide dismutase (*sod2*) caused hypersensitivity towards reactive oxygen species in *sod2*$^{\Delta}$ [28]. For yeast strains, synthetic defined medium without uracil (SD-Ura) was used which is a minimal medium supplemented with 2% glucose (Hi-media laboratories, India) and drop out supplement without uracil (Takara-Clontech, Japan). Luria broth (LB) (Hi-media laboratories, India) and One Shot® TOP10 Electrocomp™ *E. coli* (Thermofisher, USA) was used for bacterial transformations.

2.3. Size fractionation and cDNA library construction

Total RNA was extracted from collected soil using the RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories, Carlsbad, CA) according to the manufacturer's protocol. RNA integrity was checked by Bioanalyzer 2100 (Agilent Technologies, USA) electrophoresis and quantity and purity were determined by nanodrop spectrophotometry (Thermo Fisher, USA). cDNA synthesis and size fractionation of cDNA libraries were performed as described by Yadav et al. [11]. Briefly, cDNAs were synthesized from total RNA by using the Mint-2 cDNA synthesis kit (Evrogen, Moscow, Russia) as per the instructions manual. The cDNA size fractions A (0.1–0.5 kb), B (0.5–1 kb) and C (1–4 kb) were ligated downstream of PGK1 promoter of *S. cerevisiae* shuttle plasmid vector pFL61 modified with SfiA and SfiB restriction sites [29]. These recombinant plasmids were introduced into DH10 β electrocompetant *E. coli* cells (One Shot® TOP10 Electrocomp™ *E. coli*, Invitrogen). More than 10^6 colonies growing on ampicillin-containing Luria agar plates (100 μg ampicillin per ml of media) were pooled to represent each of the cDNA size libraries.

2.4. Screening of libracy C for copper tolerant genes

The cDNA library C (1.0 kb- 4.0 kb) was screened for copper tolerance genes by transforming into copper sensitive yeast mutant strain *cup1*$^{\Delta}$ [30] and Cu tolerant yeast colonies were selected on SD-Ura medium supplemented with 150 μM $CuSO_4$ after incubating them at 30 °C for 48–72 h. Colonies growing on SD-Ura plates with copper were selected and were subjected to 5- Fluoroorotic Acid (5-FOA) treatment to remove false positives. One of the clones PLCc38 was selected for its characterization and potential to tolerate high concentrations of Cu and other metals such as Cd, Zn and Co as well.

2.5. Sequence analysis

Plasmids were isolated from PLCc38 by using Zymoprep yeast plasmid Miniprep II (Zymo Research) according to manufacturer's protocol and were introduced into DH10 β electrocompetant *E. coli* cells. Vector specific forward primer Nf (5′-CAGATCATCAAGGAAGTAATTATCTAC-3′) and reverse primer Nr (5′-CAGAAAAGCAGGCTGGGAAGC-3′) were used to sequence the insert. Sequence homology search was performed using BLASTX tool of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple sequence alignment of homologous sequences along with PLCc38 was performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo). The tools and databases of CBS Prediction Servers (http://www.cbs.dtu.dk/services/) were used for glycosylation (NetOGlyc 4.0 Server) and phosphorylation sites prediction (NetPhos 3.1 Server). For phylogenetic analysis, aldehyde dehydrogenase sequences from various taxa were aligned using Clustal W program in MEGA version 7.0

[31], and the phylogenetic tree was reconstructed using neighbor-joining algorithm employing 1000 bootstrap replicates.

2.6. Metal tolerance potential of PLCc38

Yeast mutant strains *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ susceptible towards Cu, Cd, Zn and Co, respectively were transformed by the recombinant plasmid pFL61 carrying PLCc38 to investigate the potential metal tolerance of the cDNA. Transformants *cup1*$^{\Delta}$/PLCc38, *ycf1*$^{\Delta}$/PLCc38, *zrc1*$^{\Delta}$/PLCc38 and *cot1*$^{\Delta}$/PLCc38 were assessed for metal tolerance on SD-Ura agar by drop assay. All mutant strains and wild type BY4741 transformed with empty vector pFL61 were served as negative and positive controls. All yeast cultures were adjusted to $OD_{600} = 1.0$ and the dilutions of 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} were prepared. Five μl from all the dilutions were spotted on SD-Ura medium agar plates containing 150 μM $CuSO_4$, 40 μM $CdCl_2$, 10 mM $ZnCl_2$ and 2 mM $CoCl_2$, respectively. The growth assay of *cup1*$^{\Delta}$/PLCc38, *ycf1*$^{\Delta}$/PLCc38, *zrc1*$^{\Delta}$/PLCc38 and *cot1*$^{\Delta}$/PLCc38 in SD-Ura broth was performed in presence of varying concentrations of metals. For growth assay, 20 ml of SD-Ura broth was inoculated with mid-log precultures of yeast cells to achieve $OD_{600} = 0.02$ and were allowed to grow at 30 °C at 220 rpm for 3 h. Then, varying concentration of $CuSO_4$ (0–1000 μM), $CdCl_2$ (0–100 μM), $ZnCl_2$ (0–13 mM) and $CoCl_2$ (0–50 mM) were supplemented separately to the respective yeast mutant cells carrying PLCc38 and the growth was measured till 40 h at an interval of 3 h. For comparative study, wild type strain BY4741 transformed with empty pFL61vector was used.

2.7. Yeast metal uptake

Yeast transformants were grown in presence of different concentration of metals as mentioned above. After 24 h of growth, the cells were harvested and washed thrice with 10 mM EDTA solution. Finally cells were washed with Milli-Q water and dried at 80 °C for 24 h. Dry yeast biomass was digested with nitric acid and perchloric acid (2:1, v/v) [32] and the content of metals were determined by using ICP-MS (Element XR, Thermo Fisher Scientific, Germany).

2.8. Oxydative stress tolerance potential of PLCc38

To further confirm the role of PLCc38 in oxidative stress tolerance, yeast mutant *sod2*$^{\Delta}$ was transformed with plasmid pFL61 carrying PLCc38. Transformed *sod2*$^{\Delta}$/PLCc38 cells were selected on SD-Ura medium supplemented with 1 mM H_2O_2. Functional complementation by drop assay on SD-Ura agar plates and study of growth response in SD-Ura broth supplemented with 0 mM, 1 mM, 2 mM and 3 mM of H_2O_2 was performed as described above. Wild type BY4742 and *sod2*$^{\Delta}$ transformed with empty pFL61 served as positive and negative controls, respectively. Also, BY4742 was transformed with PLCc38 to understand the combined effect of PLCc38 and the wild type BY4742 in response to H_2O_2 stress. Growth response was recorded till 30 h at an interval of 5 h.

3. Results and discussion

Soil analysis for various chemical properties determined were as follows: pH 7.15; organic carbon 1.6%; total P 291 mg/kg; available P 14.2 mg/kg and total nitrogen 0.12%. The metal contents such as Cd 2.5, Cu 64 and Zn 385 mg/kg were determined in the soil.

3.1. Screening of cDNA library for copper tolerance

Copper sensitive *cup1*$^{\Delta}$ yeast mutant was employed for screening cDNA library C synthesized from total soil RNA. Cu tolerant phenotype of mutant strain was restored due to the transformed environmental cDNAs responsible for metal tolerance. Various clones were rescued from Cu tolerance indicating different genes providing tolerance to Cu. One of the cDNA clones from library C, PLCc38 was selected for further studies based on its high tolerance to Cu.

3.2. Sequence analysis

Plasmids were isolated from *cup1*$^{\Delta}$/PLCc38 cells and the insert was sequenced by using Nf and Nr primers. The length of the insert was 1667 bp long with an ORF of 720 bp nucleotides encoding a polypeptide of 239 amino acids with a predicted molecular mass of 26.4 kDa and pI of 6.9. The amino acid sequence was 62% identical to aldehyde dehydrogenase (ALDH) family 7 member of *Pseudocohnilembus persalinus* phylum Ciliophora having a similarity mean of 75.6%. Highly conserved regions were observed based on the alignment of the deduced amino acid sequence of PLCc38 with other closely related ALDH from same phylum and also from some plants (Fig. 1). Sequence obtained in this study was submitted to NCBI database with an accession number MK077675.

ALDH functions primarily as an effective enzyme to metabolize aldehydes to its corresponding carboxylic acid. This occurs via the reduction of NAD^+ whose binding is coordinated by glutamic acid in the active site and with the cysteine that provides the catalytic thiol to make a nucleophilic attack on the carbonyl carbon of the aldehyde [16]. In this study, the conserved signature sequence of the PR box PAPNRGEIVRQIGD (residues 12–25) and NAD binding sites TAFN (residues 98–101) and FTGS (residues 179–182) confirmed the cDNA as a NAD^+ dependent ALDH. A variant of ALDH glutamic acid active site signature sequence LELSGNNA was identified (residues 202–209). The catalytic thiol (Cys 228) and neighbor Arg are conserved in all ALDHs [16]. A glycine-rich sequence $Gly_1XXXXGly_2$ confirmed to a known sequence equivalent to Gly-X_4-Gly (residues 181–186) as it involves in an interaction with nicotinamide ring of NAD (Fig. 2). Further analysis of the 239 aa long polypeptide also portrays 16 potential sites for phosphorylation but no O-linked glycosylation sites.

Phylogenetic analysis of ALDH from various taxons have clearly grouped into different lineages such as Animalia, Plantae, Fungi, Protista and Prokaryota (Fig. 3). In the present study ALDH was clustered with Protista lineage confirming its origin might be from the protists.

3.3. Metal tolerance potential of PLCc38

Drop assay of PLCc38 transformed *cup1*$^{\Delta}$ *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ *and cot1*$^{\Delta}$ mutant strains of *S. cerevisiae* were able to grow on SD-Ura containing 150 μM of Cu, 40 μM Cd, 10 mM Zn and 2 mM Co respectively. Complete retardation of growth was observed for yeast mutants transformed only with empty vector pFL61. The wild type strain BY4741 transformed with empty vector pFL61 was able to grow in all metal amended media while *cup1*^{s} transformed with pFL61 failed to grow on 150 μM of Cu amended medium. In general, with increase in dilution, the decrease in the growth was observed for all yeast strains in metal amended medium (Fig. 4).

To study the metal tolerant profile of cDNA PLCc38 against Cu, Cd, Zn and Co, the transformants were allowed to grow in broth supplemented with a wide range of metal concentrations. The transformant *cup1*$^{\Delta}$/PLCc38 was capable of growing in presence of high concentrations of Cu. With the increase in external Cu concentration, inhibition of growth was observed although it was not affected majorly up till 500 μM Cu. However, when compared to wild type strain BY4741, the growth was significantly higher. The growth was significantly affected at 1000 μM Cu both in *cup1*$^{\Delta}$/

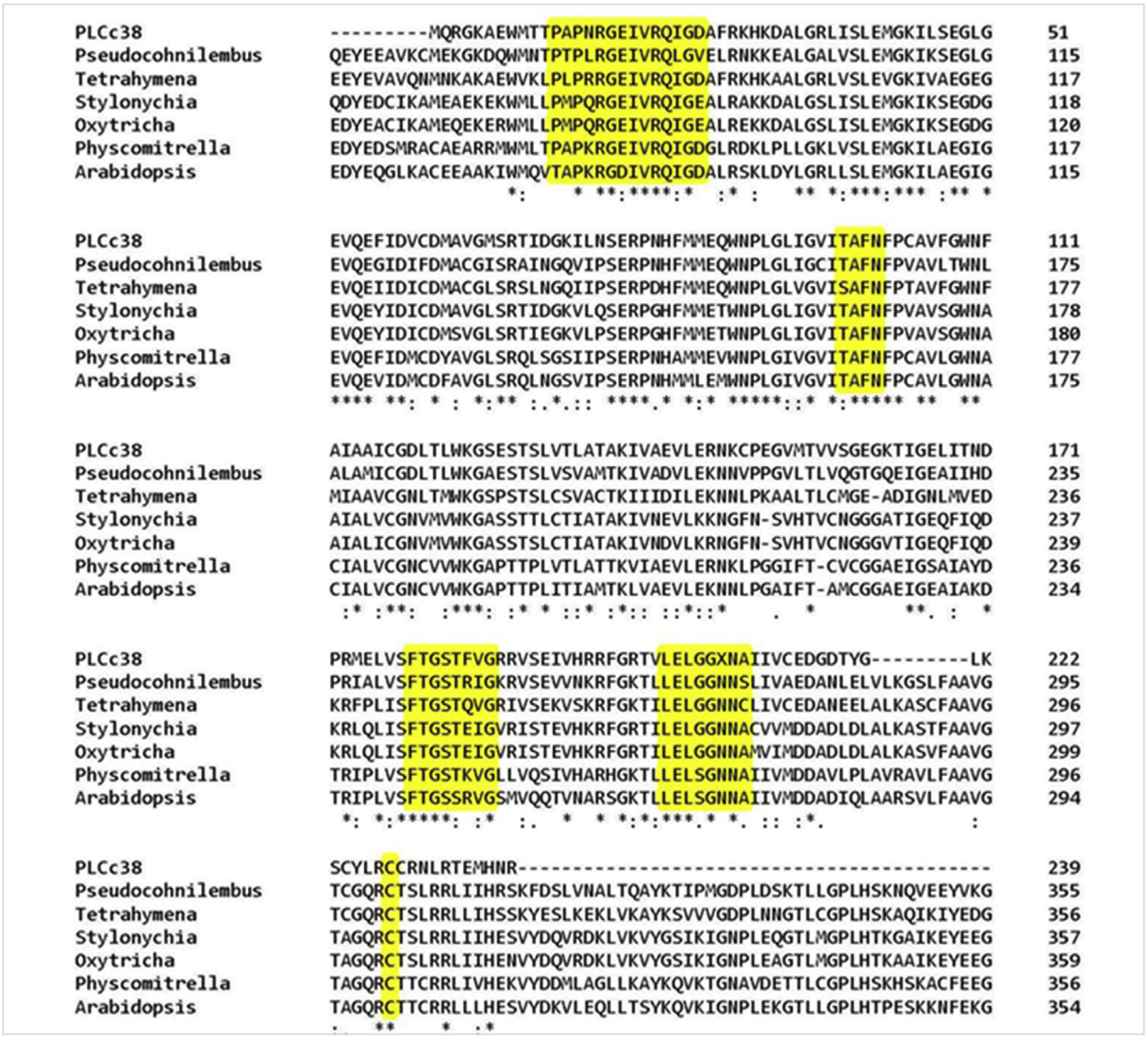


Fig. 1. Multiple sequence alignment of PLCc38 gene with other homologous sequences retrieved by BLASTp analysis. Accession numbers are *Pseudocohnilembus persalinus* (KRX07720) *Tetrahymena thermophila* SB210 (XP001027798), *Stylonychia lemnae* (CDW90834), *Oxytricha trifallax* (EJY81836), *Physcomitrella patens* (XP024368243) and *Arabidopsis thaliana* (AAK55676). Important conserved regions are highlighted.

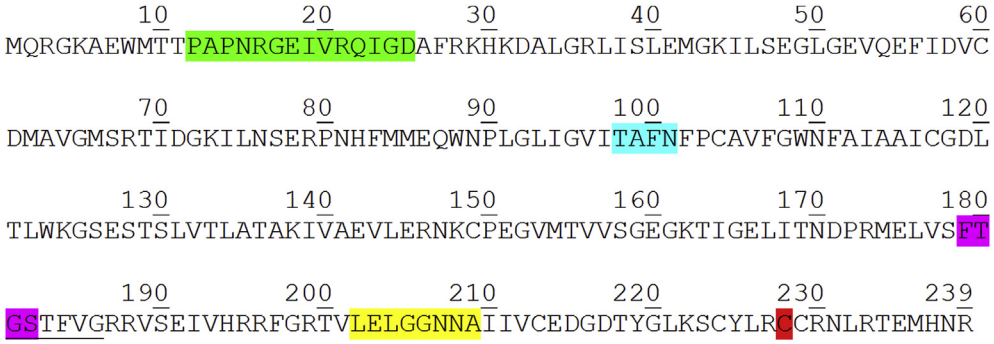


Fig. 2. Amino acid sequence of PlCc38. PR box is highlighted in green, NAD^+ Binding sites are highlighted in blue and pink. Glycine-rich region is underlined. Glutamic acid active site is in yellow. Catalytic thiol is represented in red.

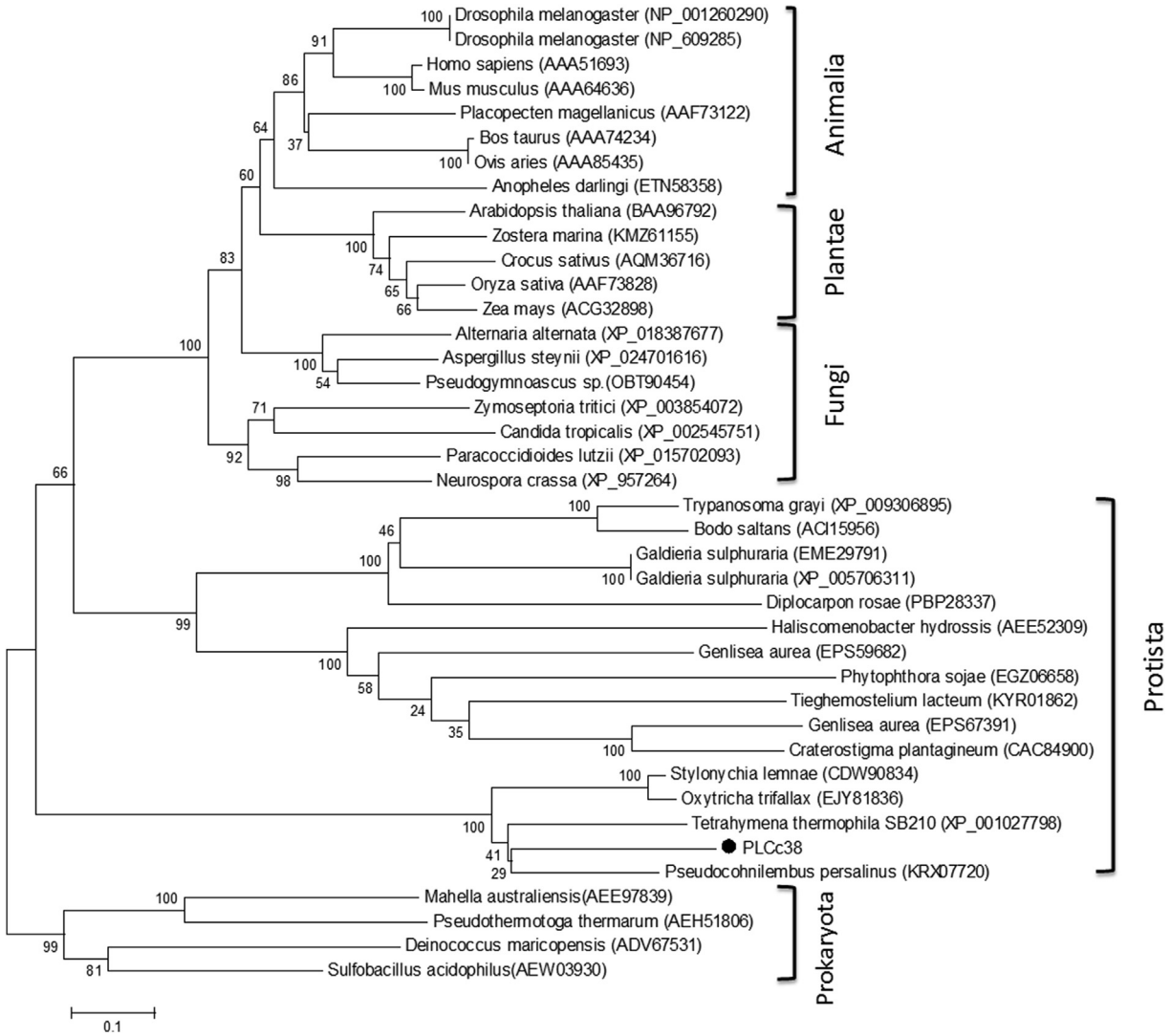


Fig. 3. Phylogenetic relation of the transcript PLCc38 with ALDHs of other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Kingdom names are denoted on the right side. GenBank Accession numbers are denoted in parentheses. Prokaryota ALDH sequences were included as outgroup. Branch lengths are proportional to evolutionary distances.

PLCc38 and wild type yeast strains. Mutant *cup1*$^{\Delta}$ transformed with empty pFL61 recorded no growth in presence of Cu (Fig. 5a). The transformant *ycf1*$^{\Delta}$/PLCc38 grew significantly well up to 60 μM Cd and decreased thereafter at higher concentration. The growth was comparable with that of the wild type strain BY4741 in all the concentrations (Fig. 5b). The growth of *zrc1*$^{\Delta}$/PLCc38 cells were able to grow well up to 12 mM, with maximum growth recorded at 10 mM of Zn. The growth was significantly decreased at 13 mM of Zn concentration. The growth of *zrc1*$^{\Delta}$/PLCc38 cells was higher than the wild type strain BY4741 at all the concentrations of Zn used in the study (Fig. 5c). The growth of PLCc38 transformed *cot1*$^{\Delta}$ cells decreased with increase in concentration of Co with maximum growth recorded at 2 mM of Co concentration. The growth was significantly higher in *cot1*$^{\Delta}$/PLCc38 cells compared to BY4741 transformed with empty vector in all concentrations of Co (Fig. 5d). Heavy metals like Cu and Cd can participate in oxidative modifications and redox cycles resulting in hyperaccumulation of extremely toxic Reactive Oxygen Species (ROS) [33,34]. There is increasing evidence which proves that increased level of ROS species can cause lipid peroxidation which generates large amounts of harmful aldehydes in the process [35] which are eliminated by ALDH. Presence of excess Cadmium, Cobalt and Zinc can easily result in the inactivation of enzymes important for life processes and also cause disruption of key enzymes required for defense against superoxides [35–38].

3.4. Yeast metal uptake

The metal content was determined in yeast transformants grown in different concentrations of Cu, Cd, Zn and Co by using ICP-MS. Uptake of Cu in *cup1*$^{\Delta}$/PLCc38 increased with increase in Cu concentration in the growth medium up to 500 μM and decreased significantly thereafter (Fig. 5e). The Cd content in the *ycf1*$^{\Delta}$/PLCc38 cells decreased with increase in Cd concentration in the medium with maximum being observed at 40 μM of Cd. The Cd content was significantly higher in *ycf1*$^{\Delta}$/PLCc38 cells compared to the wild type strain BY4741 transformed with empty vector in all concentrations (Fig. 5f). The *zrc1*$^{\Delta}$/PLCc38 cells accumulated higher content of Zn up to 11 mM and decreased at higher concentrations. The Zn content was significantly higher in these cells compared to the wild type strain in all the concentrations of Zn (Fig. 5g). In presence of Co, *cot1*$^{\Delta}$/PLCc38 cells recorded highest metal uptake at

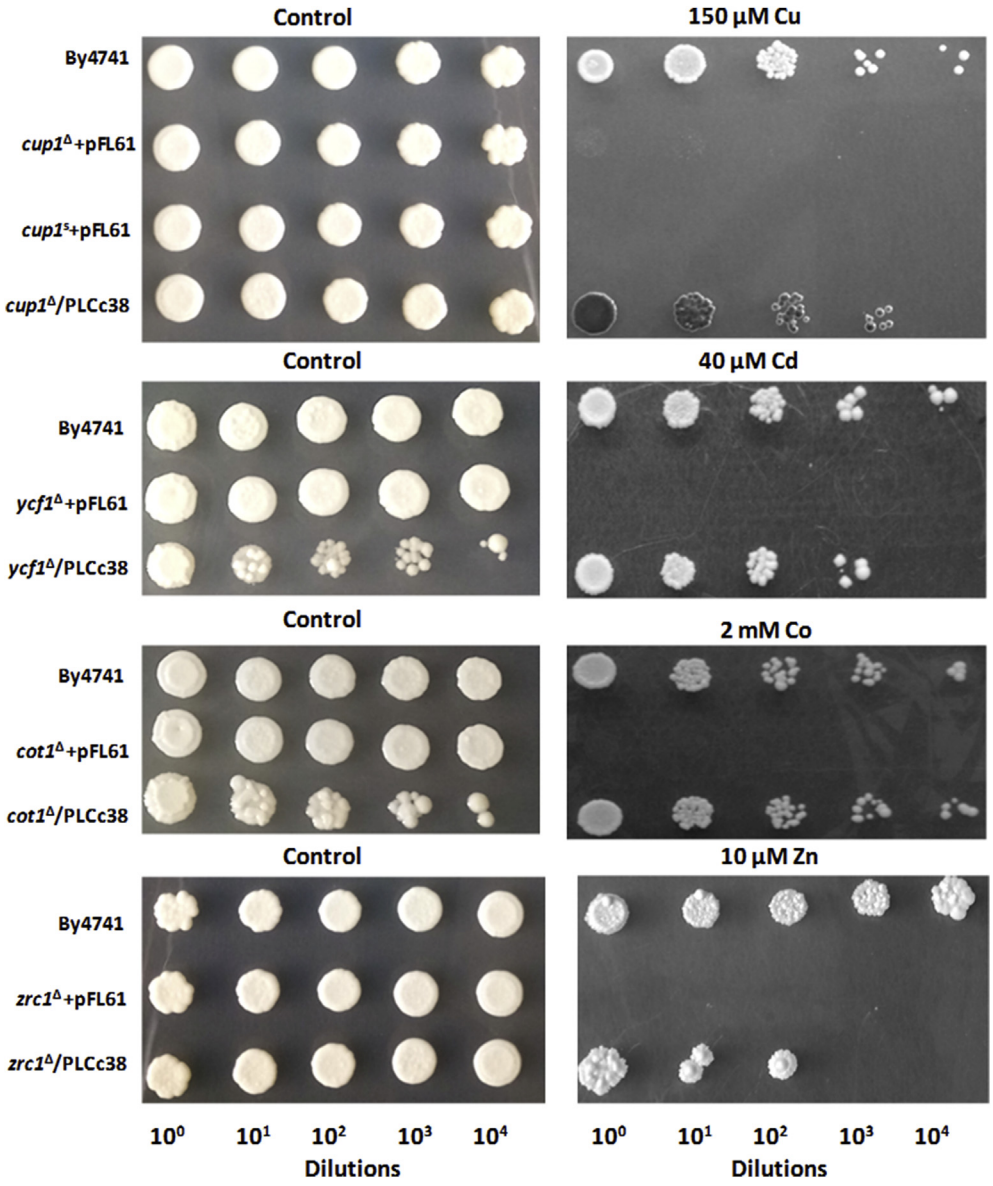


Fig. 4. Drop assay of PlCc38 transformed *Saccharomyces cerevisae* mutants, *cup1*$^{\Delta}$, *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ *and cot1*$^{\Delta}$ on SD-Ura agar. BY4741wild type and mutant strains transformed with empty vector pFL61 for positive and negative control. Serially diluted cultures were spotted on SD-Ura medium with or without metal supplement as indicated. Yeast strain *cup1*^{S} was also used as control.

10 mM Co which is comparable to wild type strain BY4741 (Fig. 5h). In a study, the Cu concentration of 175 µM and Cd concentration of 300 µM resulted in retardation of growth in wild type whereas transgenic plants with ALDH gene was less affected as far as germination, growth and lipid peroxidation are concerned [19]. The results in the present study are in agreement with these results and have depicted tolerance towards high concentrations of all four heavy metals; Cu, Cd, Zn and Co.

3.5. Oxydative stress tolerance by PLCc38

To confirm the role of the transcript PLCc38 in providing oxidative stress tolerance to the superoxide dismutase yeast mutant *sod2*$^{\Delta}$, the transformant *sod2*$^{\Delta}$/PLCc38 was tested with 0–3 mM of H_2O_2. Under control condition (0 mM H_2O_2) growth was similar in wild type BY4742, BY4742 transformed with PLCc38, *sod2*$^{\Delta}$ transformed with empty pFL61 and *sod2*$^{\Delta}$/PLCc38. In case of media supplemented with 1 mM H_2O_2, growth of BY4742, BY4742/PLCc38 and *sod2*$^{\Delta}$/PLCc38 showed similar level of growth but the growth of *sod2*$^{\Delta}$ transformed with empty vector was significantly retarded after 20 h. The transcript could confer tolerance to *sod2*$^{\Delta}$ till 2 mM H_2O_2 but at this concentration, growth of *sod2*$^{\Delta}$ with empty vector was significantly hindered. The growth of BY4742/PLCc38 was observed to be slightly higher than BY4742 alone at this concentration. At 3 mM H_2O_2, only the wild type and BY4742/PLCc38 could survive till 15 h and thereafter the growth declined. These results suggested that the wild type strain BY4742 transformed with PLCc38 showed higher tolerance to H_2O_2 than the other transformants. At this concentration neither *sod2*$^{\Delta}$/PLCc38 nor *sod2*$^{\Delta}$ transformed with empty pFL61 could survive (Fig. 6).

Heavy metals can cause cellular damage in two ways; firstly by binding to important functional domains of active biomolecules and thereby inactivating them, secondly by stimulating the formation of reactive oxygen species (ROS) and free radicals either by

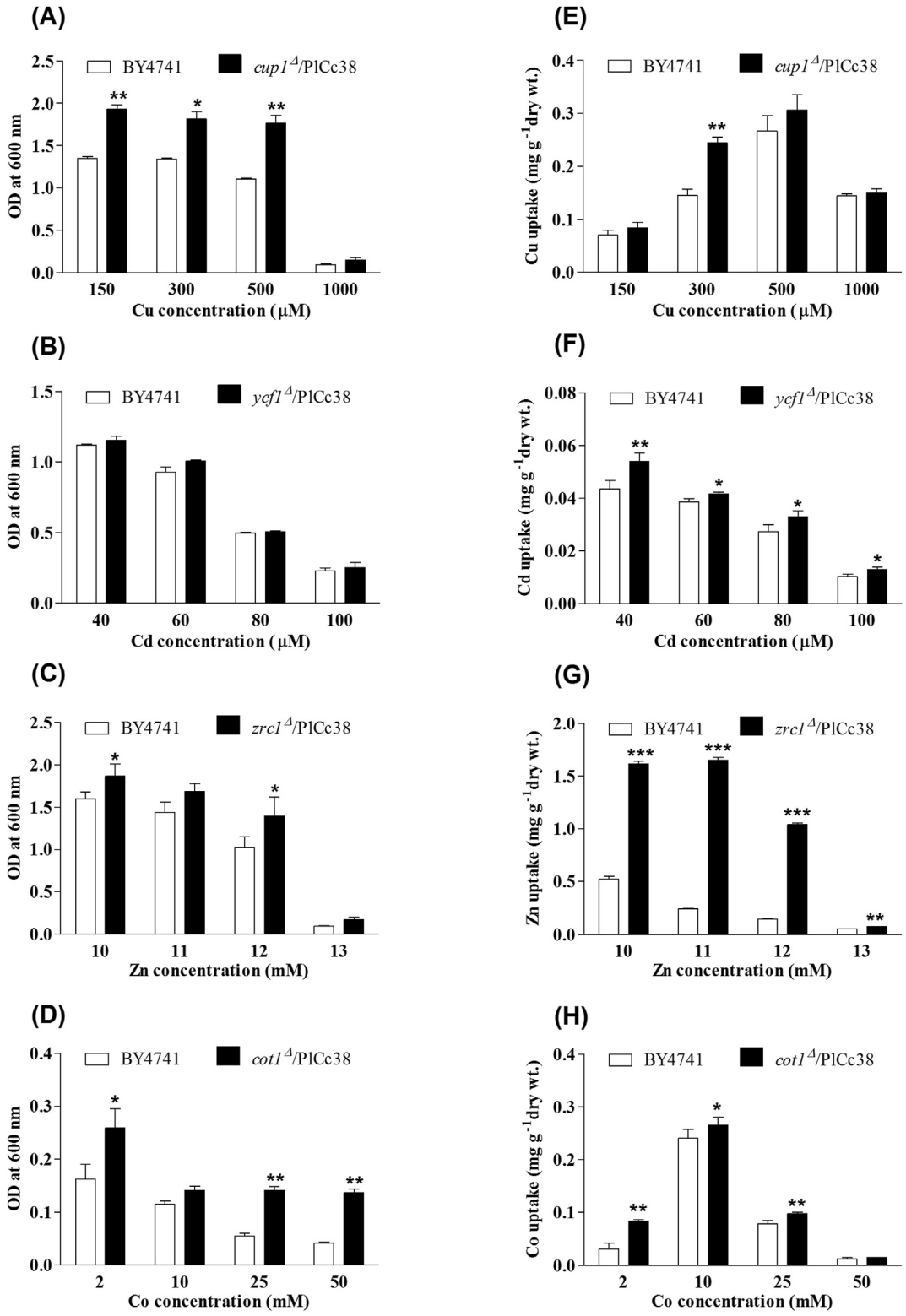


Fig. 5. Growth response of (A) *cup1*$^{\Delta}$ (B) *ycf1*$^{\Delta}$ (C) *zrc1*$^{\Delta}$ and (D) *cot1*$^{\Delta}$ transformed with PlCc38 and wild type BY4741 transformed with Empty Vector pFL61 in SD-Ura broth supplemented with different metals at different concentrations as shown. Effect of different concentration of (E) Copper (F) Cadmium (G) Zinc (H) Cobalt on metal uptake by PlCc38 when compared to wild type BY4741 after 40 h of growth. Error bars are ±SD ($n = 3$). (*, ** and *** are significant at $P < 0.05$, $P < 0.01$ and $P < 0.001$ respectively).

direct electron transfer involving metal cations, or as a consequence of metal mediated inhibition of metabolic reactions [39]. If the uncontrolled oxidation process is not mediated it would result in an increased rate of ROS formation and accumulation which would ultimately result in oxidative stress. It has been observed that the most reactive aldehydes are usually generated due to ROS-induced post-translational modifications or "protein carbonylation". ALDH helps in the oxidation which converts free aldehydes into less toxic

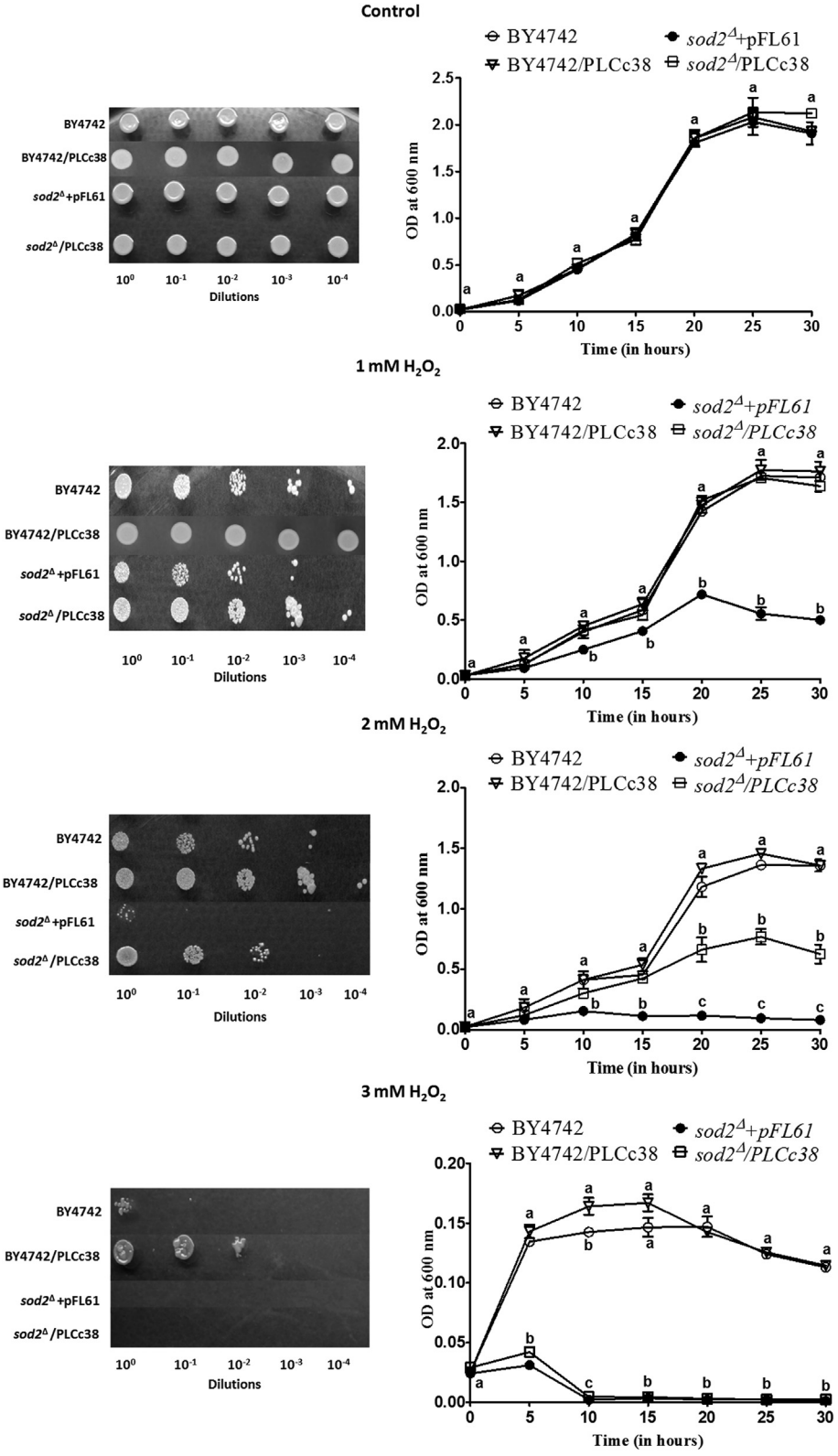


Fig. 6. Drop assay and growth response of yeast mutant *sod2*$^{\Delta}$ and wild type BY4742 transformed with PLCc38 in presence of different concentrations of H_2O_2 in the growth medium. Wild type BY4742 and *sod2*$^{\Delta}$ transformed with empty pFL61 were used as controls. Growth response was recorded till 30 h at an interval of 5 h. Error bars are ±SD. Values sharing common letters on error bars are not significant among the treatments at $P < 0.05$ ($n = 3$).

molecules and thus helps in detoxification [39]. Damage caused by abiotic stresses has been studied mostly in plants where transgenic technologies have confirmed the induced role of ALDH in response to drought stress, ROS-producing chemical treatment and heavy metal exposure. Reports suggest that oxidative stress due to lipid peroxydation in transgenic plants [40] and animal tissues [41] decreased considerably when the ALDH gene was expressed in the hosts. In another study, the presence of heavy metal cadmium caused significant proteomic changes in a fungal pathogen *Trichosporon asahii* as a result of metal-induced oxidative stress which increased the abundance of aldehyde dehydrogenase-like protein [42]. A wide array of metals was also analyzed in yeast for the response to oxidative test and was found to be regulated by aldehyde metabolism [43].

In the present study, experimental data strongly suggest that PLCc38 showing homology and having characteristic features of ALDH is responsible for metal induced oxidative stress which illustrates the active role of the transcriptome in maintaining cellular activity. This study is the first to report the characteristics of an environmental ALDH isolated using metatranscriptomics approach from polluted soil capable to tolerate and accumulate heavy metals inside cells at high metal concentrations.

4. Conclusions

From the results obtained in this study, we can conclude that the cDNA PLCc38 has the potential to tolerate toxic levels of metals when overexpressed in yeast. The translated protein contained all the characteristic features of aldehyde dehydrogenase which signified its role as a stress response protein having important contribution in dealing with abiotic stress in metal contaminated soil. The results suggest that metatranscriptomics can be employed as a modern-day 'omics' technique for the discovery of known and unknown genes with their mechanisms and can reveal a high degree of diversity of natural phenomenon in any kind of ecosystems. Heterologous expression of this gene proved that it could be utilized for a number of biotechnological applications like revegetation of metal contaminated sites, a biomarker for contamination studies and planning effective bioremediation techniques.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

Acknowledgements

Authors are thankful to Council of Scientific and Industrial Research, India, Govt. of India for providing financial support to carry out research in the CSIR project no. 38(1425)/16/EMR-II. The initial work was also supported by Indo-French Centre for the Promotion of Advanced Research (IFCPAR) under the project No. 4709-1.

References

[1] C. Garbisu, I. Alkorta, Basic concepts on heavy metal soil bioremediation, Eur. J. Miner. Process. Environ. Prot. 3 (1) (2003) 58–66.
[2] I. Nunes, S. Jacquiod, A. Brejnrod, P.E. Holm, A. Johansen, K.K. Brandt, et al., Coping with copper: legacy effect of copper on potential activity of soil bacteria following a century of exposure, FEMS Microbiol. Ecol. 92 (11) (2016) fiw175.
[3] J. Berg, K.K. Brandt, W.A. Al-Soud, P.E. Holm, L.H. Hansen, S.J. Sørensen, et al., Selection for Cu-tolerant bacterial communities with altered composition, but unaltered richness, via long-term Cu exposure, Appl. Environ. Microbiol. 78 (20) (2012) 7438–7446.
[4] L. Dziewit, L. Drewniak, Heavy metals resistance, metabolism and transformation—genomic, metagenomic and metatranscriptomic studies, in: J. Dlugonski (Ed.), Microbial Biodegradation: from Omics to Function and Application, Caister Academic Press, Poole, 2016, pp. 13–26.
[5] S. Clemens, Molecular mechanisms of plant metal tolerance and homeostasis, Planta 212 (4) (2001) 475–486.
[6] R. Marmeisse, H. Kellner, L. Fraissinet-Tachet, P. Luis, Discovering protein-coding genes from the environment: time for the eukaryotes? Trends Biotechnol. 35 (9) (2017) 824–835.
[7] J. Bailly, L. Fraissinet-Tachet, M.C. Verner, J.C. Debaud, M. Lemaire, M. Wésolowski-Louvel, Soil eukaryotic functional diversity, a metatranscriptomic approach, ISME J. 1 (7) (2007) 632–642.
[8] B. Thakur, R. Yadav, L. Fraissinet-Tachet, R. Marmeisse, M.S. Reddy, Isolation of multi-metal tolerant ubiquitin fusion protein from metal polluted soil by metatranscriptomic approach, J. Microbiol. Methods 152 (2018) 119–125.
[9] S. Bashiardes, G. Zilberman-Schapira, E. Elinav, Use of metatranscriptomics in microbiome research, Bioinf. Biol. Insights 10 (2016) 19–25.
[10] V. Aguiar-Pulido, W. Huang, V. Suarez-Ulloa, T. Cickovski, K. Mathee, G. Narasimhan, Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis, Evol. Bioinf. 12 (S1) (2016) 5–16.
[11] R.K. Yadav, F. Barbi, A. Ziller, P. Luis, R. Marmeisse, M.S. Reddy, et al., Construction of sized eukaryotic cDNA libraries using low input of total environmental metatranscriptomic RNA, BMC Biotechnol. 14 (1) (2014) 80–86.
[12] F. Lehembre, D. Doillon, E. David, S. Perrotto, J. Baude, J. Foulon, et al., Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes, Environ. Microbiol. 15 (10) (2013) 2829–2840.
[13] H. Kellner, P. Luis, D. Portetelle, M. Vandenbol, Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants, Microbiol. Res. 166 (5) (2011) 360–368.
[14] Z.S. Hua, Y.J. Han, L.X. Chen, J. Liu, M. Hu, S.J. Li, et al., Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics, ISME J. 9 (6) (2015) 1280–1294.
[15] M. Liljeqvist, F.J. Ossandon, C. González, S. Rajan, A. Stell, J. Valdes, et al., Metagenomic analysis reveals adaptations to a cold-adapted lifestyle in a low-temperature acid mine drainage stream, FEMS Microbiol. Ecol. 91 (4) (2015) fiv011.
[16] Z.J. Liu, Y.J. Sun, J. Rose, Y.J. Chung, C.D. Hsiao, W.R. Chang, et al., The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, Nat. Struct. Mol. Biol. 4 (4) (1997) 317–326.
[17] H.H. Kirch, A. Nair, D. Bartels, Novel ABA-and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*, Plant J. 28 (5) (2001) 555–567.
[18] S.M. Rodrigues, M.O. Andrade, A.P. Gomes, F.M. DaMatta, M.C. Baracat-Pereira, E.P. Fontes, Arabidopsis and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress, J. Exp. Bot. 57 (9) (2006) 1909–1918.
[19] R. Sunkar, D. Bartels, H.H. Kirch, Overexpression of a stress-inducible aldehyde dehydrogenase gene from Arabidopsis thaliana in transgenic plants improves stress tolerance, Plant J. 35 (4) (2003) 452–464.
[20] A. Walkley, A critical examination of a rapid method for determining organic carbon in soils—effect of variations in digestion conditions and of inorganic soil constituents, Soil Sci. 63 (4) (1947) 251–264.
[21] R.E. Kitson, M.G. Mellon, Colorimetric determination of phosphorus as molybdivanadophosphoric acid, Ind. Eng. Chem. 16 (6) (1944) 379–383.
[22] S.R. Olsen, Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate, United States Department of Agriculture, Washington, 1954.
[23] C.S. Piper, Soil and Plant Analysis; A Laboratory Manual of Methods for the Examination of Soils and the Determination of the Inorganic Constitutents of Plants, Hans Publishers, Bombay, 1966.
[24] V.D. Longo, E.B. Gralla, J.S. Valentine, Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae* mitochondrial production of toxic oxygen species in vivo, J. Biol. Chem. 271 (21) (1996) 12275–12280.
[25] Z.S. Li, Y.P. Lu, R.G. Zhen, M. Szczypka, D.J. Thiele, P.A. Rea, A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis (glutathionato) cadmium, Proc. Natl. Acad. Sci. Unit. States Am. 94 (1) (1997) 42–47.
[26] L. Li, J. Kaplan, The yeast gene MSC2, a member of the cation diffusion facilitator family, affects the cellular distribution of zinc, J. Biol. Chem. 276 (7) (2001) 5036–5043.
[27] D.S. Conklin, J.A. McMaster, M.R. Culbertson, C.H. Kung, COT1, a gene involved in cobalt accumulation in *Saccharomyces cerevisiae*, Mol. Cell Biol. 12 (9) (1992) 3678–3688.
[28] J.M. Van Raamsdonk, S. Hekimi, Deletion of the mitochondrial superoxide dismutase sod-2 extends lifespan in *Caenorhabditis elegans*, PLoS Genet. 5 (2) (2009), e1000361.
[29] M. Minet, M.E. Dufour, F. Lacroute, Complementation of Saccharomyces cerevisiae auxotrophic mutants by *Arabidopsis thaliana* cDNAs, Plant J. 2 (3) (1992) 417–422.
[30] R.D. Gietz, R.H. Schiestl, High-efficiency yeast transformation using the LiAc/

SS carrier DNA/PEG method, Nat. Protoc. 2 (1) (2007) 31–34.
[31] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (7) (2016) 1870–1874.
[32] Z.Y. Hseu, Evaluating heavy metal contents in nine composts using four digestion methods, Bioresour. Technol. 95 (1) (2004 1) 53–59.
[33] M.C. Romero-Puertas, J.M. Palma, M. Gómez, L.A. Del Rio, L.M. Sandalio, Cadmium causes the oxidative modification of proteins in pea plants, Plant Cell Environ. 25 (5) (2002) 677–686.
[34] T. Remans, K. Opdenakker, K. Smeets, D. Mathijsen, J. Vangronsveld, A. Cuypers, Metal-specific and NADPH oxidase dependent changes in lipoxygenase and NADPH oxidase gene expression in Arabidopsis thaliana exposed to cadmium or excess copper, Funct. Plant Biol. 37 (6) (2010) 532–544.
[35] A. Chaoui, S. Mazhoudi, M.H. Ghorbal, E. El Ferjani, Cadmium and zinc induction of lipid peroxidation and effects on antioxidant enzyme activities in bean (Phaseolus vulgaris L.), Plant Sci. 127 (2) (1997) 139–147.
[36] C. Geslin, J. Llanos, D. Prieur, C. Jeanthon, The manganese and iron superoxide dismutases protect Escherichia coli from heavy metal toxicity, Res. Microbiol. 152 (10) (2001) 901–905.
[37] Ö. Atici, G. Ağar, P. Battal, Changes in phytohormone contents in chickpea seeds germinating under lead or zinc stress, Biol. Plant. 49 (2) (2005) 215–222.
[38] K.J. Dietz, M. Baier, U. Krämer, Free radicals and reactive oxygen species as mediators of heavy metal toxicity in plants, in: M.N.V. Prasad, J. Hagemeyer (Eds.), Heavy Metal Stress in Plants: from Molecules to Ecosystems, Springer-Verlag, Berlin, 1999, pp. 73–97.
[39] P.A. Grimsrud, H. Xie, T.J. Griffin, D.A. Bernlohr, Oxidative stress and covalent modification of protein with bioactive aldehydes, J. Biol. Chem. 283 (32) (2008) 21837–21841.
[40] F. Liu, P.S. Schnable, Functional specialization of maize mitochondrial aldehyde dehydrogenases, Plant Physiol. 130 (4) (2002) 1657–1674.
[41] M. Hlavacova, J. Gumulec, T. Stracina, M. Fojtu, M. Raudenská, M. Masarik, et al., Different doxorubicin formulations affect plasma 4-hydroxy-2-nonenal and gene expression of aldehyde dehydrogenase 3A1 and thioredoxin reductase 2 in rat, Physiol. Res. 64 (2015) S653.
[42] S. Ilyas, A. Rehman, A.C. Varela, D. Sheehan, Redox proteomics changes in the fungal pathogen *Trichosporon asahii* on arsenic exposure: identification of protein responses to metal-induced oxidative stress in an environmentally-sampled isolate, PLoS One 9 (7) (2014) e102340.
[43] D. Hosiner, S. Gerber, H. Lichtenberg-Frate, W. Glaser, C. Schüller, E. Klipp, Impact of acute metal stress in Saccharomyces cerevisiae, PLoS One 9 (1) (2014), e83330.